Model Confidence Sets in Multivariate Systems*

Job Market Paper

Florian Richard
Carleton University†

October 13th, 2021
[Click here for the latest version]

Abstract

This paper provides a generalization of a model confidence set (MCS) procedure, as originally introduced by Hansen et al. (2011) for univariate models, to systems of $P > 1$ dependent variables. A $(1-\alpha)$ level MCS collects the set of models with equal predictive ability, based on a sequential elimination procedure that relies on an equivalence test. I introduce supremum-type t and Hotelling-type T^2 statistics which account for correlation between loss differentials. I assess the performance of 14 candidate asset pricing models using the Fama and French research portfolios, with monthly data for the period 1972-2013. Under quadratic loss, I find that for out-of-sample tests with the T^2 statistic using 12, 18 and 25 portfolios, the prominent Fama and French (2015) model is the only selected model at the 1-year prediction horizon, but the MCS often includes multiple competing models at the 2- and 5-year horizons, featuring liquidity and mispricing factors. For in-sample tests, models are much harder to distinguish, particularly when the number of test assets is small. Overall, out-of-sample tests and a larger number of more heterogenous test assets provide more information to disentangle models. The market-based capital asset pricing model is never included in the MCS. The procedure shows good size and power properties in simulations.

Keywords: Equal predictive ability, confidence set, factor models, multiple testing.
JEL Codes: C52, C53, G12.

*I am thankful to the participants of the Carleton University Ph.D. Workshop, the CIREQ Ph.D. Student's Conference, the Annual Conference of the CEA, the International Conference CEF, the Annual Conference of the IAAE, and the Annual Meeting of the Société canadienne de science économique for comments. I am grateful to Kevin Sheppard for providing for the code for the univariate model confidence set procedure.

†Email: florian.richard@carleton.ca. Website: https://sites.google.com/view/florianrichard. Carleton University, 1125 Colonel By Drive, D-886 Loeb, Ottawa ON, Canada K1S 5B6.

1 Introduction

In this paper, I provide a generalization of Hansen et al. (2011)'s model confidence set (MCS) procedure to systems of $P > 1$ dependent variables, and apply it to asset pricing factor models. Formally, a $(1-\alpha)$ level MCS can be defined as the set of models with equal predictive ability, where the latter is assessed in- or out-of-sample through a sequential elimination procedure that uses an equivalence test. Systems of $P > 1$ predictive equations raise dimensionality problems that have not been addressed to date. This paper is motivated by this fact, particularly because empirical asset pricing models commonly aim to explain returns on many test assets.

The usefulness of the MCS procedure can best be illustrated through the many published empirical models in the beta pricing context. Lewellen et al. (2010) qualify these findings as "an embarrassment of riches", and in his seminal contribution, Harvey (2017) draws serious attention to the underlying multiple testing problems, and suggests raising the hurdle for the discovery of new factors: "our standard testing methods are often ill-equipped to answer the questions that we pose".

In this paper, I propose to use the MCS procedure, which controls the asymptotic family-wise error rate at level α, to address these issues. Specifically, in view of the dimensionality, I propose supremum t and Hotelling T^2 statistics, which account for potential correlations across the P-variate loss differentials. Statistical properties are illustrated in simulations. The empirical analysis provides an alternative perspective on underlying asset pricing issues, in particular, on the relevance of various anomalies, the information content of in- and out-of-sample assessments, and of predictions across short and long term horizons.

The MCS, denoted $\widehat{\mathcal{M}}^*_{1-\alpha}$, is the set of *best* models from a collection of models $\mathcal{M}^0$, i.e., the set of models that survived a sequential selection procedure based on an equivalence test $\delta_{\mathcal{M}}$ and an elimination rule $e_{\mathcal{M}}$, both determined by the user. Similarly to confidence intervals for point estimates, the MCS selects a set of models for some coverage probability $(1-\alpha)$: $\widehat{\mathcal{M}}^*_{1-\alpha}$ covers the set of models with equal predictive ability with probability $(1-\alpha)$. I apply this procedure to a set of popular multivariate asset pricing factor models.

