# Multi-Factor Asset Pricing via Model Averaging

Moyu Liao* Fengli Ma† Wenyu Zhou‡

April 18, 2026

**Abstract**

Empirical asset pricing routinely conditions on a single benchmark factor model, yet no specification commands consensus and the choice materially affects estimated alphas and anomaly significance. We propose a frequentist Model-Averaging Multi-Factor (MAMF) framework that replaces benchmark selection with a data-driven combination of standard specifications, balancing pricing fit against model complexity. MAMF is asymptotically optimal under $\alpha$-mixing dependence, but its data-dependent weights yield a non-pivotal limiting distribution that invalidates plug-in standard errors. We therefore develop a simulation-based inference procedure with asymptotically valid confidence intervals and show in calibrated simulations that it attains near-oracle accuracy and nominal coverage. Applied to 148 anomalies from the Global Factor Data of Jensen, Kelly, and Pedersen (2023), MAMF classifies 53.4% as having statistically significant alphas, a rejection rate between those of parsimonious and richly parameterized benchmarks, obtained with the tightest confidence intervals among all estimators. Roughly half of the factor zoo survives once benchmark uncertainty is formally incorporated into inference.

**Keywords:** Multi-factor Models, Model Averaging, Asset Pricing, Model Uncertainty, Factor Zoo

**JEL Codes:** G12, C12, C52

*School of Economics, The University of Sydney, NSW, Australia. Email: moyu.liao@sydney.edu.au.

†School of Economics, Zhejiang University, Hangzhou, Zhejiang 310058, China. Email: mafl@zju.edu.cn.

‡Corresponding author. International Business School, Zhejiang University, Hangzhou, Zhejiang 310058, China. Email: wenyuzhou@zju.edu.cn.

# 1 Introduction

Multi-factor models are a workhorse of empirical asset pricing. They underpin the detection and validation of return anomalies, the benchmarking and attribution of fund performance, and the construction of risk-adjusted measures of expected returns. Over the past several decades, researchers have proposed a large menu of "standard" factor specifications—some of which have become canonical benchmarks in applied work. Most of these models take a linear form: returns on test assets (or portfolios) are regressed on a set of common factors intended to proxy for systematic risks. In this framework, the key objects are the intercept (the abnormal, or risk-adjusted, return) and the slope coefficients (the factor loadings that reflect exposures to the underlying risks).

These applications confront a practical and sharply posed question: which factor model should serve as the benchmark? Despite the prominence of multi-factor models, there remains limited consensus on a single best specification across samples, asset classes, and research settings. This model uncertainty is especially salient in the anomaly literature, where studies often adopt a pre-selected benchmark with limited justification. Such flexibility is not innocuous: because inference on abnormal performance is inherently model-dependent, a researcher who surveys several benchmarks and reports the one delivering the largest alpha engages in implicit specification search, inflating rejection rates and compounding concerns about "alpha hacking."

We address this with frequentist model averaging. Beginning with a menu of well-established factor specifications, we estimate model-specific alphas and betas and combine them using a Mallows-type criterion that explicitly trades off in-sample fit against effective complexity. The resulting Model-Averaging Multi-Factor (MAMF) estimator replaces discrete benchmark selection with a transparent, data-driven combination of the candidate models. Weights are asset-specific and chosen from a convex simplex, so the averaged alphas and betas retain their usual economic interpretation as abnormal returns and factor exposures, now robust to the particular benchmark used.

We establish two theoretical results. First, the MAMF estimator satisfies an oracle-type property: its fitted-value loss is asymptotically as small as the minimum loss achievable over feasible convex combinations of the candidate models. Second, because the weights depend on the data, the limiting distribution of the averaged alpha and beta is nonstandard and non-pivotal, which invalidates conventional plug-in standard errors. We therefore develop a simulation-based inference procedure that consistently approximates this limiting distribution and delivers confidence intervals with asymptotically correct coverage under $\alpha$-mixing dependence.

Calibrated Monte Carlo experiments verify the theory in realistic sample sizes. Factor realizations are anchored to observed monthly data to preserve empirical magnitudes and time-series dependence, and a standard six-factor benchmark serves as the oracle data-generating model. Two findings emerge. First, MAMF yields materially more stable alpha estimates than any single benchmark, with mean absolute errors close to those of the oracle specification and substantially smaller than those of misspecified alternatives. Second, the simulation-based confidence intervals attain near-nominal coverage with moderate length, accommodating the additional uncertainty from data-dependent weighting without undue loss of precision. Importantly, MAMF correctly concentrates weight on the oracle specification rather than on the largest candidate, confirming that the averaging is driven by fit rather than dimensionality.

Applied to 148 anomaly portfolios from the Global Factor Data of Jensen et al. (2023), MAMF produces three empirical findings. First, benchmark disagreement is first-order: across seven widely used specifications, 58% of anomalies change significance status at the 5% level, and economic magnitudes differ markedly across benchmarks. Second, MAMF classifies 53.4% of the 148 anomalies as having statistically significant alphas, with non-significant estimates concentrated near zero. Third, and most informatively, MAMF delivers the tightest confidence intervals of any estimator considered—narrower than those under every single benchmark—reflecting the efficiency gain from reallocating weight toward well-

fitting specifications. The MAMF rejection rate sits between those implied by parsimonious and richly parameterized benchmarks, moderating the high rates from the former while avoiding the benchmark-specific swings of the latter. Roughly half of the factor zoo thus survives once benchmark uncertainty is formally incorporated into inference.

Our study contributes to three primary streams of literature. First, we add to research on multi-factor model specification. Since the seminal development of the CAPM (Sharpe, 1964; Lintner, 1965), the field has progressed from the three-factor model (Fama and French, 1993) and momentum factor (Carhart, 1997) to specifications incorporating investment, profitability, and intangibles (Cooper et al., 2008; Fama and French, 2015, 2018; Hou et al., 2015, 2021). Other studies highlight the roles of liquidity (Pástor and Stambaugh, 2003), gross profitability (Novy-Marx, 2013), mispricing (Stambaugh and Yuan, 2017), human capital (Eiling, 2013), and the low-beta anomaly (Frazzini and Pedersen, 2014). While recent work employs machine learning to extract factors from high-dimensional pools (Lettau and Pelger, 2020; Gu et al., 2020; Giglio et al., 2022), we offer a more conservative but transparent framework that integrates information across well-acknowledged specifications to construct a model-averaged benchmark suitable for formal inference.

Second, our study speaks to the debate over the "factor zoo" and the robustness of return anomalies. Following Cochrane (2011)'s critique of the exploding number of factors, researchers have scrutinized reported predictors along several dimensions (Harvey et al., 2016; McLean and Pontiff, 2016; Hou et al., 2020; Chen and Zimmermann, 2022a). Recent efforts to identify valid factors extend beyond conventional $t$-statistic thresholds to include Bayesian inference (Jensen et al., 2023; Avramov et al., 2023b), shrinkage methods (Kelly et al., 2019; Kozak et al., 2020), and machine learning architectures (Gu et al., 2020; Leippold et al., 2022; Avramov et al., 2023a). We contribute in two ways. We confirm that model uncertainty materially affects perceived factor validity (Avramov et al., 2023b), potentially reconciling mixed findings in the literature (Novy-Marx and Velikov, 2016; Chen and Zimmermann, 2022a,b; Jensen et al., 2023; Hirshleifer and Ma, 2024). And our model-averaged inference

reveals that approximately 50% of factors in our sample remain significant, aligning with the conservative estimates of Hou et al. (2020) and offering a principled middle ground in the replication debate.

Finally, we connect to the econometric literature on model averaging as an alternative to discrete model selection. Building on frequentist methods (Hansen, 2007; Wan et al., 2010; Zhu et al., 2019; Wang et al., 2024; Peng et al., 2025; Hansen and Racine, 2012; Zhang et al., 2013), we adapt them to the dependence structures typical of asset-pricing data.[1] Our contribution brings a Mallows-type criterion to canonical factor specifications (Qiu et al., 2019), drawing on broader developments in model averaging for non-nested settings (Fang and Liu, 2020; Gao et al., 2023), heteroskedastic environments (Liu and Okui, 2013; Zhao et al., 2020), and varying-coefficient structures (Li et al., 2018; Sun et al., 2021, 2023), together with asymptotically optimal weighting schemes (Liang et al., 2011; Zhang et al., 2013, 2016; Li et al., 2018; Zhang, 2021; Peng et al., 2025). The framework thereby delivers reliable post-estimation inference in applications where benchmark uncertainty is central (Zhang and Liu, 2019; Yu et al., 2024).

The remainder of the paper is organized as follows. Section 2 develops the MAMF framework, establishes its asymptotic theory, and details the simulation-based inference procedure. Section 3 assesses finite-sample performance. Section 4 presents the empirical application, and Section 5 concludes. The Appendix contains formal derivations and proofs.

# 2 Theory

## 2.1 Setup

Consider a standard empirical asset-pricing setting with $K \geq 1$ factors and $N$ individual assets. Let $\mathcal{N} = \{1, \ldots, N\}$ and $\mathcal{K} = \{1, \ldots, K\}$ denote the index sets for assets and factors,

[1] Related work uses Bayesian criteria to pool information across benchmark models; see, for example, Avramov (2002), Cremers (2002), Wright (2008), and O'Doherty et al. (2016).

respectively. For each asset $i \in \mathcal{N}$, we observe a time series of gross returns $\{r_{it}\}_{t=1}^{T}$ and the risk-free rate $r_f$, and define excess returns $R_{it} = r_{it} - r_f$ for $t = 1, \ldots, T$. For each factor $k \in \mathcal{K}$, we likewise observe a time series of factor returns $\{F_{kt}\}_{t=1}^{T}$. For simplicity, assume all series share the same sample size $T$.

Let $\mathcal{M} = \{1, \ldots, M\}$ denote the set of candidate multi-factor asset-pricing models. Some candidates in $\mathcal{M}$ may include only a subset of the available factors. Without loss of generality, we assume that model $M$ contains all $K$ factors and refer to it as the full factor model. For any $m \in \mathcal{M}$, let $\mathcal{K}_m \subseteq \mathcal{K}$ be the index set of factors included in model $m$, with cardinality $K_m = |\mathcal{K}_m|$. For each asset $i$ and model $m$, we estimate the following time-series regression:

$$R_{it} \;=\; \alpha_i^{(m)} \;+\; \mathbf{F}_t^{(m)\top} \boldsymbol{\beta}_i^{(m)} \;+\; e_{it}^{(m)}, \tag{1}$$

where $\mathbf{F}_t^{(m)} = (F_{kt})_{k \in \mathcal{K}_m} \in \mathbb{R}^{K_m}$ collects the factor returns used by model $m$ at time $t$, and $\boldsymbol{\beta}_i^{(m)} = (\beta_{ik}^{(m)})_{k \in \mathcal{K}_m} \in \mathbb{R}^{K_m}$ is the corresponding loading vector for asset $i$.[2] The disturbance $e_{it}^{(m)}$ denotes the regression error in model $m$ for asset $i$ at time $t$.

The baseline exposition assumes that the candidate models satisfy a nesting structure (formalized in Assumption 1 below). This assumption, however, is without loss of generality for linear factor models. Any collection of non-nested models can be embedded into a nested structure by augmenting the factor space. To illustrate, consider two non-nested models $A$ and $B$ with factor sets $\mathcal{K}_A$ and $\mathcal{K}_B$. Define an augmented model $C$ with factor set $\mathcal{K}_C = \mathcal{K}_A \cup \mathcal{K}_B$. Then both models $A$ and $B$ can be represented as restricted versions of model $C$, obtained by imposing zero restrictions on the corresponding coefficients. This construction shows that MAMF can accommodate arbitrary collections of linear factor models, including those that are not nested in their original formulation.

In most empirical asset-pricing applications, the intercept $\alpha_i^{(m)}$ and the coefficient vector

[2] Let $\mathbf{F}_t = (F_{1t}, \ldots, F_{Kt})^\top$ denote the full factor vector. Then there exists a selection-and-permutation matrix $S^{(m)} \in \{0,1\}^{K_m \times K}$ such that $\mathbf{F}_t^{(m)} = S^{(m)} \mathbf{F}_t$. Hence the ordering of factors within $\mathbf{F}_t^{(m)}$ may differ from that of the full set without affecting the specification.

$\boldsymbol{\beta}_i^{(m)}$ are the primary objects of interest in the multi-factor regression. The intercept $\alpha_i^{(m)}$, often called the asset's "alpha" (abnormal or risk-adjusted return), captures the part of the excess return that is not explained by the included risk factors. In contrast, $\boldsymbol{\beta}_i^{(m)}$ is the vector of factor loadings that measures the asset's exposure to the systematic risks associated with each factor.

Current literature offers a wide range of multi-factor asset-pricing models and an extensive set of risk factors. As noted in the introduction, widely used specifications such as the CAPM, the Fama and French three-factor model, the Carhart four-factor model, and the Fama and French five-factor model are now standard in empirical work. In practice, researchers often estimate the intercept ("alpha") from a chosen specification to test new risk factors or to evaluate asset performance after controlling for systematic risk. However, these estimates are inherently model specific: an asset's alpha or beta is meaningful only relative to the selected specification. As a result, researchers must commit to a model before estimation, and the flexibility to search across specifications raises concerns about data snooping.

Motivated by this gap in the literature, we propose a frequentist model-averaging procedure to estimate the alpha and beta vectors. This approach aggregates information across all candidate models in a transparent, data-driven way and avoids ad hoc model selection. Let $w_{im}$ denote the weight assigned to candidate model $m \in \mathcal{M}$ for asset $i$, and let $\mathbf{w}_i = (w_{i1}, \ldots, w_{iM})^\top$ be the corresponding weight vector. The weights lie in the simplex

$$\mathcal{W} = \left\{ \mathbf{w} \in [0,1]^M : \sum_{m=1}^{M} w_m = 1 \right\}.$$

Given $K$ factors in total and $K_m$ factors included in model $m$, we define the $K$-dimensional augmentation $\boldsymbol{\beta}_i^{(m),*} \in \mathbb{R}^K$ by placing zeros in the coordinates corresponding to factors excluded from model $m$, so that $\boldsymbol{\beta}_i^{(m),*}$ embeds $\boldsymbol{\beta}_i^{(m)} \in \mathbb{R}^{K_m}$ in the full factor space.[3] For a

[3]For example, suppose there are five risk factors in total and model $m$ includes only the first three. Then $\boldsymbol{\beta}_i^{(m)} = (\beta_{i1}, \beta_{i2}, \beta_{i3})^\top$ and $\boldsymbol{\beta}_i^{(m),*} = (\beta_{i1}, \beta_{i2}, \beta_{i3}, 0, 0)^\top$. Embedding loadings in a common $K$-dimensional space allows direct comparison and averaging across models.

given asset $i$, the frequentist model-averaging estimators of alpha and beta are

$$\widehat{\alpha}_i(\mathbf{w}_i) \;=\; \sum_{m=1}^{M} w_{im}\,\widehat{\alpha}_i^{(m)} \qquad \text{and} \qquad \widehat{\boldsymbol{\beta}}_i(\mathbf{w}_i) \;=\; \sum_{m=1}^{M} w_{im}\,\widehat{\boldsymbol{\beta}}_i^{(m),*}, \tag{2}$$

where $\widehat{\alpha}_i^{(m)}$ and $\widehat{\boldsymbol{\beta}}_i^{(m),*}$ are the ordinary least-squares estimators from model $m$, with $\widehat{\boldsymbol{\beta}}_i^{(m),*}$ denoting the $K$-dimensional augmentation that zero-pads coordinates for factors excluded by $m$. Two features of equation (2) deserve emphasis. First, the zero-padding ensures that every candidate estimator $\widehat{\boldsymbol{\beta}}_i^{(m),*}$ lives in the same $\mathbb{R}^K$ space. Consequently, the averaged estimator $\widehat{\boldsymbol{\beta}}_i(\mathbf{w}_i)$ is a $K$-dimensional vector for every asset $i$, corresponding to the full set of factors. Second, the weight vector $\mathbf{w}_i$ controls how aggressively the estimator shrinks certain factor loadings toward zero. Assets with high signal-to-noise ratios in certain factor dimensions may place weight on richer specifications, while noisier assets may favor parsimony. Crucially, the factor space itself is invariant across assets. The central design choice is the weight vector $\mathbf{w}_i = (w_{i1}, \ldots, w_{iM})^\top$ on the unit simplex, which governs the bias–variance trade-off and thus the performance of the averaged estimators.

## 2.2 Common Factor Structure with Asset-specific Estimation

A potential concern is that asset-specific weights $\boldsymbol{w}_i$ depart from the standard asset-pricing paradigm of a common factor structure. They do not. For every asset $i$, the MAMF estimator admits the representation

$$R_{it} = \alpha_i(\boldsymbol{w}_i) + \boldsymbol{\beta}_i(\boldsymbol{w}_i)^\top \boldsymbol{F}_t + \varepsilon_{it}, \tag{3}$$

a standard linear factor pricing equation with the same factor vector $\boldsymbol{F}_t \in \mathbb{R}^K$ for all assets. Asset-specific weights translate into heterogeneous factor *loadings*, not heterogeneous factor *structures*.

The distinction is between the pricing model and the estimation strategy. The model

is (3): a common set of $K$ factors with asset-specific loadings, identical in structure to any standard multi-factor specification. The weights $\boldsymbol{w}_i$ belong to the estimator rather than the model, governing how information is pooled across candidate specifications to recover $\boldsymbol{\beta}_i$. A familiar analogy clarifies this role: in ridge regression, the penalty parameter $\lambda_i$ is routinely chosen by cross-validation separately for each asset, yet no one reads asset-specific penalties as imposing different factor structures. MAMF weights play the same role, controlling how aggressively loadings are shrunk toward the estimates implied by sparser nested specifications. Asset-specific tuning is an efficiency device, not a structural assumption.

Such tuning is not merely permissible but desirable. Assets differ in residual volatility, exposure magnitudes, and signal-to-noise ratios, and imposing a common $\boldsymbol{w}$ would require every asset to benefit equally from the same degree of model simplification—sacrificing efficiency without structural gain. Section 2.3 details the construction of $\boldsymbol{w}_i$ and the inference procedure, and Section 2.4 establishes the corresponding theoretical properties.

## 2.3 Implementation Details

### 2.3.1 Construction of the Model-Averaging Estimators

The performance of the model-averaging estimators for alpha and beta is mainly driven by the choice of weights. We therefore adopt a Mallows-type criterion to determine the model-averaging weights, whose objective is an estimate of the mean-squared prediction error and thus transparently balances in-sample fit against model complexity. To set notation, let $X_t^{(m)} = \left(1, F_{1t}^{(m)}, \ldots, F_{K_m t}^{(m)}\right)^\top$ denote the regressor vector for model $m$ at time $t$, and stack these regressors as $\mathbf{X}^{(m)} = \left(X_1^{(m)}, X_2^{(m)}, \ldots, X_T^{(m)}\right)^\top \in \mathbb{R}^{T\times(K_m+1)}$. For asset $i \in \mathcal{N}$, let $\mathbf{R}_i = (R_{i1}, \ldots, R_{iT})^\top \in \mathbb{R}^T$ and define the parameter vector $\boldsymbol{\gamma}_i^{(m)} = \left(\alpha_i^{(m)}, \boldsymbol{\beta}_i^{(m)\top}\right)^\top \in \mathbb{R}^{K_m+1}$. The construction of the weights proceeds in three steps, as detailed in Algorithm 1.

Two features of the MAMF estimator deserve emphasis. First, (8) averages fitted values across several plausible factor specifications to avoid committing to a single, possibly misspecified, model. Each candidate model $m$ defines a linear smoother $\mathbf{P}^{(m)}$, and the averaged

**Algorithm 1** Computing the MAMF Estimators

**Step 1.** For each $m \in \mathcal{M}$ and asset $i$, estimate the coefficients by ordinary least squares and compute the projection matrix:

$$\widehat{\boldsymbol{\gamma}}_i^{(m)} = \begin{pmatrix} \widehat{\alpha}_i^{(m)} \\ \widehat{\boldsymbol{\beta}}_i^{(m)} \end{pmatrix} = \arg\min_{\boldsymbol{\gamma} \in \mathbb{R}^{K_m+1}} \left\| \mathbf{R}_i - \mathbf{X}^{(m)} \boldsymbol{\gamma} \right\|_2^2, \tag{4}$$

$$\mathbf{P}^{(m)} = \mathbf{X}^{(m)} \big( (\mathbf{X}^{(m)})^\top \mathbf{X}^{(m)} \big)^{-1} (\mathbf{X}^{(m)})^\top. \tag{5}$$

**Step 2.** Let $\mathbf{w} = (w_1, \ldots, w_M)^\top$ be a generic weight vector in the unit simplex $\mathcal{W} = \{\mathbf{w} \in [0,1]^M : \sum_{m=1}^M w_m = 1\}$. Define the averaged projection matrix

$$\mathbf{P}(\mathbf{w}) = \sum_{m=1}^{M} w_m \, \mathbf{P}^{(m)}. \tag{6}$$

Choose the asset-specific weight vector $\mathbf{w}_i$ by minimizing the Mallows-type criterion

$$\mathcal{C}_i(\mathbf{w}) = \left\| \big( \mathbf{I}_T - \mathbf{P}(\mathbf{w}) \big) \mathbf{R}_i \right\|_2^2 + 2 \, \widehat{\sigma}_i^2 \, \mathbf{w}^\top \mathbf{p}, \tag{7}$$

where

$$\widehat{\sigma}_i^2 = \big[ T - (K_M + 1) \big]^{-1} \left\| \mathbf{R}_i - \mathbf{X}^{(M)} \widehat{\boldsymbol{\gamma}}_i^{(M)} \right\|_2^2$$

is the residual variance from the full model $M$ for asset $i$, and $\mathbf{p} = (K_1 + 1, \ldots, K_M + 1)^\top$ records the number of parameters in each candidate model, including the intercept.[4] Let

$$\widehat{\mathbf{w}}_i^{\mathrm{MAMF}} = \arg\min_{\mathbf{w} \in \mathcal{W}} \mathcal{C}_i(\mathbf{w})$$

denote the MAMF weight vector.

