

THE CROSS-SECTIONAL INTERCEPT IN ASSET PRICING TESTS

March 15, 2026

Abstract

We study the cross-sectional intercept in two-pass asset pricing tests with excess returns. When omitted priced factors exist, allowing a free intercept in the second-pass regression absorbs common pricing errors into fitted values, inflates conventional fit measures, and distorts estimated risk prices. We propose R_{pen}^2 , a fit measure for unrestricted cross-sectional regressions that treats the intercept as pricing error, and derive large-sample tests for both R_{pen}^2 and the no-intercept fit measure $R_{\text{ols}*}^2$. We apply the framework to the CAPM and 12 macro-factor models using anomaly-sorted portfolio cross sections. Across 78 specifications, R_{pen}^2 is negative in 76 cases and never significantly positive, whereas the conventional residual-based R^2 is significantly positive in 15 cases; $R_{\text{ols}*}^2$ is significantly positive in only one case. Allowing an intercept also often attenuates or reverses estimated factor premia. The evidence supports evaluating excess-return models primarily with no-intercept cross-sectional regressions and $R_{\text{ols}*}^2$, while using R_{pen}^2 to interpret unrestricted specifications.

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

JEL CLASSIFICATION: G12, C12, C58.

1 Introduction

Cross-sectional asset-pricing tests often estimate zero-beta rates that exceed 10% per year and imply negative (aggregate) equity premia. Taken literally, these estimates suggest that investors pay to bear systematic risk rather than demand compensation for it. Yet the cross-sectional intercepts that generate such implications are frequently treated as nuisance parameters rather than as informative diagnostics of model specification. This practice deserves closer scrutiny: When test assets are excess returns (the common empirical setting), any linear factor model’s pricing restriction mandates a zero intercept in the second-pass CSR. Why, then, do empirical implementations so often estimate a free intercept, and what are the consequences for model evaluation and risk-price inference?

The two-pass regression approach of [Black, Jensen, and Scholes \(1972\)](#) and [Cochrane \(2005\)](#) is the workhorse method for estimating linear factor models when at least one factor is not traded (i.e., does not represent an excess return).¹ In the canonical application, the second-pass CSR is run on average excess returns. In that setting, theory restricts the intercept to zero, suggesting a natural baseline specification in which the CSR is estimated without an intercept. In practice, however, many studies estimate an unrestricted intercept, motivated by the possibility that the observed risk-free rate imperfectly proxies the true zero-beta rate due to money-market frictions such as liquidity premia or borrowing–lending wedges ([Brennan, 1971](#); [Black, 1972](#); [Burnside, 2011](#); [Frazzini and Pedersen, 2014](#)). Yet as [Lewellen, Nagel, and Shanken \(2010\)](#) emphasize, plausible wedges are small (perhaps 1–2% annually) and cannot justify the much larger intercepts commonly estimated. More generally, asymptotic theory for two-pass cross-sectional regressions shows that standard inference can be misleading under heteroskedasticity and model misspecification, reinforcing the need to interpret unrestricted intercepts with caution ([Jagannathan and Wang, 1998](#)). This tension raises a basic question: When test assets are excess returns, what is the economic and econometric content of the cross-sectional intercept, and how should it enter model evaluation?²

¹The procedure of [Fama and MacBeth \(1973\)](#) is equivalent when factor loadings are estimated from the full sample period.

²Some studies impose the zero-intercept restriction ([Brennan, Wang, and Xia, 2004](#); [Burnside, 2011](#); [Savov, 2011](#); [Maio and Santa-Clara, 2012, 2017](#); [Maio and Philip, 2018](#); [Cooper and Maio, 2019](#); [Giglio and Xiu, 2021](#); [Cooper, Ma, and Maio, 2022](#)), while others report unrestricted CSRs as the baseline ([Gospodinov, Kan, and Robotti, 2019](#); [Raponi, Robotti, and Zaffaroni, 2020](#); [Gospodinov and Robotti, 2021](#); [Liao, Liu, and Xie, 2024](#)) or incorporate the restriction only in secondary analyses ([Kan, Robotti, and Shanken, 2013](#); [Lewellen, Nagel, and Shanken, 2010](#)). Recent work by [Maio \(2024\)](#) and [Maio and Min \(2025\)](#) explicitly treats the intercept as part of pricing errors.

This paper clarifies the role of the cross-sectional intercept in two-pass tests with excess returns. Within a broad linear asset pricing framework, an estimated intercept in an excess-return CSR represents a common pricing offset relative to the tested factor span and the chosen excess-return normalization. One source of such an offset is omitted priced components: Under omitted priced factors, the intercept equals the risk-price-weighted mean of omitted exposures that are not spanned by the included loadings. Treating this common offset as a free fitted parameter mechanically reallocates pricing failures into fitted values, overstating conventional fit measures and distorting risk-price inference. For model testing under the maintained normalization that excess returns are formed with the relevant (benchmark) risk-free rate, the null restriction is therefore a CSR without an intercept.

We make four contributions to how cross-sectional factor models are tested and evaluated. First, we characterize how omitted priced factors generate nonzero intercepts and prove that the intercept equals the risk-price-weighted mean of factor exposures that are not spanned by the included loadings. This result provides a structural foundation for treating the intercept as a common pricing offset (a pricing-error component) rather than as explained risk premia.

Second, we derive the exact mapping between restricted (no-intercept) and unrestricted (with-intercept) risk-price estimates. The key insight is that the distortion depends on the degree to which factor loadings are collinear with the constant vector. When loadings have high cross-sectional means relative to their dispersion (a condition satisfied empirically by market betas; see [Ahn, Perez, and Gadarowski 2013](#)), even moderate average pricing errors (in the specification with no intercept) generate large changes in estimated risk prices, including sign reversals through a collinearity amplification channel.

Third, we formalize the identification cost of including an intercept: The determinant of the cross-sectional design matrix equals the determinant of the factor-loading covariance matrix, so adding a constant regressor degrades rank conditions whenever loadings approach collinearity with the constant. In other words, it is more likely that the risk prices are not identified when the intercept is unrestricted, especially in settings where the betas display little cross-sectional dispersion.

Fourth, we develop large- T inference for the penalized and no-intercept R^2 measures. A key simplification relative to existing inference for the conventional R^2 ([Kan, Robotti, and Shanken , 2013](#)) is that both measures admit negative values, so the null $R^2 = 0$ is generically an interior

point. Standard delta-method Wald statistics with HAC standard errors therefore apply, and we provide closed-form null gradients that make the tests straightforward to implement while fully accounting for first-pass estimation error in the factor loadings.

Our analysis complements recent work on omitted-factor bias by [Giglio and Xiu \(2021\)](#), who propose a three-pass procedure to span the latent factor space, and by [Dello-Preite et al. \(2025\)](#), who interpret cross-sectional alphas as compensation for unsystematic risk. It also relates to [Kan and Zhang \(1999\)](#), who show that factors entirely uncorrelated with returns, “useless factors,” can nonetheless appear priced in two-pass tests because sampling error in estimated betas generates spurious cross-sectional explanatory power. Importantly, our setting differs: The omitted factors are priced and the included factors may be correlated with returns, yet the intercept still distorts inference by absorbing a common pricing offset. Unlike these approaches, we focus on how a free intercept mechanically reallocates pricing errors into fitted values within the classical two-pass regression framework.

We evaluate these implications using the CAPM and 12 prominent macro-factor models estimated on decile portfolios sorted by investment, profitability, momentum, and book-to-market. Including both single-anomaly and joint 40- and 17-portfolio cross-sections yields 78 cross-sectional estimations. The empirical results illustrate the quantitative importance of the intercept mechanism. When we treat the intercept as part of total pricing error, R^2_{pen} is negative in 76 of the 78 cases and positive in only two. Under the one-sided alternative of a positive fit, neither of these positive estimates is statistically distinguishable from zero. This contrasts sharply with conventional assessments: The traditional R^2 based only on CSR residuals is significantly positive for 15 of these 78 estimations. The discrepancy arises because estimated intercepts are economically large, often exceeding 2% per quarter in the two augmented estimations (associated with all four portfolio groups), and because many factor loadings exhibit precisely the pattern that amplifies distortions: High cross-sectional means and low dispersion. A complementary pricing-error metric yields a similar conclusion: when the intercept is included in the definition of pricing errors, total pricing error exceeds the raw average risk premia in 54 of the 78 estimations. The restricted CSR yields qualitatively similar evidence, with a significantly positive fit in only one of the 78 cases.

The distortion in fit measures has a direct counterpart in risk-price estimation. Consistent with the theoretical mapping, including an intercept frequently flips the sign of estimated risk

prices. The market factor illustrates the mechanism sharply: Across the nine 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 negative market price of risk is difficult to reconcile with standard economic priors. In our common design, this outcome arises when the constant is nearly in the span of the betas, a direct consequence of the collinearity amplification channel.

Rank tests confirm that identification is weaker when the intercept is included, indicating that unrestricted risk-price estimates are less precisely identified. These findings are consistent with a pattern observed across several studies: estimated risk prices for macro factors can change substantially, including sign reversals and losses of significance, when the zero-intercept restriction is imposed.

The economic content of the intercept can also be assessed through simple plausibility checks. Many of our intercept estimates imply economically large differences between the model-implied zero-beta rate and the observed risk-free rate used to form excess returns, and can imply very small or even negative implied (aggregate) equity premia if taken literally. We emphasize that these calculations are not intended as structural estimates of equilibrium objects; rather, they highlight that intercept magnitudes of the size documented above are difficult to reconcile with modest money-market frictions and point to model misspecification and/or fragile estimation when the intercept is left unrestricted.

Monte Carlo simulations reinforce this interpretation: 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 factors.

We stress that our analysis targets the mechanical consequences of including a free intercept, not the motivations for doing so, which are often well-founded (e.g., concern about the appropriate risk-free benchmark). These results suggest a simple reporting and evaluation practice in cross-sectional asset pricing tests. For excess-return CSRs under the maintained benchmark-risk-free-rate normalization, the second-pass regression should be estimated without an intercept. Researchers may still report the unrestricted intercept as a descriptive diagnostic of a common pricing offset (e.g., from omitted priced components or normalization differences), but model evaluation and

risk-price inference should be anchored on the restricted specification.

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 dispersion fit under the maintained normalization, whereas R^2_{pen} (which imposes the restriction that the intercept is the average pricing error) evaluates total pricing error 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 sensitive to collinearity amplification and, as the rank tests show, are often weakly identified. The restricted specification therefore provides a more reliable basis for inference.

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 addresses factor measurement error. Section 9 develops asymptotic inference for the penalized and no-intercept R^2 measures. Section 10 concludes.

2 Theory

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

We develop four sets of results. In Subsection 2.1, we characterize how omitted factors generate a nonzero cross-sectional intercept and relate the baseline (unrestricted) cross-sectional regression (CSR, henceforth) coefficients to the coefficients from an augmented CSR. In Subsection 2.2, we analyze the implications of the intercept for pricing errors and common performance metrics, including the cross-sectional R^2 . In Subsection 2.3, we derive the mapping between risk price estimates from restricted (no-intercept) and unrestricted (with-intercept) CSRs. Finally, Subsection 2.4 discusses identification and collinearity issues that arise when an intercept is included.

2.1 The Cross-Sectional Intercept and Missing Factors

In empirical cross-sectional 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 $i = 1, \dots, n$ (relative to the risk-free rate), with sample analogue given by $\mu_i = (1/T) \sum_{t=1}^T R_{i,t} > 0$.³ Let $\bar{\mu} = \mathbb{E}_n(\mu_i) > 0$ denote the (cross-sectional) average risk premium among the test assets. The associated $n \times 1$ vector of sample risk premia on all assets is $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$. Let $\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{i,1}, \dots, \hat{\beta}_{i,K})'$ denote the first-stage betas (for asset i) estimated from time-series multivariate regressions, and let $\hat{\boldsymbol{\beta}}$ be the corresponding $n \times K$ matrix of factor loadings. The $K \times 1$ vector of factor risk prices is $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_K)'$. We work with cross-sectional operators, $\mathbb{E}_n(\cdot)$, $\text{Var}_n(\cdot)$, and $\text{Cov}_n(\cdot, \cdot)$, taken over $i = 1, \dots, n$. To simplify the notation, write $\mathbf{m} = \mathbb{E}_n(\hat{\boldsymbol{\beta}}_i)$ as the $K \times 1$ vector of (cross-sectionally) average betas and $\mathbf{M} = \mathbb{E}_n(\hat{\boldsymbol{\beta}}_i \hat{\boldsymbol{\beta}}_i')$ as the $K \times K$ matrix of betas' second moments. It follows that the covariance matrix of the factor loadings is given by $\boldsymbol{\Sigma} = \text{Var}_n(\hat{\boldsymbol{\beta}}_i) = \mathbf{M} - \mathbf{m}\mathbf{m}'$. A “hat” denotes a sample-based estimate or statistic, typically obtained by OLS (either in the time-series or in the cross-section).⁴

³The assumption of positive risk premia facilitates the interpretation of the theoretical results presented in this section, while being consistent with most empirical setups in cross-sectional asset pricing tests. Yet, such an assumption is not necessary for the derivations in this section.

⁴The 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 $(\mu_i, \hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\rho}}_i)$ with $(\mathbb{E}(R_{i,t}), \boldsymbol{\beta}_i, \boldsymbol{\rho}_i)$.

Consider the second-stage OLS CSR with an intercept,⁵

$$\mu_i = \hat{\lambda}_0 + \sum_{j=1}^K \hat{\lambda}_j \hat{\beta}_{i,j} + \hat{\omega}_i = \hat{\lambda}_0 + \hat{\boldsymbol{\lambda}}' \hat{\boldsymbol{\beta}}_i + \hat{\omega}_i, \quad i = 1, \dots, n, \quad (1)$$

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

$$\boldsymbol{\mu} = \hat{\lambda}_0 \mathbf{1} + \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\lambda}} + \hat{\boldsymbol{\omega}}, \quad (2)$$

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

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

$$\mu_i = \hat{\theta}_0 + \sum_{j=1}^K \hat{\theta}_j \hat{\beta}_{i,j} + \sum_{l=1}^L \hat{\gamma}_l \hat{\rho}_{i,l} + \hat{\omega}_i = \hat{\theta}_0 + \hat{\boldsymbol{\theta}}' \hat{\boldsymbol{\beta}}_i + \hat{\boldsymbol{\gamma}}' \hat{\boldsymbol{\rho}}_i + \hat{\omega}_i, \quad (3)$$

where $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_K)'$ 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.

⁵We do not consider the case in which the second-step CSR is estimated by GLS, an approach that is defended by Kandel and Stambaugh (1995) and Lewellen, Nagel, and Shanken (2010), among others. The reason is that (following the bulk of the literature) we are interested in assessing the capacity of a given model in explaining the original portfolios, which have a fundamental interest, rather than explaining a mean-variance transformation of those same portfolios (“repackaged portfolios”). Importantly, it is often the case that the repackaged portfolios assume extreme weights on the original portfolios, which makes the former difficult to be implemented in practice. Hence, despite the prescription from Lewellen, Nagel, and Shanken (2010), the bulk of the literature still focuses nowadays on the OLS CSR when estimating and evaluating linear models that contain macro factors.

Accordingly, $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\gamma}}$ 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, $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \dots, \hat{\gamma}_L)'$ are the OLS coefficients on the additional factors in the augmented CSR, and $\hat{\boldsymbol{\rho}}_i$ denotes the corresponding first-pass loading estimates for asset i (i.e., the i th row of $\hat{\boldsymbol{\rho}}$) in that augmented specification.

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

$$\boldsymbol{\mu} = \hat{\theta}_0 \mathbf{1} + \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\rho}} \hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\omega}}, \quad (4)$$

where $\hat{\boldsymbol{\omega}} = (\hat{\omega}_1, \dots, \hat{\omega}_n)'$ 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 $l = 1, \dots, L$. These restrictions are not required for the algebraic results below.⁶

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

Assumption 3. The first-stage betas on the included factors, $\hat{\boldsymbol{\beta}}$, and (when they appear) the first-stage loadings on the omitted factors, $\hat{\boldsymbol{\rho}}$, 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 $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\rho}})$. When the analysis is embedded in standard two-pass asymptotics with $T \rightarrow \infty$ (and fixed n), the first-pass estimates satisfy $\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i + O_p(T^{-1/2})$ under regularity conditions (e.g., [Shanken \(1992\)](#); [Kan, Robotti, and Shanken \(2013\)](#)). Hence, the same identities characterize probability limits of the corresponding

⁶These two assumptions on the omitted factors make economic sense: The omitted factors should be priced to be relevant in the first place. Moreover, given the assumption of positive risk premia for each asset, $\mu_i > 0$, it makes sense that all omitted factors generate a positive risk premium (on average), otherwise the omitted factors would do worse than a constant when it comes to pricing the test assets.

statistics by continuous mapping. Section 9 then develops large- T inference for the cross-sectional explanatory ratios presented below as a generated regressor and applying the delta method with heteroskedasticity and autocorrelation consistent (HAC) standard errors.

Consider the auxiliary cross-sectional projection of omitted loadings onto the included betas,

$$\hat{\boldsymbol{\rho}} = \mathbf{1}\delta_0' + \hat{\boldsymbol{\beta}}\boldsymbol{\Delta} + \mathbf{U},$$

where $\delta_0 \in \mathbb{R}^L$ is the intercept coefficient vector, $\boldsymbol{\Delta} \in \mathbb{R}^{K \times L}$ is the slope coefficient matrix, and $\mathbf{U}$ is an $n \times L$ matrix of residuals orthogonal (in sample) to both $\mathbf{1}$ and $\hat{\boldsymbol{\beta}}$. For an arbitrary omitted loading, $l = 1, \dots, L$, we have:

$$\hat{\rho}_{i,l} = \delta_{0,l} + \sum_{j=1}^K \delta_{j,l} \hat{\beta}_{i,j} + u_{i,l}, \quad i = 1, \dots, n.$$

The OLS estimates are:

$$\hat{\boldsymbol{\Delta}} = \boldsymbol{\Sigma}^{-1} \text{Cov}_n(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\rho}}_i), \quad \hat{\boldsymbol{\delta}}_0 = \mathbb{E}_n(\hat{\boldsymbol{\rho}}_i) - \hat{\boldsymbol{\Delta}}' \mathbf{m}. \quad (5)$$

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

Proposition 1. *Let $(\hat{\theta}_0, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}})$ be the OLS coefficients from the augmented CSR (4). Under Assumptions 1–3, the baseline second-pass OLS coefficients from (2) satisfy*

$$\hat{\lambda}_0 = \hat{\theta}_0 + \hat{\boldsymbol{\delta}}_0' \hat{\boldsymbol{\gamma}}, \quad \hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\Delta}}' \hat{\boldsymbol{\gamma}}, \quad (6)$$

where

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

is the OLS coefficient matrix from the cross-sectional projection of $\hat{\boldsymbol{\rho}}$ on $\mathbf{W} = [\mathbf{1} : \hat{\boldsymbol{\beta}}]$. 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 $\widehat{\Delta}\widehat{\gamma}$, 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.⁷

Two special cases of Proposition 1 are worth highlighting, which imply the same restricted version of $\widehat{\lambda}_0$. For intuition, we take $\theta_0 = 0$ in population, so that $\widehat{\theta}_0$ should be small in large samples.