Factor models are widely used empirical finance to relate expected asset or portfolios returns to exposure to risk factors. Examples of early factor models are the Capital Asset Pricing Model (CAPM), resulting from the independent work of Treynor (1961, 1962), Sharpe (1964), Lintner (1965), and Mossin (1966), where the factor is the return of the market portfolio over the risk-free rate. The Fama and French (1993, 2015) three- and five-factor models include the CAPM factor along with the 2 and 4 novel factors. In recent years, Pástor and Stambaugh (2003), Moskowitz et al. (2012), and Asness et al. (2013) introduced liquidity and momentum factors. Harvey et al. (2016) documents more than 316 possible factors since 1964. Given this abundance of potential factors, numerous papers attempts to decipher which factors best explain asset returns. Recently, Barillas and Shanken (2017), Gu et al. (2018), Hou et al. (2018), Huang et al. (2018), Pukthuanthong et al. (2018) designed methods for factor selection, including machine learning techniques. The MCS has been applied extensively to univariate macroeconomic models, as Samuels and Sekkel (2017) uses the out-of-sample performance and Aslanidis et al. (2018) uses the RMSE criterion to determine the MCS. Hansen et al. (2003) applies the MCS to stochastic volatility models (prior to the publication of Hansen et al. (2011)) and compares it to Bonferroni-type bounds. They find that the MCS surpassed the Bonferroni method. Diebold and Mariano (2002)'s criteria for comparing forecasts is also used in the t-statistics detailed in the methodology section. Barillas and Shanken (2018) compares asset pricing models use a Bayesian methodology but requires assumptions on the tradability of the tested factors. In contrast, a confidence set aims to control coverage, i.e., the probability of including the true unknown set of superior models.

First, I provide an extension to multivariate losses of the model confidence set procedure, using a supremum t (or sup t) statistic and a Hotelling T^2 statistic. The latter is adapted from the multivariate test of equal predictive ability of Mariano and Preve (2012). Both statistics accounts for the correlations between loss differentials via a moving block bootstrap. Second, I present simulation results that reflects the conventional size and power properties. I present a design with dependent losses drawn from a multivariate normal distribution, with varying parameter values for between-model and within-model correlations. The procedure works well in terms of both size and power. In the case of a single true model, the MCS behaves as predicted theoretically in Corollary 1 of Hansen et al. (2011). For multiple true models, the procedure attains the conventional coverage probability for reasonable sample sizes, and in most cases, all inferior models are eliminated for sample sizes smaller than 1,000. Empirically, I provide an analysis of a large set of candidate factor models using the model confidence set approach, using the Fama and French research portfolios as dependent variables with monthly data for the period 1972-2013. Using the T^2 statistic, I find that the Fama and French (2015) is only the surviving model for out-of-sample predictions at the 1-year horizon for 12, 18, and 25 portfolios. Multiple candidate models are selected at longer horizons, including the Stambaugh and Yuan (2017) and Liu (2006) models. For in-sample predictions, candidate models are not easily differentiable, especially for 5 test portfolios. The MCS never selects the capital asset pricing model for the out-of-sample tests. All things considered, using out-of-sample predictions and a greater number of dependent variables helps in distinguishing models. Section 2 outlines the framework and the MCS procedure. I present simulation results in Section 3, and the empirical analysis in Section 4. Section 5 concludes.

2 Framework

This section details the model confidence set procedure for multivariate loss functions, and presents the proposed sup t and Hotelling T^2 statistics. Section 2.1 details the econometric framework of the MCS. In Section 2.2, I outline the MCS procedure and its asymptotic properties. I describe the multivariate statistics and the moving block bootstrap procedure in Section 2.3.

2.1 MCS Framework

Let $\{Y_{s,t}\}_{t\geq 1}$ denote a sequence of n time series observations for the dependent variable s, for $s = 1, \ldots, P$. Let $\boldsymbol{Y}$ be the matrix of observations

$$\boldsymbol{Y} = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \ldots & Y_{1,P} \\ Y_{2,1} & Y_{2,2} & \ldots & Y_{2,P} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n,1} & Y_{n,2} & \ldots & Y_{n,P} \end{bmatrix},$$

and let $\boldsymbol{Y}_t = [Y_{t,1} \ldots Y_{t,P}]$ denote the $(1 \times P)$ row vector of observations at time t of each variable s, and $\boldsymbol{Y}_s = [Y_{1,s}, \ldots, Y_{n,s}]'$ the $(P \times 1)$ column vector of observations for the random variables s for $t = 1, \ldots, n$. Let $\widehat{\boldsymbol{Y}}_i$ denote the $(n \times P)$ matrix of fitted values produced by model i:

$$\widehat{\boldsymbol{Y}}_i = \begin{bmatrix} \widehat{\boldsymbol{Y}}_1^i & \ldots & \widehat{\boldsymbol{Y}}_s^i & \ldots & \widehat{\boldsymbol{Y}}_P^i \end{bmatrix}.$$