**Step 3.** Compute the corresponding MAMF estimators of alpha and beta by

$$\widehat{\alpha}_i^{\mathrm{MAMF}} = \sum_{m=1}^{M} \widehat{w}_{im}^{\mathrm{MAMF}} \, \widehat{\alpha}_i^{(m)} \qquad \text{and} \qquad \widehat{\boldsymbol{\beta}}_i^{\mathrm{MAMF}} = \sum_{m=1}^{M} \widehat{w}_{im}^{\mathrm{MAMF}} \, \widehat{\boldsymbol{\beta}}_i^{(m),*}, \tag{8}$$

where $\widehat{w}_{im}^{\mathrm{MAMF}}$ is the $m$th element of $\widehat{\mathbf{w}}_i^{\mathrm{MAMF}}$.

operator $\mathbf{P}(\mathbf{w}_i) = \sum_m w_{im}\mathbf{P}^{(m)}$ blends these smoothers. The Mallows objective (7) tilts the weights toward specifications that explain the data without unnecessary complexity: if one or more models are close to the truth, the weights concentrate on them; if all are imperfect, averaging balances bias and variance. Zero-padding the loadings into the common $K$-dimensional factor space preserves the usual interpretation of the averaged alpha and betas.

Second, (7) is a convex quadratic program on the unit simplex, with a global minimum that is unique whenever the candidates' fitted-value vectors are not collinear. Under nesting, $M \leq K + 1$, so the program remains small even for moderately large factor sets. Relaxing nesting via the augmentation device in Section 2.1 can make $M$ grow combinatorially with $K$, but in practice $\mathcal{M}$ is restricted to a curated set of economically motivated specifications. Each additional candidate adds one OLS regression and an update to the $M \times M$ inner-product matrix; the QP solver itself is negligible for the model counts considered here.

### 2.3.2 Simulation-Based Inference for the MAMF Estimators

Beyond point estimation, it is also crucial to assess the statistical significance of alpha and beta. To formalize the inferential problem, we posit a generic factor model that describes the true relationship between asset $i$'s excess return and the factor returns, which will be discussed in greater detail in Section 2.4. Let the associated true parameter vector be

$$\boldsymbol{\gamma}_i = (\alpha_i, \boldsymbol{\beta}_i^\top)^\top \in \mathbb{R}^{K+1}.$$

For any factor that does not affect the excess return, the corresponding element of $\boldsymbol{\beta}_i$ is equal to zero. Based on the MAMF estimator $\widehat{\boldsymbol{\gamma}}_i^{\text{MAMF}} = (\widehat{\alpha}_i^{\text{MAMF}}, \widehat{\boldsymbol{\beta}}_i^{\text{MAMF}\top})^\top$ introduced in Section 2.3.1, we now develop an inference procedure for $\boldsymbol{\gamma}_i$.

As shown in Theorem A in Section 2.4, the MAMF estimator $\widehat{\boldsymbol{\gamma}}_i^{\text{MAMF}}$ has a nonstandard asymptotic distribution that depends on the long-run covariance structure of the factor

innovations. Because this limiting distribution is non-pivotal, it cannot be directly used to construct standard test statistics or confidence intervals. Following Zhang and Liu (2019), we therefore adopt a simulation-based approach that approximates the limiting distribution of the MAMF estimator to arbitrary accuracy. The key idea is to replace the unknown nuisance quantities in the limiting distribution with consistent estimators and then simulate from the resulting approximate distribution of $\widehat{\boldsymbol{\gamma}}_i^{\mathrm{MAMF}}$. These simulated draws are used to obtain critical values and to construct confidence intervals for $\boldsymbol{\gamma}_i$.

Before describing the algorithm, we introduce some additional notation. Recall that model $M$ contains all $K$ factors by construction. Let

$$\widehat{\sigma}_i^2 = \big[T - (K+1)\big]^{-1} \left\| \mathbf{R}_i - \mathbf{X}^{(M)} \widehat{\boldsymbol{\gamma}}_i^{(M)} \right\|_2^2,$$

where $\mathbf{X}^{(M)}$ is the regressor matrix for the full model $M$. Let

$$\widehat{\mathbf{Q}} = T^{-1} \sum_{t=1}^{T} X_t X_t^\top$$

be a consistent estimator of $\mathbf{Q} = \mathbb{E}[X_t X_t^\top]$, and let $\widehat{\boldsymbol{\Omega}}_\infty$ denote a heteroskedasticity- and autocorrelation-consistent (HAC) estimator of the long-run covariance matrix

$$\boldsymbol{\Omega}_\infty = \mathrm{Var}\left( T^{-1/2} \sum_{t=1}^{T} X_t e_t \right),$$

such as the Newey–West estimator (Newey and West, 2023). Moreover, let $\boldsymbol{\Pi}_m \in \mathbb{R}^{(K_m+1)\times(K+1)}$ be the selection matrix such that $X_t^{(m)} = \boldsymbol{\Pi}_m X_t$ and $\mathbf{X}^{(m)} = \mathbf{X}^{(M)} \boldsymbol{\Pi}_m^\top$. Algorithm 2 describes the simulation-based construction of confidence intervals for alpha and betas.

Algorithm 2 implements a plug-in Monte Carlo approximation to the nonstandard limit in Theorem A in Appendix A.1. The limit distribution of the MAMF estimator is non-standard because it can be viewed as a weighted average of correlated normal distributions, where each normal distribution corresponds to a just- or over-specified factor model esti-

**Algorithm 2** Simulation-Based Confidence Interval for $\gamma = (\alpha, \beta)$

**Step 1.** For each asset $i$, estimate the full factor model $M$ and obtain the consistent estimators $\widehat{\sigma}_i^2$, $\widehat{\mathbf{Q}}$, and $\widehat{\mathbf{\Omega}}_\infty$ as defined above. For notational brevity, we suppress the asset index $i$ in what follows.

**Step 2.** Solve for the penalized weight vector $\breve{\mathbf{w}} \in \mathcal{W}$:

$$\breve{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \breve{\mathcal{C}}(\mathbf{w}), \qquad \breve{\mathcal{C}}(\mathbf{w}) \;=\; \big\|\big(\mathbf{I}_T - \mathbf{P}(\mathbf{w})\big)\mathbf{R}_i\big\|_2^2 \;+\; 2\,\phi_T\,\widehat{\sigma}_i^2\,\mathbf{w}^\top\mathbf{p},$$

where $\phi_T$ is a tuning parameter satisfying $\phi_T \to \infty$ and $\phi_T/\sqrt{T} \to 0$. In practice, we set $\phi_T = \log(T)$. Define the estimated true model index as $\widehat{M}_0 + 1 = \arg\max_{m=1,\ldots,M} \breve{w}_m$, i.e., the model receiving the largest weight in $\breve{\mathbf{w}}$.

**Step 3.** For $s = 1, \ldots, M - \widehat{M}_0$, compute

$$\widehat{\mathbf{Q}}_{\widehat{M}_0+s} \;=\; \mathbf{\Pi}_{\widehat{M}_0+s}\,\widehat{\mathbf{Q}}\,\mathbf{\Pi}_{\widehat{M}_0+s}^\top, \qquad \widehat{\mathbf{V}}_{\widehat{M}_0+s} \;=\; \mathbf{\Pi}_{\widehat{M}_0+s}^\top\,\widehat{\mathbf{Q}}_{\widehat{M}_0+s}^{-1}\,\mathbf{\Pi}_{\widehat{M}_0+s}. \tag{9}$$

Then, for each simulation draw $r = 1, \ldots, R$, construct the matrix $\widehat{\mathbf{\Gamma}}^{(r)} \in \mathbb{R}^{(M-\widehat{M}_0)\times(M-\widehat{M}_0)}$ with $(s, j)$-th entry

$$\widehat{\mathbf{\Gamma}}_{sj}^{(r)} \;=\; 2\,\widehat{\sigma}_i^2\,(K_{\widehat{M}_0+s} + 1) \;-\; Z^{(r)\top}\widehat{\mathbf{V}}_{\max\{s,j\}}\,Z^{(r)}, \qquad 1 \le s, j \le M - \widehat{M}_0. \tag{10}$$

Next, compute the simulated counterpart of the limiting quantity in Theorem A:

$$\Lambda^{(r)}(\widehat{M}_0 + 1) \;=\; \sum_{s=1}^{M-\widehat{M}_0} \widehat{\lambda}_s^{(r)}(\widehat{M}_0 + 1)\;\widehat{\mathbf{V}}_{\widehat{M}_0+s}\,Z^{(r)}, \tag{11}$$

where the simplex-constrained weights are given by

$$\widehat{\lambda}^{(r)}(\widehat{M}_0{+}1) \;=\; \arg\min_{\boldsymbol{\lambda}\in\mathcal{L}(\widehat{M}_0+1)} \boldsymbol{\lambda}^\top\widehat{\mathbf{\Gamma}}^{(r)}\,\boldsymbol{\lambda}, \qquad \mathcal{L}(\widehat{M}_0{+}1) \;=\; \Big\{\boldsymbol{\lambda}\in[0,1]^{M-\widehat{M}_0} : \textstyle\sum_{s=1}^{M-\widehat{M}_0}\lambda_s = 1\Big\}.$$

Denote by $\Lambda_j^{(r)}(\widehat{M}_0 + 1)$ the $j$-th element of $\Lambda^{(r)}(\widehat{M}_0 + 1)$.

**Step 4.** From the $R$ simulation draws, obtain the empirical quantiles $q_j(\tau/2)$ and $q_j(1 - \tau/2)$ of $\big\{\Lambda_j^{(r)}(\widehat{M}_0 + 1)\big\}_{r=1}^R$. The $(1 - \tau)$ simulation-based confidence interval for $\gamma_j$ is

$$\mathrm{CI}_j \;=\; \big[\widehat{\gamma}_j\big(\widehat{\mathbf{w}}^{\mathrm{MAMF}}\big) - T^{-1/2}\,q_j(1-\tau/2),\;\; \widehat{\gamma}_j\big(\widehat{\mathbf{w}}^{\mathrm{MAMF}}\big) - T^{-1/2}\,q_j(\tau/2)\big]\,,$$

where $\gamma_1$ corresponds to the intercept (alpha) and $\gamma_j$ for $j \ge 2$ correspond to the factor loadings (betas).

mated via OLS.[5] Because the model weights are estimated, they correlate with the limiting normal vector associated with each factor model, so the weights and the normal vector must be simulated jointly. Two challenges therefore arise in simulating the limit distribution: (1) how to distinguish under-specified models ($m = 1, \ldots, M_0$) from just- and over-specified models ($m = M_0 + 1, \ldots, M$), since under-specified models do not participate in the limit distribution; and (2) how to generate the correlation between the MAMF weights and the limiting normal vector, given that both are jointly determined by the same sample realizations and cannot be simulated independently.

Step 2 addresses the first challenge. Because the true index $M_0 + 1$ is unknown, the penalized criterion $\breve{\mathcal{C}}(\mathbf{w})$ is constructed to mimic the asymptotic behavior of the Mallows objective: it heavily penalizes under-specified models, which suffer from systematic misspecification error, and shrinks the weights on over-specified models, which add noise without improving fit. As a result, the minimizer $\breve{\mathbf{w}}$ concentrates its mass near $M_0 + 1$ at an appropriate rate, so the estimated index $\widehat{M}_0 + 1$ equals $M_0 + 1$ with probability approaching one.

Step 3 addresses the second challenge by coupling the simulated weights and the Gaussian vector within each replication. The first-order behavior of the MAMF estimator is driven by the long-run covariance of the score process $X_t e_t$, conditional on the estimated true model index $\widehat{M}_0 + 1$. Rather than resampling the original time series, we approximate this source of sampling variation by drawing Gaussian vectors $Z^{(r)} \sim N(0, \widehat{\mathbf{\Omega}})$, where $\widehat{\mathbf{\Omega}}$ is a HAC estimate of the long-run covariance matrix. The matrices $\widehat{\mathbf{Q}}_{\widehat{M}_0+s}$ and $\widehat{\mathbf{V}}_{\widehat{M}_0+s}$ summarize the regression geometry of the just- and over-specified models, and the simulated quadratic form $\widehat{\mathbf{\Gamma}}^{(r)}$ encodes, for a given draw $Z^{(r)}$, the Mallows trade-off between fit and complexity. Minimizing $\boldsymbol{\lambda}^\top \widehat{\mathbf{\Gamma}}^{(r)} \boldsymbol{\lambda}$ over $\mathcal{L}(\widehat{M}_0+1)$ produces a simulated analogue of the population-optimal weights,[6] and the resulting quantities $\Lambda^{(r)}(\widehat{M}_0 + 1)$ behave like draws from the limiting distribution that would arise if $\widehat{M}_0 + 1$ were the true model index. Because $\widehat{\mathbf{\Gamma}}^{(r)}$ and the

[5] We leave the details for deriving the limit distribution of the MAMF estimator to Appendix A.1.

[6] The padded vector $(0, \ldots, 0, \widehat{\boldsymbol{\lambda}}(\widehat{M}_0 + 1))$ mimics the distribution of the estimated weights $\widehat{\mathbf{w}}^{\mathrm{MAMF}}$.

solution $\widehat{\boldsymbol{\lambda}}^{(r)}$ are computed from the same $Z^{(r)}$, the coupling reproduces the joint dependence between weights and score in the limiting distribution.

Step 4 uses the empirical quantiles of these draws as critical values for the scaled MAMF estimator, delivering confidence intervals that account simultaneously for model uncertainty and serial dependence in returns.

## 2.4 Theoretical Properties of the MAMF Estimators

In this section, we present the theoretical properties of the proposed MAMF estimators of alpha and factor loadings in the model-averaging multi-factor asset-pricing framework. Specifically, we study their optimality and limiting distribution as the time dimension $T$ grows. For notational simplicity, throughout this subsection we fix an asset $i \in \mathcal{N}$ and suppress the index $i$. We begin by stating the regularity conditions on the candidate models and the data-generating process under which the theoretical results are derived.

**Assumption 1** *The candidate factor models in $\mathcal{M}$ satisfy the following conditions: (i) the $M$ candidate factor models exhibit a nesting structure such that $\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \cdots \subseteq \mathcal{K}_M$; and (ii) there exists a candidate model $M_0 + 1 \in \mathcal{M}$ that includes exactly $K_{M_0+1}$ tradable factors with non-zero factor loadings.*

**Assumption 2** *The factor returns and the error term satisfy the following regularity conditions: (i) the process $\{(X_t, e_t)\}_{t=1}^{T}$ is weakly stationary; (ii) it is $\alpha$-mixing with $\sum_{\ell=1}^{\infty} \alpha(\ell)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$; and (iii) $\mathbb{E}\|X_t\|^{2+\delta} < \infty$.*

**Assumption 3** *The error term admits the linear representation $e_t = \epsilon_t + \sum_{k=1}^{\infty} \phi_k \epsilon_{t-k}$, where: (i) $\mathbb{E}[\epsilon_t \mid \mathcal{F}_{t-1}] = 0$; (ii) $\mathbb{E}[\epsilon_t^2 \mid \mathcal{F}_{t-1}] = \sigma_\epsilon^2$; (iii) there exists a positive integer $N$ and a constant $S > 4N$ such that*

$$\sup_{-\infty<t<\infty} \mathbb{E}\big[|\epsilon_t|^S \mid \mathcal{F}_{t-1}\big] \leq C_S$$

*for some finite constant $C_S > 0$; (iv) the spectral density of $\{\epsilon_t\}$, denoted by $f_e(\lambda)$, is nonzero for all $\lambda \in (-\pi, \pi]$; and (v) the coefficients are absolutely summable, $\sum_{j=1}^{\infty} |\phi_j| < \infty$.*

Assumption 1(i) imposes a nested structure on the candidate model space: each specification adds factors to the previous one, so that $\mathcal{K}_1 \subseteq \cdots \subseteq \mathcal{K}_M$. This requirement mirrors the historical development of the factor literature, in which the CAPM is nested in the Fama–French three-factor model, which in turn is nested in the Fama–French five-factor model, and empirical studies often work with an "all-factor" model that nests its submodels. Hence, the nesting requirement is both empirically realistic and conceptually natural in multi-factor asset-pricing applications.[7]

As noted in Section 2.1, this nesting assumption is without loss of generality for linear factor models: any collection of non-nested specifications can be embedded into a nested hierarchy by augmenting the factor space with the union of all factor sets. The theoretical results below are stated under nesting for notational convenience, but they extend directly to the augmented setting. We note, however, that augmentation increases the dimension of the full model $M$, which raises the effective number of parameters in the largest specification. When the number of candidate non-nested models is large, the augmented factor space can become high-dimensional, and the researcher may wish to pre-screen factors or group specifications into families before applying MAMF. We discuss the computational implications in more detail in Section 2.3.1.

Assumption 1(ii) complements this by anchoring the underlying true model parameters. It requires that, after reindexing if necessary, there exists a candidate model $M_0 + 1$ whose factor set coincides with the set of factors that have nonzero loadings in the true pricing relation. In other words, model $M_0 + 1$ is correctly specified in the sense that it includes all and only the priced factors. Together with the nesting in part (i), this implies that models $1, \ldots, M_0$ are under-specified because they omit at least one priced factor, while models $m > M_0 + 1$ contain all priced factors but may include redundant ones. This "true-model-in-

[7] We illustrate this nesting structure in more detail in the empirical applications in Section 4.

the-union" condition is standard in the model selection and model averaging literature and underlies the asymptotic validity of our inference procedure.

Assumption 1 requires that factors be tradable, i.e., that they correspond to returns on investable portfolios. This condition serves two purposes. First, it ensures that the intercept $\alpha_i$ has a clear economic interpretation as an investable abnormal return: it measures the risk-adjusted payoff from holding asset $i$ after hedging out exposures to the benchmark factors. If factors were non-tradable (e.g., macroeconomic growth rates), the intercept would not correspond to a feasible trading strategy, and statements about "significant alpha" would lack a direct portfolio interpretation. Second, tradability guarantees that the factor returns $\{F_{kt}\}$ are observed at the same frequency and in the same units as asset returns, which simplifies the regression framework and avoids the errors-in-variables complications that arise with estimated or proxy factors. The assumption is satisfied by all candidate models considered in our empirical application (Section 4), which use long-short portfolio returns as factors.

Assumption 2 imposes mild regularity on the joint dynamics of the regressors and disturbances. Specifically, weak stationarity in part (i) is a baseline requirement satisfied by excess returns and traded-factor returns over any fixed sample window. The $\alpha$-mixing condition in part (ii) is among the weakest measures of temporal dependence in the time-series literature (Bradley, 2005), accommodating conditional heteroskedasticity, GARCH-type volatility clustering, and other forms of serial dependence routinely observed in financial data, and is substantially weaker than the independence or martingale-difference assumptions imposed in many asset-pricing studies. The finite $(2+\delta)$-th moment condition in part (iii) is equally unrestrictive, requiring only slightly more than finite variance. Collectively, these conditions ensure that the sample second-moment matrix $\widehat{\mathbf{Q}}$ and the HAC estimator $\widehat{\mathbf{\Omega}}_\infty$ are consistent (Newey and West, 2023; Andrews, 1991), which are the two key inputs to the simulation-based inference in Algorithm 2.

Assumption 3 refines the structure of the regression error by imposing a linear represen-

tation $e_t = \epsilon_t + \sum_{k=1}^{\infty} \phi_k \epsilon_{t-k}$ driven by martingale-difference innovations $\epsilon_t$ with bounded higher moments and absolutely summable coefficients $\{\phi_k\}$. Together with the requirement that the spectral density $f_e(\lambda)$ be nonzero at all frequencies, this guarantees a well-behaved short-memory error process with a finite, invertible long-run covariance matrix that can be consistently estimated by HAC methods. These assumptions are widely used in the time-series and long-run variance estimation literature and appear empirically reasonable in asset-pricing applications.

### 2.4.1 Asymptotic Optimality

Based on Assumptions 1–3, we first study the optimality properties of the MAMF estimator introduced in Section 2.3.1. For a given weight vector $\mathbf{w} = (w_1, \ldots, w_M)^\top \in \mathbb{R}^M$, define the sample prediction loss, conditional on the regressor matrix $\mathbf{X}$, as

$$L_T(\mathbf{w}) \;=\; \big(\mathbf{X}\widehat{\boldsymbol{\gamma}}(\mathbf{w}) - \mathbf{X}\boldsymbol{\gamma}\big)'\big(\mathbf{X}\widehat{\boldsymbol{\gamma}}(\mathbf{w}) - \mathbf{X}\boldsymbol{\gamma}\big), \tag{12}$$

where

$$\widehat{\boldsymbol{\gamma}}(\mathbf{w}) = \begin{pmatrix} \sum_{m=1}^{M} w_m \, \widehat{\alpha}^{(m)} \\ \sum_{m=1}^{M} w_m \, \widehat{\boldsymbol{\beta}}^{(m),*} \end{pmatrix} \in \mathbb{R}^{K+1}$$

is the model-averaged coefficient vector obtained by zero-padding the factor loadings to the full $K$-dimensional factor space as in (2), $\widehat{\alpha}^{(m)}$ and $\widehat{\boldsymbol{\beta}}^{(m),*}$ are the ordinary least squares estimators from candidate model $m$, and $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}^\top)^\top$ denotes the true parameter vector defined in Section 2.3.2. The loss $L_T(\mathbf{w})$ measures the sample squared distance between the fitted conditional mean $\mathbf{X}\widehat{\boldsymbol{\gamma}}(\mathbf{w})$ and the oracle mean $\mathbf{X}\boldsymbol{\gamma}$.