(i) If the included betas have zero cross-sectional means, $\mathbf{m} = \mathbf{0}$, then $\widehat{\delta}_0 = \mathbb{E}_n(\widehat{\rho}_i)$ and therefore

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

Thus, up to the sampling term $\widehat{\theta}_0$, the intercept in the baseline CSR equals the risk-price-weighted average omitted exposure. Intuitively, when the included betas are mean-zero across assets, shifting their prices cannot absorb any part of the average omitted risk premia, so the entire amount shows up in the intercept.

(ii) If the omitted loadings have zero cross-sectional covariance with the included betas after demeaning across assets, then $\text{Cov}_n(\widehat{\beta}_i, \widehat{\rho}_i) = \mathbf{0}$ and (5) implies $\widehat{\Delta} = \mathbf{0}$. In that case, the included and omitted loadings have no residual cross-sectional comovement, so OLS does not tilt the included risk prices: $\widehat{\lambda} = \widehat{\theta}$ and

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

The following proposition provides a useful interpretation of the cross-sectional intercept.

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

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

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, $\widehat{\gamma}'\mathbb{E}_n(\widehat{\rho}_i)$; and (ii) the part of that average risk premia that the baseline model reallocates to the

⁷This omitted-factor bias aligns with Giglio and Xiu (2021), who show similar mappings in population and use rotation invariance for identification regardless of unobserved factors. Our bias-adjustment complements Balduzzi and Robotti (2008), who derive relations between premia in traditional CSRs and time-series formulations using mimicking portfolios. Our results extend these insights by treating the intercept as the risk price-weighted unspanned omitted loadings, providing a rationale for penalized fit measures in standard two-pass settings, as discussed below.

included factors by tilting their prices from $\hat{\theta}$ to $\hat{\lambda}$, namely $\mathbf{m}'(\hat{\lambda} - \hat{\theta})$. 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. A priori, this second component is likely to have a smaller (in magnitude) contribution (than the first component) to the intercept estimate: The average loadings are scaled by differences in risk price estimates for the included factors, $\hat{\lambda} - \hat{\theta}$, while in the first component the average loadings are scaled by the omitted risk prices, $\hat{\gamma}$.

Importantly, Proposition 2 provides guidance on the actual intercept estimates obtained in cross-sectional tests (under the assumption that $\hat{\theta}_0 \approx 0$). 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 $-\mathbf{m}'(\hat{\lambda} - \hat{\theta}) > 0$, that is, when the tilt in the included-factor risk prices has a negative projection on the mean-loading direction $\mathbf{m}$, which implies $\hat{\lambda}_0 \approx \hat{\gamma}'\mathbb{E}_n(\hat{\rho}_i) - \mathbf{m}'(\hat{\lambda} - \hat{\theta}) > \hat{\gamma}'\mathbb{E}_n(\hat{\rho}_i)$; ⁸ and (ii) when $\mathbf{m}'(\hat{\lambda} - \hat{\theta}) \approx 0$, which implies $\hat{\lambda}_0 \approx \hat{\gamma}'\mathbb{E}_n(\hat{\rho}_i) > 0$, and can arise from either small average betas for the included factors, $\mathbf{m} \approx \mathbf{0}$, or an approximate zero bias in risk-price estimates, $\hat{\lambda} - \hat{\theta} \approx \mathbf{0}$, with this second case being discussed further below. Second, given the assumption that $\hat{\gamma}'\mathbb{E}_n(\hat{\rho}_i) > 0$, the rarer case of a negative intercept estimate ($\hat{\lambda}_0 < 0$) requires $-\mathbf{m}'(\hat{\lambda} - \hat{\theta}) < 0$, equivalently a positive projection of the tilt $\hat{\lambda} - \hat{\theta}$ on the mean-loading direction $\mathbf{m}$. ⁹

Another important implication from Proposition 2 is that $\hat{\lambda}_0$ can be close to zero even if there are omitted factors ($\hat{\gamma} \neq \mathbf{0}$) whenever $\hat{\gamma}'\mathbb{E}_n(\hat{\rho}_i) \approx \mathbf{m}'(\hat{\lambda} - \hat{\theta}) > 0$ (assuming that $\hat{\theta}_0$ is negligible). 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 a non-zero intercept implies omitted priced risks. Therefore, a zero cross-sectional 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., $\hat{\Delta}$ 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 $\hat{\theta}_0$) to $\hat{\gamma}'\mathbb{E}_n(\hat{\rho}_i)$. ¹⁰ From

⁸In the single-factor case ($K = 1$), this condition means that the average beta (m) and the tilt in the risk price estimate, $\hat{\lambda} - \hat{\theta}$, have opposite signs.

⁹In the single-factor case ($K = 1$), this condition means that the average beta (m) and the tilt in the risk price estimate, $\hat{\lambda} - \hat{\theta}$, have the same sign.

¹⁰A similar expression for the intercept is obtained in [Maio and Min \(2025\)](#) by relying on a simplified framework: $K = L = 1$ combined with the assumption that the betas for the observed factor have a zero mean.

Proposition 1, we have,

$$\hat{\lambda} = \hat{\theta} \iff \hat{\Delta}\hat{\gamma} = \mathbf{0},$$

which holds if $\hat{\Delta} = \mathbf{0}$, i.e., if the omitted exposures have zero cross-sectional covariance with the included betas after demeaning across assets. Therefore, a model with a large intercept estimate (which is approximately equal to the price-weighted average omitted exposure) is likely to have included betas that have nearly zero cross-sectional covariance with the omitted loadings.

The results above are purely algebraic and require no time-series orthogonality restrictions. The main conclusion is that large cross-sectional intercepts predominantly reflect 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; Lioui and Maio 2014), 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.¹¹

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 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, whose intercept satisfies

$$\hat{\lambda}_0 = \hat{\theta}_0 + \hat{\delta}'_0 \hat{\gamma}.$$

The same logic applies to conditional tests of the Consumption–CAPM (e.g., Lettau and Ludvigson

¹¹ 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.

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 Section 8, 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 Cross-Sectional Intercept and Pricing Errors

Building on the omitted-factor characterization in Subsection 2.1, we now examine how the cross-sectional 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 $s = \mathbf{m}'\mathbf{M}^{-1}\mathbf{m}$ denote the R^2 from the auxiliary (no-intercept) regression $1 = \hat{\beta}_i'\mathbf{u} + \varepsilon_i$, with $\mathbf{u} = \mathbf{M}^{-1}\mathbf{m}$ (share of the constant vector $\mathbf{1}$ explained by the span of $\hat{\beta}$).¹²

Under Assumption 2, $s \in [0, 1]$. In addition to **Assumptions 1 – 3**, the following assumption holds:

Assumption 4. Let μ_i denote the mean (excess) return of asset i . Then, $\text{Var}_n(\mu_i) > 0$.

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

$$\hat{\alpha}_i = \mu_i - \hat{\mathbf{X}}'\hat{\beta}_i = \hat{\omega}_i + \hat{\lambda}_0, \quad (7)$$

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

¹²When $K = 1$, 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.

Given the baseline cross-sectional OLS R^2 ,

$$R_{\text{ols}}^2 = 1 - \frac{\mathbb{E}_n(\hat{\omega}_i^2)}{\text{Var}_n(\mu_i)}, \quad (8)$$

we defined a similar metric that is based on the “total” pricing errors (rather than the regression residuals):

$$R_{\text{pen}}^2 = 1 - \frac{\mathbb{E}_n(\hat{\alpha}_i^2)}{\text{Var}_n(\mu_i)}. \quad (9)$$

R_{ols}^2 credits the intercept as part of the fitted risk premia from the CSR, while R_{pen}^2 treats it as part of total pricing error.¹³ We therefore interpret R_{pen}^2 as a dispersion-scaled total-pricing-error diagnostic for the unrestricted CSR rather than as a pure dispersion-fit statistic. Keeping $\text{Var}_n(\mu_i)$ in the denominator preserves comparability with the conventional cross-sectional R^2 and yields the exact wedge decomposition below. The complementary pure dispersion-fit question is answered by $R_{\text{ols}*}^2$ from the restricted CSR, which is presented below. In light of the discussion in the previous subsection, the risk premia associated with the missing factors (that originate an intercept) are “discounted” by R_{pen}^2 , while credited as part of the model’s performance by R_{ols}^2 . The (negative) intercept’s contribution to the model’s fit (the *wedge*) is given by

$$\text{Wedge} = \frac{\hat{\lambda}_0^2}{\text{Var}_n(\mu_i)},$$

as shown in the following proposition.

Proposition 3. *Assuming $\text{Var}_n(\mu_i) > 0$ (Assumption 4) so that the R^2 ratios are well-defined, with $\hat{\alpha}_i = \hat{\lambda}_0 + \hat{\omega}_i$ and $\mathbb{E}_n(\hat{\omega}_i) = 0$,*

$$R_{\text{pen}}^2 = R_{\text{ols}}^2 - \frac{\hat{\lambda}_0^2}{\text{Var}_n(\mu_i)}, \quad \text{equivalently} \quad \mathbb{E}_n(\hat{\alpha}_i^2) = \mathbb{E}_n(\hat{\omega}_i^2) + \hat{\lambda}_0^2. \quad (10)$$

Moreover, $R_{\text{pen}}^2 = R_{\text{ols}}^2$ iff $\hat{\lambda}_0 = 0$.

¹³ Another frequently used R^2 relies on the cross-sectional variance (rather than the cross-sectional second moment) of the pricing errors, $\text{Var}_n(\hat{\alpha}_i)$. However, this alternative measure is not impacted by the new definition of the pricing errors (in the CSR with an intercept),

$$R^2 = 1 - \frac{\text{Var}_n(\hat{\alpha}_i)}{\text{Var}_n(\mu_i)} = 1 - \frac{\text{Var}_n(\hat{\omega}_i)}{\text{Var}_n(\mu_i)} = 1 - \frac{\mathbb{E}_n(\hat{\omega}_i^2)}{\text{Var}_n(\mu_i)},$$

given that $\hat{\lambda}_0$ is common across test assets.

This proposition makes clear that the intercept has a negative effect on the model's fit, and that such 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 cross-sectional fit. The following corollary provides further intuition on the new cross-sectional R^2 .

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

$$R_{\text{pen}}^2 \in [-\text{Wedge}, 1], \quad R_{\text{pen}}^2 < 0 \iff \hat{\lambda}_0^2 > \text{Var}_n(\mu_i) R_{\text{ols}}^2 \iff \hat{\lambda}_0^2 > \text{Var}_n(\mu_i) - \mathbb{E}_n(\hat{\omega}_i^2).$$

Thus, the lower bound on the new metric equals $-\hat{\lambda}_0^2 / \text{Var}_n(\mu_i)$, which coincides with $R_{\text{ols}}^2 = 0$. This implies that models producing very large intercept estimates (in magnitude) will, in general, be associated with an extreme negative fit in sample. Moreover, models that produce large intercept estimates and/or large cross-sectional residuals (in magnitude) are more likely to have a negative fit compared to models with small intercepts and/or residuals. As usually, a small cross-sectional dispersion in risk premia, $\text{Var}_n(\mu_i)$, is more likely to lead to a negative fit. In this context, a negative cross-sectional R^2 means that the squared intercept is large enough relative to the cross-sectional variation in the explained risk premia, $\text{Var}_n(\mu_i) - \mathbb{E}_n(\hat{\omega}_i^2)$.

It is helpful to compare the new R^2 metric from the unrestricted CSR (i.e., with intercept) with the standard OLS metric associated with the restricted CSR (i.e., without intercept). The reason is that each of these two explanatory ratios represents the appropriate cross-sectional R^2 under the respective setting. Define the OLS CSR without intercept,

$$\mu_i = \tilde{\boldsymbol{\lambda}}' \hat{\boldsymbol{\beta}}_i + \hat{\nu}_i, \tag{11}$$

where $\tilde{\boldsymbol{\lambda}}$ are the OLS coefficients and $\hat{\nu}_i$ are the corresponding OLS cross-sectional residuals. The following corollary provides the relationship of the fit obtained from both types of CSRs.

Corollary 2. *Given*

$$R_{\text{ols}*}^2 = 1 - \frac{\mathbb{E}_n(\hat{\nu}_i^2)}{\text{Var}_n(\mu_i)},$$

we have:

$$R_{\text{pen}}^2 \leq R_{\text{ols*}}^2, \quad R_{\text{ols*}}^2 - R_{\text{pen}}^2 = s \frac{\hat{\lambda}_0^2}{\text{Var}_n(\mu_i)}.$$

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

This corollary shows that, in general, the unrestricted CSR will produce a lower fit than the restricted counterpart when fit is evaluated using the appropriate pricing errors for the unrestricted specification (i.e., when the intercept is treated as part of total pricing error). Intuitively, the gap in fit between the two CSRs increases with the magnitude of the intercept estimate. Thus, models with large intercept estimates are more likely to also produce large gaps in fit across the two CSRs. The discrepancy in fit also increases with s , which means that models in which the factors have betas with high cross-sectional means (in magnitude) and/or low cross-sectional dispersion are more likely to produce a large difference in pricing performance between the two CSRs.

There are two extreme scenarios in which the gap in fit is zero, so that estimating the unrestricted or restricted CSR becomes largely indifferent (with respect to the fit obtained): (i) The obvious case where the intercept is zero, thus implying that both OLS CSRs produce the same risk prices and residuals, which in turn implies the same explanatory ratios. In other words, models with small intercept estimates are more likely to deliver very similar cross-sectional R^2 under both CSRs. (ii) If all betas have a (cross-sectional) mean of zero ($s = 0$), it follows that both regressions produce the same risk price estimates, while the intercept estimate coincides with the average raw risk premia, $\hat{\lambda}_0 = \bar{\mu}$. This implies that the difference in residuals from the two CSRs exactly equals the intercept estimate, $\hat{\nu}_i - \hat{\omega}_i = \hat{\lambda}_0$, which in turn implies that $R_{\text{pen}}^2 = R_{\text{ols*}}^2$. The empirical implication is that the two regressions are largely equivalent for models in which all factors have betas with a small mean (in magnitude).¹⁴

Another widely used performance metric is the root mean squared pricing error as a fraction of

¹⁴In the single-factor case ($K = 1$), this equivalence is immediate because $s = m^2/M$.

the raw average risk premia (e.g., [Lettau, Ludvigson, and Ma 2019](#)):¹⁵

$$RRMSE = \sqrt{\frac{\mathbb{E}_n(\hat{\omega}_i^2)}{\mathbb{E}_n(\mu_i^2)}}. \quad (12)$$

Similarly to the R^2 metrics, the pricing errors should contain the intercept estimate, and failure to do so might imply (substantially) lower values of the (relative) average pricing error. This is summarized in the following proposition.

Proposition 4. *The correct (relative) average pricing error in the unrestricted CSR is*

$$RRMSE_{\text{pen}} = \sqrt{\frac{\hat{\lambda}_0^2}{\mathbb{E}_n(\mu_i^2)} + \frac{\mathbb{E}_n(\hat{\omega}_i^2)}{\mathbb{E}_n(\mu_i^2)}} \geq \sqrt{\frac{\mathbb{E}_n(\hat{\omega}_i^2)}{\mathbb{E}_n(\mu_i^2)}}, \quad (13)$$

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

Similarly to the case of the cross-sectional R^2 , the intercept introduces a wedge in the (relative) average pricing error, which equals $\hat{\lambda}_0^2/\mathbb{E}_n(\mu_i^2)$. This term is typically ignored in the standard definition of the (relative) average pricing error, which is based on the “wrong” pricing errors (i.e., the cross-sectional residuals, $\hat{\omega}_i$). Thus, models with large intercept estimates are more likely to have large (relative) average pricing errors. Trivially, when the intercept is zero it follows that the two metrics coincide, $RRMSE_{\text{pen}} = RRMSE$.

The magnitude of $\hat{\lambda}_0$ matters for pricing-error measures even when it is statistically insignificant. Just as all cross-sectional residuals enter the computation of the average pricing error and the cross-sectional 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.

¹⁵ A closely related version is based on the mean absolute pricing error (e.g., [Maio 2024](#)):

$$RMAE = \frac{\sum_{i=1}^n |\hat{\omega}_i|}{\sum_{i=1}^n |\mu_i|}.$$

2.3 The Cross-Sectional 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 CSR in (1) and its restricted counterpart (without intercept) in (11). Recall that $\mathbf{u} = \mathbf{M}^{-1}\mathbf{m}$. The following proposition relates the corresponding risk price estimates.

Proposition 5. *Given Assumption 2, both the restricted and unrestricted CSRs are identified (so that $\mathbf{M} \succ 0$ and $\mathbf{\Sigma} \succ 0$, equivalently $s < 1$). Then*

$$\hat{\lambda}_0 = \bar{\mu} - \mathbf{m}'\hat{\boldsymbol{\lambda}} = \frac{\bar{\mu} - \mathbf{m}'\tilde{\boldsymbol{\lambda}}}{1 - s}, \quad (14)$$

$$\hat{\boldsymbol{\lambda}} = \mathbf{\Sigma}^{-1}(\mathbf{M}\tilde{\boldsymbol{\lambda}} - \mathbf{m}\bar{\mu}) = \tilde{\boldsymbol{\lambda}} + \frac{\mathbf{m}'\tilde{\boldsymbol{\lambda}} - \bar{\mu}}{1 - s} \mathbf{u}. \quad (15)$$

The slope shift is thus given by

$$\hat{\boldsymbol{\lambda}} - \tilde{\boldsymbol{\lambda}} = -\hat{\lambda}_0 \mathbf{u} = \frac{\mathbf{m}'\tilde{\boldsymbol{\lambda}} - \bar{\mu}}{1 - s} \mathbf{u}. \quad (16)$$

Both the direction and magnitude of the shift are governed by $\mathbf{m}'\tilde{\boldsymbol{\lambda}} - \bar{\mu}$ (the negative of the mean pricing error from the restricted CSR) and by $\mathbf{u}$.

Recall that s is the R^2 from the auxiliary (no-intercept) regression $1 = \hat{\beta}'_i \mathbf{u} + \varepsilon_i$ with $\mathbf{u} = \mathbf{M}^{-1}\mathbf{m}$. 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 $\mathbf{W} = [\mathbf{1} : \hat{\boldsymbol{\beta}}]$ is to being rank deficient; Section 2.4 formalizes this link.

Geometrically, near-collinearity arises when the mean loading vector $\mathbf{m}$ lies predominantly along directions of $\mathbf{\Sigma}$ with small eigenvalues, so that $\|\mathbf{\Sigma}^{-1/2}\mathbf{m}\|^2$ is large and $s \approx 1$. In this regime, Eq. (16) shows that a modest mean pricing error in the restricted CSR can be magnified by the factor $(1 - s)^{-1}$ in both the intercept and the slope coefficients, and in the limit the shift can diverge. We refer to this familiar ill-conditioning effect as the *collinearity amplification channel* for brevity.