In the P-equation multivariate framework, let $\boldsymbol{L}_i = \boldsymbol{L}(\boldsymbol{Y}, \widehat{\boldsymbol{Y}}_i)$ denote the $(n \times P)$ matrix of losses associated with model i. Define the sequence of random vectors $\{\boldsymbol{d}_{ij,t}\}_{i,j\in\mathcal{M}^0}$ as the loss differentials between models i and j at time t as

$$\boldsymbol{d}_{ij,t} = \boldsymbol{L}_{i,t} - \boldsymbol{L}_{j,t}, \tag{2.1}$$

and the loss differential across all time periods as

$$\boldsymbol{d}_{ij} = \boldsymbol{L}_i - \boldsymbol{L}_j. \tag{2.2}$$

If model i's loss is statistically smaller than model j's, model i is superior to model j. I assume that the expectation of (2.1) over t is a finite P-dimensional vector and time-invariant:

Assumption 1. $\boldsymbol{\mu}_{ij} = \mathbb{E}[\boldsymbol{d}_{ij,t}] < \infty \qquad \forall\, i, j \in \mathcal{M}^0.$

Suppose that:

Assumption 2. $\{\boldsymbol{d}_{ij,t}\}_{i,j\in\mathcal{M}^0}$ *is covariance stationary.*

Then, the multivariate counterpart of Hansen et al. (2011)'s "set of superior objects" is

$$\mathcal{M}^* \equiv \left\{ i \in \mathcal{M}^0 : \boldsymbol{\mu}_{ij} \leq \mathbf{0}_P \quad \text{for all} \quad j \in \mathcal{M}^0 \right\}, \tag{2.3}$$

where $\mathbf{0}_p$ is the P-dimensional vector of zeros. The corresponding null hypothesis is

$$H_{0,\mathcal{M}} : \boldsymbol{\mu}_{ij} = \mathbf{0}_P \quad \text{for all} \quad i, j \in \mathcal{M}. \tag{2.4}$$

The number of models in $\mathcal{M}$ is m, such that the elements in $\mathcal{M}$ are $i_1, \ldots, i_m$. In the following section, I describe the MCS procedure.

2.2 MCS Procedure

Following the notation of Hansen et al. (2011), the MCS procedure relies on an equivalence test $\delta_{\mathcal{M}}$ to test $H_{0,\mathcal{M}}$ and an elimination rule $e_{\mathcal{M}}$ to eliminate model $\mathcal{M}$. The equivalence test takes on values $\delta_{\mathcal{M}} = 0$ if $H_{0,\mathcal{M}}$ is not rejected, and $\delta_{\mathcal{M}} = 1$ if $H_{0,\mathcal{M}}$ is rejected. The elimination rule $e_{\mathcal{M}}$ determines the model removed from $\mathcal{M}$ when $\delta_{\mathcal{M}} = 1$. Hansen et al. (2011) outlines the procedure for determining $\widehat{\mathcal{M}}^*_{1-\alpha}$ as follows:[1]

Step 0: Initially set $\mathcal{M} = \mathcal{M}^0$.

Step 1: Test $H_{0,\mathcal{M}}$ using $\delta_{\mathcal{M}}$ at level α.

Step 2: If $H_{0,\mathcal{M}}$ is accepted, define $\widehat{\mathcal{M}}^*_{1-\alpha} = \mathcal{M}$ otherwise, use $e_{\mathcal{M}}$ to eliminate an object from $\mathcal{M}$ and repeat the procedure from Step 1.