To describe the comparison class for the MAMF estimator, let $\delta_T > 0$ be a sequence with $\delta_T \to 0$ and define

$$\mathcal{H}_T = \bigcup_{l=1}^{M} \left\{ \mathbf{w} \in \mathcal{W} : \delta_T \le w_m \mathbb{I}_{\{w_m \neq 0\}} \le 1, \; \sum_{m=1}^{M} \mathbb{I}_{\{w_m \neq 0\}} = l \right\},$$

where $\mathcal{W} = \{\mathbf{w} \in [0,1]^M : \sum_{m=1}^{M} w_m = 1\}$ is the unit simplex. The set $\mathcal{H}_T$ collects weight vectors that place nonnegligible mass on a finite number of candidate models and is the usual benchmark class in the Mallows model-averaging literature.

**Theorem 1** *Under Assumptions 1(i), 2, and 3, the MAMF weight vector* $\widehat{\mathbf{w}}^{\mathrm{MAMF}}$ *satisfies*

$$\frac{L_T\big(\widehat{\mathbf{w}}^{\mathrm{MAMF}}\big)}{\inf\limits_{\mathbf{w}\in\mathcal{H}_T} L_T(\mathbf{w})} \xrightarrow{p} 1,$$

*as* $T \to \infty$*, where* $L_T(\mathbf{w})$ *is defined in* (12)*.*

Theorem 1 shows that the proposed MAMF estimator is asymptotically optimal in the sense that its prediction loss $L_T(\widehat{\mathbf{w}}^{\mathrm{MAMF}})$ is asymptotically indistinguishable from the smallest achievable loss over the class $\mathcal{H}_T$. Equivalently, there is no first-order loss from estimating the weights $\mathbf{w}$ rather than using the infeasible weight vector that minimizes (12) ex post. This ratio optimality is an oracle property for model averaging: in large samples, the MAMF estimator behaves as if it knew in advance which convex combination of candidate models delivers the best mean-squared fit to the conditional mean of returns.

In the multi-factor asset-pricing context, this result implies that the MAMF alphas and betas are, asymptotically, as informative as those obtained from the best convex combination of benchmark factor models for explaining expected returns. If one specification is close to, or coincides with, the true pricing relation, the optimal weights concentrate on it and the MAMF estimator matches its performance; if all models are misspecified, the optimal combination balances their biases and variances. Thus, the model-averaged alphas and betas produced by the MAMF procedure are, in a rigorous sense, as informative for pricing tests and performance evaluation as those from the best factor combination in hindsight, while avoiding the need to commit ex ante to a single multi-factor specification.

### 2.4.2 Validity of the Inference Procedure

We now justify the simulation-based inference procedure described above. The derivation of the validity of the inference method requires first deriving the asymptotic distribution for the MAMF estimator, which we leave to Appendix A.1 to avoid heavy math notations. Recall that Step 3 of Algorithm 2 introduces the penalized objective

$$\breve{\mathcal{C}}(\mathbf{w}) = \big\| \big(\mathbf{I}_T - \mathbf{P}(\mathbf{w})\big)\mathbf{R}_i \big\|_2^2 + 2\phi_T \, \widehat{\sigma}_i^2 \, \mathbf{w}^\top \mathbf{p},$$

where $\phi_T$ is a tuning parameter satisfying $\phi_T \to \infty$ and $\phi_T/\sqrt{T} \to 0$ as $T \to \infty$. Let $\breve{\mathbf{w}} = (\breve{w}_1, \ldots, \breve{w}_M)^\top$ denote the minimizer of $\breve{\mathcal{C}}(\mathbf{w})$ over the simplex $\mathcal{W}$, and recall that $M_0$ indexes the largest under-specified model so that models $1, \ldots, M_0$ are under-specified and models $M_0+1, \ldots, M$ are just- or over-specified. We now show that using the estimated true model $\widehat{M}_0+1$ to simulate the limit distribution will yield a valid inference for the parameters.

**Theorem 2 (Validity of the simulation-based confidence intervals)** *Suppose Assumptions 1–3 hold, the long-run covariance matrices and second-moment matrices are consistently estimated, and the tuning parameter $\phi_T$ satisfies $\phi_T \to \infty$ and $\phi_T/\sqrt{T} \to 0$. Then the penalized estimator $\breve{\mathbf{w}}$ satisfies $\breve{w}_m = O_p(\phi_T/T)$ for each under-specified model $m \le M_0$ and $\breve{w}_m = O_p(\phi_T^{-1})$ for each over-specified model $m > M_0$ that includes all relevant factors. As a result, for every coordinate $j$ and simulation replication $r$, we have*

$$\big|\Lambda_j^{(r)}(\widehat{M}_0 + 1) - \Lambda_j^{(r)}(M_0 + 1)\big| \xrightarrow{p} 0,$$

*where $\Lambda_j^{(r)}(M_0+1)$ is the simulated limiting quantity associated with the (unknown) true model index $M_0 + 1$. Consequently, the confidence interval constructed in Step 4 of Algorithm 2 satisfies*

$$\Pr\big\{\gamma_j \in \mathrm{CI}_j\big\} \;\longrightarrow\; 1 - \tau \quad \textit{a.s.} \quad T, R \to \infty,$$

*so that the simulation-based procedure delivers asymptotically valid $(1-\tau)$-level inference for*

*each component* $\gamma_j$ *of* $\boldsymbol{\gamma}$.

Theorem 2 clarifies how the penalized criterion $\breve{\mathcal{C}}(\mathbf{w})$ links the nonstandard limiting distribution in Theorem A to the practical simulation algorithm. By driving the weights on under-specified models to zero at rate $\phi_T/T$ and shrinking the weights on over-specified models at rate $\phi_T^{-1}$, the procedure implicitly concentrates on the model index $M_0+1$ that best captures the true factor structure, even though $M_0$ is unknown. By choosing the index that maximizes the weights, we estimate the true model $M_0+1$ with probability approaching 1. The simulated mixtures $\Lambda_j^{(r)}(\widehat{M_0}+1)$ therefore approximate the correct limiting law of the MAMF estimator, and the resulting confidence intervals for alphas and betas achieve asymptotically correct coverage in the presence of both model uncertainty and serial correlation in returns. This guarantees that hypothesis tests about alphas and betas constructed from the MAMF estimator have asymptotically correct size, making the proposed procedure reliable for empirical multi-factor asset-pricing applications.

# 3 Monte Carlo Simulations

## 3.1 Setting

We study the finite-sample performance of our model-averaging procedure in a simulation design that is anchored in observed factor dynamics. The data-generating process uses realized monthly factor returns from the Global Factor Data website and Kenneth French's Factor Data Library.[8] The sample runs from January 1985 to December 2024, so each factor series has $T = 480$ observations. Since multi-factor models are primarily used to assess risk-adjusted performance, we focus on the accuracy of alpha estimation and the reliability of inference on alpha.

[8]The Global Factor Data website provides monthly returns for 153 factors grouped into 13 themes and has been used in Jensen et al. (2023), while Kenneth French's Factor Data Library provides monthly returns for the Fama–French five factors and the momentum factor.

Our candidate set $\mathcal{M}$ contains $M = 7$ nested factor-pricing models built from eleven widely used factors, ranging from the single-factor CAPM to an eleven-factor specification. The eleven factors are the excess market return (MKT), the size factor (SMB), the value factor (HML), the momentum factor (MOM), the profitability factor (RMW), the investment factor (CMA), the liquidity factor (LIQ), the management mispricing factor (MGMT), the performance mispricing factor (PERF), the quality-minus-junk factor (QMJ), and the betting-against-beta factor (BAB). We start with the CAPM, which includes MKT only. We then add SMB and HML to obtain the Fama–French three-factor model (FF3), add MOM to obtain the Carhart four-factor model (Carhart4), and add RMW and CMA to obtain the Fama–French six-factor model (FF6). Next we include LIQ to obtain a seven-factor model (7F), add MGMT and PERF to obtain a nine-factor model (9F), and finally add QMJ and BAB to obtain the eleven-factor model (Full11). This sequence matches how benchmark models are expanded in practice and provides a natural environment for studying model uncertainty.

For each replication, excess returns follow a linear factor model

$$R_{it} = \alpha_i + \mathbf{F}_t^\top \boldsymbol{\beta}_i + e_{it}, \qquad 1 \le t \le T, \tag{13}$$

where $\mathbf{F}_t$ collects factor returns and $e_{it}$ is an idiosyncratic disturbance. We take FF6 as the true pricing relation and therefore as the oracle model in this simulation. The simulated asset loads on MKT, SMB, HML, RMW, CMA, and MOM, and the loadings on LIQ, MGMT, PERF, QMJ, and BAB are set to zero.

Factor loadings are calibrated from the data rather than chosen ad hoc. For each testing-factor portfolio in Jensen et al. (2023) that is not among the eleven factors above, we treat the portfolio as a test asset and estimate FF6 betas and the residual standard deviation from an FF6 regression. Let $J$ denote the number of such portfolios. We construct two exposure designs. In the high-beta design, the population beta vector is set to the 90% quantile of

the estimated betas across testing portfolios. In the low-beta design, it is set to the 10% quantile. We set the idiosyncratic volatility to the cross-sectional average of the estimated residual standard deviations, $\sigma_i = J^{-1} \sum_{j=1}^{J} \widehat{\sigma}_j$, expressed in percent per month.

We consider three monthly alpha scenarios, $\alpha_i \in \{2\%, 1\%, 0\%\}$. The 2% case is a stress test—annualizing to roughly 24%, it exceeds almost all documented anomaly alphas and probes model averaging under large abnormal returns. The 1% case represents a moderately strong anomaly, and the 0% case isolates size control under the null.

To preserve time-series dependence in factor returns, we generate factor paths via block bootstrap resampling with replacement. In each replication, we resample blocks of 12 consecutive months from the observed factor series and concatenate them to length $T$. Conditional on the bootstrapped factor path, we draw $e_{it}$ independently from $N(0, \sigma_i^2)$ and form $\{R_{it}\}_{t=1}^{T}$ using (13). We report results for $T = 480$, corresponding to a 40-year sample length typical of empirical multi-factor asset-pricing studies.

We compare our model-averaging estimator constructed over $\mathcal{M}$ with single-model OLS estimators under each candidate specification. This comparison mirrors common empirical reporting under benchmark models and isolates the incremental value of averaging across plausible specifications.

We evaluate performance for alpha using both accuracy and inference diagnostics. All accuracy and inference measures below are expressed in percent per month, consistent with the scaling of the simulated returns. For method $m$ and replication $b$, let $\widehat{\alpha}_b^{(m)}$ denote the alpha estimate. Mean absolute error is

$$\mathrm{MAE}^{(m)}(\alpha) = \frac{1}{B} \sum_{b=1}^{B} \left| \widehat{\alpha}_b^{(m)} - \alpha_i \right|. \tag{14}$$

Root mean squared error is

$$\mathrm{RMSE}^{(m)}(\alpha) = \left( \frac{1}{B} \sum_{b=1}^{B} \left( \widehat{\alpha}_b^{(m)} - \alpha_i \right)^2 \right)^{1/2}. \tag{15}$$

We also report the standard deviation of the alpha estimator across replications,

$$\mathrm{SD}^{(m)}(\alpha) = \left( \frac{1}{B-1} \sum_{b=1}^{B} \left( \widehat{\alpha}_b^{(m)} - \overline{\widehat{\alpha}}^{(m)} \right)^2 \right)^{1/2}, \qquad \overline{\widehat{\alpha}}^{(m)} = \frac{1}{B} \sum_{b=1}^{B} \widehat{\alpha}_b^{(m)}. \tag{16}$$

Inference is evaluated using nominal 95% confidence intervals for $\alpha_i$. Let $\mathrm{CI}_b^{(m)}(\alpha) = [L_b^{(m)}(\alpha), U_b^{(m)}(\alpha)]$ denote the interval produced by method $m$ in replication $b$. Coverage is

$$\mathrm{Coverage}^{(m)}(\alpha) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\Big\{ L_b^{(m)}(\alpha) \le \alpha_i \le U_b^{(m)}(\alpha) \Big\}, \tag{17}$$

and average interval length is

$$\mathrm{AvgLen}^{(m)}(\alpha) = \frac{1}{B} \sum_{b=1}^{B} \Big( U_b^{(m)}(\alpha) - L_b^{(m)}(\alpha) \Big). \tag{18}$$

We set $\tau = 0.05$ and target 95% confidence intervals throughout.

## 3.2 Simulation Results

Tables 1 and 2 report the simulation results. The model-averaging estimator delivers stable alpha estimates across all designs: its MAE and RMSE remain close to the oracle FF6 benchmark in both the high-beta and low-beta cases, and the accuracy measures change little as $\alpha_i$ varies across $\{2\%, 1\%, 0\%\}$. This stability is important because researchers in applications typically face uncertainty about which benchmark specification to condition on. The omitted-factor bias of parsimonious benchmarks is economically large: in the high-beta, $\alpha = 2\%$ design, the CAPM MAE of 0.77% per month exceeds the oracle FF6 MAE by more than an order of magnitude, and the Carhart4 MAE of 0.35% per month remains five times that of FF6. MAMF largely eliminates this sensitivity while keeping dispersion comparable to the best-performing specifications.

The confidence intervals from the model-averaging procedure are similarly well behaved.

Table 1: Simulation performance for alpha estimation

| | | High Beta Case | | | Low Beta Case | | |
|---|---|---|---|---|---|---|---|
| | Method | MAE (%) | RMSE (%) | SD (%) | MAE (%) | RMSE (%) | SD (%) |
| $\alpha = 2\%$ | MAMF | 0.0676 | 0.0850 | 0.0849 | 0.0679 | 0.0853 | 0.0851 |
| | CAPM | 0.7683 | 0.7955 | 0.2059 | 0.3817 | 0.4029 | 0.1293 |
| | FF3 | 0.5114 | 0.5273 | 0.1284 | 0.2363 | 0.2515 | 0.0866 |
| | Carhart4 | 0.3496 | 0.3682 | 0.1159 | 0.1067 | 0.1274 | 0.0826 |
| | FF6 | 0.0643 | 0.0808 | 0.0808 | 0.0644 | 0.0809 | 0.0809 |
| | 7F | 0.0714 | 0.0898 | 0.0898 | 0.0714 | 0.0897 | 0.0897 |
| | 9F | 0.0734 | 0.0923 | 0.0923 | 0.0734 | 0.0922 | 0.0922 |
| | Full11 | 0.0749 | 0.0940 | 0.0940 | 0.0748 | 0.0939 | 0.0939 |
| $\alpha = 1\%$ | MAMF | 0.0663 | 0.0832 | 0.0832 | 0.0665 | 0.0836 | 0.0832 |
| | CAPM | 0.7639 | 0.7910 | 0.2052 | 0.3816 | 0.4025 | 0.1282 |
| | FF3 | 0.5097 | 0.5261 | 0.1303 | 0.2378 | 0.2525 | 0.0850 |
| | Carhart4 | 0.3471 | 0.3665 | 0.1179 | 0.1074 | 0.1277 | 0.0814 |
| | FF6 | 0.0637 | 0.0800 | 0.0800 | 0.0635 | 0.0799 | 0.0799 |
| | 7F | 0.0701 | 0.0879 | 0.0879 | 0.0700 | 0.0878 | 0.0878 |
| | 9F | 0.0719 | 0.0902 | 0.0902 | 0.0718 | 0.0900 | 0.0901 |
| | Full11 | 0.0729 | 0.0915 | 0.0915 | 0.0727 | 0.0914 | 0.0914 |
| $\alpha = 0\%$ | MAMF | 0.0680 | 0.0854 | 0.0853 | 0.0684 | 0.0859 | 0.0856 |
| | CAPM | 0.7684 | 0.7959 | 0.2074 | 0.3820 | 0.4030 | 0.1288 |
| | FF3 | 0.5122 | 0.5283 | 0.1294 | 0.2373 | 0.2525 | 0.0870 |
| | Carhart4 | 0.3492 | 0.3684 | 0.1173 | 0.1070 | 0.1277 | 0.0825 |
| | FF6 | 0.0648 | 0.0812 | 0.0812 | 0.0649 | 0.0813 | 0.0813 |
| | 7F | 0.0719 | 0.0901 | 0.0901 | 0.0720 | 0.0903 | 0.0903 |
| | 9F | 0.0737 | 0.0924 | 0.0924 | 0.0739 | 0.0926 | 0.0926 |
| | Full11 | 0.0753 | 0.0944 | 0.0944 | 0.0755 | 0.0945 | 0.0946 |

*Notes.* The table reports simulation-based accuracy measures for estimating monthly alpha, expressed in percent per month. MAE is the mean absolute error, RMSE is the root mean squared error, and SD is the standard deviation of the estimated alpha across replications. Results are reported separately for the high-beta and low-beta designs. MAMF denotes the proposed model-averaging estimator. CAPM, FF3, Carhart4, FF6, 7F, 9F, and Full11 denote OLS estimates from the corresponding benchmark factor specifications. Each entry is computed from $B = 10{,}000$ simulation replications with $T = 480$.

Table 2: Simulation inference performance for alpha

| | | High Beta Case | | Low Beta Case | |
|---|---|---|---|---|---|
| | Method | Coverage (%) | AvgLen (%) | Coverage (%) | AvgLen (%) |
| $\alpha = 2\%$ | MAMF | 94.39 | 0.3294 | 94.26 | 0.3301 |
| | CAPM | 1.03 | 0.6954 | 10.98 | 0.4537 |
| | FF3 | 1.29 | 0.4813 | 21.01 | 0.3337 |
| | Carhart4 | 13.73 | 0.4505 | 76.89 | 0.3164 |
| | FF6 | 94.21 | 0.3093 | 94.20 | 0.3093 |
| | 7F | 94.19 | 0.3422 | 94.23 | 0.3423 |
| | 9F | 94.24 | 0.3502 | 94.28 | 0.3502 |
| | Full11 | 94.05 | 0.3559 | 94.07 | 0.3559 |
| $\alpha = 1\%$ | MAMF | 94.80 | 0.3289 | 94.67 | 0.3296 |
| | CAPM | 0.87 | 0.6922 | 10.56 | 0.4523 |
| | FF3 | 1.42 | 0.4812 | 20.12 | 0.3334 |
| | Carhart4 | 14.46 | 0.4496 | 76.47 | 0.3161 |
| | FF6 | 94.35 | 0.3089 | 94.38 | 0.3089 |
| | 7F | 94.55 | 0.3422 | 94.56 | 0.3421 |
| | 9F | 94.60 | 0.3501 | 94.58 | 0.3501 |
| | Full11 | 94.71 | 0.3558 | 94.72 | 0.3558 |
| $\alpha = 0\%$ | MAMF | 94.52 | 0.3293 | 94.30 | 0.3300 |
| | CAPM | 0.95 | 0.6944 | 10.54 | 0.4541 |
| | FF3 | 1.20 | 0.4815 | 21.25 | 0.3337 |
| | Carhart4 | 13.67 | 0.4503 | 76.57 | 0.3164 |
| | FF6 | 94.21 | 0.3092 | 94.13 | 0.3092 |
| | 7F | 94.00 | 0.3421 | 93.97 | 0.3421 |
| | 9F | 93.91 | 0.3501 | 93.91 | 0.3501 |
| | Full11 | 93.80 | 0.3557 | 93.78 | 0.3557 |

*Notes.* The table reports simulation-based inference measures for monthly alpha. Coverage is the empirical coverage rate in percent for the nominal 95% confidence interval. AvgLen is the average confidence-interval length, expressed in percent per month. Results are reported separately for the high-beta and low-beta designs. MAMF denotes the proposed model-averaging procedure. CAPM, FF3, Carhart4, FF6, 7F, 9F, and Full11 denote confidence intervals constructed from the corresponding benchmark factor specifications. Each entry is computed from $B = 10{,}000$ simulation replications with $T = 480$.

Coverage rates are close to the nominal 95% level across both the high-beta and low-beta designs and remain stable across the three alpha scenarios. Average interval lengths are moderate: relative to the oracle FF6 intervals, the model-averaging intervals are slightly wider, reflecting the expected cost of accommodating model uncertainty, yet maintain near-nominal coverage. Relative to over-parameterized specifications that include irrelevant factors, the model-averaging intervals are shorter while achieving comparable coverage. In short, the proposed procedure produces confidence intervals that are sufficiently conservative to remain reliable when the correct specification is unknown, yet sufficiently tight to preserve statistical power.

The baseline design sets the loadings on the five excluded factors to exactly zero, which represents a sharp null. In practice, however, assets may have small but nonzero exposures to factors outside the true specification (Onatski, 2012; Anatolyev and Mikusheva, 2022; Bai and Ng, 2023). To assess sensitivity, we run a supplementary simulation in which the excluded factors receive small nonzero loadings, with magnitudes set to the 25th percentile of estimated absolute exposures across test portfolios and signs set to the cross-sectional median. Appendix C reports the full results. The main conclusions carry over: MAMF continues to achieve near-oracle accuracy and nominal coverage, confirming that the method is not sensitive to mild departures from the sharp zero-loading null.