Equation (16) implies that

$$\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = -\widehat{\lambda}_0 \mathbf{u}, \quad \mathbf{u} = \mathbf{M}^{-1} \mathbf{m}, \quad \widehat{\lambda}_0 = \frac{\mathbb{E}_n(\widehat{\nu}_i)}{1-s}.$$

Hence, when $\mathbb{E}_n(\widehat{\nu}_i) = \bar{\mu} - \mathbf{m}'\widetilde{\boldsymbol{\lambda}} > 0$, we have $\widehat{\lambda}_0 > 0$, so the sign of each coordinate 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 $u_j < 0$. In the single-factor case with $m > 0$, this reduces to $\widehat{\lambda} - \widetilde{\lambda} < 0$. More generally, the componentwise gap satisfies

$$|\widehat{\lambda}_j - \widetilde{\lambda}_j| = \frac{|\mathbb{E}_n(\widehat{\nu}_i)|}{1-s} |u_j|,$$

so larger mean pricing errors in the restricted CSR, especially when $s \approx 1$, amplify the difference between unrestricted and restricted risk-price estimates.¹⁶

Unspanned omitted loadings therefore contribute to the intercept estimate, which, by (14), is proportional to the mean residual in the restricted CSR, $\mathbb{E}_n(\widehat{\nu}_i) = \bar{\mu} - \mathbf{m}'\widetilde{\boldsymbol{\lambda}}$, scaled by $(1-s)^{-1}$. Equation (16) then implies that a larger intercept mechanically amplifies the slope shift between the unrestricted and restricted CSRs via $-\widehat{\lambda}_0 \mathbf{u}$.

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

Corollary 3. *Under $\mathbf{M} \succ 0$ and $\boldsymbol{\Sigma} \succ 0$,*

$$\widehat{\boldsymbol{\lambda}} - \widetilde{\boldsymbol{\lambda}} = \mathbf{0} \iff \left(\mathbb{E}_n(\widehat{\nu}_i) = 0 \right) \text{ or } \left(\mathbf{m} = \mathbf{0} \right).$$

Thus, if the residual from the restricted CSR has zero mean, it follows that the two sets of risk prices exactly coincide. The same holds when the average loading for all factors are zero. Empirically, models in which either the restricted CSR's residual or the factors loadings have a small mean are more likely to have approximate risk prices from both CSRs.

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

Corollary 4. *For $j = 1, \dots, K$,*

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

¹⁶These slope-shift identities parallel the relation in [Balduzzi and Robotti \(2008\)](#) between traditional beta models and mimicking-portfolio formulations. Our contribution is to derive the explicit shift, show how near-collinearity ($s \approx 1$) amplifies it, and connect it to the penalized R^2 decomposition and the omitted-factor concerns emphasized by [Giglio and Xiu \(2021\)](#).

where u_j is the j th element of $\mathbf{u} = \mathbf{M}^{-1}\mathbf{m}$ and $\hat{\lambda}_0 = (\bar{\mu} - \mathbf{m}'\tilde{\lambda})/(1-s)$ is the intercept estimate in the unrestricted CSR (cf. (14)). It follows:

- (i) Attenuation without sign change. If $\tilde{\lambda}_j > 0$, attenuation ($0 < \hat{\lambda}_j < \tilde{\lambda}_j$) occurs iff $0 < \hat{\lambda}_0 u_j < \tilde{\lambda}_j$; if $\tilde{\lambda}_j < 0$, attenuation ($\tilde{\lambda}_j < \hat{\lambda}_j < 0$) occurs iff $\tilde{\lambda}_j < \hat{\lambda}_0 u_j < 0$.
- (ii) Sign flip. $\text{sign}(\hat{\lambda}_j) \neq \text{sign}(\tilde{\lambda}_j)$ iff

$$(\tilde{\lambda}_j - \hat{\lambda}_0 u_j) \tilde{\lambda}_j < 0, \quad \text{equivalently} \quad \begin{cases} \hat{\lambda}_0 u_j > \tilde{\lambda}_j, & \text{if } \tilde{\lambda}_j > 0, \\ \hat{\lambda}_0 u_j < \tilde{\lambda}_j, & \text{if } \tilde{\lambda}_j < 0. \end{cases}$$

Therefore, for fixed u_j and mean pricing error, a rise in s increases the componentwise gap $|\hat{\lambda}_j - \tilde{\lambda}_j| = |\mathbb{E}_n(\hat{\nu}_i)| |u_j|/(1-s)$. Thus, near-collinearity ($s \approx 1$) amplifies the discrepancy between unrestricted and restricted risk-price estimates. To obtain a better intuition on the decoupling among both sets of risk price estimates, we assume the single-factor model. The following corollary holds as a special case of Corollary 4 under $K = 1$.

Corollary 5. Suppose $K = 1$. Let $m = \mathbb{E}_n(\hat{\beta}_i)$, $M = \mathbb{E}_n(\hat{\beta}_i^2)$, $\sigma^2 = \text{Var}_n(\hat{\beta}_i) = M - m^2 > 0$, and $\bar{\mu} = \mathbb{E}_n(\mu_i)$. Then

$$\hat{\lambda} = \frac{M\tilde{\lambda} - \bar{\mu}m}{\sigma^2}, \quad \hat{\lambda} - \tilde{\lambda} = \frac{m(\tilde{\lambda}m - \bar{\mu})}{\sigma^2}.$$

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

- (i) Sign flip. $\hat{\lambda} < 0$ occurs iff $\tilde{\lambda} < \bar{\mu}m/M$.
- (ii) Attenuation without sign change. $0 < \hat{\lambda} < \tilde{\lambda}$ occurs iff $\bar{\mu}m/M < \tilde{\lambda} < \bar{\mu}/m$. Analogous conditions obtain for $\tilde{\lambda} < 0$ and $m < 0$ by symmetry.

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

$$\hat{\lambda} < 0 \iff \tilde{\lambda} < \frac{\bar{\mu}m}{M} = \frac{\bar{\mu}}{m}s, \quad 0 < \hat{\lambda} < \tilde{\lambda} \iff \frac{\bar{\mu}}{m}s < \tilde{\lambda} < \frac{\bar{\mu}}{m},$$

where $s = m^2/M$ and $\sigma^2 = M - m^2$. A higher s shrinks the attenuation region and, for fixed $(\tilde{\lambda}, \bar{\mu}, m)$, raises the threshold for a sign flip. Separately, a small σ^2 amplifies the magnitude of the

unrestricted estimate through

$$\hat{\lambda} = \frac{M\tilde{\lambda} - \bar{\mu}m}{\sigma^2}.$$

Equivalently, since $v = M\tilde{\lambda}$, a sign flip occurs iff $v < \bar{\mu}m$.

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 $\mathbf{v} = \mathbb{E}_n(\hat{\beta}_i \mu_i)$ and $\mathbf{c} = \text{Cov}_n(\hat{\beta}_i, \mu_i) = \mathbf{v} - \mathbf{m}\bar{\mu}$. Identification of the unrestricted CSR requires

$$\mathbf{\Sigma} = \mathbf{M} - \mathbf{m}\mathbf{m}' \succ \mathbf{0},$$

since $\hat{\lambda} = \mathbf{\Sigma}^{-1}\mathbf{c}$. By contrast, the restricted CSR requires $\mathbf{M} \succ \mathbf{0}$ because $\tilde{\lambda} = \mathbf{M}^{-1}\mathbf{v}$ (Assumption 2). Both conditions are primitive properties of the regressors in the CSRs, $\{\hat{\beta}_i\}_{i=1}^n$.

We note that

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

so that $s = \|\mathbf{\Sigma}^{-1/2}\mathbf{m}\|^2 / (1 + \|\mathbf{\Sigma}^{-1/2}\mathbf{m}\|^2)$ and $(1 - s)^{-1} = 1 + \|\mathbf{\Sigma}^{-1/2}\mathbf{m}\|^2$. 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 6. *Let $\mathbf{W} = [\mathbf{1} : \hat{\beta}] \in \mathbb{R}^{n \times (K+1)}$ and let $\mathbf{M}_1 = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$. Then the unrestricted CSR is identified (i.e., the OLS coefficients on $[\mathbf{1}, \hat{\beta}]$ are unique) if and only if*

$$\text{rank}(\mathbf{M}_1\hat{\beta}) = K \quad (\text{equivalently, } \hat{\beta}'\mathbf{M}_1\hat{\beta} \text{ is nonsingular}).$$

Since $\frac{1}{n}\hat{\beta}'\mathbf{M}_1\hat{\beta} = \mathbb{E}_n[(\hat{\beta}_i - \mathbf{m})(\hat{\beta}_i - \mathbf{m})'] = \mathbf{\Sigma}$, the condition above is equivalent to $\mathbf{\Sigma} \succ \mathbf{0}$. Because identification in the unrestricted CSR requires $\mathbf{\Sigma} \succ \mathbf{0}$ rather than $\mathbf{M} \succ \mathbf{0}$, adding an intercept strictly tightens the rank condition. Empirically, market betas are often close to constant across assets, pushing s toward one (Ahn, Perez, and Gadarowski, 2013; Ahn and Horenstein, 2020), 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; including a free intercept further degrades rank conditions via the determinant identity in [Corollary 6](#). In contrast, the restricted CSR requires only $\mathbf{M} \succ 0$, thus avoiding such amplification of risk price estimates.

A common normalization is to center the regressors across assets by replacing $\hat{\beta}_i$ with $\hat{\beta}_i - \mathbf{m}$. Centering orthogonalizes the intercept and the regressors in sample and can improve numerical stability and exposition. However, centering does not change $\Sigma = \text{Var}_n(\hat{\beta}_i)$ and therefore does not resolve weak identification when Σ is ill-conditioned; it only reparameterizes the regression. In particular, when factor loadings exhibit little cross-sectional dispersion (so that Σ has small eigenvalues), the unrestricted CSR remains ill-conditioned regardless of centering.¹⁷

3 Data and Variables

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

3.1 Models

Our choice of models is guided by two considerations. First, the original factor data are publicly available from the authors' web pages, or alternatively the factors can be constructed from publicly available data. Second, the models span a range of factor structures (single-factor, multi-factor, intermediary-based, and consumption-based) that are representative of the macro-factor approach to cross-sectional asset pricing, and their original implementations typically include an unrestricted intercept in the second-pass CSR, making them natural candidates for examining how the intercept affects inference. Our goal is not a comprehensive horse race or a replication of the original studies. Rather, we use a deliberately small but heterogeneous set of models to illustrate how the cross-

¹⁷[Kleibergen and Zhan \(2020\)](#) emphasize that identification with a zero-beta return requires $[\mathbf{1} : \hat{\beta}]$ to have full rank, whereas without the zero-beta return it suffices that $[\hat{\beta}]$ be full rank. Our contribution is to restate these rank conditions as $\Sigma = \mathbf{M} - \mathbf{m}\mathbf{m}'$ (unrestricted) versus $\mathbf{M}$ (restricted) and to quantify intercept-induced near-collinearity via $s = \mathbf{m}'\mathbf{M}^{-1}\mathbf{m}$, implying the amplification factor $(1 - s)^{-1}$ for shifts in estimated risk prices.

sectional intercept affects pricing performance under a common sample period and a common set of test assets.

Each specification implies the standard no-intercept beta-pricing relation for excess returns and can be written as

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

Table 1 summarizes the baseline CAPM and 12 macro-factor specifications used in the empirical analysis. The models are ordered by family, and nested variants are indicated explicitly.

Table 1: Factor models

Label	Reference	K	Factors
CAPM	Sharpe (1964), Lintner (1965), Mossin (1966)	1	RM
JW3	Jagannathan and Wang (1996)	3	$RM, \Delta y, def$
JW2	JW3 without def	2	$RM, \Delta y$
SV4	Santos and Veronesi (2006)	4	$RM, \Delta y, RM_{sw}, Y_{sw}$
SV3	SV4 without RM_{sw}	3	$RM, \Delta y, Y_{sw}$
SV2	SV4 without $\Delta y, Y_{sw}$	2	RM, RM_{sw}
LL2	Lettau and Ludvigson (2001)	2	RM, RM_{cay}
AEM1	Adrian, Etula, and Muir (2014)	1	LEV
AEM2	AEM1 + RM	2	RM, LEV
HKM1	He, Kelly, and Manela (2017)	1	IC
HKM2	HKM1 + RM	2	RM, IC
CCAPM	Lucas (1978), Breeden (1979)	1	Δc
LVN3	Lustig and Van Nieuwerburgh (2005)	3	$\Delta c, my, Cmy$

The factors are defined as follows. RM is the excess market return. Δy denotes per capita real labor income growth, used as a proxy for human capital returns (Jagannathan and Wang, 1996); def is the lagged default spread (Baa minus Aaa yield). The Santos–Veronesi models scale RM and Δy by the lagged labor income-to-consumption ratio (sw_{t-1}), producing the scaled factors $RM_{sw} = RM_t sw_{t-1}$ and $Y_{sw} = \Delta y_t sw_{t-1}$; Lettau and Ludvigson (2001) scale RM by the log consumption-to-wealth ratio (cay), producing $RM_{cay} = RM_t cay_{t-1}$. LEV and IC denote the financial intermediary leverage factor of Adrian, Etula, and Muir (2014) and the intermediary capital factor of He, Kelly, and Manela (2017), respectively. Δc is log real per capita consumption growth, and LVN3 conditions it on the lagged housing collateral ratio (my_{t-1}), yielding the scaled factor $Cmy = \Delta c_t my_{t-1}$.

3.2 Data

In this subsection, we describe the construction of the factors and test assets. Consumption and wage data are obtained from the National Income and Product Account (NIPA) database. In line with the macro-finance literature, consumption is the sum of personal expenditure on nondurable goods and services. Following [Jagannathan and Wang \(1996\)](#), aggregate labor income represents total personal income minus dividend income. We deflate consumption and labour 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.¹⁸ 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. 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. Our sample period is 1970:I to 2017:III, where the starting (ending) quarter is constrained by the availability of data on *IC* (*LEV*).

The testing assets employed in the asset pricing tests are associated with prominent CAPM anomalies, that is, patterns in cross-sectional risk premia left unexplained by the baseline CAPM. Concretely, we consider decile portfolios sorted on four stock characteristics—book-to-market ratio (BM10, [Rosenberg, Reid, and Lanstein 1985](#) and [Fama and French 1992](#)); investment (INV10, [Titman, Wei, and Xie 2004](#), [Fama and French 2008](#), and [Cooper, Gulen, and Schill 2008](#)); operating profitability (OP10, [Novy-Marx 2013](#) and [Fama and French 2015](#)); and price momentum (M10, [Jegadeesh and Titman 1993](#) and [Fama and French 1996](#)). We do not consider portfolios sorted on market capitalization ([Banz 1981](#) and [Fama and French 1992](#)), 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

¹⁸For convenience, we use the same price deflator for both macro variables.

“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 cross-sectional intercept on pricing errors and model fit. The data on all monthly portfolio returns are retrieved from Kenneth French’s data 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, TB_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 benchmark equal to $TB_t/4$. Thus, $TB_t/4$ 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.

Table 2 contains (cross-sectional) summary statistics for portfolio risk premia (i.e., average excess portfolio returns). We can see that the BM deciles have the highest average risk premium (2.08% per quarter), while both the profitability and momentum deciles exhibit the lowest average risk premium (1.69% per quarter). The average portfolio in the cross-section has a risk premium of 1.85%. Regarding the cross-sectional volatility in risk premia, it follows that the momentum portfolios show the largest dispersion in risk premia by a good margin in comparison with the other three groups (0.89%). At the other extreme, the profitability portfolios have the lowest cross-sectional volatility in average excess returns (0.33%). The average portfolio has a cross-sectional volatility of 0.57%. 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%).

4 Pricing Errors and Performance Metrics

In this section, we evaluate the impact of the cross-sectional intercept in the pricing performance of the various factor models described in the last section.

Before turning to the results, we note that the empirical performance of any factor model depends on the choice of test assets, sample period, and data frequency. Accordingly, the specifications below are not intended to provide a comprehensive assessment of each model’s strengths and weaknesses. Rather, they form a representative set of prominent macro-factor models that allows us to illustrate the systematic effect of the CSR intercept on pricing diagnostics. When particular models perform notably well or poorly in our common 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.

To address the statistical uncertainty associated with the different sample cross-sectional R^2 — R^2_{pen} , R^2_{ols} , and $R^2_{\text{ols*}}$ —we rely on asymptotic distributions. Namely, to test the null of a zero fit (against the alternative of a positive fit) when the performance metric is either R^2_{pen} or $R^2_{\text{ols*}}$, we compute one-sided p -values for statistics that are asymptotically normally distributed. Section 9 provides the details on the derivation of these two new statistics. In the case of R^2_{ols} , we rely on the test-statistic derived by [Kan, Robotti, and Shanken \(2013\)](#) (which follows a weighted chi-squared distribution) when testing the null of zero pricing performance.¹⁹

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—by assessing the capacity of a given model in explaining all portfolio groups or anomalies on a joint basis—it can provide a largely biased picture. The reason is that a given model might produce a large fit simply 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. Hence, it is also important to provide the asset pricing results for the individual portfolio groups or anomalies.

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 factor (for a total of 17 portfolios). This choice of removing the intermediate deciles from the

¹⁹In alternative, [Guo and Maio \(2020\)](#) and [Maio and Min \(2025\)](#) employ bootstrap simulations to assess the statistical significance of the cross-sectional OLS R^2 .

estimation stems from the abundant evidence showing that most of the cross-sectional dispersion in risk premia is associated with the “extreme portfolios” within a given portfolio sort, thus enabling a more powerful asset pricing test.

The results for the performance evaluation metrics are shown in Tables 3 and 4. We have a total of 78 estimations, which stems from 13 models and six different cross-sections. When the intercept is restricted to zero, most models fail to achieve statistically significant dispersion fit, as the null of a zero R^2_{ols*} is not rejected. The sole exception holds for LVN3 in the estimation with the 17 portfolios (which means 1.3%=1/78 of all the estimations), in which case the null of zero fit is rejected (at the 5% level). Even if we relax the significance level toward 10%, it follows that the null is reject in only four additional cases, all in the estimation with the momentum deciles (JW3, SV3, SV4, and again LVN3), which still amounts to only 6.4%=5/78 of the 78 estimations. In fact, an insignificant explanatory ratio (even at the 10% level) actually holds for models with a seemingly good sample fit (explanatory ratios around or above 50%), such as AEM2 (in the estimation with INV10, M10, BM10, or All*), JW2 (M10), AEM1 (M10), CCAPM (M10 and All*), LL2 (BM10), HKM1 (BM10), or HKM2 (BM10). These results reveal the importance of accounting for the sampling error associated with the cross-sectional R^2 when evaluating linear factor models.

Critically, there are several models that mimic the baseline CAPM in terms of producing a negative explanatory ratio, which implies that these models do worse than a constant when it comes to explaining cross-sectional dispersion in risk premia. Concretely, there are 39 of such cases, which represents exactly half of the 78 estimations being conducted, indicating limited dispersion fit across these specifications. This pattern is especially evident in the estimations with INV10 or OP10, as 10 models (out of 13) produce a negative explanatory ratio (within each of these two portfolio groups). In the case of the two more challenging cross-sections (All and All*), there are five models producing a negative R^2_{ols*} in both settings (SV2, LL2, HKM1, HKM2, in addition to the baseline CAPM). Overall, the main exception is LVN3 in the estimation with the 17 portfolios, as referred above, but this stems mainly from the capacity of such a model in explaining momentum portfolios.

When the intercept is added to the second-pass CSR, at a first sight one would be tempted to conclude for a 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 rejected (at the 5%

level) only in 15 of 78 the estimations being conducted (which means about 19% of the cases). In the remaining 63 estimations (or 81% of all the estimations)—including many cases with a sample explanatory ratio around or above 50%—the null of no pricing power is not rejected at the 5% level. This pattern is especially notable in the estimations with OP10—where there is not a single case of rejection of the null—and with INV10 or BM10—where there is only one model for which the null is rejected. All* represents by far the more “friendly” cross-section, as the null $R_{ols}^2 = 0$ is rejected for eight of the 13 models. In contrast, when all 40 portfolios are included in the estimation, it follows that the null is rejected (at the 5% level) only in three cases: CAPM, AEM1, and AEM2. Again, these results illustrate the importance of accounting for the estimating error in the sample explanatory ratios.