The output of this algorithm is $\widehat{\mathcal{M}}^*_{1-\alpha}$, the model confidence set. The assumptions of Hansen et al. (2011) with regards to asymptotic level and power apply to the multivariate case, and are stated in Appendix A.1 for completeness. The MCS procedure also produces p-values. The p-value $\hat{p}_i$ for model i defined as the smallest p-value such that model i belongs to the MCS. Thus, a model with $\hat{p}_i = 1$ will be included in the confidence set. This p-value is given by $\hat{p}_{e_{\mathcal{M}_j}} = \max_{i \leq j} P_{H_0,\mathcal{M}_i}$, $P_{H_0,\mathcal{M}_i}$, for the corresponding null hypothesis $H_{0,\mathcal{M}}$. In a multivariate setting, we can test $H_{0,\mathcal{M}}$ using a sup t and a Hotelling T^2 statistic.

2.3 Statistics and Bootstrap Procedure

Both supremum- and Hotelling-type statistic can be used to test hypothesis $H_{0,\mathcal{M}}$. One advantage of the sup t statistic is that it does not require the inversion of a possibly high dimensional covariance matrix. The sup t statistics are written as follows:

$$t^p_{\mathcal{M},\sup} = \sup_{i,j} t^p_{ij} = \sup_{i,j} \left[\bar{d}^p_{ij} / \mathrm{var}(\bar{d}^p_{ij})\right] \qquad \text{for } p = 1, \ldots, P, \tag{2.5}$$

and

$$t_{\mathcal{M},\sup} = \sup_p t^p_{\mathcal{M},\sup}, \tag{2.6}$$

where $\mathrm{var}(\bar{d}^p_{ij})$ estimated via bootstrap. The proof of bootstrap validity of Hansen et al. (2011) applies to the supremum-type t statistics. Additionally, I use the Hotelling-type T^2 statistic based

[1] Hansen et al. (2011), p. 459, using their exact wording.

on Mariano and Preve (2012)'s test for equal predictive ability. Define the vector of losses $\boldsymbol{L}_t \equiv (\boldsymbol{L}_{i_1,t} \dots \boldsymbol{L}_{i_m,t})'$, where $\boldsymbol{L}_{i_j,t} = [l^1_{i_j,t} \dots l^P_{i_j,t}]'$, for $j = 1, \dots, m$. The vector sample averages over t is

$$\bar{\boldsymbol{L}} \equiv n^{-1} \sum_{t=1}^{n} \boldsymbol{L}_t. \tag{2.7}$$

Consider the two statistics

$$\bar{\boldsymbol{d}}_{ij} = n^{-1} \sum_{t=1}^{n} \boldsymbol{d}_{ij,t} \tag{2.8}$$

and

$$\bar{\boldsymbol{d}}_{i\cdot} = m^{-1} \sum_{j=1}^{m} \bar{\boldsymbol{d}}_{ij}, \tag{2.9}$$

corresponding to the sample counterpart of $\boldsymbol{\mu}_{ij}$ and the multivariate sample loss across models, respectively. From equations (2.8) and (2.9), we can construct the Hotelling T^2 statistics

$$T^2_{ij} = n(\bar{\boldsymbol{d}}_{ij} - \boldsymbol{\mu}^0_{ij})'\boldsymbol{\Sigma}^{-1}_{ij}(\bar{\boldsymbol{d}}_{ij} - \boldsymbol{\mu}^0_{ij}), \tag{2.10}$$

and

$$T^2_{i\cdot} = n(\bar{\boldsymbol{d}}_{i\cdot} - \boldsymbol{\mu}^0_{i\cdot})'\boldsymbol{\Sigma}^{-1}_{i\cdot}(\bar{\boldsymbol{d}}_{i\cdot} - \boldsymbol{\mu}^0_{i\cdot}), \tag{2.11}$$

where $\boldsymbol{\Sigma}_{ij} = n^{-1}(\boldsymbol{d}_{ij} - \bar{\boldsymbol{d}}_{ij})(\boldsymbol{d}_{ij} - \bar{\boldsymbol{d}}_{ij})'$, $\boldsymbol{\Sigma}_{i\cdot} = n^{-1}(\boldsymbol{d}_{i\cdot} - \bar{\boldsymbol{d}}_{i\cdot})(\boldsymbol{d}_{i\cdot} - \bar{\boldsymbol{d}}_{i\cdot})'$, and $\boldsymbol{\mu}^0_{ij}$ and $\boldsymbol{\mu}^0_{i\cdot}$ are the value of $\boldsymbol{\mu}_{ij}$ under $H_{0,\mathcal{M}}$. In our case, this value is the P-dimensional vector $\mathbf{0}_P$. The resulting can be written as $T_{R,\mathcal{M}} \equiv \max_{i,j\in\mathcal{M}} \left|T^2_{ij}\right|$. Similarly, the multivariate sample loss across models is given by the following a $(1 \times P)$ row vector