Overall, the simulation supports the main use case of the method. When the analyst does not want to treat a single benchmark model as given, model averaging provides alpha confidence intervals that remain close to nominal in coverage and remain reasonably tight, with point estimates that stay close to the oracle benchmark.

# 4 Empirical Applications

## 4.1 Background

We revisit the statistical significance of a broad set of pricing factors, commonly referred to as anomalies, that have been proposed in prior research. From an econometric perspective, this exercise is a spanning test: each candidate factor is treated as a test asset, and the question is whether it delivers a statistically and economically significant alpha after controlling for a set of benchmark risk factors. The primary goal of this section is to illustrate the practical value of the proposed model-averaging multi-factor framework in settings where the choice of benchmark specification is itself uncertain.

Our test assets are drawn from the Global Factor Dataset of Jensen et al. (2023), which provides monthly returns for 153 factor portfolios.[9] Five of the eleven factors in our largest benchmark model (LIQ, MGMT, PERF, QMJ, and BAB) are obtained from this dataset, and the remaining 148 factors serve as anomaly-based test assets. The other six benchmark factors (MKT, SMB, HML, MOM, RMW, and CMA), together with the risk-free rate, are taken from Kenneth French's Data Library. The sample runs from January 1970 to December 2024, yielding 660 monthly observations, and the baseline analyses are conducted in-sample over the full period.

To reflect the historical development of the empirical factor-model literature, we specify a nested sequence of benchmark multi-factor models. The sequence begins with the CAPM and progressively adds factors to obtain the Fama–French three-factor model (FF3), the Carhart four-factor model (Carhart4), and the Fama–French six-factor model (FF6). Three richer specifications extend FF6: a seven-factor model (7F) that adds the liquidity factor (LIQ); a nine-factor model (9F) that further adds the management and performance mispricing factors (MGMT and PERF); and an eleven-factor model (Full11) that additionally incorporates the quality-minus-junk (QMJ) and betting-against-beta (BAB) factors. Table 3 summarizes the

[9]The data are available at `https://www.jkpfactors.com/`. We thank the authors for making the factor data publicly available.

composition of each benchmark.

Table 3: Benchmark Factor Model Specifications

| Model | Model Name | Abbrev. | Factor Composition |
|---|---|---|---|
| 1 | CAPM | CAPM | MKT |
| 2 | Fama–French 3-Factor | FF3 | MKT, SMB, HML |
| 3 | Carhart 4-Factor | Carhart4 | MKT, SMB, HML, MOM |
| 4 | Fama–French 6-Factor | FF6 | MKT, SMB, HML, MOM, RMW, CMA |
| 5 | 7-Factor Model | 7F | MKT, SMB, HML, MOM, RMW, CMA, LIQ |
| 6 | 9-Factor Model | 9F | MKT, SMB, HML, MOM, RMW, CMA, LIQ, MGMT, PERF |
| 7 | 11-Factor Model | Full11 | MKT, SMB, HML, MOM, RMW, CMA, LIQ, MGMT, PERF, QMJ, BAB |

*Notes*: The table reports the set of benchmark factor-pricing models used in the model-averaging analysis. MKT denotes the excess market return; SMB and HML denote the size and value factors; MOM denotes the momentum factor; RMW and CMA denote the profitability and investment factors; LIQ denotes the liquidity factor; MGMT and PERF denote the management and performance mispricing factors; QMJ and BAB denote the quality-minus-junk and betting-against-beta factors, respectively.

## 4.2 Main Findings

### 4.2.1 Alpha Dispersion Across Benchmarks

Empirical asset pricing offers no universally accepted benchmark for risk adjustment. Because different factor models absorb different components of the cross section, both point estimates and inference for a given test asset can vary materially with the benchmark specification. This subsection documents the extent of this benchmark sensitivity and motivates an approach that explicitly accounts for it.

Figure 1 reports alpha estimates under the seven candidate benchmarks in Table 3 for 50 anomalies randomly drawn from the 148 test assets. Marker shape identifies the benchmark, and color shading indicates the $p$-value. Appendix D reports the analogous evidence for the full anomaly set.

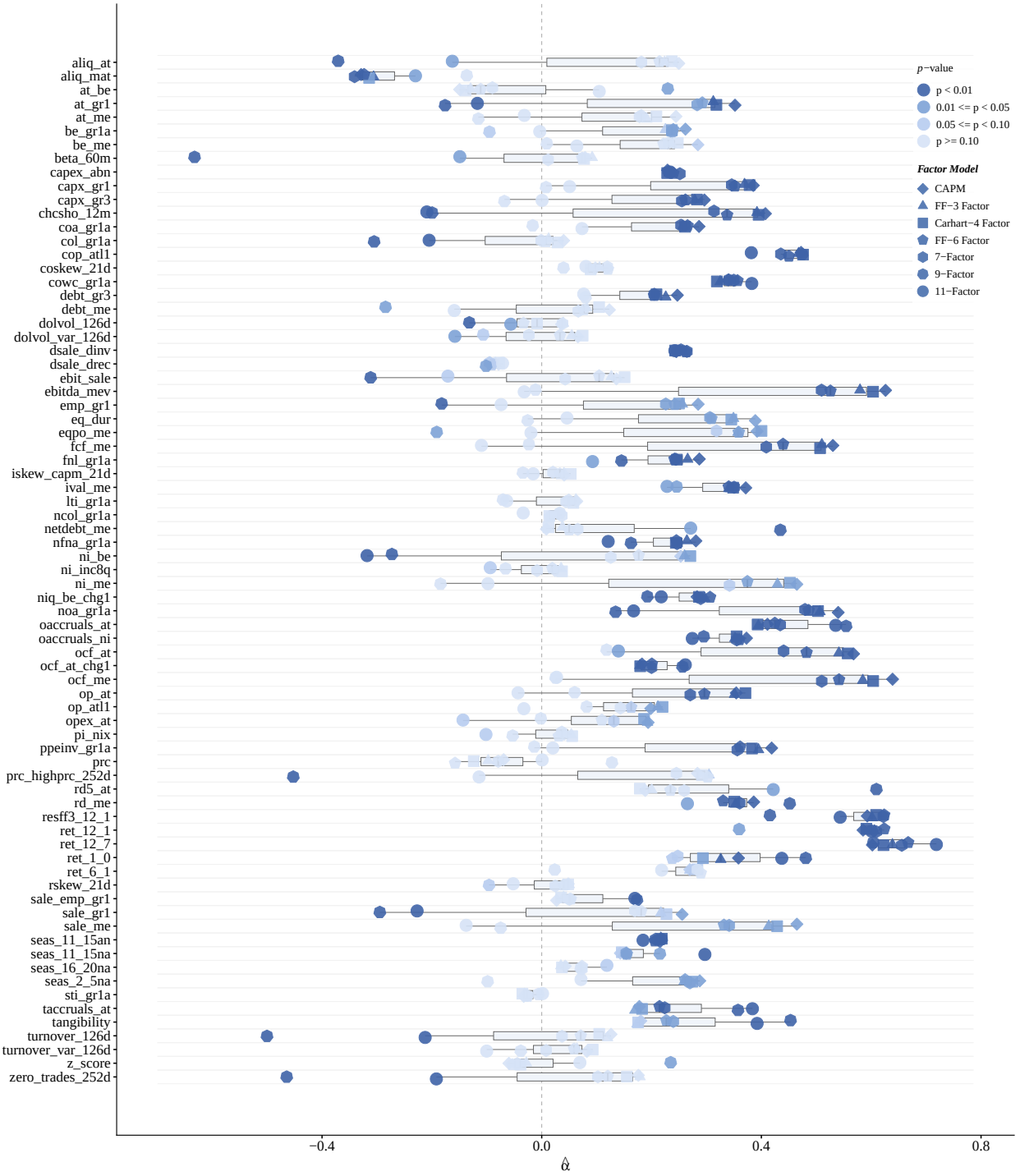



Figure 1: Distribution of Estimated Alphas. Alpha estimates across the seven candidate benchmark models for 50 anomalies randomly sampled from the 148 test assets. Marker shape identifies the benchmark; color shading indicates the $p$-value.

Three patterns stand out. First, the dispersion of alpha estimates across benchmarks is economically large: for assets such as `ebitda_mev`, `fcf_me`, and `ocf_me`, the spread across the seven benchmarks exceeds 0.3% per month. Second, the sign of alpha frequently switches across benchmarks; `sale_me`, for example, delivers a large positive alpha under the CAPM but a negative alpha under the nine- and eleven-factor models. Third, statistical significance is itself benchmark dependent: many anomalies move between the $p < 0.01$ and $p \geq 0.10$ bands as the benchmark is expanded, so conclusions about which factors survive a spanning test depend materially on the specification. These patterns motivate an inferential procedure that integrates over the candidate model set to address model uncertainty.

### 4.2.2 Cross-model Comparisons and Model Averaging

This subsection compares inference under MAMF with inference under each single benchmark. MAMF combines the seven specifications in Table 3 using Mallows-type weights and produces a single set of alphas and confidence intervals that integrate evidence across the candidate models. Alphas are estimated from monthly returns over January 1970 to December 2024, and inference uses the simulation-based procedure in Algorithm 2 with 1,000 draws and $\phi_T = \log T$. An asset is classified as significant when its 95% confidence interval excludes zero.

We begin at the asset level. Figure 2 plots alphas and confidence intervals for all 148 test assets under MAMF and under each single benchmark. Assets are sorted by the magnitude of their MAMF alphas and sign-normalized together with their confidence intervals, so the vertical axis reads as the absolute magnitude of abnormal returns. Blue intervals are significant at the 5% level; grey intervals are not.

Two patterns are visible. Under parsimonious benchmarks (CAPM, FF3, Carhart4), alphas are large in absolute value and widely dispersed, confidence intervals are wide, and insignificant assets are scattered across the range of point estimates rather than concentrated near zero. Under richer benchmarks (9F and Full11) and under MAMF, insignificant alphas

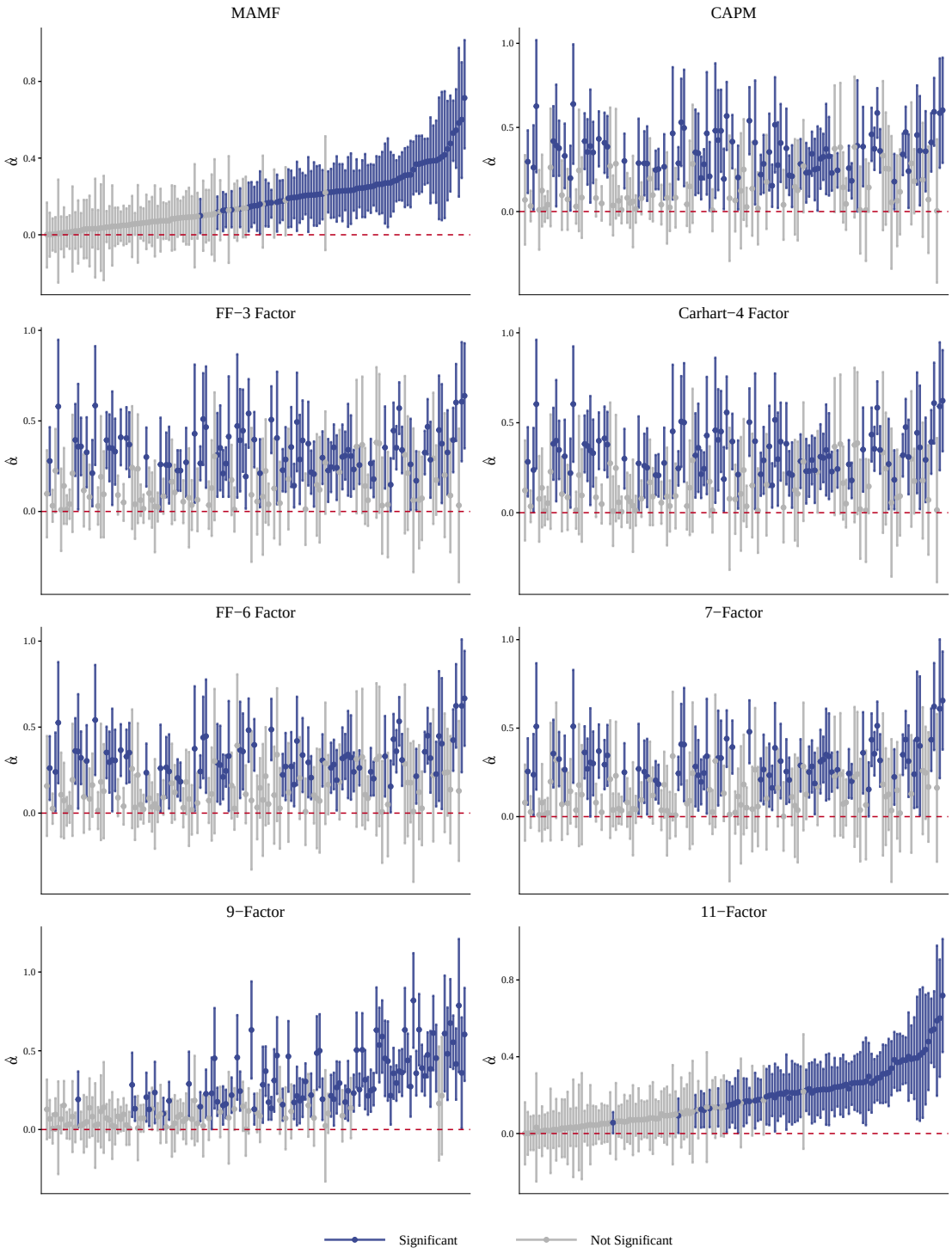



Figure 2: Estimated Alphas and 95% Confidence Intervals across MAMF and Individual Benchmarks. All 148 test assets are sorted by their MAMF alpha magnitudes and sign-normalized. Blue intervals indicate significance at the 5% level; grey intervals denote non-significance.

cluster tightly around zero and significant alphas rise smoothly along the sorted axis. The MAMF panel and the Full11 panel are visually nearly indistinguishable, a similarity we quantify below.

Table 4 summarizes the cross-sectional distribution of alphas under each estimator and highlights three facts. First, the significant share is non-monotonic in benchmark richness: it starts at 56.1% under CAPM, declines through the Fama–French family to 47.3% under 7F as additional factors absorb cross-sectional variation, then rebounds to 55.4% under 9F once the mispricing factors enter and stays at 54.1% under Full11. Aggregate conclusions about the factor zoo thus depend materially on the chosen benchmark, even within a standard menu of nested models. Second, MAMF occupies an intermediate position, with a significant share of 53.4% and a mean alpha of 0.044, close to Full11's 0.042. Third, MAMF delivers the tightest confidence intervals in the table: the widths under CAPM, FF3, FF6, and 7F are roughly 1.49 to 1.53 times the MAMF width, and even Full11 intervals are marginally wider (1.01×). That MAMF intervals are narrower than every single-benchmark interval, despite incorporating uncertainty across seven specifications, reflects the efficiency gain from reallocating weight toward well-fitting models.

Table 4: Cross-Sectional Distribution of Estimated Alphas

| Model | $N$ | $T$ | Mean | Std. | $Q_{25}$ | Median | $Q_{75}$ | Sig. (%) | CI Width |
|---|---|---|---|---|---|---|---|---|---|
| MAMF | 148 | 660 | 0.044 | 0.219 | -0.122 | 0.014 | 0.209 | 53.4 | 1.000 |
| CAPM | 148 | 660 | 0.233 | 0.190 | 0.081 | 0.251 | 0.373 | 56.1 | 1.534 |
| FF3 | 148 | 660 | 0.226 | 0.182 | 0.087 | 0.234 | 0.354 | 54.7 | 1.494 |
| Carhart4 | 148 | 660 | 0.225 | 0.187 | 0.085 | 0.236 | 0.361 | 54.7 | 1.519 |
| FF6 | 148 | 660 | 0.205 | 0.175 | 0.076 | 0.222 | 0.323 | 48.6 | 1.512 |
| 7F | 148 | 660 | 0.195 | 0.167 | 0.066 | 0.210 | 0.308 | 47.3 | 1.435 |
| 9F | 148 | 660 | 0.006 | 0.293 | -0.142 | 0.009 | 0.192 | 55.4 | 1.151 |
| Full11 | 148 | 660 | 0.042 | 0.221 | -0.126 | 0.015 | 0.203 | 54.1 | 1.006 |

*Notes*: Descriptive statistics of estimated alphas across the 148 test assets over the full sample. $T$ is the number of monthly observations. $Q_{25}$, Median, and $Q_{75}$ denote the 25th, 50th, and 75th percentiles of the cross-sectional alpha distribution. "Sig. (%)" is the share of assets with alphas significant at the 5% level. "CI Width" is the average ratio of each model's 95% confidence interval width to that of MAMF; MAMF is normalized to 1.000.

To see how the estimators relate to one another, Figure 3 reports, for each pair, the share of assets on which they agree in their significance classification (upper entries) and the mean alpha difference between them (lower entries). The estimators fall into three groups. The Fama–French family (CAPM through 7F) is internally very consistent, with pairwise agreement between 88% and 99% and mean alpha differences below 0.04. The 9F specification stands apart: it agrees with the Fama–French family on only about 50% of assets and produces mean alphas roughly 0.20 lower, reflecting the cross-sectional content of the MGMT and PERF factors. MAMF is essentially indistinguishable from Full11, agreeing on 99.3% of assets with a mean alpha difference of 0.001.

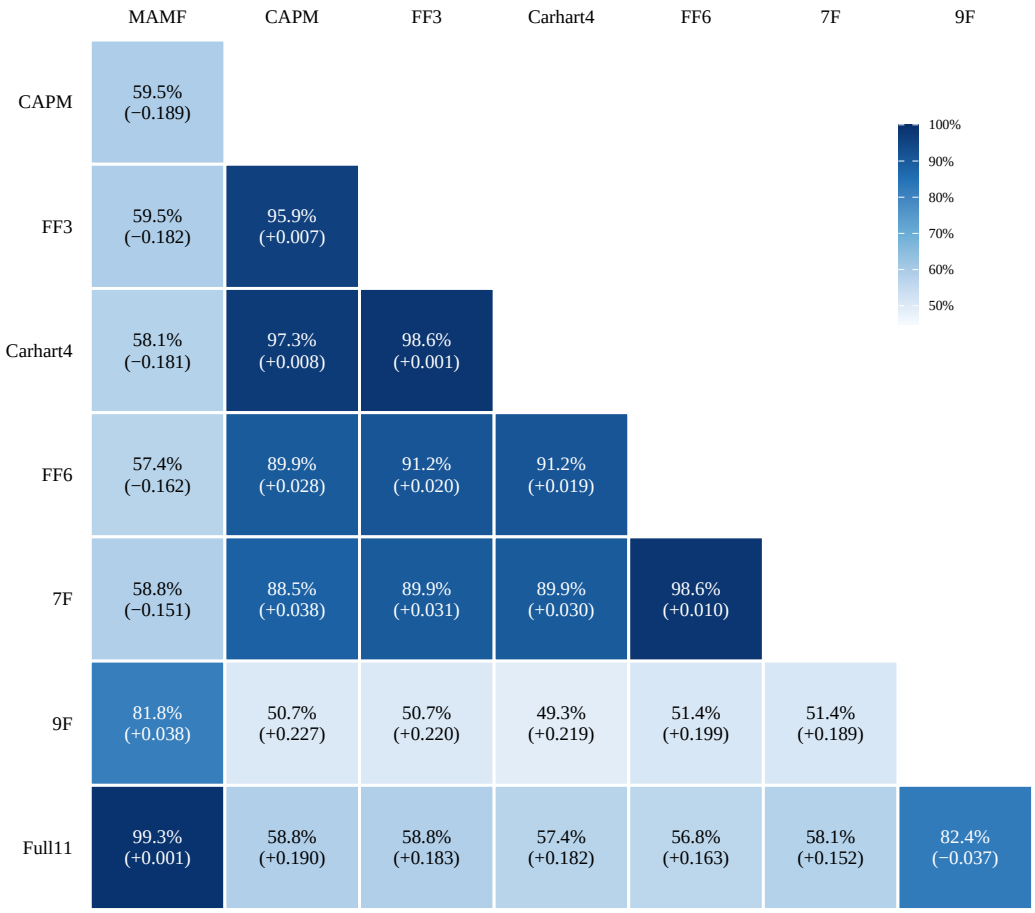



Figure 3: Pairwise Model Comparison of Factor Consistency and Alpha Differences. Upper entries report the share of assets with consistent significance classifications across model pairs. Lower entries report mean alpha differences defined as column minus row. Darker shading indicates higher consistency.

This near-equivalence between MAMF and Full11 has a substantive rather than mechanical interpretation, and it is the main reason why MAMF is useful in practice. In our sample, the richest specification happens to be close to the best-fitting model, and the Mallows criterion correctly concentrates weight on it. Were MAMF mechanically biased toward the largest candidate, it would also track the largest specification in samples where a smaller benchmark is the true DGP. The simulation evidence in Section 3 confirms that it does not: when FF6 is the oracle model, MAMF concentrates weight on FF6 rather than on the eleven-factor candidate, and its alpha estimates track FF6 rather than Full11. The empirical near-equivalence between MAMF and Full11 therefore reflects the Mallows criterion selecting a well-fitting specification in this sample, not a preference for dimensionality.

More importantly, the eleven-factor specification is rarely adopted as a maintained benchmark in applied work: the overwhelming majority of empirical studies condition on the CAPM, FF3, Carhart4, or FF6, and Table 4 shows that these benchmarks deliver significant shares between 47% and 56% and produce substantially looser confidence intervals than MAMF. A researcher who adopts MAMF obtains inference close to what Full11 would deliver in this sample, but without committing ex ante to any single specification, and with the smallest confidence intervals in the table. This is precisely how a principled response to benchmark uncertainty should behave: it matches the best-fitting model when the data support it, while accommodating the possibility that the best specification is smaller in other samples.