Abstracting from statistical uncertainty, the cases of CAPM, SV2, LL2, and HKM2 raise particular attention, as the fit moves from being negative to values around 30% (with significance in the case of the CAPM) 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 explanatory ratios associated with these models register a large rise in magnitude (relative to the estimation with all 40 portfolios) and become statistically significant. These results illustrate the quantitative effect of relaxing the zero-intercept restriction and using R_{ols}^2 as a performance metric. In particular, the baseline CAPM produces a large and statistically significant R_{ols}^2 for three of the six cross-sections (M10, All, and All*), which reflects the mechanical inflation documented in Proposition 3: the intercept absorbs the common pricing offset and credits it as explained variation.

When one uses the more appropriate R_{pen}^2 metric, the overall picture becomes much clearer, even without looking at the corresponding statistical inference. Concretely, 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 being conducted) refer to AEM1 (in the estimation with the momentum deciles) and AEM2 (BM10), where the sample R_{pen}^2 is positive. Nonetheless, there is no statistical significance in these two cases, as indicated by p -values way above 10%. Hence, in the two more challenging estimations associated with all portfolio groups, there is not a single model that delivers a positive R_{pen}^2 , and the same pattern holds in the estimations with the investment or profitability deciles.

As noted above, such a pattern compares with a scenario in which the R_{ols}^2 is significantly

positive for eight of the 13 models over the cross-section of 17 portfolios. In other words, the penalized metric yields a qualitatively different assessment of pricing performance across all models under consideration. Particularly striking are the cases of the CAPM, JW2, SV2, LL2, AEM1, AEM2, HKM2, and CCAPM, in which cases the model produces simultaneously a significantly large R_{ols}^2 (i.e., around or higher than 50%) jointly with a negative R_{pen}^2 in the second augmented estimation (All*). A similar pattern applies to both AEM1 and AEM2 in the estimation with all 40 portfolios. This further illustrates the extent to which R_{ols}^2 overstates pricing performance when the cross-sectional intercept is not restricted to zero.

The magnitude of the intercept's contribution to pricing errors, when λ_0 is not restricted to zero, is also apparent from the sample relative average pricing errors. By using the wrong definition of pricing errors (that is, by ignoring the intercept and considering only the cross-sectional residuals), one obtains average pricing errors that represent at most 36% of the raw average 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 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 regression without an intercept ($RRMSE^*$), thus suggesting a reasonable pricing performance.

However, by relying on the more appropriate $RRMSE_{pen}$ metric, which imposes the restriction that the intercept represents the average pricing error, the assessment changes materially: the average pricing errors exceed the average risk premia that one seeks to explain (i.e., $RRMSE_{pen}$ above one) in the majority of the cases (i.e., 54 out of 78 or nearly 70% of the estimations being conducted). 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 $RRMSE_{pen}$ 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 $RRMSE_{pen}$ is clearly above the corresponding estimate of $RRMSE$, something that stems from large intercept estimates. In sum, the picture provided by $RRMSE_{pen}$ is qualitatively similar to that provided by R_{pen}^2 , as discussed above: The intercept estimates have a very large contribution to the pricing errors, confirming that standard residual-based metrics substantially understate total pricing error when the intercept is economically large.

Consistent with this argument, and as shown in Table 5 below, the estimates of λ_0 for most

models are above 2% per quarter (or 8% per year) in the augmented estimation with all 40 portfolios. These estimates are economically large and point to substantial misspecification (i.e., missing factors) in most specifications, consistent with Propositions 1 and 2. Indeed, the three models producing intercept estimates below 2% (AEM1, CCAPM, and LVN3) are still associated with a high average pricing error (i.e., λ_0 varying between 0.46% and 1.20% per quarter).

Therefore, we see that *Wedge* (which equals minus the lower bound for R_{pen}^2 based on Corollary 1) is above one in nearly all cases, often with very large magnitudes, as a consequence of the large estimates for λ_0 . The very few exceptions hold for AEM1 (estimation with M10 and All*), AEM2 (BM10), and CCAPM (All and All* portfolios). Yet, three of these five cases point to a large magnitude of *Wedge* (between 0.67 and 0.99). These estimates drive a large wedge between the penalized and conventional fit measures (i.e., R_{pen}^2 falls well below the corresponding $R_{\text{ols}^*}^2$), consistent with Proposition 3 and Corollary 1. The very large estimates of *Wedge* also explain why $RRMSE_{\text{pen}}$ is substantially above $RRMSE$ in nearly all 78 estimations (see Proposition 4). The above mentioned case of AEM2 represents indeed the main outlier to such a pattern, as *Wedge* assumes a value around zero (0.02). This explains the corresponding similar estimates of $R_{\text{ols}^*}^2$ and R_{pen}^2 (83% versus 81%), which is consistent with Corollary 2.

Moreover, the scalar s also plays an important role in explaining the discrepancy between the (appropriate) fit among the two types of CSR, $R_{\text{ols}^*}^2 - R_{\text{pen}}^2$, as described in Corollary 2. This pattern is the fit-metric manifestation of the collinearity amplification channel. The estimates are very close to one in nearly all cases, thus justifying why we obtain R_{pen}^2 that are substantially below the $R_{\text{ols}^*}^2$ counterparts (i.e., more negative or negative versus positive). In other words, the large values of s reinforce the large values of *Wedge* in terms of producing a substantial discrepancy in fit between the two CSRs. 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_{pen}^2 for AEM1 in the first of these estimations is positive and not extremely far apart the corresponding $R_{\text{ols}^*}^2$ (16% versus 46%). The large values of *Wedge* in the remaining three cases, especially in the estimation with BM10 or All, more than offset the contribution from s , thus producing a large gap between $R_{\text{ols}^*}^2$ and R_{pen}^2 .

One important finding from the results in Tables 3 and 4 is that using “correct performance metrics” associated with the unrestricted CSR, which impose the restriction that the intercept

represents a pricing error, tends to produce qualitative evidence that is on par with that produced by the corresponding restricted CSR. That is, both approaches (e.g., $R_{\text{ols}*}^2$ and R_{pen}^2) tend to be equally effective in discarding “bad models”. However, the penalized metrics can occasionally *overshoot*: models that fit well under the restricted CSR may appear to perform poorly under R_{pen}^2 when the unrestricted intercept is large. The clearest example is LVN3 estimated on the 17-portfolio cross-section. The restricted CSR yields an $R_{\text{ols}*}^2$ above 70%, statistically significant at the 5% level, indicating that the model explains a substantial share of the cross-sectional dispersion in risk premia. Yet the unrestricted CSR delivers $R_{\text{pen}}^2 = -2.10$ and $\text{RRMSE}_{\text{pen}} = 73\%$, implying a sharp deterioration in fit. This divergence is informative rather than paradoxical. It arises because R_{pen}^2 penalizes the level of the common pricing offset, in addition to the model’s ability to match dispersion across assets. When s is close to one, as in LVN3 for this cross-section where $s = 0.98$, the collinearity-amplification channel inflates the intercept and the associated wedge, generating a large penalty even when the model fits dispersion well. Put differently, R_{pen}^2 and $R_{\text{ols}*}^2$ answer different questions: the former evaluates total pricing error, including the common offset, whereas the latter evaluates dispersion fit under the maintained normalization. This distinction reinforces our main recommendation: The restricted CSR and $R_{\text{ols}*}^2$ should be the primary specification for evaluating macro models. When unrestricted results are also reported, R_{pen}^2 remains a useful diagnostic of the model’s total pricing-error burden, but its interpretation requires care, especially when s is high and amplification is likely to be important.

5 Risk Price Estimates and Identification

Beyond its impact on pricing errors and model fit, the cross-sectional intercept also significantly influences risk price estimates and their identification. This section delves into these effects, evaluating how the intercept alters the magnitude, sign, and statistical significance of estimated factor risk premia.

5.1 Risk Price Estimates

The risk price estimates for both types of CSR are reported in Table 5. 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 both the popular Shanken’s standard errors (which correct for the error-in-variables in the first-pass betas, as shown in [Shanken \(1992\)](#)), as well as more robust standard errors that account for model misspecification (see [Kan, Robotti, and Shanken \(2013\)](#)).²⁰

Several patterns emerge from Table 5. First, we observe that most of the intercept estimates are statistically significant (at the 5% level) by relying on both t -ratios. The main exceptions hold for AEM1, CCAPM, and LVN3, in which cases the estimates of λ_0 are insignificant by using any of the two types of standard error. On the other hand, the corresponding estimate associated with either JW3 or AEM2 achieves significance by relying exclusively on Shanken’s standard errors.

Second, and regarding the risk price estimates, one clear pattern stands out: All risk price estimates that are significant under the restricted CSR (based on both types of standard error) and that preserve the sign, cease to be significant once we include the intercept in the CSR. This includes the cases of λ_Y (within JW2, SV3, and SV4), λ_{LEV} within AEM2, and λ_C within LVN3. We only observe a single case in the opposite direction: The estimate of λ_{LEV} within AEM1 becomes significant (based on both types of standard error) in the unrestricted CSR, at odds with the corresponding estimate in the restricted CSR. Therefore, including the intercept has generally a relevant downward effect in the statistical significance of the risk price estimates, that is, the statistical uncertainty increases.

Third, a notable finding is that none of the nine market risk price estimates remains positive in the unrestricted CSRs, something that compares with significantly (using both types of inference) positive estimates in all the corresponding restricted CSRs. Indeed, some of these negative estimates of λ_M (CAPM and HKM2) are significant based on both types of standard errors. If we relax the criteria and focus only on the Shanken’s standard errors, it follows that the negative market prices of risk associated with both SV2 and LL2 are also significant.

Fourth, 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 (based on Shanken’s standard errors) in the restricted CSR to negative (albeit insignificant) estimates in the corresponding unrestricted CSR.

²⁰For comparability with the Shanken standard errors, the misspecification-robust standard errors are computed using a spectral density estimate with zero lags.

Fifth, another robust pattern observed in Table 5 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 $\lambda_{Y_{sw}}$ (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 discussed in Section 2, confirming the quantitative importance of the intercept for risk price estimates. As shown in Proposition 5, the slope shift satisfies

$$\hat{\lambda} - \tilde{\lambda} = -\hat{\lambda}_0 \mathbf{u}, \quad \hat{\lambda}_0 = \frac{\mathbb{E}_n(\hat{\nu}_i)}{1 - s}.$$

Hence, when the average pricing error of the restricted CSR is positive, the unrestricted risk prices move in the direction $-\mathbf{u}$: Coordinates with $u_j > 0$ fall, whereas coordinates with $u_j < 0$ rise (for example, becoming less negative). Table 5 shows that $\mathbb{E}_n(\hat{\nu}_i) > 0$ for all 13 models, so the relevant empirical implication is not a uniformly negative componentwise gap, but rather that the componentwise shift magnitude scales with $|\mathbb{E}_n(\hat{\nu}_i)| |u_j| / (1 - s)$, often producing attenuation and, in extreme cases, sign reversals.²¹

AEM1 represents an interesting case. Its restricted-CSR mean pricing error is the largest in Table 5 ($\mathbb{E}_n(\hat{\nu}_i) = 0.15\%$), yet the corresponding intercept estimate is smaller than in most other models because the amplification term $(1 - s)^{-1}$ is less extreme here: with $s = 0.84$, Eq. (14) implies $\hat{\lambda}_0 \approx 0.15 / (1 - 0.84) = 0.94\%$, which matches closely the reported estimate of 0.95% (up to rounding). Thus, s is not offsetting the mean pricing error; rather, relative to the models with s much closer to one, AEM1 is subject to less collinearity amplification. Even so, the positive mean pricing error together with a large value of $u = m/M$ in the single-factor case still generates a substantial decline in the LEV risk price, from 20.87% to 11.07%.

To illustrate the amplification mechanism in a familiar benchmark, consider the CAPM estimated on the 40-portfolio cross-section. The restricted CSR yields $\tilde{\lambda} = 1.76\%$ per quarter, while

²¹By Corollary 5, in the single-factor case ($K = 1$) we have $u = m/M$. Hence, if $m > 0$ and $\mathbb{E}_n(\hat{\nu}_i) > 0$ (equivalently, $\hat{\lambda}_0 > 0$), then $\hat{\lambda} - \tilde{\lambda} < 0$, implying attenuation and, if the shift is large enough, sign reversal.

the average portfolio mean return is $\bar{\mu} = 1.85\%$. Because $m = \mathbb{E}_n(\hat{\beta}_{i, RM})$ is close to one, the restricted-CSR mean pricing error, $\mathbb{E}_n(\hat{\nu}_i) = \bar{\mu} - m\tilde{\lambda}$, is small but positive. With $s \approx 0.98$, Eq. (14) implies

$$\hat{\lambda}_0 \approx \frac{\bar{\mu} - m\tilde{\lambda}}{1 - s} \approx 3.99\%.$$

Since $u = m/M$ is also close to one in the single-factor case, the induced slope shift is $\hat{\lambda} - \tilde{\lambda} = -u\hat{\lambda}_0 \approx -3.89\%$, so the unrestricted estimate falls to approximately -2.13% . The sign reversal therefore reflects the amplification channel: a modest 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 slope.

The gap in risk prices (from the two CSRs) is also magnified when s is close to one, something that applies to most models under consideration, as discussed in the previous section. In the notation of Section 2, this is the collinearity amplification channel implied by Eq. (16). This is especially true for the models including the market factor, as the corresponding s estimates are always very close to one, which implies substantial changes among the two sets of risk prices (often with sign flip). Recall that a high s stems from a high (cross-sectionally) average beta and/or a low cross-sectional 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 cross-sectional 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.

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 documented in Section 2. While the pricing-error distortion can be corrected by including the intercept in the definition of total pricing error (Section 4), the risk-price distortion has no analogous remedy within the unrestricted specification. The practical implication is that the unrestricted CSR is best viewed as a supplementary diagnostic, with the restricted specification serving as the primary basis for risk-price inference.

Table 6 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 5. Concretely, we find that (i) apart from the cases of AEM1 and CCAPM, the intercept estimates assume very large magnitudes, varying between 1.41% (LVN3) and 6.16% per quarter (SV2); (ii) all the market risk prices become negative (with statistical significance in a few cases) in the unrestricted CSRs; (iii) the magnitudes of the risk prices (that preserve sign) tend to be lower under the unrestricted CSRs; (iv) there is a generalized decline in the statistical significance of risk prices under the unrestricted CSRs; and (v) 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. 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 average pricing error produced by the restricted CSR (-0.04%). AEM1 again stands out: despite the large average pricing error from the restricted CSR (0.24%), the intercept is only 0.83% because s is relatively low (0.71). Even so, the large average beta (large m) still generates substantial attenuation in the corresponding risk price estimate.

We also present the risk price estimates when BM10 are the test assets, whose results appear in Table 7. These portfolios (or the closely related 25 portfolios sorted on both size and BM) represented the standard test assets in the related literature for a long time. One key difference relative to the results in Tables 5 and 6 is that none of the intercept estimates is statistically significant despite the large magnitudes in the majority of the cases (i.e., estimates of λ_0 around or above 1% per quarter for 11 of the 13 models being estimated). One common misconception in the literature is that the intercept estimates “can be ignored if statistically insignificant”. This argument is fallacious, as it confuses statistical significance with economic significance. 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. 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 from Table 7 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). Yet, the corresponding positive estimates of λ_{RM} are always insignificant (even based on the Shanken’s standard errors), which compares with strongly positive estimates (based on both standard errors) 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.

In fact, there is one pattern 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 nearly all cases, with the sole exception holding for λ_{IC} within HKM1. In comparison, all the risk price estimates associated with RM (as mentioned above), LEV , and IC are always significant under the restricted CSR (based at least on Shanken’s standard error). This illustrates the substantial effect of the intercept on the statistical significance of the risk price estimates.

Another relevant finding is that AEM2 produces the smallest intercept (only 0.06%) and the smallest $\mathbb{E}_n(\hat{\nu}_i)$ (around 0%). Therefore, and in line with both Proposition 5 and Corollary 3, this is among the models with the lowest gap in risk prices from the two CSRs (these almost coincide). In comparison, the corresponding single-factor model (AEM1) produces the largest positive mean pricing error in Table 7, $\mathbb{E}_n(\hat{\nu}_i) = 0.22\%$ per quarter. The key point is not that this number is large in isolation, but that Eq. (16) magnifies even moderate mean pricing errors when $u = m/M$ is large and s is still close to one. Indeed, with $s = 0.83$, Eq. (14) implies an intercept of about $0.22/(1 - 0.83) \approx 1.29\%$, close to the reported 1.26% up to rounding, and the resulting slope shift is economically large: the unrestricted estimate of λ_{LEV} falls to 7.97%, versus 18.06% in the restricted CSR. This is the same collinearity-amplification mechanism observed in Table 5.

We can also see that the intercept estimate is negative (-0.48%) in the case of HKM1, in line with the estimate for the CCAPM over the cross-section of 17 portfolios discussed above. Again, this stems from the negative average pricing error produced by the restricted CSR (-0.01%). Because that average pricing error is tiny, the resulting intercept is also small, so the risk-price estimates from the two CSRs are nearly identical.

5.2 Risk Price Identification

In this subsection, we assess whether the risk prices in the different models are identified. Following Burnside (2011) and Kleibergen and Zhan (2015, 2020), the risk price estimates, and

associated pricing errors, are unreliable when the risk prices are not identified, that is, the $n \times K$ factor loading matrix (β) does not have full column rank (considering the case with $\lambda_0 = 0$), where K denotes the number of factors. When the cross-sectional regression includes an intercept, the rank condition is that the matrix containing the factor loadings and a vector of ones ($\mathbf{W} = [\mathbf{1}; \beta]$) has full column rank.

We employ the rank tests on the factor loadings proposed by Kleibergen and Zhan (2020). The null hypothesis is that the risk prices in a given model are not identified, that is, either β or $\mathbf{W}$ has a reduced rank. When $\lambda_0 = 0$, the statistic admits an $F(n + 1 - K, T - n - 1)$ bounding distribution (in the sense of Kleibergen and Zhan (2020)). When there is an intercept, it admits an $F(n - K, T - n)$ bounding distribution.

The p -values for the Kleibergen–Zhan rank tests are reported in Table 8. Focusing on the unrestricted CSR (the specification with an intercept; the p -value reported before the comma in the table), we fail to reject at the 5% level in most of our 78 estimations.²² We do not interpret non-rejection as evidence of exact rank deficiency, which is a knife-edge event. Rather, since rank tests have limited power against severe near-collinearity in the second-pass matrix $\mathbf{W} = [\mathbf{1} : \beta]$, we treat non-rejection as a diagnostic of weak identification: small perturbations in (μ, β) can translate into large changes in $(\hat{\lambda}_0, \hat{\lambda})$.

This interpretation is supported by two additional diagnostics. First, the collinearity index s in Tables 3–4 is close to one in those same cases, directly indicating severe near-collinearity. Second, Eq. (16) shows that $s \approx 1$ generates an amplification channel that destabilizes risk-price estimates. Accordingly, when the rank test fails to reject and s approaches one, we regard the unrestricted CSR estimates $(\hat{\lambda}_0, \hat{\lambda})$ as unreliable.