The cross-sectional statistics mask disagreement at the asset level that illustrates why averaging matters for individual inference. Table 5 classifies each asset by the number of specifications under which its alpha is significant. Only 23.6% of assets are significant under all eight estimators and 18.2% are insignificant under all eight; the remaining 58.2% switch significance status across benchmarks. A researcher reporting results from any single specification therefore makes an implicit, undisclosed choice about risk adjustment for more than half of the anomaly cross section. MAMF resolves this choice in a transparent, data-driven

way: for example, `dolvol_126d` is classified as significant under Full11 but insignificant under MAMF, indicating that the Full11 rejection is not robust to the broader benchmark set and is attenuated once model uncertainty is formally incorporated.

Table 5: Agreement in Alpha Significance across Benchmark Models

| # Benchmarks with Significant Alpha | # Test Assets | Share (%) |
|---|---|---|
| 8 | 35 | 23.6 |
| 7 | 7 | 4.7 |
| 6 | 7 | 4.7 |
| 5 | 28 | 18.9 |
| 4 | 6 | 4.1 |
| 3 | 26 | 17.6 |
| 2 | 3 | 2.0 |
| 1 | 9 | 6.1 |
| 0 | 27 | 18.2 |
| Total | 148 | 100.0 |

*Notes*: Each asset is classified by the number of benchmark specifications under which its alpha is significant at the 5% level. The benchmark set contains eight estimators: MAMF, CAPM, FF3, Carhart4, FF6, 7F, 9F, and Full11. Asset-level significance outcomes are reported in Table A.3 in the appendix.

### 4.2.3 Stability in the Nested Structure

A natural concern with the baseline construction is that the nested ordering may mechanically tilt MAMF weights toward the largest specification. The candidate sequence begins with CAPM and adds Fama–French factors before introducing the anomaly-related factors (LIQ, MGMT, PERF, QMJ, BAB), so the intermediate models are severely underspecified relative to the full eleven-factor model. If the Mallows criterion concentrates weight on Full11 simply because the intermediate models are poor fits, the near-equivalence between MAMF and Full11 documented in Section 4.2.2 would be an artifact of the ordering rather than a feature of the data.

To test this, we construct an alternative nesting that preserves the same eleven factors but reverses the order of introduction. The sequence begins with the anomaly-related factors

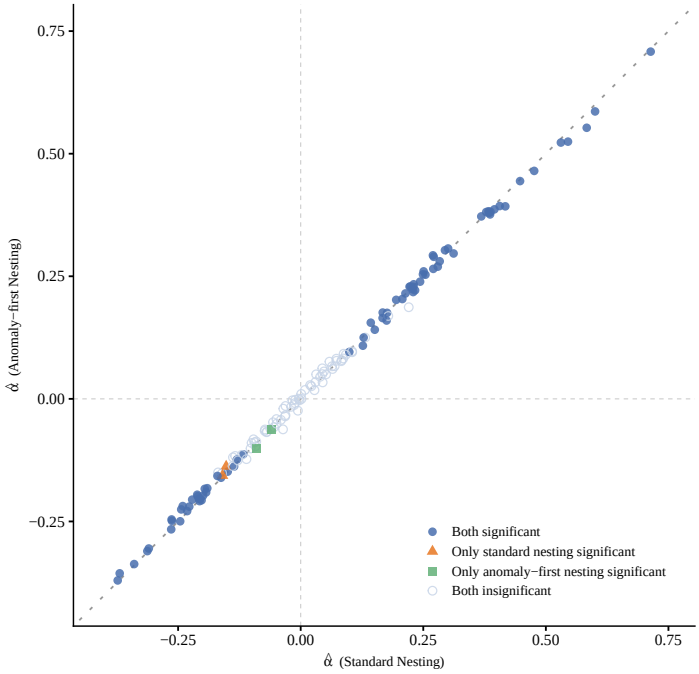
(BAB, QMJ, MGMT, PERF, LIQ) and subsequently adds the Fama–French factors. The model space is unchanged, but the intermediate candidates now capture distinct, economically meaningful sources of variation before being subsumed. If the baseline result reflects the ordering, the anomaly-first nesting should shift MAMF weights and produce materially different alpha estimates and confidence intervals.

Figure 4 shows that it does not. In panel (a), alpha estimates under the two nesting schemes align tightly along the 45-degree line across the full cross section, with no systematic deviation for either positive or negative alphas. Panel (b) extends this invariance to inference: confidence interval widths also track the 45-degree line closely, so the precision of the MAMF estimator is similarly insensitive to nesting order. Significance classifications are almost perfectly preserved, and the few assets that switch status are isolated cases near the significance boundary rather than a coherent pattern.

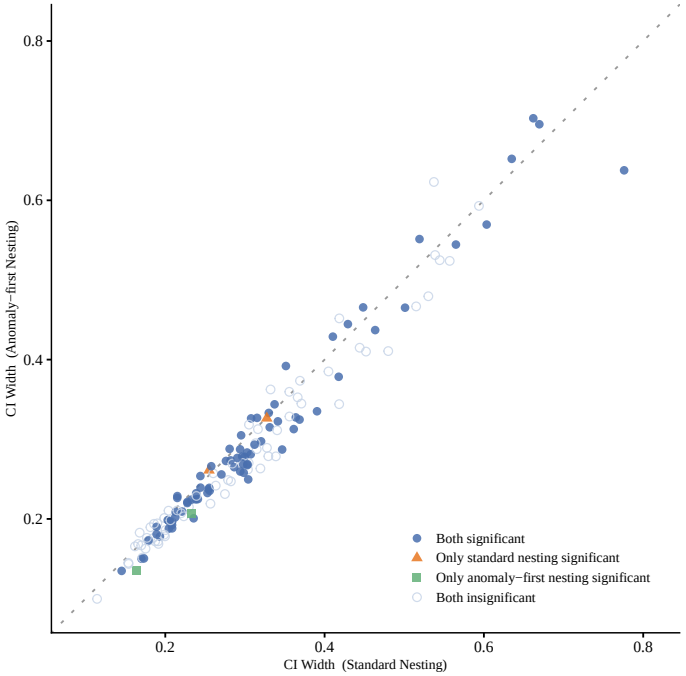
These results establish that MAMF weights are driven by the data rather than by the ordering of the candidate sequence. Combined with the simulation evidence of Section 3, where MAMF correctly tracks the FF6 oracle rather than the larger Full11 candidate, they rule out two distinct mechanical explanations for the baseline result: concentration induced by ordering and concentration induced by dimensionality. The near-equivalence between MAMF and Full11 documented in Section 4.2.2 therefore reflects the Mallows criterion selecting a well-fitting specification, not a preference for the largest model.

# 5 Conclusions

This paper develops the Model-Averaging Multi-Factor (MAMF) framework to address benchmark uncertainty in empirical asset pricing. Rather than committing to a single factor specification, MAMF uses a Mallows-type criterion to combine evidence across a prespecified set of benchmarks, balancing pricing fit against model complexity. The framework treats the benchmark as an object of uncertainty and reports a single set of risk-adjusted returns




(a) Alpha estimates under alternative nesting structures




(b) Confidence interval widths under alternative nesting structures

Figure 4: Comparison of pricing estimates under different nesting structures. Panel (a) reports asset-level alphas and panel (b) the corresponding 95% confidence interval widths under the standard and anomaly-first nesting specifications.

that is explicit about its role.

Theoretically, MAMF attains the performance of the best feasible convex combination of candidate models as the time dimension grows. Because the averaging weights are data-dependent, the estimator has a non-pivotal limiting distribution further shaped by the serial dependence typical of return data. We develop a simulation-based inference procedure that delivers asymptotically valid confidence intervals for alphas and betas under the same dependence conditions used for estimation.

Empirically, MAMF achieves near-oracle accuracy and near-nominal coverage in calibrated simulations. Applied to 148 anomalies, it classifies 53.4% as significant at the 5% level, a share stable across rolling windows that sits between the rejection rates of parsimonious and richly parameterized benchmarks. The MAMF verdict is close to what the eleven-factor specification would deliver in this sample, but is obtained without requiring ex ante commitment to any single benchmark and with the tightest confidence intervals among all estimators considered. MAMF thus offers a middle ground in the replication debate: more conservative than single-CAPM tests yet more disciplined than benchmark-by-benchmark reporting.

Several extensions are natural. Relaxing the nested-model structure would broaden applicability to settings where competing model families coexist and the augmentation device becomes high-dimensional, raising weak-factor considerations. Integrating averaging into conditional asset-pricing settings with time-varying exposures and state-dependent premia would allow benchmark uncertainty to interact with dynamic risk adjustment. Connecting MAMF to machine-learning pipelines for factor construction, while maintaining valid post-selection inference, is another promising direction. We leave these for future research.

# References


Anatolyev, S., Mikusheva, A. 2022. Factor models with many assets: Strong factors, weak factors, and the two-pass procedure. Journal of Econometrics. 229 (1), 103–126.

Andrews, D. W. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica: Journal of the Econometric Society. 817–858.

Avramov, D. 2002. Stock return predictability and model uncertainty. Journal of Financial Economics. 64 (3), 423–458.

Avramov, D., Cheng, S., Metzker, L. 2023a. Machine learning vs. economic restrictions: Evidence from stock return predictability. Management Science. 69 (5), 2587–2619.

Avramov, D., Cheng, S., Metzker, L., Voigt, S. 2023b. Integrating factor models. The Journal of Finance. 78 (3), 1593–1646.

Bai, J., Ng, S. 2023. Approximate factor models with weaker loadings. Journal of Econometrics. 235 (2), 1893–1916.

Bradley, R. C. 2005. Basic properties of strong mixing conditions. a survey and some open questions.

Carhart, M. M. 1997. On persistence in mutual fund performance. The Journal of Finance. 52 (1), 57–82.

Chen, A. Y., Zimmermann, T. 2022a. Open source cross-sectional asset pricing. Critical Finance Review. 27 (2), 207–264.

Chen, A. Y., Zimmermann, T. 2022b. Publication bias in asset pricing research. arXiv preprint arXiv:2209.13623.

Cheng, T.-C. F., Ing, C.-K., Yu, S.-H. 2015. Toward optimal model averaging in regression models with time series errors. Journal of Econometrics. 189 (2), 321–334.

Cochrane, J. H. 2011. Presidential address: Discount rates. The Journal of Finance. 66 (4), 1047–1108.

Cooper, M. J., Gulen, H., Schill, M. J. 2008. Asset growth and the cross-section of stock returns. The Journal of Finance. 63 (4), 1609–1651.

Cremers, K. J. M. 2002. Stock return predictability: A bayesian model selection approach. The Review of Financial Studies. 15 (4), 1223–1249.

Eiling, E. 2013. Industry-specific human capital, idiosyncratic risk, and the cross-section of expected stock returns. The Journal of Finance. 68 (1), 43–84.

Fama, E. F., French, K. R. 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics. 33 (1), 3–56.

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

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

Fang, F., Liu, M. 2020. Limit of the optimal weight in least squares model averaging with non-nested models. Economics Letters. 196, 109586.

Findley, D. F., Wei, C.-Z. 1993. Moment bounds for deriving time series clt's and model selection procedures. Statistica Sinica. 453–480.

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

Gao, Y., Xie, T., Zou, G. 2023. Least squares model averaging for two non-nested linear models. Journal of Systems Science and Complexity. 36 (1), 412–432.

Giglio, S., Kelly, B., Xiu, D. 2022. Factor models, machine learning, and asset pricing. Annual Review of Financial Economics. 14 (1), 337–368.

Gu, S., Kelly, B., Xiu, D. 2020. Empirical asset pricing via machine learning. The Review of Financial Studies. 33 (5), 2223–2273.

Hansen, B. E. 2007. Least squares model averaging. Econometrica. 75 (4), 1175–1189.

Hansen, B. E., Racine, J. S. 2012. Jackknife model averaging. Journal of Econometrics. 167 (1), 38–46.

Harvey, C. R., Liu, Y., Zhu, H. 2016. ... and the cross-section of expected returns. The Review of Financial Studies. 29 (1), 5–68.

Hirshleifer, D., Ma, L. 2024. The effect of new information technologies on asset pricing anomalies.Technical report, National Bureau of Economic Research.

Hou, K., Mo, H., Xue, C., Zhang, L. 2021. An augmented q-factor model with expected growth. Review of Finance. 25 (1), 1–41.

Hou, K., Xue, C., Zhang, L. 2015. Digesting anomalies: An investment approach. The Review of Financial Studies. 28 (3), 650–705.

Hou, K., Xue, C., Zhang, L. 2020. Replicating anomalies. The Review of Financial Studies. 33 (5), 2019–2133.

Jensen, T. I., Kelly, B., Pedersen, L. H. 2023. Is there a replication crisis in finance? The Journal of Finance. 78 (5), 2465–2518.

Kelly, B. T., Pruitt, S., Su, Y. 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics. 134 (3), 501–524.

Kozak, S., Nagel, S., Santosh, S. 2020. Shrinking the cross-section. Journal of Financial Economics. 135 (2), 271–292.

Leippold, M., Wang, Q., Zhou, W. 2022. Machine learning in the chinese stock market. Journal of Financial Economics. 145 (2), 64–82.

Lettau, M., Pelger, M. 2020. Factors that fit the time series and cross-section of stock returns. The Review of Financial Studies. 33 (5), 2274–2325.

Li, C., Li, Q., Racine, J. S., Zhang, D. 2018. Optimal model averaging of varying coefficient models. Statistica Sinica. 28 (4), 2795–2809.

Liang, H., Zou, G., Wan, A. T., Zhang, X. 2011. Optimal weight choice for frequentist model average estimators. Journal of the American Statistical Association. 106 (495), 1053–1066.

Lintner, J. 1965. Security prices, risk, and maximal gains from diversification. The Journal of Finance. 20 (4), 587–615.

Liu, Q., Okui, R. 2013. Heteroscedasticity-robust cp model averaging. The Econometrics Journal. 16 (3), 463–472.

McLean, R. D., Pontiff, J. 2016. Does academic research destroy stock return predictability?

The Journal of Finance. 71 (1), 5–32.

Newey, W. K., West, K. D. 2023. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica. 55 (3), 703–708.

Novy-Marx, R. 2013. The other side of value: The gross profitability premium. Journal of financial economics. 108 (1), 1–28.

Novy-Marx, R., Velikov, M. 2016. A taxonomy of anomalies and their trading costs. The Review of Financial Studies. 29 (1), 104–147.

Onatski, A. 2012. Asymptotics of the principal components estimator of large factor models with weakly influential factors. Journal of Econometrics. 168 (2), 244–258.

O'Doherty, M. S., Savin, N. E., Tiwari, A. 2016. Evaluating hedge funds with pooled benchmarks. Management Science. 62 (1), 69–89.

Pástor, L., Stambaugh, R. F. 2003. Liquidity risk and expected stock returns. Journal of Political Economy. 111 (3), 642–685.

Peng, J., Li, Y., Yang, Y. 2025. On optimality of mallows model averaging. Journal of the American Statistical Association. 120 (550), 1152–1163.

Qiu, Y., Ren, Y., Xie, T. 2019. Weighing asset pricing factors: a least squares model averaging approach. Quantitative Finance. 19 (10), 1673–1687.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance. 19 (3), 425–442.

Stambaugh, R. F., Yuan, Y. 2017. Mispricing factors. The Review of Financial Studies. 30 (4), 1270–1315.

Sun, Y., Hong, Y., Lee, T.-H., Wang, S., Zhang, X. 2021. Time-varying model averaging. Journal of Econometrics. 222 (2), 974–992.

Sun, Y., Hong, Y., Wang, S., Zhang, X. 2023. Penalized time-varying model averaging. Journal of Econometrics. 235 (2), 1355–1377.

Wan, A. T., Zhang, X., Zou, G. 2010. Least squares model averaging by mallows criterion. Journal of Econometrics. 156 (2), 277–283.

Wang, M., You, K., Zhu, L., Zou, G. 2024. Robust model averaging approach by mallows-type criterion. Biometrics. 80 (4), ujae128.

Wei, C.-Z. 1987. Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. The Annals of Statistics. 1667–1682.

Wright, J. H. 2008. Bayesian model averaging and exchange rate forecasting. Journal of Econometrics. 146 (1), 6–15.

Yu, D., Lian, H., Sun, Y., Zhang, X., Hong, Y. 2024. Post-averaging inference for optimal model averaging estimator in generalized linear models. Econometric Reviews. 43 (2-4), 98–122.

Zhang, X. 2021. Optimal model averaging based on generalized method of moments. Statistica Sinica. 31 (4), 2103–2122.

Zhang, X., Liu, C.-A. 2019. Inference after model averaging in linear regression models. Econometric Theory. 35 (4), 816–841.

Zhang, X., Wan, A. T., Zou, G. 2013. Model averaging by jackknife criterion in models with dependent data. Journal of Econometrics. 174 (2), 82–94.

Zhang, X., Yu, D., Zou, G., Liang, H. 2016. Optimal model averaging estimation for generalized linear models and generalized linear mixed-effects models. Journal of the American Statistical Association. 111 (516), 1775–1790.

Zhao, S., Ma, Y., Wan, A. T., Zhang, X., Wang, S. 2020. Model averaging in a multiplicative heteroscedastic model. Econometric Reviews. 39 (10), 1100–1124.

Zhu, R., Wan, A. T., Zhang, X., Zou, G. 2019. A mallows-type model averaging estimator for the varying-coefficient partially linear model. Journal of the American Statistical Association. 114 (526), 882–892.

# Online Appendix to

# "Multi-Factor Asset Pricing via Model-Averaging"

# A Additional Results for Limit Distribution

## A.1 Limiting Distribution of the MAMF Estimators

We now derive the limiting distribution of the MAMF estimator. To facilitate the discussion, we first introduce some notation. Let $X_t$ denote the $(K+1)$-dimensional regressor vector in the full model (intercept plus $K$ factors), and set

$$Z_T = T^{-1/2} \sum_{t=1}^{T} X_t e_t.$$

Under Assumptions 2–3, we have $Z_T \xrightarrow{d} Z \sim N\big(0, \mathbf{\Omega}_\infty\big)$, where $\mathbf{\Omega}_\infty$ is the long-run covariance matrix of the process $\{X_t e_t\}_{t=1}^T$. For each candidate model $m$, let $X_t^{(m)}$ be its regressor vector, $\mathbf{Q}^{(m)} = \mathbb{E}[X_t^{(m)} X_t^{(m)\top}]$, and $\mathbf{Q}_T^{(m)} = T^{-1} \sum_{t=1}^T X_t^{(m)} X_t^{(m)\top}$. Let $\mathbf{\Pi}_m$ be the $(K_m+1)\times(K+1)$ selection matrix such that $X_t^{(m)} = \mathbf{\Pi}_m X_t$, and define

$$\mathbf{V}^{(m)} = \mathbf{\Pi}_m^\top \big(\mathbf{Q}^{(m)}\big)^{-1} \mathbf{\Pi}_m, \qquad m = 1, \ldots, M.$$

Recall that the first $M_0$ models are under-specified (they omit at least one relevant factor), so that $S = M - M_0$ models are just-specified or over-specified. The MAMF weights on the latter models are collected in

$$\lambda = \big(\widehat{w}_{M_0+1}^{\mathrm{MAMF}}, \ldots, \widehat{w}_M^{\mathrm{MAMF}}\big)^\top \in \mathcal{L} := \big\{\lambda \in [0,1]^S : \sum_{s=1}^{S} \lambda_s = 1\big\}.$$

Let $\Gamma$ be the $S \times S$ matrix with entries

$$\Gamma_{sj} = 2\sigma^2 K_{M_0+s} - \mathbb{E}\big[Z^\top \mathbf{V}^{(\max\{s,j\})} Z\big], \qquad s, j = 1, \ldots, S,$$

where $\sigma^2$ is the innovation variance in Assumption 3, and $K_m$ denotes the number of regressors (including the intercept) in model $m$.

**Theorem A** *Suppose Assumptions 1–3 hold. Then*

$$\sqrt{T}\begin{pmatrix} \widehat{\alpha}^{\mathrm{MAMF}} - \alpha \\ \widehat{\boldsymbol{\beta}}^{\mathrm{MAMF}} - \boldsymbol{\beta} \end{pmatrix} = \sum_{m=M_0+1}^{M} \widehat{w}_m^{\mathrm{MAMF}} \, \boldsymbol{\Pi}_m^\top \big(\mathbf{Q}_T^{(m)}\big)^{-1} \boldsymbol{\Pi}_m \, Z_T + o_p(1) \xrightarrow{d} \sum_{s=1}^{S} \tilde{\lambda}_s^{\mathrm{MAMF}} \, \mathbf{V}^{(M_0+s)} Z,$$

*where $Z \sim N(0, \boldsymbol{\Omega}_\infty)$, and the limiting weight vector*

$$\tilde{\lambda}^{\mathrm{MAMF}} = \arg\min_{\lambda \in \mathcal{L}} \lambda^\top \boldsymbol{\Gamma} \lambda$$

*is the solution of a population quadratic program based on $\boldsymbol{\Gamma}$.*

Theorem A shows that the properly scaled MAMF estimator has a nonstandard limiting distribution. Instead of converging to a single Gaussian law, the limit is a random linear combination of Gaussian terms $\mathbf{V}^{(m)} Z$, where the combination weights $\tilde{\lambda}^{\mathrm{MAMF}}$ solve a quadratic optimization problem that trades off model fit and complexity in the population. Intuitively, the first-order behavior of $(\widehat{\alpha}^{\mathrm{MAMF}}, \widehat{\boldsymbol{\beta}}^{\mathrm{MAMF}})$ is equivalent to that of an infeasible "oracle" average using these population-optimal weights, but the resulting limit is a mixture of Gaussian components indexed by the just-specified and over-specified models.

From the perspective of multi-factor asset pricing, the theorem highlights that the sampling variation of the MAMF alphas and betas is driven jointly by the long-run covariance structure of factor innovations and by the way the Mallows criterion reallocates weight across competing factor specifications. A key implication is that the limiting distribution is non-pivotal: it depends on unknown quantities such as $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}_\infty$, so conventional plug-in

standard errors and $t$-statistics are not valid. This nonstandard limit justifies the simulation-based inference procedure in Section 2.3.2, which replaces the nuisance parameters with consistent estimators and uses Monte Carlo approximation to construct confidence intervals and test statistics for MAMF-based alphas and betas.

# B Proofs

### Proof of Theorem 1

Let $\mu_T = \mathbf{X}\gamma$ denote the target mean value at all periods of $t$, and let $\widehat{\mu}_T = \mathbf{X}\widehat{\gamma}(\mathbf{w})$ denote the $\mathbf{w}$-weighted OLS prediction. Let

$$R_T(\mathbf{w}) = \mathbb{E}[L_n(\mathbf{w})|\mathbf{X}] \equiv \mathbb{E}_{\mathbf{X}}[L_T(\mathbf{w})]$$

**Lemma A** *For any weights* $\mathbf{w}$*, we can write*

$$R_T(\mathbf{w}) = \sum_{m=1}^{M}\sum_{l=1}^{M} w_m w_l \left[\mu_T' \left(\mathbf{I} - \mathbf{P}_{\max\{m,l\}}\right)\mu_T + \min\{k_m, k_l\}\right].$$

**Proof.** Note first that

$$R_T(\mathbf{w}) = \mathbb{E}_{\mathbf{X}}(L_T(\mathbf{w})) = \mathbb{E}_{\mathbf{X}}(\mathbf{e}_T'\mathbf{P}(\mathbf{w})'\mathbf{P}(\mathbf{w})\mathbf{e}_T) + \mu_T'(\mathbf{I} - \mathbf{P}(\mathbf{w}))'(\mathbf{I} - \mathbf{P}(\mathbf{w}))\mu_T.$$

Since

$$\mathbf{P}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{P}_m,$$

it follows that

$$\mathbb{E}_{\mathbf{X}}(\mathbf{e}_T'\mathbf{P}^*(\mathbf{w})'\mathbf{P}^*(\mathbf{w})\mathbf{e}_T) = \mathbb{E}_{\mathbf{X}}\left(\sum_{m=1}^{M}\sum_{l=1}^{M} w_m w_l \mathbf{e}_T'\mathbf{P}_l\mathbf{P}_m\mathbf{e}_T\right) = \sum_{m=1}^{M}\sum_{l=1}^{M} w_m w_l \min\{k_m, k_l\}.$$

Similarly,

$$\mu_T'(\mathbf{I}-\mathbf{P}(\mathbf{w}))'(\mathbf{I}-\mathbf{P}(\mathbf{w}))\mu_T = \sum_{m=1}^{M}\sum_{l=1}^{M} w_m w_l \mu_T'(\mathbf{I}-\mathbf{P}_{\max\{m,l\}})\mu_T.$$

Consequently, the desired conclusion in Lemma A follows. ■

**Main Proof.** Our proof follows Cheng et al. (2015) closely. We omit the dependence on the asset index $i$ whenever there is no confusion, and we use $C_T$ to denote the MAMF criteria function. Define

$$\mathbf{w}_T^* = \arg\min_{\mathbf{w}\in\mathcal{H}_N} L_T(\mathbf{w})$$

$$\widehat{\mathbf{w}}_T = \arg\min_{\mathbf{w}\in\mathcal{H}_N} C_T(\mathbf{w})$$

which are the optimal weight that minimizes the deviation from the target, and the optimal weight that minimizes the MAMF criteria, repsectively.

By noticing

$$C_T(\mathbf{w}) - L_T(\mathbf{w}) = \mathbf{e}_T'\mathbf{e}_T + 2\mathbf{e}_T'(\mathbf{I}-\mathbf{P}(\mathbf{w}))\mu_T - 2\left[\mathbf{e}_T'\mathbf{P}(\mathbf{w})\mathbf{e}_T - \sum_{m=1}^{M} w_m k_m\right],$$

we get

$$\begin{aligned}
0 \geq C_T(\widehat{\mathbf{w}}_T) - C_T(\mathbf{w}_T^*) &= L_T(\widehat{\mathbf{w}}_T) - L_T(\mathbf{w}_T^*) + 2\mathbf{e}'(\mathbf{I}-\mathbf{P}(\widehat{\mathbf{w}}_T))\mu_T \\
&\quad - 2\left[\mathbf{e}'\mathbf{P}(\widehat{\mathbf{w}}_T)\mathbf{e} - \sum_{m=1}^{M}\hat{w}_m k_m\right] - 2\mathbf{e}'(\mathbf{I}-\mathbf{P}(\mathbf{w}_T^*))\mu_T \\
&\quad + 2\left[\mathbf{e}'\mathbf{P}(\mathbf{w}_T^*)\mathbf{e} - \sum_{m=1}^{M} w_m^* k_m\right] \\
&= L_T(\widehat{\mathbf{w}}_T) - L_T(w_T) + 2A_T(\widehat{\mathbf{w}}_T^*) - 2B_T(\widehat{\mathbf{w}}_T^*) - 2A_T(\mathbf{w}_T^*) + 2B_T(\mathbf{w}_T^*),
\end{aligned} \tag{19}$$

where

$$A_T(\mathbf{w}) = \mathbf{e}_T'(\mathbf{I} - \mathbf{P}(\mathbf{w}))\mu_T, \quad B_T(\mathbf{w}) = \mathbf{e}_T'\mathbf{P}(\mathbf{w})\mathbf{e}_T - \sum_{m=1}^{M} w_m k_m.$$

In view of (A.2) and $L_T(\mathbf{w}_T^*) \leq L_T(\widehat{\mathbf{w}}_T)$, it suffices for (2.11) to show that

$$\sup_{\mathbf{w}\in\mathcal{H}_N} \left| \frac{A_T(\mathbf{w})}{R_T(\mathbf{w})} \right| = o_p(1), \tag{20}$$

$$\sup_{\mathbf{w}\in\mathcal{H}_N} \left| \frac{B_T(\mathbf{w})}{R_T(\mathbf{w})} \right| = o_p(1), \tag{21}$$

$$\sup_{\mathbf{w}\in\mathcal{H}_N} \left| \frac{L_T(\mathbf{w})}{R_T(\mathbf{w})} - 1 \right| = o_p(1), \tag{22}$$

where $\xrightarrow{p}$ denotes convergence in probability.

To show , first note that

$$\mathcal{H}_{(l)} = \bigcup_{1\leq j_1<\cdots<j_l\leq M} \mathcal{H}_{j_1,\ldots,j_l},$$

where for $1 \leq j_1 < \cdots < j_l \leq M$, $\mathcal{H}_{j_1,\ldots,j_l} = \{\mathbf{w} : \mathbf{w} \in \mathcal{H}_{(l)}, \omega_{j_k} \neq 0, k \leq l\}$. Hence, for any $\varepsilon > 0$,

$$\begin{aligned} P_{\mathbf{X}}\left( \sup_{\mathbf{w}\in\mathcal{H}_{(l)}} \left| \frac{A_T(\mathbf{w})}{R_T(\mathbf{w})} \right| > \varepsilon \right) &\leq \sum_{l=1}^{N} \sum_{j_l=l}^{M} \cdots \sum_{j_1=1}^{j_2-1} P_{\mathbf{X}}\left( \sup_{\mathbf{w}\in\mathcal{H}_{j_1,\ldots,j_l}} \left| \frac{A_T(\mathbf{w})}{R_T(\mathbf{w})} \right| > \varepsilon \right) \\ &\leq \sum_{l=1}^{N} \sum_{j_l=l}^{M} \cdots \sum_{j_1=1}^{j_2-1} P_{\mathbf{X}}\left( \frac{\sum_{m\in[j_1,\ldots,j_l]} |\mu_T'(\mathbf{I} - \mathbf{P}_m)e_T|}{\delta^2 \max_{m\in[j_1,\ldots,j_l]} D_T(m)} > \varepsilon \right) \\ &\equiv \sum_{l=1}^{N} Q_l, \end{aligned}$$

where $P_{\mathbf{X}}(\cdot) = \mathbf{P}(\cdot \mid \mathbf{X})$, and the second inequality follows from

$$\inf_{w\in\mathcal{H}_{j_1,\ldots,j_l}} R_T(w) \geq \delta^2 \max_{m\in[j_1,\ldots,j_l]} D_T(m)$$

$$D_T(m) = \mu_T'(1 - \mathbf{P}_m)\mu_T + K_m$$

which is ensured by Lemma 1 and the definition of $\mathcal{H}_T$. Let $S_1 = S/2$. Then, it holds that

$$
\begin{aligned}
&P_{\mathbf{X}}\left( \sum_{m\in[j_1,\ldots,j_l]} \frac{|\mu_T'(\mathbf{I}-\mathbf{P}_m)\mathbf{e}_T|}{\delta^2 \max_{m\in[j_1,\ldots,j_l]} D_T(m)} > \varepsilon \right) \\
&\leq_{(1)} C_1 \sum_{m\in[j_1,\ldots,j_l]} \mathbb{E}_{\mathbf{X}}\left(|\mu_T'(\mathbf{I}-\mathbf{P}_m)\mathbf{e}_T|^{S_1}\right) \left\{ \frac{1}{\max_{m\in[j_1,\ldots,j_l]} D_T(m)} \right\}^{S_1} \\
&\leq_{(2)} C_2 \sum_{m\in[j_1,\ldots,j_l]} (\mu_T'(\mathbf{I}-\mathbf{P}_m)\mu_T)^{S_1/2} \frac{1}{\max_{m\in[j_1,\ldots,j_l]} D_T(m)^{S_1}} \\
&\leq_{(3)} C_3 \sum_{m\in[j_1,\ldots,j_l]} \frac{1}{(\max_{m\in[j_1,\ldots,j_l]} D_T(m))^{S_1/2}} \leq \frac{C_4 l}{(D_T(j_l))^{S_1/2}}
\end{aligned}
$$

where (1) follows by the Chevyshev's inequality, (2) follows by the assumption (*iii*) in the Theorem and Lemma 2 of Wei (1987), (3) follows by the definition of the $D_T$, and $C_1 \sim C_4$ denote constants thatdoes not depend on $T$.

Now, define $K_T^* = \min_{1\leq m\leq M} D_T(m)$, and we decompose the summation $\sum_{j_l=l}^{M}$ into the summation of $\sum_{j_l=l}^{K_T^*}$ and $\sum_{j_l=K_T^*+1}^{M}$, we can have

$$
\begin{aligned}
Q_l &\leq C \left\{ \sum_{j_l=l}^{K_T^*} \cdots \sum_{j_1=1}^{j_2-1} \frac{1}{(K_T^*)^{S_1/2}} + \sum_{j_l=K_T^*+1}^{M} \cdots \sum_{j_1=1}^{j_2-1} \frac{1}{(D_n(j_l))^{S_1/2}} \right\} \\
&\leq C \left\{ (K_T^*)^{-(S_1/2-l)} + \sum_{j_l=K_T^*+1}^{\infty} \frac{j_l^{l-1}}{j_l^{S_1/2}} \right\},
\end{aligned}
$$

Note that $K_T^* \to \infty$ almost surely because the matrix $\mathbf{X}$ is full rank. As a result,

$$
\mathbb{P}_{\mathbf{X}}\left( \sup_{w\in\mathcal{H}_N} \left| \frac{A_T(w)}{R_T(w)} \right| > \varepsilon \right) \to 0, \quad \text{a.s.}
$$

This and the dominated convergence theorem together imply (B).

Similarly,

$$
\mathbb{P}_{\mathbf{X}}\left( \sup_{w\in\mathcal{H}_N} \left| \frac{B_T(w)}{R_T(w)} \right| > \varepsilon \right) \leq C \sum_{l=1}^{N} E_l,
$$

where

$$E_l = \sum_{j_l=l}^{M} \cdots \sum_{j_1=1}^{j_2-1} \mathbb{P}_{\mathbf{X}} \left( \frac{\sum_{m\in[j_1,\ldots,j_l]} |\mathbf{e}_T' \mathbf{P}_m \mathbf{e}_T - K_m|}{\delta^2 \max_{m\in[j_1,\ldots,j_l]} D_T(m)} > \varepsilon \right).$$

By $(iii)$ in the Theorem assumption and the first moment bound theorem of Findley and Wei (1993), it follows that

$$\mathbb{P}_{\mathbf{X}} \left( \frac{\sum_{m\in[j_1,\ldots,j_l]} |e_T' \mathbf{P}_m e_T - k_m|}{\delta^2 \max_{m\in[j_1,\ldots,j_l]} D_T(m)} > \varepsilon \right) \leq C \sum_{m\in[j_1,\ldots,j_l]} \frac{K_m^{S_1/2}}{(D_T(j))^{S_1/2}},$$

for some constant $C$ that does not depend on T. Therefore, (21) follows immediately from an argument similar to that used to prove (B).

Last, we prove (22). We note that

$$\frac{L_T(w)}{R_T(w)} - 1 = \frac{E_{\mathbf{X}}(\mathbf{e}_T' \mathbf{P}(\mathbf{w})' \mathbf{P}(\mathbf{w}) \mathbf{e}_T)}{E_{\mathbf{X}}(\mathbf{e}_T' \mathbf{P}(\mathbf{w})' \mathbf{P}(\mathbf{w}) \mathbf{e}_T) + \mu_T' (\mathbf{I} - \mathbf{P}(\mathbf{w}))' (\mathbf{I} - \mathbf{P}(\mathbf{w})) \mu_T}$$

The numerator, $E_{\mathbf{X}}(\mathbf{e}_T' \mathbf{P}(\mathbf{w})' \mathbf{P}(\mathbf{w}) \mathbf{e}_T) = O_p(1)$ uniformly over the $\mathbf{w}$, while the denominator term $\mu_T' (\mathbf{I} - \mathbf{P}(\mathbf{w}))' (\mathbf{I} - \mathbf{P}(\mathbf{w})) \mu_T / T \to_p Constant$ uniformly over all $\mathbf{w}$. As a result, the (22) holds. ■

## Proof of Theorem A

The following Lemma will be useful when we prove Theorem A.

**Lemma A** *Under Assumption 1-3, for any fixed $i$ and any under-specified model $m$, the weights $\hat{w}_m^{MAMF} = o_p(1/T)$.*

**Proof.** Fix an asset $i$ and suppress the asset index in notation. For model $m$, recall the projection matrix $\mathbf{P}^{(m)} := \mathbf{X}^{(m)} \big( (\mathbf{X}^{(m)})' \mathbf{X}^{(m)} \big)^{-1} (\mathbf{X}^{(m)})'$, and define

$$a_m := \mathbf{R}' (\mathbf{I}_T - \mathbf{P}^{(m)}) \mathbf{R}, \qquad K_m := \text{dim of model } m, \qquad K := (K_1, \ldots, K_M)'.$$

Define the alternative weight vector $\widetilde{\mathbf{w}}$ by setting the $m$-th weight to zero and redistribut-

ing it to the largest model $M$:

$$\widetilde{\mathbf{w}} := \big(\widehat{w}_1, \ldots, \widehat{w}_{m-1}, 0, \widehat{w}_{m+1}, \ldots, \widehat{w}_{M-1},\ \widehat{w}_M + \widehat{w}_m\big)'.$$

By optimality, $0 \le C(\widetilde{\mathbf{w}}) - C(\widehat{\mathbf{w}})$. A quadratic expansion identical in spirit to the standard argument yields

$$\begin{aligned} 0 &\le C(\widetilde{\mathbf{w}}) - C(\widehat{\mathbf{w}}) \\ &= \widehat{w}_m^2\,(a_m - a_M) \;+\; 2\widehat{w}_m \sum_{j=1}^{M} \widehat{w}_j \big\{\mathbf{R}'\mathbf{P}^{(M)}\mathbf{P}^{(j)}\mathbf{R} - \mathbf{R}'\mathbf{P}^{(m)}\mathbf{P}^{(j)}\mathbf{R}\big\} \;+\; 2\sigma^2\widehat{w}_m(K_M - K_m). \end{aligned}$$

Therefore, whenever $\widehat{w}_m \neq 0$ we have the bound

$$\widehat{w}_m \;\le\; \frac{2\sigma^2(K_M - K_m) + 2\sum_{j=1}^{M} \widehat{w}_j \{\mathbf{R}'\mathbf{P}^{(m)}\mathbf{P}^{(j)}\mathbf{R} - \mathbf{R}'\mathbf{P}^{(M)}\mathbf{P}^{(j)}\mathbf{R}\}}{a_m - a_M}. \tag{23}$$

We next control the numerator and the denominator stochastically under our time–series conditions. First, by Assumption 1-3, for every fixed $j$, $\mathbf{R}'\mathbf{P}^{(j)}\mathbf{R} = \mathbf{e}'\mathbf{P}^{(j)}\mathbf{e} + 2\mathbf{e}'\mathbf{P}^{(j)}\mathbf{X}^{(j)}\boldsymbol{\gamma}^{(j)} + \{\mathbf{X}^{(j)}\boldsymbol{\gamma}^{(j)}\}'\mathbf{P}^{(j)}\{\mathbf{X}^{(j)}\boldsymbol{\gamma}^{(j)}\}$, and, using the fixed $K_j$, Assumption 1-3 imply the convergence of the score function to a normal distribution, i.e., $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t e_t \;\Rightarrow_d\; \mathcal{N}\Big(0,\ \mathbf{\Omega}_{\infty}^{(m)}\Big)$. As a result, by noting that $\mathbf{P}^{(j)}$ is the projection to column space of $\mathbf{X}^{(m)}$, we have

$$\mathbf{e}'\mathbf{P}^{(j)}\mathbf{e} = O_p(1), \;\; \mathbf{e}'\mathbf{P}^{(j)}\mathbf{X}^{(j)}\boldsymbol{\gamma}^{(j)} = O_p(1),$$

hence $\mathbf{R}'\mathbf{P}^{(m)}\mathbf{P}^{(j)}\mathbf{R} - \mathbf{R}'\mathbf{P}^{(M)}\mathbf{P}^{(j)}\mathbf{R} = O_p(1)$ uniformly over fixed $m, j$. Thus the entire second term in the numerator of (23) is $O_p(1)$ (recall $\sum_j \widehat{w}_j = 1$), and so the numerator is $O_p(1)$.

For the denominator, under–specification means that model $m$ omits at least one relevant regressor/lag with nonzero coefficient. Let $\boldsymbol{\gamma} := (\alpha, \boldsymbol{\beta})$ collect intercept and loadings in the *full* factor–lag space, and write $\boldsymbol{\gamma}^{(m)}$ for the subvector corresponding to model $m$ and $\boldsymbol{\gamma}^{(m)c}$

for the omitted coordinates. A standard algebra gives

$$a_m - a_M = (\mathbf{e} + \mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c})'(\mathbf{I}_n - \mathbf{P}^{(m)})(\mathbf{e} + \mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c}) - \mathbf{e}'(\mathbf{I}_n - \mathbf{P}^{(M)})\mathbf{e}$$
$$= (\mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c})'(\mathbf{I}_n - \mathbf{P}^{(s)})(\mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c}) + 2\mathbf{e}'(\mathbf{I}_n - \mathbf{P}^{(m)})\mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c} - \mathbf{e}'(\mathbf{P}^{(s)} - \mathbf{P}^{(M)})\mathbf{e}.$$

By noting that $\mathbf{P}^{(j)}$ is the projection matrix on the column space of $\mathbf{X}$, and the convergence $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t e_t \ \Rightarrow_d \ \mathcal{N}\Big(0,\ \boldsymbol{\Omega}_{\infty}^{(m)}\Big)$, $2\mathbf{e}'(\mathbf{I}_n - \mathbf{P}^{(m)})\mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c} = O_p(1)$ and $\mathbf{e}'(\mathbf{P}^{(s)} - \mathbf{P}^{(M)})\mathbf{e} = O_p(1)$.

We then show that the $T^{-1}(\mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c})'(\mathbf{I}_n - \mathbf{P}^{(s)})(\mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c}) \to_p$ constant. First, we note that

$$T^{-1}[\mathbf{X}^{(m)}, \mathbf{X}^{(m)^c}]'[\mathbf{X}^{(m)}, \mathbf{X}^{(m)^c}] \to_p \mathbf{Q} = \begin{pmatrix} \mathbf{Q}^{(m)} & \mathbf{Q}^{(m)(m)^c} \\ \mathbf{Q}^{(m)^c(m)} & \mathbf{Q}^{(m)^c} \end{pmatrix}.$$

Since $\mathbf{Q}$ is of full rank, we have

$$\det(\mathbf{Q}) = |\mathbf{Q}^{(m)}||\mathbf{Q}^{(m)^c} - \mathbf{Q}^{(m)^c(m)}(\mathbf{Q}^{(m)})^{-1}\mathbf{Q}^{(m)(m)^c}| > 0,$$

and $|\mathbf{Q}^{(m)}| > 0$. As a result, $|\mathbf{Q}^{(m)^c} - \mathbf{Q}^{(m)^c(m)}(\mathbf{Q}^{(m)})^{-1}\mathbf{Q}^{(m)(m)^c}| > 0$.