Importantly, when the intercept is restricted to zero, we obtain a clearly more favorable picture. Concretely, the null is now rejected (at the 5% level) in the cases of SV2 (estimation with M10), AEM2 (BM10), and CCAPM (INV10 and OP10), which contrasts with the respective tests for the unrestricted CSR. The same pattern holds for AEM1, AEM2, and CCAPM in the estimation with all 40 portfolios and for the latter model in the estimation with the 17 portfolios. In the opposite direction, we detect only one case (estimation of HKM2 with OP10) in which the null is

²²Rejections are concentrated in CAPM and HKM1 (all six cross-sections), HKM2 (all but M10), SV2 (OP10, All, and All*), LL2 (OP10, M10, All, and All*), AEM1 (BM10 and All*), and AEM2 (All*).

not rejected in the restricted CSR, but becomes rejected in the unrestricted counterpart. Still, the non-rejection of the null is marginal, as indicated by the p -value of 5.7%.

Therefore, these results indicate that a situation of weakly identified factors is more likely to occur, thus making the respective risk price estimates less credible, when the intercept is included in the CSR. The reason is that it is not enough that the factor loadings are jointly different from zero; instead, enough dispersion in betas (among the test assets) is necessary to reject the null of a reduced rank, in line with Corollary 6.²³

Our analysis focuses on classical two-pass regressions applied to portfolio-level cross-sections with fixed risk premia. Complementary econometric frameworks address identification and inference in much larger asset panels and allow for time-varying risk premia. In particular, [Gagliardini, Ossola, and Scaillet \(2016\)](#) develop weighted two-pass procedures with formal large-sample theory in settings with many assets and unbalanced panels. While their focus is distinct, the present results highlight that even with a relatively small cross-section, allowing a free intercept can substantially distort inference and weaken identification.

6 Intercept-Implied Zero-Beta Rates and Implied Equity Premia

In this section, we provide a simple economic plausibility diagnostic for the magnitude of the cross-sectional intercept. A common motivation for estimating an intercept in the CSR is that the observed risk-free rate used to form excess returns may differ from the relevant pricing benchmark (e.g., a model-implied zero-beta return) due to money-market frictions such as liquidity premia or lending–borrowing spreads. Within a linear pricing framework, this interpretation amounts to viewing λ_0 as the difference between the model-implied benchmark return and the reference rate used in the data.

Concretely, consider the average risk-free return of 1.2% per quarter in our sample period. By using the lowest intercept estimate (which represents the excess benchmark return relative to the reference rate) from Table 5 (around 0.5% per quarter in the estimation of CCAPM), this would imply a benchmark return around 1.7% ($= 1.2\% + 0.5\%$) per quarter or about 6.8% per

²³Often, the non-rejection of the null under $\lambda_0 \neq 0$ arises from the existence of (near) multicollinearity in the cross-sectional regression, that is, the betas are approximately constant, thus being proportional to a vector of ones. Notably, the same near-collinearity that drives weak identification here is also what drives the collinearity amplification channel in risk-price estimates.

year. Such a positive difference of about two percentage points per year is already sizable relative to common discussions of U.S. money-market imperfections. However, if we pick the other (and larger) intercept estimates from Table 5, then such type of justification becomes much more difficult. This simple example illustrates that modest money-market frictions are unlikely to account for the magnitudes of λ_0 typically obtained in our unrestricted CSRs. This type of interpretation also does not naturally explain the existence of negative intercept estimates that occur occasionally (see the example of HKM1 in the estimation with BM10 in Table 7), as pointed out by Burnside (2011).

The same logic applies to implications for the average equity premium. With an average market return of 2.9% per quarter or 11.6% per year in our sample period, the benchmark return of 6.8% per year computed above would imply an annual equity premium of 4.8%. This compares with the sample mean of the market factor of 1.7% per quarter or 6.8% per year. In other words, the implied equity premium is too small. In fact, it follows that the quarterly equity premium is negative,

$$2.9\% - (\lambda_0 + 1.2\%) < 0 \Leftrightarrow \lambda_0 > 1.7\%,$$

whenever the intercept is above 1.7% per quarter. As shown in Table 5, this holds for 10 of the 13 models estimated (or 77% of the cases). The same conclusion is achieved if we rely instead on the intercept estimates from Table 6. We emphasize that these computations are not structural estimates of equilibrium objects. Rather, they serve as absurdity-reduction arguments: if one were to interpret the intercept literally as a zero-beta rate premium, the implied benchmark returns and equity premia would be difficult to reconcile with standard economic priors. This highlights that large intercept estimates are more naturally interpreted as evidence of model misspecification (via omitted priced factors) than as reflections of money-market frictions.

Many empirical applications that include λ_0 in the CSR emphasize the resulting increase in cross-sectional “fit” without reporting the implied benchmark return or implied equity premium. Reporting these implications is informative because it makes transparent that a large intercept reflects a failure of the model to match the level of the cross section (and, for equity test assets, the aggregate equity premium) under the maintained normalization used to form excess returns.

That a large intercept represents a failure of a given model in matching the aggregate equity premium can actually be seen by the own definition of the intercept, thus avoiding the need of

pursuing the “back-of-the-envelope” calculation above. Specifically, we have

$$\hat{\lambda}_0 = \bar{\mu} - \sum_{j=1}^K \hat{\lambda}_j m_j, \quad (17)$$

where $m_j = \mathbb{E}_n(\hat{\beta}_{i,j})$.

Hence, a large intercept means that the model cannot match the cross-sectional average risk premium, $\bar{\mu}$, that is, there is a large gap between the two terms on the right hand side of (17). When the test assets represent equity portfolios, as in our empirical setting and in most of the empirical asset pricing literature, it follows that the cross-sectional average risk premium approximately equals (by construction) the sample equity premium, $\bar{\mu} \approx \overline{RM}$, where $\overline{RM}$ denotes the sample mean of the market factor (1.7% in our case).²⁴ This can be confirmed from the summary statistics in Table 2, as the average equity premium (among the 40 portfolios) is about 1.85% per quarter. Therefore, a large $\hat{\lambda}_0$ automatically means that the model cannot price the market factor. This implication is often overlooked when evaluating the pricing performance for the cross-section of stock returns.

Equation (17) also helps describing why large intercept estimates can be associated with (economically difficult to interpret) negative risk-price estimates. In particular, if the dominant factors have positive average betas and the weighted sum $\sum_{j=1}^K \hat{\lambda}_j m_j$ becomes negative (due to negative risk prices), then (17) implies $\hat{\lambda}_0 > \bar{\mu}$. The most frequent example is the case of the market risk price, which is consistently negatively estimated, as discussed in Section 5.

7 Monte-Carlo Simulations

We conduct two experiments to illustrate the relationship between model misspecification and the cross-sectional intercept, as predicted by Propositions 1 and 2.

We employ the 40 portfolios as the test assets and either the 6-factor model of Fama and French (2018) (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 draws)

²⁴This stems from the fact that a given portfolio group (e.g., the BM deciles) includes all stocks in the market in a given period. However, it holds as an approximation (rather than an equality), since an equal-weighted average of a group of value-weighted portfolio returns (e.g., BM10) does not exactly coincide with the value-weighted market return (the market factor).

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, whose data are retrieved from Kenneth French's data library.

We compute the residual returns from the auxiliary regression,

$$R_{i,t} = \mu_i + \varsigma_{i,t}, \quad (18)$$

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

$$\begin{aligned} & \{RM_t^m, SMB_t^m, HML_t^m, RMW_t^m, CMA_t^m, UMD_t^m, \varsigma_{i,t}^m\}, \\ & \{RM_t^m, \varsigma_{i,t}^{*m}\}, m = 1, \dots, 10000. \end{aligned}$$

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, $\{\hat{\lambda}_0^m\}, m = 1, \dots, 10000$. We then compute the average pseudo intercept estimate across the 10000 replications:

$$E^*(\hat{\lambda}_0) = \frac{1}{10000} \sum_{m=1}^{10000} \hat{\lambda}_0^m. \quad (19)$$

The first experiment is based on the CAPM. The corresponding cross-sectional regression is given by

$$\mu_i = \lambda_0 + \lambda_{RM} \hat{\beta}_{i, RM} + \omega_i, \quad (20)$$

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

$$R_{i,t}^m = \hat{\lambda}_{RM} \hat{\beta}_{i, RM} + \varsigma_{i,t}^{*m}, \quad (21)$$

where the beta and risk price estimates are obtained from the estimation of the single-factor over the original sample. It follows that the average CAPM intercept estimate (across the 10000 replications)

is very small (-0.05%).

The second experiment is based on FF6, with cross-sectional regression given by

$$\mu_i = \lambda_0 + \lambda_{RM}\hat{\beta}_{i,RM} + \lambda_{SMB}\hat{\beta}_{i,SMB} + \lambda_{HML}\hat{\beta}_{i,HML} + \lambda_{RMW}\hat{\beta}_{i,RMW} + \lambda_{CMA}\hat{\beta}_{i,CMA} + \lambda_{UMD}\hat{\beta}_{i,UMD} + \omega_i. \quad (22)$$

We now assume that FF6 holds in the population, with the return generating process given by

$$R_{i,t}^m = \hat{\lambda}_{RM}\hat{\beta}_{i,RM} + \hat{\lambda}_{SMB}\hat{\beta}_{i,SMB} + \hat{\lambda}_{HML}\hat{\beta}_{i,HML} + \hat{\lambda}_{RMW}\hat{\beta}_{i,RMW} + \hat{\lambda}_{CMA}\hat{\beta}_{i,CMA} + \hat{\lambda}_{UMD}\hat{\beta}_{i,UMD} + \varsigma_{i,t}^m, \quad (23)$$

where the factor loadings and risk prices are obtained from the estimation of FF6 over the original sample. 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 other words, the misspecified model (CAPM) produces an average intercept that is about 16 times as large (in magnitude) the corresponding estimate for the “true” model.

These experiments confirm that model misspecification is an important driver of large cross-sectional intercept estimates, consistent with Propositions 1 and 2.

8 The Cross-Sectional Intercept and Factor Measurement Error

A natural concern is that a nonzero estimate of the CSR intercept $\hat{\lambda}_0$ in (1) might reflect measurement error in the factor(s), rather than omitted priced components. This section 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 2.1.²⁵

²⁵To isolate the role of factor mismeasurement, we work in population throughout this section 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 specification.

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

$$E(R_{i,t}) = \lambda_h \beta_{i,h} + \alpha_i, \quad \beta_{i,h} = \frac{\text{Cov}(R_{i,t}, h_t)}{\text{Var}(h_t)}, \quad i = 1, \dots, n, \quad (24)$$

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

$$f_t = h_t + \eta_t, \quad E(\eta_t) = 0, \quad \text{Cov}(h_t, \eta_t) = 0. \quad (25)$$

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

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

$$\beta_{i,f} = \frac{\text{Cov}(R_{i,t}, h_t)}{\text{Var}(h_t) + \text{Var}(\eta_t)} = \kappa \beta_{i,h}, \quad \kappa = \frac{\text{Var}(h_t)}{\text{Var}(h_t) + \text{Var}(\eta_t)} \in (0, 1). \quad (26)$$

Substituting (26) into (24) yields

$$E(R_{i,t}) = \lambda_f \beta_{i,f} + \alpha_i, \quad \lambda_f = \frac{\lambda_h}{\kappa} = \lambda_h \left[1 + \frac{\text{Var}(\eta_t)}{\text{Var}(h_t)} \right], \quad (27)$$

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 cross-sectional regression (best linear projection)

$$E(R_{i,t}) = \lambda_0^{\text{pop}} + b^{\text{pop}} \beta_{i,f} + \omega_i^{\text{pop}}, \quad \mathbb{E}_n(\omega_i^{\text{pop}}) = 0, \quad \mathbb{E}_n(\beta_{i,f} \omega_i^{\text{pop}}) = 0, \quad (28)$$

where ω_i^{pop} denotes the population CSR residual. If the factor model is correctly specified ($\alpha_i = 0$), then (27) implies $E(R_{i,t}) = \lambda_f \beta_{i,f}$ for all i , so $\lambda_0^{\text{pop}} = 0$ and $b^{\text{pop}} = \lambda_f$, and therefore $\omega_i^{\text{pop}} = 0$. More generally, when an intercept is included, the orthogonality conditions in (28) imply that any nonzero intercept reflects the mean component of the structural pricing errors α_i (after partialling

out $\beta_{i,f}$), not factor noise.

When can an intercept appear? Only when the “error” behaves like an omitted factor. If $\text{Cov}(R_{i,t}, \eta_t) \neq 0$ for some assets (while maintaining $\text{Cov}(h_t, \eta_t) = 0$ from (25)), the proxy beta $\beta_{i,f}$ mixes exposure to h_t and exposure to η_t . Define $\beta_{i,\eta} = \text{Cov}(R_{i,t}, \eta_t) / \text{Var}(\eta_t)$. A simple covariance decomposition gives:

$$\beta_{i,h} = \pi \beta_{i,f} - \phi_i, \quad \pi = \frac{\text{Var}(h_t) + \text{Var}(\eta_t)}{\text{Var}(h_t)} > 1, \quad \phi_i = \beta_{i,\eta} \frac{\text{Var}(\eta_t)}{\text{Var}(h_t)}. \quad (29)$$

Combining (24) and (29) implies

$$\mathbb{E}(R_{i,t}) = (\pi \lambda_h) \beta_{i,f} - \lambda_h \phi_i + \alpha_i. \quad (30)$$

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 $\mathbb{E}(R_{i,t})$ 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 $\beta_{i,f}$. To isolate the measurement-error channel, suppose $\mathbb{E}_n(\alpha_i) = 0$; allowing $\mathbb{E}_n(\alpha_i) \neq 0$ simply adds the usual mean pricing-error term to the intercept. In particular, if $\alpha_i \approx 0$ and ϕ_i is not systematically explained by $\beta_{i,f}$, then the intercept is approximately

$$\lambda_0^{\text{pop}} \approx -\lambda_h \mathbb{E}_n(\phi_i) = -\lambda_h \mathbb{E}_n(\beta_{i,\eta}) \frac{\text{Var}(\eta_t)}{\text{Var}(h_t)}. \quad (31)$$

Moreover, averaging (30) across i gives the identity

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

where $\bar{\mu}$ denotes the cross-sectional average expected excess return. Since $\lambda_0^{\text{pop}} = \bar{\mu} - b^{\text{pop}} \mathbb{E}_n(\beta_{i,f})$ in (28), it follows that if $\mathbb{E}_n(\alpha_i) = 0$ and $b^{\text{pop}} = \lambda_f$ (e.g., if $-\lambda_h \phi_i + \alpha_i$ is cross-sectionally uncorrelated with $\beta_{i,f}$), then

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

With $\lambda_h > 0$, this channel yields a positive intercept only if $\mathbb{E}_n(\beta_{i,\eta}) < 0$. Its magnitude is approximately

$$|\lambda_0^{\text{pop}}| \approx \lambda_h |\mathbb{E}_n(\beta_{i,\eta})| \frac{\text{Var}(\eta_t)}{\text{Var}(h_t)},$$

so an economically large positive intercept further requires either a sufficiently negative average loading on the measurement-error component and/or a large variance ratio $\text{Var}(\eta_t)/\text{Var}(h_t)$. Both features are hard to reconcile with a proxy f_t that contains meaningful signal about 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 $\mathbb{E}_n(\beta_{i,\eta}) < 0$, while the magnitude is amplified by $\text{Var}(\eta_t)/\text{Var}(h_t)$. Such a case is best interpreted as an omitted (or otherwise unmodeled) factor component rather than innocuous measurement noise.

9 Asymptotic Inference for R_{pen}^2 and $R_{\text{ols}*}^2$

This section develops large- T (time-series) inference for R_{pen}^2 and $R_{\text{ols}*}^2$. Our focus is the economically relevant null $H_0 : R_x^2 = 0$ for $x \in \{\text{ols}*, \text{pen}\}$. We exploit the fact that the lower bound for R_x^2 is negative, in particular for R_{pen}^2 when $\lambda_0 \neq 0$.

For the conventional with-intercept CSR ratio $R_{\text{ols}}^2 \in [0, 1]$, inference at the boundary points $R_{\text{ols}}^2 = 0$ and $R_{\text{ols}}^2 = 1$ is nonstandard; see [Kan, Robotti, and Shanken \(2013\)](#).

9.1 Setup

Let $\mathbf{R}_t$ be the $n \times 1$ vector of excess returns on the test assets and $\mathbf{f}_t$ the $K \times 1$ factor vector, with n and K fixed and $T \rightarrow \infty$. Define the primitive moment vector

$$\boldsymbol{\psi} = (\boldsymbol{\mu}'_0, \text{vec}(\boldsymbol{\Sigma}_{Rf})', \text{vec}(\boldsymbol{\Sigma}_f)')', \quad \boldsymbol{\mu}_0 = \mathbb{E}(\mathbf{R}_t), \quad \boldsymbol{\Sigma}_{Rf} = \text{Cov}(\mathbf{R}_t, \mathbf{f}_t), \quad \boldsymbol{\Sigma}_f = \text{Var}(\mathbf{f}_t),$$

and let $\widehat{\boldsymbol{\psi}}$ denote the corresponding sample analogues. Following [Kan, Robotti, and Shanken \(2013\)](#), we assume a time-series CLT:

$$\sqrt{T}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \quad (32)$$

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

Let $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{Rf} \boldsymbol{\Sigma}_f^{-1}$ denote the population beta matrix. Let $Q_0 = \frac{1}{n} \boldsymbol{\mu}'_0 \mathbf{M}_1 \boldsymbol{\mu}_0$ denote the cross-sectional variance of expected returns, where $\mathbf{M}_1 = \mathbf{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}'$. Let $Q^* = \frac{1}{n} \boldsymbol{\nu}' \boldsymbol{\nu}$ denote the restricted (no-intercept) CSR numerator and 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 two population ratios are

$$R_{\text{ols}*}^2 = 1 - \frac{Q^*}{Q_0}, \quad R_{\text{pen}}^2 = 1 - \frac{Q_{\text{pen}}}{Q_0}. \quad (33)$$

Unlike R_{ols}^2 , both ratios in (33) can be negative, which makes $R_x^2 = 0$ generically an interior point.

9.2 Wald inference at $R_x^2 = 0$

Proposition 6. *Maintain Assumptions A.1–A.2 in the IA. Then,*

- (a) *The null hypothesis $H_0 : R_{\text{ols}*}^2 = 0$ is an interior restriction. Under H_0 ,*

$$\sqrt{T} \widehat{R}_{\text{ols}*}^2 \Rightarrow \mathcal{N}(0, V_{\text{ols}*}), \quad V_{\text{ols}*} = D_{\text{ols}*} \boldsymbol{\Omega} D_{\text{ols}*}',$$

where $D_{\text{ols}} = \partial g_{\text{ols}*}(\boldsymbol{\psi}) / \partial \boldsymbol{\psi}'$ is evaluated at the true $\boldsymbol{\psi}$ satisfying $R_{\text{ols}*}^2 = 0$.*

- (b) *If $R_{\text{pen}}^2 = 0$ and $\lambda_0 \neq 0$, then $H_0 : R_{\text{pen}}^2 = 0$ is an interior restriction and*

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

If instead $\lambda_0 = 0$, then $R_{\text{pen}}^2 = R_{\text{ols}}^2$ and inference at $R_{\text{pen}}^2 = 0$ follows [Kan, Robotti, and Shanken \(2013\)](#).