Then, by algebra, we can write

$$\begin{aligned} & T^{-1}(\mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c})'(\mathbf{I}_n - \mathbf{P}^{(s)})(\mathbf{X}^{(m)^c}\boldsymbol{\gamma}^{(m)^c}) \\ &= T^{-1}\boldsymbol{\gamma}^{(m)^c\prime}\left(\mathbf{X}^{(m)^c\prime}\mathbf{X}^{(m)^c} - \mathbf{X}^{(m)^c\prime}\mathbf{X}^{(m)}(\mathbf{X}^{(m)\prime}\mathbf{X}^{(m)})^{-1}\mathbf{X}^{(m)\prime}\mathbf{X}^{(m)^c}\right)\boldsymbol{\gamma}^{(m)^c} \\ &\to_p \boldsymbol{\gamma}^{(m)^c\prime}\left(\mathbf{Q}^{(m)^c} - \mathbf{Q}^{(m)^c(m)}(\mathbf{Q}^{(m)})^{-1}\mathbf{Q}^{(m)(m)^c}\right)\boldsymbol{\gamma}^{(m)^c} > 0. \end{aligned}$$

Hence $a_m - a_M = O_p(T)$.

Combining the $O_p(1)$ numerator and the $O_p(T)$ denominator in (23) yields $\widehat{w}_m = O_p(T^{-1})$ for any under–specified $m \le M_0$. This completes the proof. ■

**Main Proof.** We first recall that the objective function

$$
\begin{aligned}
L_n(\mathbf{w}) &:= \left\|(\mathbf{I}_T - \mathbf{P}(\mathbf{w}))\,\mathbf{R}\right\|^2 \;+\; 2\,\hat{\sigma}^2\,\mathbf{w}'K - \|\sum_m w_m\mathbf{e}\|^2 \\
&= \left\|\sum_m w_m(\mathbf{I}_T - \mathbf{P}^{(m)})\,\mathbf{R}\right\|^2 \;+\; 2\,\hat{\sigma}^2\sum_m w_m K_m - \|\sum_m w_m\mathbf{e}\|^2 \\
&= \Big\{\|\sum_{m\le M_0} w_m(\mathbf{I}_T - \mathbf{P}^{(m)})\,\mathbf{R}\|^2 + \|\sum_{m\ge M_0+1} w_m(\mathbf{I}_T - \mathbf{P}^{(m)})\,\mathbf{R}\|^2 \\
&+ 2(\sum_{m\le M_0} w_m(\mathbf{I}_T - \mathbf{P}^{(m)})\,\mathbf{R})'(\sum_{m\ge M_0+1} w_m(\mathbf{I}_T - \mathbf{P}^{(m)})\,\mathbf{R}) \\
&+ 2\,\hat{\sigma}^2\sum_{m\le M_0} w_m K_m + 2\,\hat{\sigma}^2\sum_{m\ge M_0+1} w_m K_m \\
&- \|\sum_{m\le M_0} w_m\mathbf{e}\|^2 - \|\sum_{m\ge M_0+1} w_m\mathbf{e}\|^2 - 2(\sum_{m\le M_0} w_m\mathbf{e})'(\sum_{m\ge M_0+1} w_m\mathbf{e})\Big\}
\end{aligned}
$$

Note that $\widehat{\mathbf{w}}^{MAMF} = \arg\min_{w\in\mathcal{W}} L_n$. Then we define

$$
\begin{aligned}
\tilde{L}_n(\mathbf{w}) := \Big\{&\|\sum_{m\ge M_0+1} w_m(\mathbf{I}_T - \mathbf{P}^{(m)})\,\mathbf{R}\|^2 \\
&+ 2\,\hat{\sigma}^2\sum_{m\ge M_0+1} w_m K_m - \|\sum_{m\ge M_0+1} w_m\mathbf{e}\|^2.\Big\}
\end{aligned}
$$

Let $\tilde{\mathcal{W}} = \{(w_{i,M_0+1},...,w_{i,M}) : \sum_{m=M_0+1}^{M} w_{i,m} = 1\}$, and define $\tilde{\mathbf{w}} = \arg\max_{\mathbf{w}\in\tilde{\mathcal{W}}} \tilde{L}_n(\mathbf{w})$.

Define a weighting $\ddot{\mathbf{w}} \in \tilde{\mathcal{W}}$ such that

$$
\ddot{w}_m = \widehat{w}_m^{MAMF} + \frac{1}{M_0}\sum_{j=1}^{M_0}\widehat{w}_j^{MAMF}
$$

Since $\widehat{w}_m = O_p(1/T)$ for all $m \leq M_0$, we have

$$\Big\| \sum_{m \leq M_0} w_m (\mathbf{I}_T - \mathbf{P}^{(m)})\, \mathbf{R} \Big\|^2 = O_p(1/T)$$

$$(\sum_{m \leq M_0} w_m (\mathbf{I}_T - \mathbf{P}^{(m)})\, \mathbf{R})' (\sum_{m \geq M_0+1} w_m (\mathbf{I}_T - \mathbf{P}^{(m)})\, \mathbf{R}) = O_p(1/\sqrt{T}) \times O_p(1/\sqrt{T}) = O_p(1/T)$$

$$\hat{\sigma}^2 \sum_{m \leq M_0} w_m K_m = O_p(1/T)$$

$$\Big\| \sum_{m \leq M_0} w_m \mathbf{e} \Big\|^2 - 2(\sum_{m \leq M_0} w_m \mathbf{e})' (\sum_{m \geq M_0+1} w_m \mathbf{e}) = O_p(1/T).$$

and therefore

$$\begin{aligned} \|L_n(\widehat{\mathbf{w}}^{MAMF}) - \tilde{L}_n(\ddot{\mathbf{w}})\| &\leq \|L_n(\widehat{\mathbf{w}}^{MAMF}) - \tilde{L}_n((\widehat{w}_{M_0+1}^{MAMF}, ..., \widehat{w}_M^{MAMF}))\| \\ &\quad + \|\tilde{L}_n((\widehat{w}_{M_0+1}^{MAMF}, ..., \widehat{w}_M^{MAMF})) - \tilde{L}_n(\ddot{\mathbf{w}})\| \\ &= O_p(1/T) + \|\tilde{L}_n((\widehat{w}_{M_0+1}^{MAMF}, ..., \widehat{w}_M^{MAMF})) - \tilde{L}_n(\ddot{\mathbf{w}})\| = O_p(1/T) \end{aligned} \tag{24}$$

where in the last step we use again the fact that $\widehat{w}_m^{MAMF} = O_p(1/T)$ for under-specified models.

We can then derive the following inequalities:

$$\begin{aligned} \tilde{L}_n(\tilde{\mathbf{w}}) &\geq_{(i)} \tilde{L}_n(\ddot{\mathbf{w}}) \\ &\geq_{(ii)} L_n(\widehat{\mathbf{w}}^{MAMF}) - O_p(1/T) \\ &\geq_{(iii)} L_n((\mathbf{0}_{M_0}, \tilde{\mathbf{w}})) - O_p(1/T) \\ &=_{(iv)} \tilde{L}_n(\tilde{\mathbf{w}}) - O_p(1/T), \end{aligned}$$

where step $(i)$ follows by the optimality of $\tilde{\mathbf{w}}$, step $(ii)$ follows by our previous derivation (24), step $(iii)$ follows by the optimality of $\widehat{\mathbf{w}}^{MAMF}$ for $L_n$, and $\mathbf{0}_{M_0} = (0, ..., 0)_{1 \times M_0}$, and step $(iv)$ is by the definition of the $\tilde{L}_n$. Therefore, we show that $\ddot{\mathbf{w}}$ is the $O_p(1/T)$ optimal solution to $\tilde{L}_n(\mathbf{w})$. By the argmax theorem, since $\tilde{L}_n(\lambda) \Rightarrow_d \lambda \in \mathbf{\Gamma}\lambda$, $\ddot{\mathbf{w}} \Rightarrow_d \arg\max \lambda \in \mathbf{\Gamma}\lambda$.

Last, because $\widehat{\mathbf{w}}^{MAMF} = \ddot{\mathbf{w}} + O_p(1/T)$, we also have $\widehat{\mathbf{w}}^{MAMF} \Rightarrow_d \arg\max \lambda \in \mathbf{\Gamma}\lambda$.

The convergence of $\widehat{\alpha}(\widehat{\mathbf{w}}^{MAMF})$ and $\widehat{\boldsymbol{\beta}}(\widehat{\mathbf{w}}^{MAMF})$ follows by the continuous mapping theorem and the Slutsky's theorem. ■

## Proof of Theorem 2

**Proof.** Following the proof of Lemma A, we have

$$\widehat{w}_m \;\leq\; \frac{2\phi_T\sigma^2(K_M - K_m) + 2\sum_{j=1}^{M}\widehat{w}_j\{\mathbf{R}'\mathbf{P}^{(m)}\mathbf{P}^{(j)}\mathbf{R} - \mathbf{R}'\mathbf{P}^{(M)}\mathbf{P}^{(j)}\mathbf{R}\}}{a_m - a_M}, \tag{25}$$

where we have shown that $a_m - a_M = O_p(T)$. After multiplying $\phi_T$, the numerator is of order $O_p(\phi_T)$, so we can conclude that for under-fitted model $m$, $\breve{w}_m = O_p(\phi_T/T)$.

For the overfitted model, we similarly consider

$$\tilde{\mathbf{w}}_m = (\breve{w}_1, ..., \breve{w}_{M_0+1} + \breve{w}_m, \breve{w}_{m-1}, 0, \breve{w}_{m+1}, \breve{w}_M),$$

and following the same derivation in the proof of Proposition A, we can derive

$$\begin{aligned} &0 \leq \breve{\mathcal{C}}(\tilde{\mathbf{w}}_m) - \breve{\mathcal{C}}(\breve{\mathbf{w}}) \\ &= \breve{w}_m^2\,(a_m - a_M) \;+\; 2\breve{w}_m \sum_{j=1}^{M} \breve{w}_j\big\{\mathbf{R}'\mathbf{P}^{(M)}\mathbf{P}^{(j)}\mathbf{R} - \mathbf{R}'\mathbf{P}^{(m)}\mathbf{P}^{(j)}\mathbf{R}\big\} \;+\; 2\Phi_T\sigma^2\breve{w}_m(K_{M_0+1} - K_m). \end{aligned} \tag{26}$$

Since $m$ and $M$ are both over-specified model, we have

$$a_m - a_M = \mathbf{R}'(p^{(M)} - p^{(m)})\mathbf{R}' = O_p(1).$$

The second term

$$2\breve{w}_m \sum_{j=1}^{M} \breve{w}_j\big\{\mathbf{R}'\mathbf{P}^{(M)}\mathbf{P}^{(j)}\mathbf{R} - \mathbf{R}'\mathbf{P}^{(m)}\mathbf{P}^{(j)}\mathbf{R}\big\} \leq 2\sum_{j=1}^{M} \breve{w}_j\big\{\mathbf{R}'\mathbf{P}^{(M)}\mathbf{P}^{(j)}\mathbf{R} - \mathbf{R}'\mathbf{P}^{(m)}\mathbf{P}^{(j)}\mathbf{R}\big\} = O_p(1)$$

by the proof of Lemma A.

Last, by the (26), and the fact that $K_{M_0+1} - K_m < 0$ for over-fitted model $m$, we know that $0 \leq O_p(1) + \phi_T \breve{w}_m \times (\text{negative})$, so it must be the case that $\phi_T \breve{w}_m = O_p(1)$, which proves that $\breve{w}_m = O_p(1/\phi_T)$.

Next, we show that the confidence interval in Algorithm 2 is valid. Since the weights for the misspecified models and overspecified models (i.e., $m \neq M_0 + 1$), the weights $\breve{w}_m \to_p 0$, then with probability approaching one, $\breve{w}_{M_0+1}$ will be the largest weights and

$$\liminf \Pr(\widehat{M_0} = M_0) \to 1.$$

Therefore, with probability approaching 1, we will be simulating the limit distribution for the true model $M_0 + 1$, and $|\mathbf{\Lambda}_j^{(r)}(\widehat{M_0} + 1) - \mathbf{\Lambda}^{(r)}(M_0 + 1)_j| = o_p(1)$.

Let $q_j^0(\tau/2)$ and $q_j^0(1-\tau/2)$ be the $\tau/2$ and $1-\tau/2$ quantile of the true limit $\Lambda(M_0)^(r)_j$, then $|q_j(\tau/2) - q_j(\tau/2)| = o_p(1)$ and $|q_j(1-\tau/2) - q_j(1-\tau/2)| = o_p(1)$. Finally,

$$\begin{aligned}
\Pr(\gamma_j \in CI) &= Pr(\gamma_j \in \widehat{\gamma}_j(\widehat{\mathbf{w}}^{\text{MAMF}}) - T^{-1/2} q_j(1-\tau/2),\ \widehat{\gamma}_j(\widehat{\mathbf{w}}^{\text{MAMF}}) - T^{-1/2} q_j(\tau/2)) \\
&= \Pr(\gamma_j \in \widehat{\gamma}_j(\widehat{\mathbf{w}}^{\text{MAMF}}) - T^{-1/2} q_j^0(1-\tau/2),\ \widehat{\gamma}_j(\widehat{\mathbf{w}}^{\text{MAMF}}) - T^{-1/2} q_j^0(\tau/2)) + o_p(1) \\
&= \tau + o_p(1).
\end{aligned}$$

So, the proposed confidence interval covers the true parameter with probability $1-\tau$ asymptotically. ■

# C Weak Factor Robustness

This appendix reports the weak-factor simulation results described in Section 3. The design is identical to the baseline except that the five excluded factors (LIQ, MGMT, PERF, QMJ, BAB) receive small nonzero loadings. For each excluded factor, the loading magnitude is set to the 25th percentile of estimated absolute exposures across the $J$ testing-factor portfolios,

and the sign is set to the cross-sectional median. Tables A.1 and A.2 report accuracy and inference results, respectively.

The key patterns from the baseline carry over. MAMF delivers MAE and RMSE close to the best single-model estimator across all designs, and its coverage remains near the nominal 95% level. The main difference relative to the baseline is that the under-specified models (CAPM, FF3, Carhart4) now exhibit somewhat different bias profiles, because the omitted-factor bias depends on the nonzero loadings of the excluded factors. Among the correctly or over-specified models, FF6 coverage drops modestly (to roughly 90%) because it now omits five weakly relevant factors, while the richer specifications (9F, Full11) and MAMF maintain near-nominal coverage. This confirms that MAMF adapts to mild misspecification by reallocating weight toward specifications that absorb the weak exposures.

# D Additional Empirical Results

This appendix provides supplementary evidence that complements the results reported in the main text. In particular, we extend the analysis of benchmark sensitivity of alpha estimates to the full cross-section of test assets. While Figure 5 in the main text illustrates the dispersion of estimated alphas for a randomly selected subset of 30 anomalies, the figures in this appendix report the corresponding results for the complete set of 148 test assets.

In addition, Table A.3 summarizes the statistical significance of alpha estimates across alternative benchmark specifications for all test assets. The table documents substantial heterogeneity in significance outcomes across benchmark models, highlighting that both the magnitude and statistical significance of alphas are highly sensitive to the choice of factor specification. Taken together, the appendix evidence demonstrates that the pronounced benchmark dependence of alpha estimates is a pervasive feature of the entire cross-section, rather than an artifact of subsample selection.

Table A.1: Weak-factor simulation: alpha estimation accuracy

| | | High Beta Case | | | Low Beta Case | | |
|---|---|---|---|---|---|---|---|
| | Method | MAE | RMSE | SD | MAE | RMSE | SD |
| $\alpha = 2\%$ | MAMF | 0.0723 | 0.0901 | 0.0881 | 0.0708 | 0.0883 | 0.0881 |
| | CAPM | 0.8954 | 0.9238 | 0.2274 | 0.2571 | 0.2813 | 0.1166 |
| | FF3 | 0.6145 | 0.6298 | 0.1378 | 0.1380 | 0.1582 | 0.0833 |
| | Carhart4 | 0.4335 | 0.4505 | 0.1225 | 0.0647 | 0.0811 | 0.0802 |
| | FF6 | 0.0754 | 0.0935 | 0.0804 | 0.0754 | 0.0935 | 0.0804 |
| | 7F | 0.0778 | 0.0970 | 0.0896 | 0.0778 | 0.0970 | 0.0895 |
| | 9F | 0.0733 | 0.0917 | 0.0916 | 0.0732 | 0.0917 | 0.0915 |
| | Full11 | 0.0743 | 0.0930 | 0.0930 | 0.0743 | 0.0930 | 0.0930 |
| $\alpha = 1\%$ | MAMF | 0.0716 | 0.0896 | 0.0878 | 0.0704 | 0.0880 | 0.0878 |
| | CAPM | 0.8963 | 0.9250 | 0.2289 | 0.2598 | 0.2838 | 0.1168 |
| | FF3 | 0.6153 | 0.6310 | 0.1397 | 0.1399 | 0.1598 | 0.0831 |
| | Carhart4 | 0.4339 | 0.4511 | 0.1236 | 0.0657 | 0.0819 | 0.0806 |
| | FF6 | 0.0749 | 0.0931 | 0.0811 | 0.0749 | 0.0930 | 0.0811 |
| | 7F | 0.0768 | 0.0960 | 0.0890 | 0.0768 | 0.0959 | 0.0891 |
| | 9F | 0.0727 | 0.0911 | 0.0909 | 0.0728 | 0.0912 | 0.0909 |
| | Full11 | 0.0737 | 0.0922 | 0.0922 | 0.0737 | 0.0922 | 0.0922 |
| $\alpha = 0\%$ | MAMF | 0.0722 | 0.0904 | 0.0885 | 0.0707 | 0.0887 | 0.0884 |
| | CAPM | 0.8930 | 0.9216 | 0.2279 | 0.2563 | 0.2807 | 0.1165 |
| | FF3 | 0.6142 | 0.6297 | 0.1388 | 0.1377 | 0.1579 | 0.0828 |
| | Carhart4 | 0.4342 | 0.4516 | 0.1239 | 0.0650 | 0.0811 | 0.0800 |
| | FF6 | 0.0754 | 0.0937 | 0.0807 | 0.0754 | 0.0937 | 0.0807 |
| | 7F | 0.0775 | 0.0968 | 0.0895 | 0.0775 | 0.0968 | 0.0895 |
| | 9F | 0.0731 | 0.0917 | 0.0916 | 0.0732 | 0.0917 | 0.0916 |
| | Full11 | 0.0740 | 0.0929 | 0.0929 | 0.0741 | 0.0929 | 0.0929 |

*Notes.* The table reports simulation-based accuracy measures for estimating monthly alpha under the weak-factor design, in which the five excluded factors receive small nonzero loadings (see text for calibration details). MAE is the mean absolute error, RMSE is the root mean squared error, and SD is the standard deviation of the estimated alpha across replications. Results are reported separately for the high-beta and low-beta designs. Each entry is computed from $B = 10{,}000$ simulation replications with $T = 480$.

Table A.2: Weak-factor simulation: alpha inference performance

| | | High Beta Case | | Low Beta Case | |
|---|---|---|---|---|---|
| | Method | Coverage | AvgLen | Coverage | AvgLen |
| $\alpha = 2\%$ | MAMF | 94.30 | 0.3460 | 94.93 | 0.3478 |
| | CAPM | 0.39 | 0.7509 | 35.07 | 0.4209 |
| | FF3 | 0.26 | 0.5107 | 63.25 | 0.3242 |
| | Carhart4 | 5.06 | 0.4769 | 94.34 | 0.3109 |
| | FF6 | 90.22 | 0.3117 | 90.21 | 0.3117 |
| | 7F | 92.35 | 0.3454 | 92.37 | 0.3455 |
| | 9F | 94.25 | 0.3512 | 94.27 | 0.3513 |
| | Full11 | 94.29 | 0.3559 | 94.26 | 0.3560 |
| $\alpha = 1\%$ | MAMF | 94.53 | 0.3456 | 94.74 | 0.3471 |
| | CAPM | 0.41 | 0.7519 | 34.17 | 0.4207 |
| | FF3 | 0.24 | 0.5102 | 62.04 | 0.3239 |
| | Carhart4 | 5.11 | 0.4762 | 94.33 | 0.3104 |
| | FF6 | 90.17 | 0.3114 | 90.18 | 0.3114 |
| | 7F | 92.56 | 0.3452 | 92.57 | 0.3451 |
| | 9F | 94.53 | 0.3509 | 94.52 | 0.3509 |
| | Full11 | 94.34 | 0.3556 | 94.38 | 0.3555 |
| $\alpha = 0\%$ | MAMF | 94.05 | 0.3456 | 94.54 | 0.3471 |
| | CAPM | 0.48 | 0.7513 | 35.51 | 0.4212 |
| | FF3 | 0.33 | 0.5104 | 62.94 | 0.3242 |
| | Carhart4 | 5.33 | 0.4766 | 94.41 | 0.3107 |
| | FF6 | 90.19 | 0.3116 | 90.19 | 0.3116 |
| | 7F | 92.19 | 0.3450 | 92.18 | 0.3450 |
| | 9F | 94.00 | 0.3507 | 94.00 | 0.3507 |
| | Full11 | 94.00 | 0.3555 | 94.00 | 0.3555 |

*Notes.* The table reports simulation-based inference measures for monthly alpha under the weak-factor design. Coverage is the empirical coverage rate in percent for the nominal 95% confidence interval. AvgLen is the average confidence-interval length. Results are reported separately for the high-beta and low-beta designs. Each entry is computed from $B = 10{,}000$ simulation replications with $T = 480$.