Assume additionally that $V_x > 0$ for $x \in \{\text{ols*}, \text{pen}\}$ so that Wald inference is nondegenerate. A feasible Wald statistic is

$$\mathcal{T}_x = \frac{\widehat{R}_x^2}{\sqrt{\widehat{V}_x/T}}, \quad \widehat{V}_x = \widehat{D}_x \widehat{\Omega} \widehat{D}_x', \quad x \in \{\text{ols*}, \text{pen}\}, \quad (34)$$

where $\widehat{\Omega}$ is a HAC estimator of Ω and $\widehat{D}_x$ is the plug-in estimate of D_x using the closed-form null gradients reported in the IA. Under H_0 , $\mathcal{T}_x \Rightarrow \mathcal{N}(0, 1)$. For example, $\widehat{\Omega}$ may be a Newey–West HAC estimator with truncation lag $L_T \rightarrow \infty$ and $L_T/T \rightarrow 0$ as $T \rightarrow \infty$, under the weak dependence conditions in Assumption A.1.

For a two-sided level- α test of $H_0 : R_x^2 = 0$ against $H_1 : R_x^2 \neq 0$, reject when $|\mathcal{T}_x| > z_{1-\alpha/2}$. For the one-sided alternative $H_1 : R_x^2 > 0$, reject when $\mathcal{T}_x > z_{1-\alpha}$.²⁶

10 Conclusion

This paper studies the role of the cross-sectional intercept in two-pass regression tests when test assets are excess returns. In that setting, the linear pricing equation restricts the cross-sectional intercept to zero, yet empirical implementations often estimate an unrestricted intercept. We show that under omitted priced factors the intercept has a structural interpretation as a common pricing offset: It equals the mean pricing error arising because the included factor loadings fail to span the priced space. Allowing the intercept to absorb this common pricing offset mechanically reallocates pricing failures into fitted premia, inflating conventional measures of fit and distorting risk-price estimates.

On the econometric side, we derive an exact mapping between restricted (no-intercept) and unrestricted (with-intercept) second-pass estimates and characterize when the distortions are large. A key determinant is the degree to which the factor loadings are nearly collinear with the constant vector. In that case, even modest average pricing errors (in the restricted regression) can generate large changes in estimated risk premia, including sign reversals, through a familiar collinearity amplification channel—which stems from betas with high mean (in magnitude) and/or low cross-

²⁶Consistent with much of the related literature, our empirical implementation in Section 4 estimates the spectral density matrix Ω using zero lags. Using a Newey–West estimator with two lags yields very similar inference for the cross-sectional R^2 measures. Results are available upon request.

sectional dispersion. We also show that including a constant regressor can aggravate collinearity and weaken identification in the cross-section.

On the inferential side, we develop delta-method Wald tests for the penalized and no-intercept R^2 measures, exploiting the fact that these ratios can be negative so that $R^2 = 0$ is an interior point, which yields standard normal test statistics with HAC standard errors.

Empirically, these implications are quantitatively important in standard applications. Across the CAPM and a broad set of prominent macro-factor models applied to widely used anomaly-sorted portfolio cross-sections, intercept estimates are often economically large, and the conditions for amplification are frequently present. Once the intercept is treated as part of total pricing error, the penalized fit measure turns negative in nearly all cases, in contrast to conventional residual-based R^2 measures. Risk-price estimates are likewise sensitive: Allowing an intercept frequently attenuates or reverses estimated factor premia, including for the market factor, and rank tests indicate that identification is weaker in the unrestricted specification. Finally, intercept magnitudes in unrestricted CSRs often imply zero-beta rates and aggregate equity premia that are difficult to reconcile with standard economic priors, consistent with the interpretation of large intercepts as reflecting omitted priced factors rather than small money-market wedges.

The practical implication is simple. When the left-hand-side variable is an excess return, second-pass CSRs should be estimated without an intercept. Researchers may still report the unrestricted intercept as a diagnostic for common pricing offset (e.g., from omitted priced factors), but model evaluation and risk-price inference should be based on the restricted specification. More broadly, our results emphasize that seemingly innocuous departures from the model-implied cross-sectional restriction can materially alter empirical conclusions in linear asset-pricing tests.

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.
- Balduzzi, P., and C. Robotti, 2008, Mimicking portfolios, economic risk premia, and tests of multi-beta models, *Journal of Business and Economic Statistics* 26, 354–368.
- Banz, R.W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.
- 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.
- Burnside, C., 2011, The cross section of foreign currency risk premia and consumption growth risk: Comment, *American Economic Review* 101, 3456–3476.
- 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.
- Cooper, I., and P. Maio, 2019, Asset growth, profitability, and investment opportunities, *Management Science* 65, 3988–4010.
- Cooper, M.J., H. Gulen, and M.J. Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1651.
- 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, Working Paper.
- Fama, E.F., and K.R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.
- Fama, E.F., and K.R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.
- Fama, E.F., and K.R. French, 2008, Dissecting anomalies, *Journal of Finance* 63, 1653–1678.
- Fama, E.F., and K.R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.
- 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.
- Gagliardini, P., E. Ossola, and O. Scaillet, 2016, Time-varying risk premia in large cross-sectional equity data sets, *Econometrica* 84, 985–1046.

- Giglio, S., and D. Xiu, 2021, Asset pricing with omitted factors, *Journal of Political Economy* 129, 1947–1990.
- 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.
- Gospodinov, N., and C. Robotti, 2021, Common pricing across asset classes: Empirical evidence revisited, *Journal of Financial Economics* 140, 292–324.
- Guo, H., and P. Maio, 2020, ICAPM and the accruals Aanomaly, *Quarterly Journal of Finance*. 10: 2050014.
- 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.
- Jagannathan, R., and Z. Wang, 1998, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, *Journal of Finance* 53, 1285–1309.
- Jegadeesh, N., and S. Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.
- 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.
- Liao, Z., Y. Liu, and Z. Xie, 2024, Optimal cross-sectional regression, *Management Science* 70, 7911–7942.
- Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.
- Lioui, A., and P. Maio, 2014, Interest rate risk and the cross-section of stock returns, *Journal of Financial and Quantitative Analysis* 49, 483–511.
- 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, 2025, Revisiting Lettau and Ludvigson (2001): Does *cay* really matter for cross-sectional risk premia? *Critical Finance Review* (forthcoming).
- Maio, P., and D. Philip, 2018, Economic activity and momentum profits: Further evidence, *Journal of Banking & Finance* 88, 466–482.
- Maio, P., and P. Santa-Clara, 2012, Multifactor models and their consistency with the ICAPM, *Journal of Financial Economics* 106, 586–613.
- Maio, P., and P. Santa-Clara, 2017, Short-term interest rates and stock market anomalies, *Journal of Financial and Quantitative Analysis* 52, 927–961.

- Merton, R.C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.
- Mossin, Jan, 1966, Equilibrium in a capital asset market, *Econometrica* 34, 768–783.
- Newey, W.K., and K.D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
- Novy-Marx, R., 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.
- Raponi, V., C. Robotti, and P. Zaffaroni, 2020, Testing beta-pricing models using large cross-sections, *Review of Financial Studies* 33, 2796–2842.
- Rosenberg, B., K. Reid, and R. Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9–16.
- Santos, T., and P. Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.
- Savov, A., 2011, Asset pricing with garbage, *Journal of Finance* 66, 177–201.
- Shanken, J., 1992, On the estimation of beta pricing models, *Review of Financial Studies* 5, 1–34.
- Sharpe, W.F., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.
- Titman, S., K.J. Wei, and F. Xie, 2004, Capital investments and stock returns, *Journal of Financial and Quantitative Analysis* 39, 677–700.
- Yogo, M., 2006, A consumption-based explanation of expected stock returns, *Journal of Finance* 61, 539–580.

Table 2: 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.

	Mean (%)	Stdev. (%)	Min. (%)	Max. (%)
INV10	1.944	0.437	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

Table 3: Pricing Performance in Asset Pricing Tests I

This table presents the evaluation results for the asset pricing models. 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 $\lambda_0^2 / \text{Var}_n(\mu_i)$, where λ_0 denotes the intercept estimate and $\text{Var}_n(\mu_i)$ 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_{pen}^2 denotes the cross-sectional R^2 , while R_{ols}^2 denotes a similar metric, except that is based on pricing errors that exclude the intercept. The values in parentheses represent asymptotic p -values to test the null of a zero fit: $R_{ols}^2 = 0$ or $R_{pen}^2 = 0$. $RRMSE^*$ and R_{ols*}^2 denote the performance metrics associated with the CSR without intercept. The sample is 1970:I to 2017:III.

	Intercept						No intercept				
	Wedge	s	$RRMSE$	$RRMSE_{pen}$	R_{ols}^2	p	R_{pen}^2	p	$RRMSE^*$	R_{ols*}^2	p
Panel A: INV10											
CAPM	63.08	0.98	0.18	1.67	0.24	(0.22)	-62.85	(0.97)	0.30	-1.13	(0.98)
JW2	146.41	0.99	0.11	2.53	0.75	(0.31)	-145.66	(0.87)	0.30	-1.08	(0.95)
JW3	147.22	0.99	0.10	2.53	0.77	(0.37)	-146.45	(0.88)	0.30	-1.07	(0.92)
SV2	62.93	0.98	0.18	1.67	0.28	(0.44)	-62.66	(0.98)	0.30	-1.09	(0.98)
SV3	160.95	0.99	0.10	2.65	0.75	(0.41)	-160.20	(0.82)	0.24	-0.34	(0.61)
SV4	168.87	0.99	0.10	2.71	0.77	(0.56)	-168.10	(0.79)	0.24	-0.34	(0.60)
LL2	61.70	0.98	0.17	1.65	0.35	(0.29)	-61.35	(0.96)	0.29	-0.98	(0.93)
AEM1	3.93	0.91	0.10	0.43	0.76	(0.02)	-3.16	(0.59)	0.16	0.39	(0.43)
AEM2	3.98	0.99	0.10	0.43	0.76	(0.09)	-3.22	(0.60)	0.11	0.73	(0.12)
HKM1	54.26	0.99	0.20	1.55	0.10	(0.42)	-54.16	(0.91)	0.27	-0.64	(0.96)
HKM2	15.27	0.99	0.14	0.83	0.53	(0.11)	-14.74	(0.76)	0.16	0.38	(0.26)
CCAPM	24.03	0.98	0.21	1.04	0.00	(0.97)	-24.03	(0.69)	0.25	-0.47	(0.68)
LVN3	49.68	0.99	0.14	1.48	0.53	(0.45)	-49.15	(0.81)	0.22	-0.14	(0.56)
Panel B: OP10											
CAPM	149.97	0.99	0.10	2.21	0.69	(0.10)	-149.28	(0.97)	0.29	-1.52	(0.97)
JW2	122.21	0.99	0.10	2.00	0.71	(0.13)	-121.50	(0.88)	0.18	-0.02	(0.51)
JW3	88.68	1.00	0.09	1.70	0.74	(0.21)	-87.95	(0.77)	0.14	0.39	(0.35)
SV2	143.09	0.99	0.10	2.16	0.69	(0.12)	-142.40	(0.96)	0.22	-0.44	(0.74)
SV3	107.69	0.99	0.08	1.87	0.80	(0.18)	-106.89	(0.85)	0.16	0.18	(0.44)
SV4	121.00	0.99	0.08	1.99	0.83	(0.23)	-120.17	(0.85)	0.16	0.18	(0.44)
LL2	185.59	0.99	0.10	2.46	0.72	(0.13)	-184.87	(0.96)	0.23	-0.68	(0.79)
AEM1	57.84	0.97	0.17	1.38	0.14	(0.55)	-57.70	(0.79)	0.29	-1.56	(0.76)
AEM2	146.11	0.99	0.10	2.18	0.71	(0.17)	-145.40	(0.96)	0.28	-1.40	(0.88)
HKM1	134.59	0.98	0.09	2.09	0.75	(0.10)	-133.83	(0.97)	0.31	-1.88	(0.97)
HKM2	124.34	0.99	0.09	2.01	0.76	(0.13)	-123.58	(0.94)	0.23	-0.70	(0.73)
CCAPM	26.04	0.98	0.18	0.94	0.00	(0.94)	-26.04	(0.68)	0.23	-0.62	(0.70)
LVN3	38.03	0.98	0.17	1.12	0.14	(0.96)	-37.89	(0.69)	0.22	-0.51	(0.66)
Panel C: M10											
CAPM	29.06	0.96	0.34	2.43	0.43	(0.03)	-28.63	(0.99)	0.57	-0.63	(0.98)
JW2	5.87	0.98	0.18	1.09	0.84	(0.14)	-5.03	(0.68)	0.23	0.72	(0.10)
JW3	3.33	0.98	0.13	0.82	0.91	(0.16)	-2.42	(0.64)	0.17	0.85	(0.08)
SV2	50.97	0.97	0.23	3.19	0.74	(0.18)	-50.23	(0.88)	0.57	-0.62	(0.98)
SV3	2.67	0.98	0.13	0.74	0.92	(0.20)	-1.76	(0.59)	0.16	0.87	(0.07)
SV4	3.67	1.00	0.13	0.86	0.92	(0.21)	-2.75	(0.59)	0.14	0.90	(0.06)
LL2	17.84	0.97	0.15	1.89	0.88	(0.07)	-16.96	(0.90)	0.37	0.30	(0.33)
AEM1	0.39	0.78	0.30	0.41	0.55	(0.21)	0.16	(0.48)	0.33	0.46	(0.35)
AEM2	26.56	1.00	0.28	2.32	0.59	(0.71)	-25.97	(0.54)	0.30	0.56	(0.14)
HKM1	32.63	0.96	0.28	2.56	0.61	(0.05)	-32.02	(0.97)	0.59	-0.78	(0.98)
HKM2	24.91	0.96	0.23	2.24	0.73	(0.16)	-24.18	(0.91)	0.48	-0.16	(0.60)
CCAPM	1.56	0.93	0.22	0.60	0.76	(0.17)	-0.80	(0.53)	0.26	0.66	(0.20)
LVN3	6.84	0.99	0.14	1.17	0.90	(0.52)	-5.94	(0.58)	0.17	0.85	(0.06)

Table 4: Pricing Performance in Asset Pricing Tests II

This table presents the evaluation results for the asset pricing models. 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 3.

	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	(0.80)	-13.79	(0.65)	0.21	-0.09	(0.60)
JW2	43.84	1.00	0.19	1.34	0.14	(0.89)	-43.70	(0.62)	0.21	-0.07	(0.56)
JW3	103.56	1.00	0.15	2.04	0.42	(0.83)	-103.15	(0.72)	0.20	0.00	(0.50)
SV2	13.77	0.99	0.20	0.77	0.02	(0.98)	-13.75	(0.65)	0.21	-0.08	(0.58)
SV3	36.37	1.00	0.18	1.22	0.14	(0.99)	-36.22	(0.54)	0.19	0.06	(0.47)
SV4	33.01	1.00	0.18	1.16	0.23	(0.99)	-32.78	(0.54)	0.18	0.16	(0.46)
LL2	11.41	0.99	0.08	0.68	0.83	(0.22)	-10.58	(0.60)	0.10	0.75	(0.22)
AEM1	8.76	0.83	0.09	0.60	0.77	(0.10)	-7.99	(0.73)	0.26	-0.72	(0.58)
AEM2	0.02	0.99	0.08	0.09	0.83	(0.13)	0.81	(0.30)	0.08	0.83	(0.15)
HKM1	1.29	0.98	0.13	0.26	0.60	(0.04)	-0.69	(0.55)	0.13	0.58	(0.14)
HKM2	7.35	0.99	0.09	0.55	0.79	(0.05)	-6.56	(0.65)	0.10	0.74	(0.17)
CCAPM	8.74	0.97	0.19	0.62	0.11	(0.73)	-8.63	(0.60)	0.21	-0.13	(0.53)
LVN3	4.93	0.98	0.16	0.47	0.39	(0.74)	-4.53	(0.58)	0.17	0.28	(0.40)
Panel B: All											
CAPM	50.18	0.98	0.24	2.08	0.32	(0.05)	-49.86	(0.99)	0.39	-0.76	(0.99)
JW2	28.69	0.99	0.22	1.58	0.45	(0.09)	-28.25	(0.86)	0.29	0.03	(0.48)
JW3	25.28	0.99	0.22	1.48	0.46	(0.22)	-24.83	(0.81)	0.27	0.15	(0.40)
SV2	54.28	0.98	0.24	2.16	0.33	(0.09)	-53.95	(0.99)	0.38	-0.69	(0.99)
SV3	29.12	0.99	0.21	1.59	0.49	(0.19)	-28.63	(0.86)	0.28	0.07	(0.46)
SV4	32.06	0.99	0.21	1.66	0.49	(0.25)	-31.57	(0.84)	0.28	0.09	(0.44)
LL2	48.45	0.98	0.23	2.04	0.38	(0.09)	-48.07	(0.99)	0.37	-0.66	(0.93)
AEM1	2.84	0.84	0.21	0.53	0.49	(0.01)	-2.35	(0.64)	0.29	0.03	(0.49)
AEM2	15.89	0.99	0.19	1.18	0.55	(0.05)	-15.34	(0.75)	0.23	0.37	(0.23)
HKM1	37.98	0.98	0.26	1.81	0.21	(0.12)	-37.77	(0.98)	0.38	-0.71	(0.99)
HKM2	49.00	0.98	0.24	2.05	0.33	(0.10)	-48.67	(1.00)	0.38	-0.71	(0.99)
CCAPM	0.67	0.96	0.25	0.35	0.25	(0.06)	-0.41	(0.54)	0.26	0.23	(0.29)
LVN3	4.53	0.97	0.21	0.66	0.47	(0.18)	-4.06	(0.63)	0.23	0.35	(0.20)
Panel C: All*											
CAPM	49.59	0.98	0.29	2.95	0.51	(0.01)	-49.09	(1.00)	0.52	-0.54	(0.99)
JW2	22.75	0.99	0.26	2.00	0.60	(0.04)	-22.15	(0.79)	0.33	0.38	(0.22)
JW3	22.12	0.99	0.26	1.98	0.60	(0.17)	-21.52	(0.78)	0.32	0.40	(0.22)
SV2	54.41	0.98	0.29	3.08	0.52	(0.03)	-53.89	(0.99)	0.51	-0.49	(0.95)
SV3	23.19	0.99	0.26	2.02	0.62	(0.18)	-22.57	(0.78)	0.32	0.40	(0.21)
SV4	27.64	0.99	0.25	2.20	0.63	(0.26)	-27.00	(0.77)	0.32	0.40	(0.21)
LL2	45.47	0.98	0.29	2.82	0.53	(0.03)	-44.94	(1.00)	0.48	-0.36	(0.83)
AEM1	0.99	0.71	0.27	0.50	0.57	(0.00)	-0.42	(0.55)	0.35	0.29	(0.41)
AEM2	17.95	0.99	0.23	1.78	0.70	(0.02)	-17.25	(0.76)	0.30	0.50	(0.14)
HKM1	26.93	0.97	0.36	2.19	0.27	(0.08)	-26.66	(0.97)	0.51	-0.51	(0.98)
HKM2	50.14	0.98	0.28	2.96	0.54	(0.01)	-49.60	(1.00)	0.51	-0.51	(0.98)
CCAPM	0.88	0.95	0.29	0.48	0.52	(0.05)	-0.35	(0.52)	0.30	0.48	(0.15)
LVN3	2.86	0.98	0.20	0.73	0.76	(0.22)	-2.10	(0.58)	0.22	0.72	(0.05)