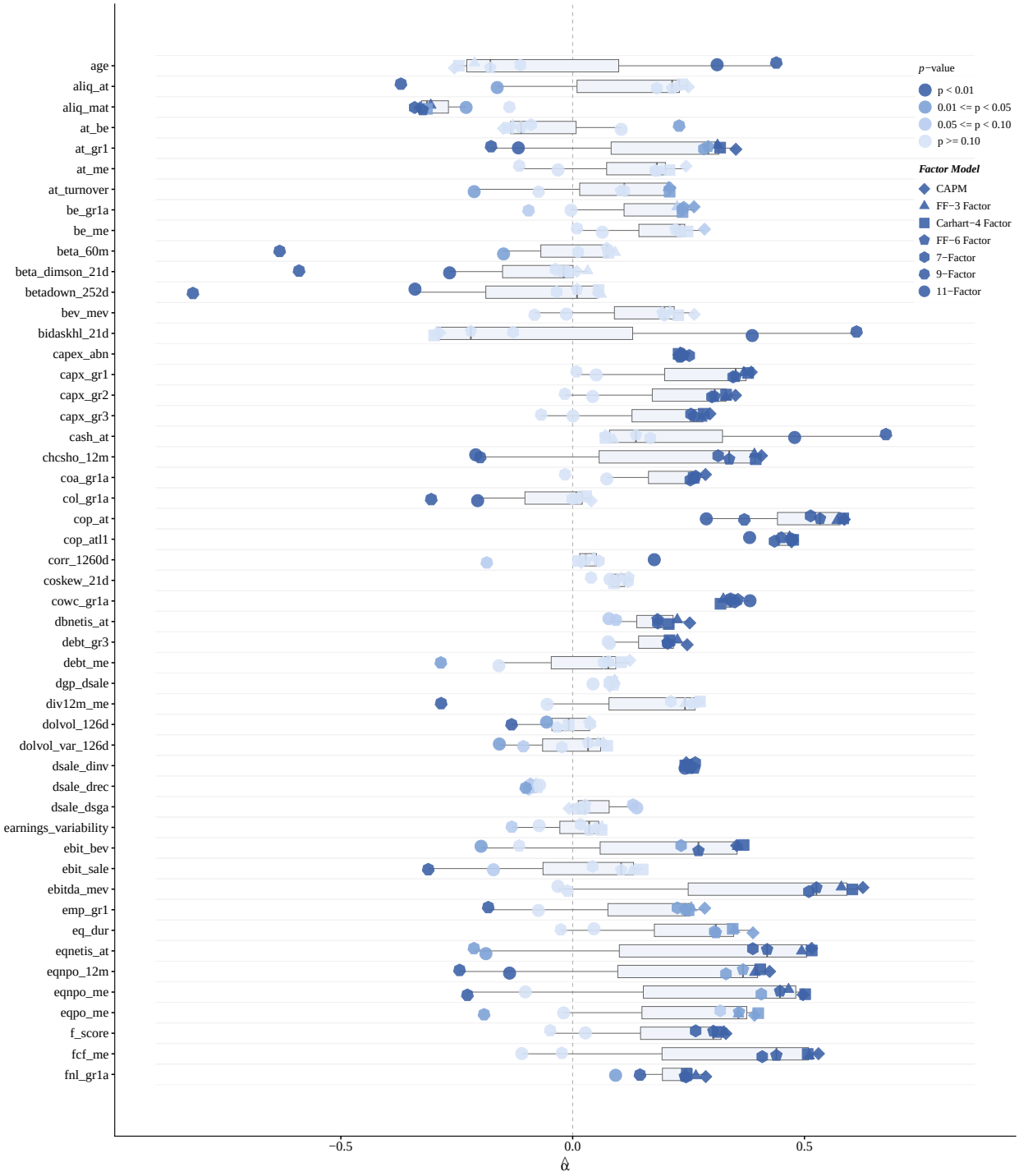



Figure 5: Distribution of Estimated Alphas. The figure reports alpha estimates across the seven candidate benchmark models for 30 anomalies randomly sampled from the 148 test assets.

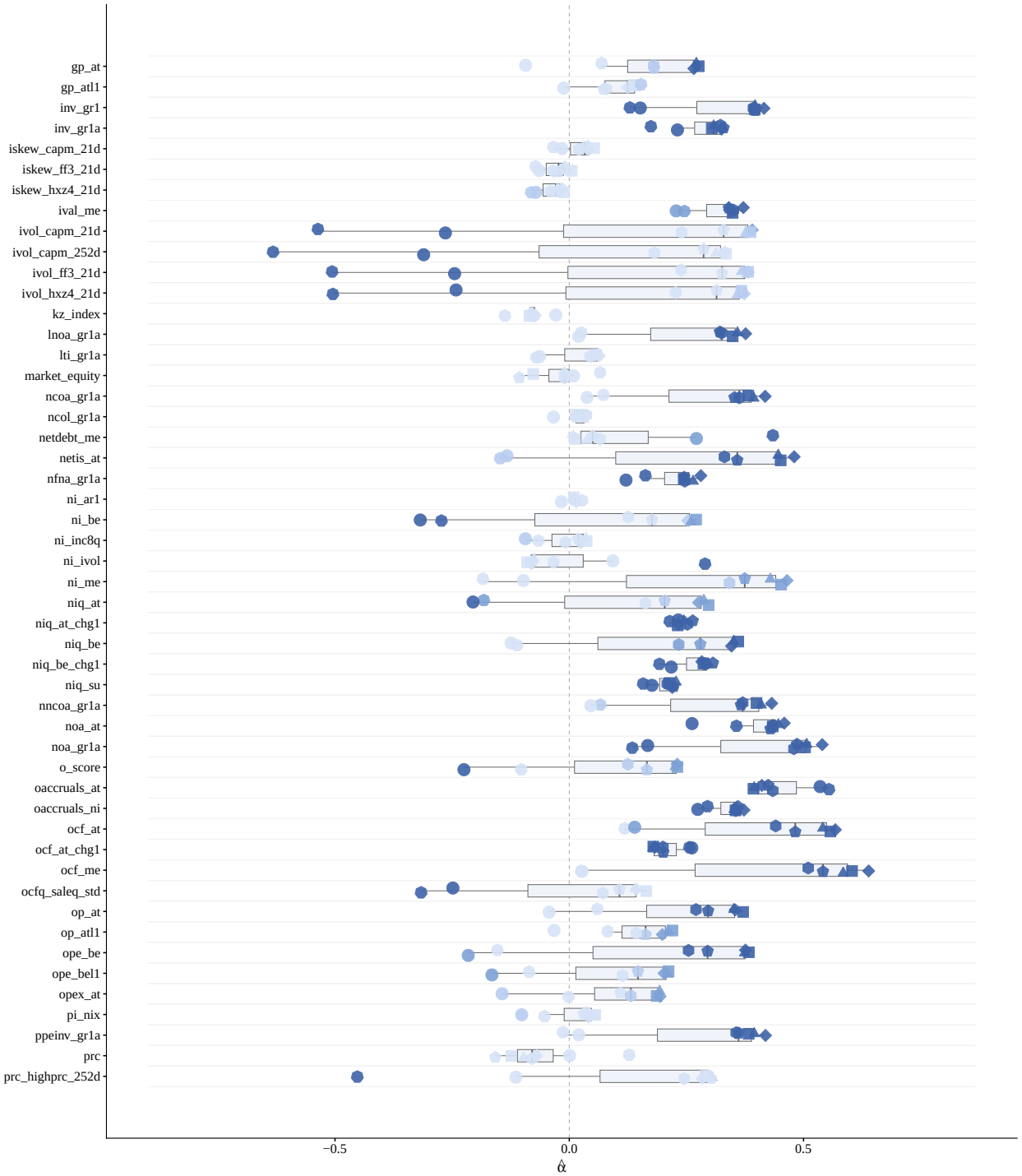



Figure 5: (continued)

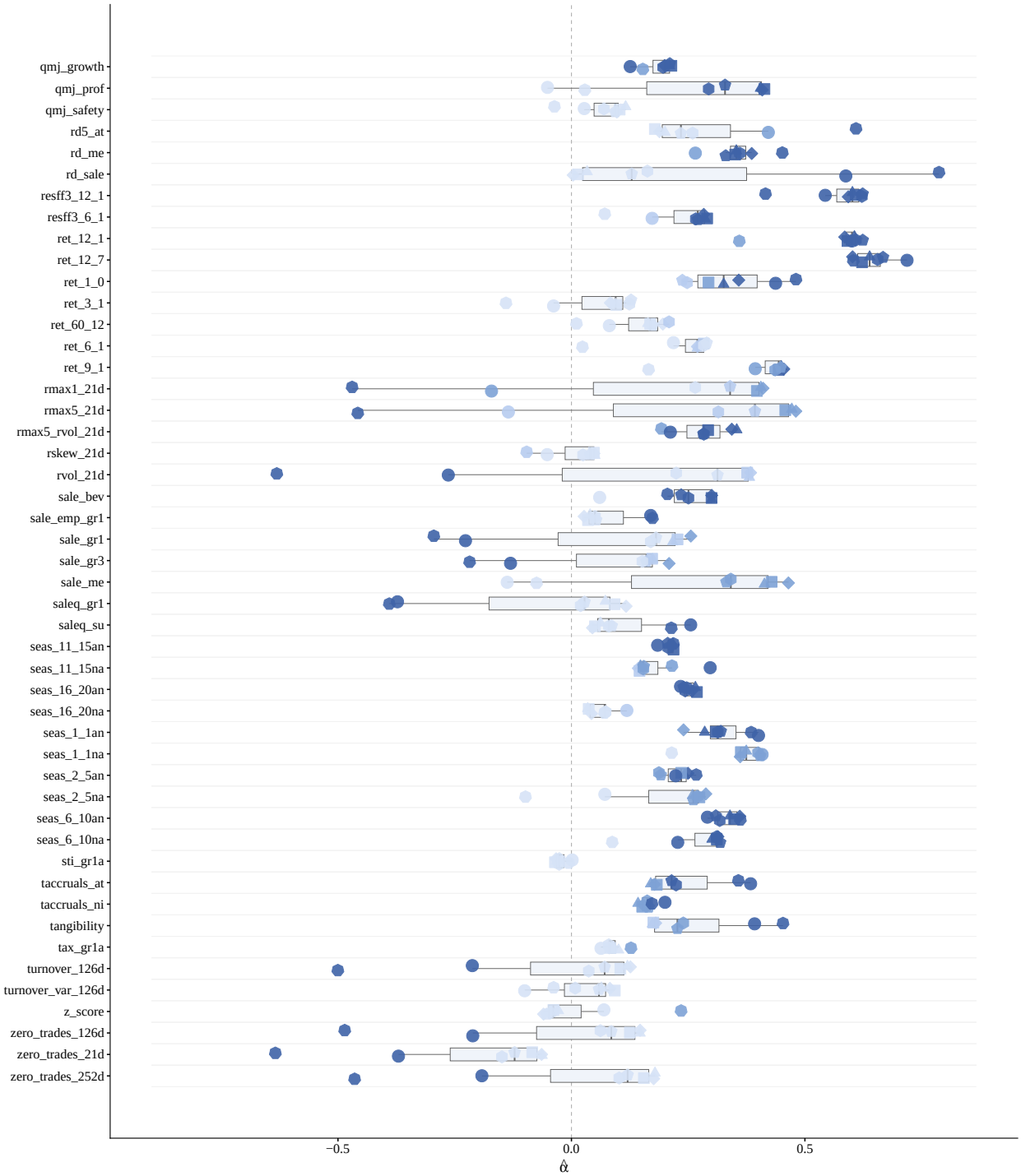



Figure 5: (continued)

Table A.3: Significance of Alpha Estimates across Benchmark Models

| Asset | MAMF | CAPM | FF3 | Carhart4 | FF6 | 7F | 9F | 11F | # Sig. |
|---|---|---|---|---|---|---|---|---|---|
| age | ✓ | | | | | | ✓ | ✓ | 3 |
| aliq_at | ✓ | | | | | | ✓ | ✓ | 3 |
| aliq_mat | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | 7 |
| at_be | | | | | | | ✓ | | 1 |
| at_gr1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| at_me | | | | | | | | | 0 |
| at_turnover | ✓ | ✓ | ✓ | ✓ | | | | ✓ | 5 |
| be_gr1a | | ✓ | | ✓ | ✓ | ✓ | | | 4 |
| be_me | | | | | | | | | 0 |
| beta_60m | ✓ | | | | | | ✓ | ✓ | 3 |
| beta_dimson_21d | ✓ | | | | | | ✓ | ✓ | 3 |
| betadown_252d | ✓ | | | | | | ✓ | ✓ | 3 |
| bev_mev | | | | | | | | | 0 |
| bidaskhl_21d | ✓ | | | | | | ✓ | ✓ | 3 |
| capex_abn | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| capx_gr1 | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| capx_gr2 | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| capx_gr3 | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| cash_at | ✓ | | | | | | ✓ | ✓ | 3 |
| chcsho_12m | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| coa_gr1a | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| col_gr1a | ✓ | | | | | | ✓ | ✓ | 3 |
| cop_at | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |

*Notes:* A checkmark (✓) indicates statistical significance at the 5% level. MAMF denotes the Mallows-type model averaging estimator. The final column reports the number of benchmark models under which the alpha is significant.

*Table A.3 (continued)*

| Asset | MAMF | CAPM | FF3 | Carhart4 | FF6 | 7F | 9F | 11F | # Sig. |
|---|---|---|---|---|---|---|---|---|---|
| cop_atl1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| corr_1260d | ✓ | | | | | | | ✓ | 2 |
| coskew_21d | | | | | | | | | 0 |
| cowc_gr1a | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| dbnetis_at | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| debt_gr3 | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| debt_me | | | | | | | ✓ | | 1 |
| dgp_dsale | | | | | | | | | 0 |
| div12m_me | | | | | | | ✓ | | 1 |
| dolvol_126d | | | | | | | ✓ | ✓ | 2 |
| dolvol_var_126d | ✓ | | | | | | | ✓ | 2 |
| dsale_dinv | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| dsale_drec | | | | | | | ✓ | | 1 |
| dsale_dsga | | | | | | | | | 0 |
| earnings_variability | | | | | | | | | 0 |
| ebit_bev | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | 7 |
| ebit_sale | | | | | | | ✓ | | 1 |
| ebitda_mev | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| emp_gr1 | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | 6 |
| eq_dur | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| eqnetis_at | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| eqnpo_12m | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| eqnpo_me | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | 6 |

*Notes:* A checkmark (✓) indicates statistical significance at the 5% level. MAMF denotes the Mallows-type model averaging estimator. The final column reports the number of benchmark models under which the alpha is significant.

*Table A.3 (continued)*

| Asset | MAMF | CAPM | FF3 | Carhart4 | FF6 | 7F | 9F | 11F | # Sig. |
|---|---|---|---|---|---|---|---|---|---|
| eqpo_me | | ✓ | ✓ | ✓ | ✓ | | ✓ | | 5 |
| f_score | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| fcf_me | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| fnl_gr1a | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| gp_at | | ✓ | ✓ | ✓ | | | | | 3 |
| gp_atl1 | | | | | | | | | 0 |
| inv_gr1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| inv_gr1a | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| iskew_capm_21d | | | | | | | | | 0 |
| iskew_ff3_21d | | | | | | | | | 0 |
| iskew_hxz4_21d | | | | | | | | | 0 |
| ival_me | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| ivol_capm_21d | ✓ | ✓ | | | | | ✓ | ✓ | 4 |
| ivol_capm_252d | ✓ | | | | | | ✓ | ✓ | 3 |
| ivol_ff3_21d | ✓ | | | | | | ✓ | ✓ | 3 |
| ivol_hxz4_21d | ✓ | | | | | | ✓ | ✓ | 3 |
| kz_index | | | | | | | | | 0 |
| lnoa_gr1a | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| lti_gr1a | | | | | | | | | 0 |
| market_equity | | | | | | | | | 0 |
| ncoa_gr1a | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| ncol_gr1a | | | | | | | | | 0 |
| netdebt_me | ✓ | | | | | | ✓ | ✓ | 3 |

*Notes:* A checkmark (✓) indicates statistical significance at the 5% level. MAMF denotes the Mallows-type model averaging estimator. The final column reports the number of benchmark models under which the alpha is significant.

*Table A.3 (continued)*

| Asset | MAMF | CAPM | FF3 | Carhart4 | FF6 | 7F | 9F | 11F | # Sig. |
|---|---|---|---|---|---|---|---|---|---|
| netis_at | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| nfna_gr1a | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| ni_ar1 | | | | | | | | | 0 |
| ni_be | ✓ | | ✓ | ✓ | | | ✓ | ✓ | 5 |
| ni_inc8q | | | | | | | | | 0 |
| ni_ivol | | | | | | | ✓ | | 1 |
| ni_me | | ✓ | ✓ | ✓ | ✓ | | | | 4 |
| niq_at | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | 6 |
| niq_at_chg1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| niq_be | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| niq_be_chg1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| niq_su | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| nncoa_gr1a | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| noa_at | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| noa_gr1a | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| o_score | ✓ | ✓ | ✓ | ✓ | | | | ✓ | 5 |
| oaccruals_at | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| oaccruals_ni | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| ocf_at | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | 7 |
| ocf_at_chg1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| ocf_me | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| ocfq_saleq_std | ✓ | | | | | | ✓ | ✓ | 3 |
| op_at | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |

*Notes:* A checkmark (✓) indicates statistical significance at the 5% level. MAMF denotes the Mallows-type model averaging estimator. The final column reports the number of benchmark models under which the alpha is significant.

*Table A.3 (continued)*

| Asset | MAMF | CAPM | FF3 | Carhart4 | FF6 | 7F | 9F | 11F | # Sig. |
|---|---|---|---|---|---|---|---|---|---|
| op_atl1 | | ✓ | ✓ | ✓ | | | | | 3 |
| ope_be | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | 7 |
| ope_bel1 | ✓ | ✓ | ✓ | ✓ | | | | ✓ | 5 |
| opex_at | | ✓ | ✓ | ✓ | | | | | 3 |
| pi_nix | | | | | | | | | 0 |
| ppeinv_gr1a | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| prc | | | | | | | | | 0 |
| prc_highprc_252d | | | | | | | ✓ | | 1 |
| qmj_growth | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| qmj_prof | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| qmj_safety | | | | | | | | | 0 |
| rd5_at | ✓ | | | | | | ✓ | ✓ | 3 |
| rd_me | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| rd_sale | ✓ | | | | | | ✓ | ✓ | 3 |
| resff3_12_1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| resff3_6_1 | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| ret_12_1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| ret_12_7 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| ret_1_0 | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | 6 |
| ret_3_1 | | | | | | | | | 0 |
| ret_60_12 | | | | | | | | | 0 |
| ret_6_1 | | | | | | | | | 0 |
| ret_9_1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | 7 |

*Notes:* A checkmark (✓) indicates statistical significance at the 5% level. MAMF denotes the Mallows-type model averaging estimator. The final column reports the number of benchmark models under which the alpha is significant.

*Table A.3 (continued)*

| Asset | MAMF | CAPM | FF3 | Carhart4 | FF6 | 7F | 9F | 11F | # Sig. |
|---|---|---|---|---|---|---|---|---|---|
| rmax1_21d | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | 6 |
| rmax5_21d | | ✓ | ✓ | ✓ | | | ✓ | | 4 |
| rmax5_rvol_21d | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| rskew_21d | | | | | | | | | 0 |
| rvol_21d | ✓ | | | | | | ✓ | ✓ | 3 |
| sale_bev | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | 6 |
| sale_emp_gr1 | ✓ | | | | | | ✓ | ✓ | 3 |
| sale_gr1 | ✓ | ✓ | | | | | ✓ | ✓ | 4 |
| sale_gr3 | ✓ | ✓ | | | | | ✓ | ✓ | 4 |
| sale_me | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| saleq_gr1 | ✓ | | | | | | ✓ | ✓ | 3 |
| saleq_su | ✓ | | | | | | ✓ | ✓ | 3 |
| seas_11_15an | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| seas_11_15na | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | 6 |
| seas_16_20an | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| seas_16_20na | | | | | | | | | 0 |
| seas_1_1an | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| seas_1_1na | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | 7 |
| seas_2_5an | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| seas_2_5na | | ✓ | ✓ | ✓ | ✓ | ✓ | | | 5 |
| seas_6_10an | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| seas_6_10na | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | 7 |
| sti_gr1a | | | | | | | | | 0 |

*Notes:* A checkmark (✓) indicates statistical significance at the 5% level. MAMF denotes the Mallows-type model averaging estimator. The final column reports the number of benchmark models under which the alpha is significant.

*Table A.3 (continued)*

| Asset | MAMF | CAPM | FF3 | Carhart4 | FF6 | 7F | 9F | 11F | # Sig. |
|---|---|---|---|---|---|---|---|---|---|
| taccruals_at | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| taccruals_ni | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 8 |
| tangibility | ✓ | | | | ✓ | ✓ | ✓ | ✓ | 5 |
| tax_gr1a | | | | | | | ✓ | | 1 |
| turnover_126d | ✓ | | | | | | ✓ | ✓ | 3 |
| turnover_var_126d | | | | | | | | | 0 |
| z_score | | | | | | | ✓ | | 1 |
| zero_trades_126d | ✓ | | | | | | ✓ | ✓ | 3 |
| zero_trades_21d | ✓ | | | | | | ✓ | ✓ | 3 |
| zero_trades_252d | ✓ | | | | | | ✓ | ✓ | 3 |