Table 5: Risk Price Estimates: All Portfolios

This table presents the 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's t -ratios (with * denoting statistical significance at the 5% level). Risk price estimates marked with * denote statistical significance (at the 5% level) based on misspecification-robust standard errors. The risk price estimates and t -ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\hat{\nu}_i)(\%)$ 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}	$\mathbb{E}_n(\hat{\nu}_i)$
CAPM	3.99* (5.11*)	-2.13*, 1.76* (-2.12*, 2.76*)						0.09
JW2	3.02* (3.24*)	-1.09, 1.95* (-0.98, 3.03*)	0.68, 1.41* (2.22*, 2.92*)					0.04
JW3	2.84 (2.63*)	-0.89, 1.98* (-0.73, 3.11*)	0.59, 0.93 (3.02*, 3.30*)	-0.13, -0.35 (-0.66, -1.70)				0.03
SV2	4.15* (5.06*)	-2.27, 1.71* (-2.19*, 2.69*)			-0.06, 0.11 (-1.05, 1.98*)			0.08
SV3	3.04* (3.03*)	-1.14, 1.93* (-0.98, 2.97*)	0.68, 1.42* (2.06*, 2.82*)		0.02, 0.02 (1.19, 0.83)			0.04
SV4	3.19* (3.27*)	-1.27, 1.89* (-1.11, 2.93*)	0.65, 1.40* (2.15*, 2.78*)		-0.00, 0.12 (-0.02, 1.40)			0.04
LL2	3.93* (4.88*)	-1.99, 1.85* (-1.97*, 2.92*)				0.05, 0.07 (1.15, 1.62)		0.08
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}	$\mathbb{E}_n(\hat{\nu}_i)$
AEM1	0.95 (0.93)		11.07*, 20.87 (2.80*, 1.70)					0.15
AEM2	2.25 (2.00*)	-0.47, 1.69* (-0.36, 2.64*)	8.63, 14.21* (2.39*, 4.02*)					0.03
HKM1	3.48* (4.41*)			-0.03, 0.03* (-1.66, 2.74*)				0.08
HKM2	3.95* (4.93*)	-2.12*, 1.69* (-2.09*, 2.67*)		-0.02, 0.03 (-0.87, 2.10*)				0.08
CCAPM	0.46 (0.36)				0.79, 1.04 (2.29*, 1.80)			0.02
LVN3	1.20 (1.11)				0.38, 0.89* (1.07, 2.25*)	-14.98, -14.11 (-2.15*, -1.75)	-0.03, -0.04 (-0.41, -0.48)	0.03

Table 6: Risk Price Estimates: 17 Portfolios

This table presents the 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's t -ratios (with * denoting statistical significance at the 5% level). Risk price estimates marked with * denote statistical significance (at the 5% level) based on misspecification-robust standard errors. The risk price estimates and t -ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\hat{\nu}_i)(\%)$ 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}	$\mathbb{E}_n(\hat{\nu}_i)$
CAPM	5.88* (5.74*)	-3.64*, 1.52* (-3.10*, 2.38*)						0.12
JW2	3.98 (3.52*)	-1.80, 1.89* (-1.42, 2.88*)	0.82, 1.79* (2.08*, 2.71*)					0.04
JW3	3.92 (3.18*)	-1.74, 1.93* (-1.27, 2.96*)	0.77, 1.53 (2.24*, 2.47*)	-0.10, -0.24 (-0.35, -0.65)				0.04
SV2	6.16* (5.30*)	-3.88*, 1.52* (-3.04*, 2.36*)			-0.09, 0.12 (-1.21, 1.91)			0.11
SV3	4.02 (3.37*)	-1.91, 1.82* (-1.44, 2.71*)	0.70, 1.69* (1.98*, 2.94*)			0.02, 0.02 (0.88, 0.57)		0.04
SV4	4.39 (3.51*)	-2.25, 1.82* (-1.63, 2.71*)	0.62, 1.68* (1.97*, 2.92*)		-0.03, 0.10 (-0.33, 0.89)	0.02, 0.02 (0.90, 0.52)		0.04
LI2	5.63* (5.92*)	-3.38*, 1.62* (-3.05*, 2.53*)					0.03, 0.11 (0.64, 1.91)	0.11
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}	$\mathbb{E}_n(\hat{\nu}_i)$
AEM1	0.83 (0.70)		13.18*, 21.03 (3.05*, 1.70)					0.24
AEM2	3.54 (2.12*)	-1.49, 1.67* (-0.87, 2.55*)	8.79, 15.60* (1.98*, 4.21*)					0.04
HKM1	4.33* (4.29*)			-0.04, 0.03* (-2.11*, 2.37*)				0.13
HKM2	5.91* (5.76*)	-3.71*, 1.49* (-3.18*, 2.33*)		-0.03, 0.03 (-1.31, 1.70)				0.12
CCAPM	-0.78 (-0.30)				1.50*, 1.07 (1.81, 1.66)			-0.04
LVN3	1.41 (0.72)				0.68, 1.14 (0.87, 2.31*)	-21.57, -18.62 (-1.77, -1.45)	0.02, -0.01 (0.22, -0.04)	0.02

Table 7: Risk Price Estimates: BM10

This table presents the 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's t -ratios (with * denoting statistical significance at the 5% level). Risk price estimates marked with * denote statistical significance (at the 5% level) based on misspecification-robust standard errors. The risk price estimates and t -ratios displayed after the comma are associated with the corresponding CSR with no intercept. $\mathbb{E}_n(\hat{\nu}_i)(\%)$ 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}	$\mathbb{E}_n(\hat{\nu}_i)$
CAPM	1.58 (1.24)	0.52, 2.15* (0.37, 3.31*)						0.01
JW2	2.82 (1.57)	-0.79, 2.14* (-0.42, 3.29*)	0.80, 0.25 (1.92, 0.69)					0.01
JW3	4.33 (1.24)	-2.38, 2.14* (-0.69, 3.24*)	0.21, -0.16 (0.26, -0.26)	0.67, 0.35 (0.78, 0.71)				0.02
SV2	1.58 (1.21)	0.57, 2.19* (0.40, 3.26*)			-0.06, -0.06 (-0.38, -0.32)			0.01
SV3	2.57 (1.26)	-0.53, 2.16* (-0.25, 3.23*)	0.75, 0.22 (1.55, 0.45)			-0.00, -0.03 (-0.19, -1.27)		0.01
SV4	2.44 (0.92)	-0.25, 2.32* (-0.09, 2.88*)	0.90, 0.40 (1.24, 0.56)		-0.26, -0.31 (-0.76, -0.74)	-0.02, -0.05 (-0.65, -1.17)		0.01
LL2	1.44 (0.57)	0.26, 1.74* (0.10, 2.63*)					-0.32, -0.32 (-1.34, -1.32)	0.01
	λ_0	λ_{RM}	λ_{LEV}	λ_{IC}	λ_C	λ_{my}	λ_{Cmy}	$\mathbb{E}_n(\hat{\nu}_i)$
AEM1	1.26 (1.55)		7.97, 18.06 (1.80, 2.04*)					0.22
AEM2	0.06 (0.04)	1.81, 1.87* (1.02, 2.93*)	8.46, 8.50 (1.82, 1.99*)					0.00
HKM1	-0.48 (-0.38)			0.05*, 0.04* (1.94, 3.31*)				-0.01
HKM2	1.15 (0.87)	0.68, 1.86* (0.46, 2.91*)		0.03, 0.05* (1.61, 3.10*)				0.01
CCAPM	1.26 (1.62)				0.45, 1.10 (1.26, 1.92)			0.03
LVN3	0.94 (0.96)				0.61, 1.18 (1.42, 1.61)	-10.95, -12.85 (-1.09, -0.95)	0.00, 0.04 (0.02, 0.31)	0.02

Table 8: Rank Tests on Quarterly Factor Loadings

This table presents the p -values associated with the F -test on the 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, 0.276	0.510, 0.336	0.120, 0.112	0.208, 0.225	0.168, 0.155	0.084, 0.097
JW3	0.507, 0.605	0.630, 0.545	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	0.589, 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, 0.000	0.018, 0.031	0.679, 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, 0.080	0.692, 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, 0.018	0.139, 0.019	0.561, 0.130	0.242, 0.124	0.107, 0.049	0.334, 0.035
LVN3	0.380, 0.426	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 Proposition 1

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

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

are

$$\mathbf{W}' \boldsymbol{\mu} = \mathbf{W}' \mathbf{W} \begin{bmatrix} \hat{\lambda}_0 \\ \hat{\boldsymbol{\lambda}} \end{bmatrix}.$$

For the augmented CSR,

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

the first $K + 1$ OLS normal equations are

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

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

$$\begin{bmatrix} \hat{\lambda}_0 \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\boldsymbol{\theta}} \end{bmatrix} + (\mathbf{W}' \mathbf{W})^{-1} \mathbf{W}' \hat{\boldsymbol{\rho}} \hat{\boldsymbol{\gamma}}. \quad (\text{A.1})$$

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

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

Substituting into (A.1) yields (6). □

A.2 Proposition 2

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

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

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

A.3 Proposition 3

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

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

since $\mathbb{E}_n(\widehat{\omega}_i) = 0$ by the OLS normal equations for the unrestricted CSR. Divide by $\text{Var}_n(\mu_i)$ to obtain (10). $\square$

A.4 Corollary 1

Proof. From Proposition 3, $R_{\text{pen}}^2 = R_{\text{ols}}^2 - \widehat{\lambda}_0^2 / \text{Var}_n(\mu_i)$. Since $R_{\text{ols}}^2 \in [0, 1]$ for the OLS regression with an intercept, it follows that $R_{\text{pen}}^2 \leq 1$ and $R_{\text{pen}}^2 \geq -\widehat{\lambda}_0^2 / \text{Var}_n(\mu_i) = -\text{Wedge}$. Moreover, $R_{\text{pen}}^2 < 0$ iff $R_{\text{ols}}^2 < \widehat{\lambda}_0^2 / \text{Var}_n(\mu_i)$, equivalently iff $\widehat{\lambda}_0^2 > \text{Var}_n(\mu_i) R_{\text{ols}}^2$. $\square$

A.5 Corollary 2

Proof. Let $\widetilde{\lambda}$ and $(\widehat{\lambda}_0, \widehat{\lambda})$ denote the restricted (no-intercept) and unrestricted (with-intercept) CSR coefficient estimates, respectively, and define

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

From the restricted CSR normal equations,

$$\mathbb{E}_n(\widehat{\beta}_i \widehat{\nu}_i) = \mathbf{0} \quad \Longleftrightarrow \quad M \widetilde{\lambda} = \mathbf{v},$$

where $\mathbf{v} = \mathbb{E}_n(\widehat{\boldsymbol{\beta}}_i \mu_i)$. For the unrestricted CSR, the slope normal equation is

$$\mathbf{v} = \mathbf{m} \widehat{\lambda}_0 + \mathbf{M} \widehat{\boldsymbol{\lambda}},$$

and therefore

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

Under $\mathbf{M} \succ 0$, let $\mathbf{u} = \mathbf{M}^{-1} \mathbf{m}$ and $s = \mathbf{m}' \mathbf{u}$. Then $\widehat{\boldsymbol{\lambda}} = \widetilde{\boldsymbol{\lambda}} - \widehat{\lambda}_0 \mathbf{u}$ and hence

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

Squaring and averaging across i gives

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

where the cross term vanishes because $\mathbb{E}_n(\widehat{\boldsymbol{\beta}}_i \widehat{\nu}_i) = \mathbf{0}$. Finally, $\mathbf{u}' \mathbf{M} \mathbf{u} = \mathbf{m}' \mathbf{M}^{-1} \mathbf{m} = s$, so

$$\mathbb{E}_n(\widehat{\alpha}_i^2) = \mathbb{E}_n(\widehat{\nu}_i^2) + s \widehat{\lambda}_0^2.$$

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

$$R_{\text{ols}*}^2 - R_{\text{pen}}^2 = \frac{\mathbb{E}_n(\widehat{\alpha}_i^2) - \mathbb{E}_n(\widehat{\nu}_i^2)}{\text{Var}_n(\mu_i)} = s \cdot \frac{\widehat{\lambda}_0^2}{\text{Var}_n(\mu_i)} \geq 0,$$

with equality if and only if $\widehat{\lambda}_0 = 0$ or $s = 0$ (equivalently $\mathbf{m} = \mathbf{0}$). □

A.6 Proposition 4

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

A.7 Proposition 5

Proof. Let $\mathbf{W} = [\mathbf{1} : \hat{\boldsymbol{\beta}}]$ and write $\bar{\mu} = \mathbb{E}_n(\mu_i)$, $\mathbf{m} = \mathbb{E}_n(\hat{\boldsymbol{\beta}}_i)$, $\mathbf{M} = \mathbb{E}_n(\hat{\boldsymbol{\beta}}_i \hat{\boldsymbol{\beta}}_i')$, and $\mathbf{v} = \mathbb{E}_n(\hat{\boldsymbol{\beta}}_i \mu_i)$.

The $K+1$ OLS normal equations for the unrestricted CSR are

$$\frac{1}{n} \mathbf{W}' \boldsymbol{\mu} = \frac{1}{n} \mathbf{W}' \mathbf{W} \begin{bmatrix} \hat{\lambda}_0 \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} \iff \begin{cases} \bar{\mu} = \hat{\lambda}_0 + \mathbf{m}' \hat{\boldsymbol{\lambda}}, \\ \mathbf{v} = \mathbf{m} \hat{\lambda}_0 + \mathbf{M} \hat{\boldsymbol{\lambda}}. \end{cases}$$

Premultiplying the first equation by $\mathbf{m}$ and subtracting from the second gives

$$\mathbf{v} - \mathbf{m} \bar{\mu} = (\mathbf{M} - \mathbf{m} \mathbf{m}') \hat{\boldsymbol{\lambda}} = \boldsymbol{\Sigma} \hat{\boldsymbol{\lambda}},$$

so

$$\hat{\boldsymbol{\lambda}} = \boldsymbol{\Sigma}^{-1}(\mathbf{v} - \mathbf{m} \bar{\mu}).$$

Using the restricted CSR normal equations $\mathbf{M} \tilde{\boldsymbol{\lambda}} = \mathbf{v}$ yields

$$\hat{\boldsymbol{\lambda}} = \boldsymbol{\Sigma}^{-1}(\mathbf{M} \tilde{\boldsymbol{\lambda}} - \mathbf{m} \bar{\mu}),$$

which proves the first equality in (15).

Next, comparing $\mathbf{M} \tilde{\boldsymbol{\lambda}} = \mathbf{v}$ with the second unrestricted normal equation implies

$$\mathbf{M} (\hat{\boldsymbol{\lambda}} - \tilde{\boldsymbol{\lambda}}) = -\mathbf{m} \hat{\lambda}_0.$$

Under $\mathbf{M} \succ 0$, let $\mathbf{u} = \mathbf{M}^{-1} \mathbf{m}$ and $s = \mathbf{m}' \mathbf{u}$. Then

$$\hat{\boldsymbol{\lambda}} - \tilde{\boldsymbol{\lambda}} = -\hat{\lambda}_0 \mathbf{u}.$$

Substitute $\hat{\boldsymbol{\lambda}} = \tilde{\boldsymbol{\lambda}} - \hat{\lambda}_0 \mathbf{u}$ into $\bar{\mu} = \hat{\lambda}_0 + \mathbf{m}' \hat{\boldsymbol{\lambda}}$ to obtain

$$\bar{\mu} = \hat{\lambda}_0 + \mathbf{m}' \tilde{\boldsymbol{\lambda}} - \hat{\lambda}_0 \mathbf{m}' \mathbf{u} = \mathbf{m}' \tilde{\boldsymbol{\lambda}} + (1 - s) \hat{\lambda}_0,$$

hence

$$\hat{\lambda}_0 = \frac{\bar{\mu} - \mathbf{m}'\tilde{\lambda}}{1 - s},$$

which is (14). Plugging this into $\hat{\lambda} = \tilde{\lambda} - \hat{\lambda}_0 \mathbf{u}$ yields the second equality in (15), and (16) follows immediately. $\square$

A.8 Corollary 3

Proof. From (16), $\hat{\lambda} - \tilde{\lambda} = -\hat{\lambda}_0 \mathbf{u}$. Under $\mathbf{M} \succ 0$, $\mathbf{u} = \mathbf{M}^{-1}\mathbf{m} = \mathbf{0}$ if and only if $\mathbf{m} = \mathbf{0}$. Hence $\hat{\lambda} = \tilde{\lambda}$ if and only if either $\hat{\lambda}_0 = 0$ or $\mathbf{m} = \mathbf{0}$. By (14), $\hat{\lambda}_0 = 0$ is equivalent to $\bar{\mu} - \mathbf{m}'\tilde{\lambda} = 0$, which in turn is equivalent to $\mathbb{E}_n(\hat{\nu}_i) = 0$ in the restricted CSR. $\square$

A.9 Corollary 4

Proof. Fix $j \in \{1, \dots, K\}$. From (16), $\hat{\lambda}_j = \tilde{\lambda}_j + \Delta_j$ with $\Delta_j = -\hat{\lambda}_0 u_j$.

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

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

A.10 Corollary 5

Proof. When $K = 1$, $\mathbf{m} = m$, $\mathbf{M} = M$, and $\Sigma = \sigma^2 = M - m^2 > 0$. The restricted CSR slope is $\tilde{\lambda} = v/M$ where $v = \mathbb{E}_n(\hat{\beta}_i \mu_i)$, while the unrestricted CSR slope is $\hat{\lambda} = c/\sigma^2$ where $c = \text{Cov}_n(\hat{\beta}_i, \mu_i) = v - m\bar{\mu}$. Since $v = M\tilde{\lambda}$, it follows that

$$\hat{\lambda} = \frac{M\tilde{\lambda} - m\bar{\mu}}{\sigma^2}, \quad \hat{\lambda} - \tilde{\lambda} = \frac{m(\tilde{\lambda}m - \bar{\mu})}{\sigma^2}.$$

Assume $\tilde{\lambda} > 0$ and $m > 0$. A sign flip ($\hat{\lambda} < 0$) occurs iff $M\tilde{\lambda} - m\bar{\mu} < 0$, i.e., iff $\tilde{\lambda} < \bar{\mu}m/M$. Attenuation without sign change ($0 < \hat{\lambda} < \tilde{\lambda}$) requires both $M\tilde{\lambda} - m\bar{\mu} > 0$ (i.e., $\tilde{\lambda} > \bar{\mu}m/M$) and

$$\frac{M\tilde{\lambda} - m\bar{\mu}}{\sigma^2} < \tilde{\lambda} \iff M\tilde{\lambda} - m\bar{\mu} < \tilde{\lambda}(M - m^2) \iff \tilde{\lambda} < \bar{\mu}/m.$$

Combining yields $\bar{\mu}m/M < \tilde{\lambda} < \bar{\mu}/m$. The remaining cases follow by symmetry. $\square$

A.11 Corollary 6

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

B Internet Appendix: Asymptotic Distribution of the Cross-Sectional R^2 Measures

This appendix provides a feasible large- T (time-series) Wald test for the null of zero explanatory power for the two fit measures that can be negative, $R_{\text{ols}*}^2$ and R_{pen}^2 . We focus on $H_0 : R_x^2 = 0$ for $x \in \{\text{ols*}, \text{pen}\}$ and report closed-form gradients evaluated at the null, which are sufficient for delta-method inference. Throughout, the cross-sectional dimension n and the number of factors K are fixed while $T \rightarrow \infty$.

B.1 Time-series sampling framework and primitive moments

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

$$\boldsymbol{\mu}_0 = \mathbb{E}(\mathbf{R}_t) \in \mathbb{R}^n, \quad \boldsymbol{\mu}_f = \mathbb{E}(\mathbf{f}_t) \in \mathbb{R}^K, \quad \boldsymbol{\Sigma}_{Rf} = \text{Cov}(\mathbf{R}_t, \mathbf{f}_t) \in \mathbb{R}^{n \times K}, \quad \boldsymbol{\Sigma}_f = \text{Var}(\mathbf{f}_t) \in \mathbb{R}^{K \times K},$$

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

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{Rf}\boldsymbol{\Sigma}_f^{-1} \in \mathbb{R}^{n \times K}. \tag{B.2}$$

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

Sample analogues are

$$\boldsymbol{\mu} = \frac{1}{T} \sum_{t=1}^T \mathbf{R}_t, \quad \bar{\mathbf{f}} = \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t,$$

$$\widehat{\Sigma}_{Rf} = \frac{1}{T} \sum_{t=1}^T (\mathbf{R}_t - \boldsymbol{\mu})(\mathbf{f}_t - \bar{\mathbf{f}})', \quad \widehat{\Sigma}_f = \frac{1}{T} \sum_{t=1}^T (\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})',$$

and

$$\widehat{\boldsymbol{\beta}} = \widehat{\Sigma}_{Rf} \widehat{\Sigma}_f^{-1}. \quad (\text{B.3})$$

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

$$\boldsymbol{\psi} = \begin{bmatrix} \boldsymbol{\mu}_0 \\ \text{vec}(\Sigma_{Rf}) \\ \text{vec}(\Sigma_f) \end{bmatrix} \in \mathbb{R}^{n+nK+K^2}, \quad \widehat{\boldsymbol{\psi}} = \begin{bmatrix} \boldsymbol{\mu} \\ \text{vec}(\widehat{\Sigma}_{Rf}) \\ \text{vec}(\widehat{\Sigma}_f) \end{bmatrix}. \quad (\text{B.4})$$

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

$$\mathbf{u}_t = \begin{bmatrix} \mathbf{R}_t - \boldsymbol{\mu}_0 \\ \text{vec}\left((\mathbf{R}_t - \boldsymbol{\mu}_0)(\mathbf{f}_t - \boldsymbol{\mu}_f)' - \Sigma_{Rf}\right) \\ \text{vec}\left((\mathbf{f}_t - \boldsymbol{\mu}_f)(\mathbf{f}_t - \boldsymbol{\mu}_f)' - \Sigma_f\right) \end{bmatrix}. \quad (\text{B.5})$$

Then the linear representation and CLT hold:

$$\sqrt{T}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{u}_t + o_p(1) \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), \quad \boldsymbol{\Omega} = \sum_{\ell=-\infty}^{\infty} \mathbb{E}(\mathbf{u}_t \mathbf{u}_{t-\ell}'). \quad (\text{B.6})$$

Because $\text{vec}(\Sigma_f)$ stacks symmetric entries twice, $\boldsymbol{\Omega}$ may be singular; this does not affect the scalar variances $V_x = D_x \boldsymbol{\Omega} D_x'$ reported below.

Assumption A.2 (Identification and nondegeneracy). Σ_f is nonsingular and the unrestricted second-pass design matrix $\mathbf{W} = [\mathbf{1} \ \boldsymbol{\beta}] \in \mathbb{R}^{n \times (K+1)}$ 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:

$$Q_0 = \frac{1}{n} \boldsymbol{\mu}_0' \mathbf{M}_1 \boldsymbol{\mu}_0 > 0, \quad \mathbf{M}_1 = \mathbf{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}'. \quad (\text{B.7})$$

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

B.2 Second-pass CSRs and the three R^2 measures

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

$$\boldsymbol{\mu} = \widehat{\lambda}_0 \mathbf{1} + \widehat{\boldsymbol{\beta}} \widehat{\boldsymbol{\lambda}} + \widehat{\boldsymbol{\omega}}, \quad \begin{bmatrix} \widehat{\lambda}_0 \\ \widehat{\boldsymbol{\lambda}} \end{bmatrix} = (\widehat{\mathbf{W}}' \widehat{\mathbf{W}})^{-1} \widehat{\mathbf{W}}' \boldsymbol{\mu}, \quad (\text{B.8})$$

and the restricted (no-intercept) CSR is

$$\boldsymbol{\mu} = \widehat{\boldsymbol{\beta}} \widetilde{\boldsymbol{\lambda}} + \widetilde{\boldsymbol{\nu}}, \quad \widetilde{\boldsymbol{\lambda}} = (\widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{\beta}})^{-1} \widehat{\boldsymbol{\beta}}' \boldsymbol{\mu}. \quad (\text{B.9})$$

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

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\mu} - \widehat{\boldsymbol{\beta}} \widehat{\boldsymbol{\lambda}} = \widehat{\lambda}_0 \mathbf{1} + \widehat{\boldsymbol{\omega}}. \quad (\text{B.10})$$

Define the common denominator

$$\widehat{Q}_0 = \frac{1}{n} \boldsymbol{\mu}' \mathbf{M}_1 \boldsymbol{\mu}, \quad (\text{B.11})$$

and the three numerator quadratic forms

$$\widehat{Q} = \frac{1}{n} \widehat{\boldsymbol{\omega}}' \widehat{\boldsymbol{\omega}}, \quad \widehat{Q}^* = \frac{1}{n} \widetilde{\boldsymbol{\nu}}' \widetilde{\boldsymbol{\nu}}, \quad \widehat{Q}_{\text{pen}} = \frac{1}{n} \widehat{\boldsymbol{\alpha}}' \widehat{\boldsymbol{\alpha}}.$$

The corresponding explanatory ratios are

$$\widehat{R}_{\text{ols}}^2 = 1 - \frac{\widehat{Q}}{\widehat{Q}_0}, \quad \widehat{R}_{\text{ols}^*}^2 = 1 - \frac{\widehat{Q}^*}{\widehat{Q}_0}, \quad \widehat{R}_{\text{pen}}^2 = 1 - \frac{\widehat{Q}_{\text{pen}}}{\widehat{Q}_0}. \quad (\text{B.12})$$

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

$$\widehat{Q}_{\text{pen}} = \frac{1}{n} (\widehat{\lambda}_0 \mathbf{1} + \widehat{\boldsymbol{\omega}})' (\widehat{\lambda}_0 \mathbf{1} + \widehat{\boldsymbol{\omega}}) = \widehat{Q} + \widehat{\lambda}_0^2, \quad \widehat{R}_{\text{pen}}^2 = \widehat{R}_{\text{ols}}^2 - \frac{\widehat{\lambda}_0^2}{\widehat{Q}_0}. \quad (\text{B.13})$$

Population counterparts satisfy the analogous identities:

$$Q_{\text{pen}} = Q + \lambda_0^2, \quad R_{\text{pen}}^2 = R_{\text{ols}}^2 - \frac{\lambda_0^2}{Q_0}.$$

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

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

where the Frisch–Waugh–Lovell formulas are

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

The population restricted (no-intercept) CSR is

$$\boldsymbol{\mu}_0 = \boldsymbol{\beta} \tilde{\boldsymbol{\lambda}} + \boldsymbol{\nu}, \quad \tilde{\boldsymbol{\lambda}} = (\boldsymbol{\beta}' \boldsymbol{\beta})^{-1} \boldsymbol{\beta}' \boldsymbol{\mu}_0.$$

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

$$Q = \frac{1}{n} \boldsymbol{\omega}' \boldsymbol{\omega}, \quad Q^* = \frac{1}{n} \boldsymbol{\nu}' \boldsymbol{\nu}, \quad Q_{\text{pen}} = \frac{1}{n} \boldsymbol{\alpha}' \boldsymbol{\alpha}, \quad R_{\text{ols}^*}^2 = 1 - \frac{Q^*}{Q_0}, \quad R_{\text{pen}}^2 = 1 - \frac{Q_{\text{pen}}}{Q_0}.$$

B.3 Wald inference at $R_x^2 = 0$: null gradients, delta-method variance, and HAC estimation

Fix $x \in \{\text{ols}^*, \text{pen}\}$ and consider the null $H_0 : R_x^2 = 0$. Under Assumptions A.1–A.2, the CLT in (B.6) and the delta method yield an asymptotic normal approximation for $\hat{R}_x^2 = g_x(\hat{\boldsymbol{\psi}})$ provided the corresponding null is an interior restriction. The key implementation objects are the gradient $D_x = \partial g_x(\boldsymbol{\psi}) / \partial \boldsymbol{\psi}'$ and the long-run covariance matrix $\boldsymbol{\Omega}$.

B.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 $\mathbf{G}_\mu(g) \in \mathbb{R}^n$ and $\mathbf{G}_\beta(g) \in \mathbb{R}^{n \times K}$ by

$$dg = \mathbf{G}_\mu(g)' d\boldsymbol{\mu} + \text{tr}(\mathbf{G}_\beta(g)' d\boldsymbol{\beta}). \quad (\text{B.14})$$

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

$$d\boldsymbol{\beta} = d\boldsymbol{\Sigma}_{Rf} \boldsymbol{\Sigma}_f^{-1} - \boldsymbol{\beta} (d\boldsymbol{\Sigma}_f) \boldsymbol{\Sigma}_f^{-1}. \quad (\text{B.15})$$

Substituting (B.15) into (B.14) and using the cyclic property of the trace yields

$$dg = \mathbf{G}_\mu(g)' d\boldsymbol{\mu} + \text{tr}(\mathbf{G}_{\Sigma_{Rf}}(g)' d\boldsymbol{\Sigma}_{Rf}) + \text{tr}(\mathbf{G}_{\Sigma_f}(g)' d\boldsymbol{\Sigma}_f), \quad (\text{B.16})$$

where

$$\mathbf{G}_{\Sigma_{Rf}}(g) = \mathbf{G}_\beta(g) \boldsymbol{\Sigma}_f^{-1} \in \mathbb{R}^{n \times K}, \quad \mathbf{G}_{\Sigma_f}(g) = -\boldsymbol{\beta}' \mathbf{G}_\beta(g) \boldsymbol{\Sigma}_f^{-1} \in \mathbb{R}^{K \times K}. \quad (\text{B.17})$$

Vectorizing (B.16) yields the gradient of g with respect to the primitive moments $\boldsymbol{\psi}$:

$$D(g) = \frac{\partial g}{\partial \boldsymbol{\psi}'} = \begin{bmatrix} \mathbf{G}_\mu(g)' & \text{vec}(\mathbf{G}_{\Sigma_{Rf}}(g))' & \text{vec}(\mathbf{G}_{\Sigma_f}(g))' \end{bmatrix} \in \mathbb{R}^{1 \times (n+nK+K^2)}. \quad (\text{B.18})$$

For $x \in \{\text{ols*}, \text{pen}\}$ we write $D_x = D(R_x^2)$ and $V_x = D_x \boldsymbol{\Omega} D_x'$.

B.3.2 Auxiliary derivatives for λ_0

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

$$\mathbf{g}_\mu = \frac{1}{n} \mathbf{1} - \mathbf{M}_1 \boldsymbol{\beta} \boldsymbol{\kappa}, \quad d\lambda_0 = \mathbf{g}_\mu' d\boldsymbol{\mu} + \text{tr}(\mathbf{G}_0' d\boldsymbol{\beta}), \quad (\text{B.19})$$

where

$$\mathbf{G}_0 = -\frac{1}{n} \mathbf{1} \boldsymbol{\lambda}' - \boldsymbol{\omega} \boldsymbol{\kappa}' + (\mathbf{M}_1 \boldsymbol{\beta}) \boldsymbol{\kappa} \boldsymbol{\lambda}'. \quad (\text{B.20})$$

(Here $\mathbf{g}_\mu$ is the gradient of the scalar λ_0 ; it is distinct from the notation $\mathbf{G}_\mu(\cdot)$ in (B.14).)

B.3.3 Closed-form null gradients for $R_{\text{ols*}}^2$ and R_{pen}^2

At the null $R_x^2 = 0$, the gradients in (B.14) simplify to closed-form expressions.

(i) **No-intercept ratio.** If $R_{\text{ols}^*}^2 = 0$ (equivalently $Q^* = Q_0$), then

$$\mathbf{G}_\mu(R_{\text{ols}^*}^2) \Big|_{R_{\text{ols}^*}^2=0} = -\frac{2}{nQ_0} \boldsymbol{\nu} + \frac{2}{nQ_0} \mathbf{M}_1 \boldsymbol{\mu}_0, \quad \mathbf{G}_\beta(R_{\text{ols}^*}^2) \Big|_{R_{\text{ols}^*}^2=0} = \frac{2}{nQ_0} \boldsymbol{\nu} \tilde{\boldsymbol{\lambda}}'. \quad (\text{B.21})$$

(ii) **Penalized ratio.** If $R_{\text{pen}}^2 = 0$ (equivalently $Q + \lambda_0^2 = Q_0$), then

$$\mathbf{G}_\mu(R_{\text{pen}}^2) \Big|_{R_{\text{pen}}^2=0} = -\frac{2}{nQ_0} \boldsymbol{\omega} - \frac{2\lambda_0}{Q_0} \mathbf{g}_\mu + \frac{2}{nQ_0} \mathbf{M}_1 \boldsymbol{\mu}_0, \quad \mathbf{G}_\beta(R_{\text{pen}}^2) \Big|_{R_{\text{pen}}^2=0} = \frac{2}{nQ_0} \boldsymbol{\omega} \boldsymbol{\lambda}' - \frac{2\lambda_0}{Q_0} \mathbf{G}_0. \quad (\text{B.22})$$

B.3.4 Null asymptotic distribution and feasible Wald test

Corollary 7. *Maintain Assumptions A.1–A.2.*

(a) *If $R_{\text{ols}^*}^2 = 0$ (equivalently $Q^* = Q_0$), then*

$$\sqrt{T} \hat{R}_{\text{ols}^*}^2 \Rightarrow \mathcal{N}(0, V_{\text{ols}^*}), \quad V_{\text{ols}^*} = D_{\text{ols}^*} \boldsymbol{\Omega} D_{\text{ols}^*}',$$

where $D_{\text{ols}^} = D(R_{\text{ols}^*}^2)$ in (B.18) uses the null gradients in (B.21).*

(b) *If $R_{\text{pen}}^2 = 0$ (equivalently $Q + \lambda_0^2 = Q_0$) and $\lambda_0 \neq 0$, then R_{pen}^2 can take negative values in a neighborhood of the null, so $R_{\text{pen}}^2 = 0$ is an interior restriction and*

$$\sqrt{T} \hat{R}_{\text{pen}}^2 \Rightarrow \mathcal{N}(0, V_{\text{pen}}), \quad V_{\text{pen}} = D_{\text{pen}} \boldsymbol{\Omega} D_{\text{pen}}',$$

where $D_{\text{pen}} = D(R_{\text{pen}}^2)$ uses the null gradients in (B.22). If instead $\lambda_0 = 0$, then $R_{\text{pen}}^2 = R_{\text{ols}}^2$ and inference at $R_{\text{pen}}^2 = 0$ follows Kan, Robotti, and Shanken (2013).

Assume additionally that $V_x > 0$ for the relevant x so that Wald inference is nondegenerate.

Feasible HAC estimation. Let $\hat{\mathbf{u}}_t$ be the sample analogue of (B.5) obtained by replacing population objects $(\boldsymbol{\mu}_0, \boldsymbol{\mu}_f, \boldsymbol{\Sigma}_{Rf}, \boldsymbol{\Sigma}_f)$ with their sample analogues $(\boldsymbol{\mu}, \bar{\mathbf{f}}, \hat{\boldsymbol{\Sigma}}_{Rf}, \hat{\boldsymbol{\Sigma}}_f)$. Given a bandwidth L , the Newey–West estimator is

$$\hat{\boldsymbol{\Omega}} = \hat{\mathbf{\Gamma}}_0 + \sum_{\ell=1}^L w_\ell (\hat{\mathbf{\Gamma}}_\ell + \hat{\mathbf{\Gamma}}_\ell'), \quad \hat{\mathbf{\Gamma}}_\ell = \frac{1}{T} \sum_{t=\ell+1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_{t-\ell}', \quad w_\ell = 1 - \frac{\ell}{L+1}.$$

Form $\widehat{D}_x$ by plugging sample analogues into (B.18), using the null gradients (B.21) (for $x = \text{ols*}$) and (B.22) (for $x = \text{pen}$). Then $\widehat{V}_x = \widehat{D}_x \widehat{\Omega} \widehat{D}_x'$.

Wald test at $R_x^2 = 0$. For $x \in \{\text{ols*}, \text{pen}\}$, the feasible Wald statistic is

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

The two-sided test rejects at asymptotic size α when $|\mathcal{T}_x| > z_{1-\alpha/2}$; the one-sided alternative $H_1 : R_x^2 > 0$ rejects when $\mathcal{T}_x > z_{1-\alpha}$.

Feasible computation

1. Compute $\boldsymbol{\mu}$, $\bar{\mathbf{f}}$, $\widehat{\Sigma}_{Rf}$, $\widehat{\Sigma}_f$, and $\widehat{\beta}$ in (B.3), and set $\widehat{Q}_0 = (1/n)\boldsymbol{\mu}'\mathbf{M}_1\boldsymbol{\mu}$.
2. Compute $(\widehat{\lambda}_0, \widehat{\lambda})$ and $\widehat{\omega}$ from (B.8), and compute $\widetilde{\lambda}$ and $\widehat{\nu}$ from (B.9).
3. Compute $\widehat{R}_{\text{ols*}}^2$ and $\widehat{R}_{\text{pen}}^2$ from (B.12).
4. Compute the auxiliary quantities for the null gradients: $\widehat{\mathbf{m}} = (1/n)\widehat{\beta}'\mathbf{1}$, $\widehat{\mathbf{A}} = \widehat{\beta}'\mathbf{M}_1\widehat{\beta}$, $\widehat{\kappa} = \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{m}}$, $\widehat{\mathbf{g}}_\mu = \frac{1}{n}\mathbf{1} - \mathbf{M}_1\widehat{\beta}\widehat{\kappa}$, and $\widehat{\mathbf{G}}_0 = -\frac{1}{n}\mathbf{1}\widehat{\lambda}' - \widehat{\omega}\widehat{\kappa}' + (\mathbf{M}_1\widehat{\beta})\widehat{\kappa}\widehat{\lambda}'$.
5. Form $\widehat{D}_{\text{ols*}}$ and $\widehat{D}_{\text{pen}}$ using (B.18) with gradients (B.21)–(B.22) (replace population objects by sample analogues).
6. Construct $\widehat{\mathbf{u}}_t$, estimate $\widehat{\Omega}$ by Newey–West, compute $\widehat{V}_x = \widehat{D}_x \widehat{\Omega} \widehat{D}_x'$, and report $\mathcal{T}_x = \widehat{R}_x^2 / \sqrt{\widehat{V}_x/T}$.