$$\begin{aligned}
\bar{\boldsymbol{d}}_{i\cdot} &= m^{-1} \sum_{j=1}^{m} \bar{\boldsymbol{d}}_{ij} = m^{-1} \left[\bar{\boldsymbol{d}}_{i1} + \dots + \bar{\boldsymbol{d}}_{im}\right] \\
&= m^{-1} \left[n^{-1} \sum_{t=1}^{n} (d^1_{i1,t} + \dots + d^1_{im,t}) \dots n^{-1} \sum_{t=1}^{n} (d^P_{i1,t} + \dots + d^P_{im,t})\right] \\
&= m^{-1} \sum_{j=1}^{m} \left[n^{-1} \sum_{t=1}^{n} (L^1_{i,t} - L^1_{j,t}) \quad \dots \quad n^{-1} \sum_{t=1}^{n} (L^P_{i,t} - L^P_{j,t})\right].
\end{aligned} \tag{2.12}$$

In practice, we can compute bootstrap critical values to circumvent the estimation of large covariance matrices. The block bootstrap procedure for multivariate loss functions is detailed below, following the notation of Hansen et al. (2011).

1. Compute the bootstrap indexes.

(a) Select the block-length bootstrap l.[2]

(b) For the first bootstrap replication $b = 1$, draw a random variable ξ_{b_1} from a uniform distribution with support $[1, n]$, and let $(\tau_{b,1}, \ldots, \tau_{b,l}) = (\xi_{b_1}, \xi_{b_1} + 1, \ldots, \xi_{b_1} + l - 1)$.

(c) For the second bootstrap replication $b = 2$, draw a random variable ξ_{b_2} from a uniform distribution with support $[1, n]$, and let $(\tau_{b,l+1}, \ldots, \tau_{b,2l}) = (\xi_{b_2}, \xi_{b_2} + 1, \ldots, \xi_{b_2} + l - 1)$.

(d) Continue until the random variable ξ_{b_Q} is generated, where $Q = n/l$ denotes the number of blocks if n/l is an integer. If n/l is not an integer, then create $\lceil n/l \rceil$ blocks, where $\lceil \cdot \rceil$ is the ceiling function, and truncate the last block to size $n - (Q - 1)l$.

2. Compute the sample and the bootstrap statistics.

(a) Compute the bootstrap equivalent of $\boldsymbol{L}_{i,t}$:

$$\boldsymbol{L}^*_{b,i,t} = \boldsymbol{L}_{i,\tau_{b,t}} \qquad \text{for} \quad b = 1, \ldots, B,\ i = 1, \ldots, m,\ t = 1, \ldots, n. \tag{2.13}$$

(b) Compute the bootstrap sample average:

$$\bar{\boldsymbol{L}}^*_{b,i} = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{L}^*_{b,i,t}. \tag{2.14}$$

3. Compute the difference between the sample and the bootstrap statistics:

$$\boldsymbol{\zeta}^*_{b,i} = \bar{\boldsymbol{L}}^*_{b,i} - \bar{\boldsymbol{L}}_i. \tag{2.15}$$

4. Test the hypotheses:

(a) Set $\mathcal{M} = \mathcal{M}_0$.

(b) Compute the average over the number of models for the sample and the bootstrap statistics:

$$\bar{\boldsymbol{L}}_{\cdot} = \frac{1}{m} \sum_{i=1}^{m} \bar{\boldsymbol{L}}_i \quad \text{and} \quad \boldsymbol{\zeta}^*_{b,\cdot} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{\zeta}^*_{b,i}. \tag{2.16}$$

(c) Compute either the sup t or the Hotelling T^2 statistic. For the Hotelling T^2 statistic:

$$T^2_{i\cdot} = n(\bar{\boldsymbol{d}}_{i\cdot} - \boldsymbol{\mu}^0_{i\cdot})' \boldsymbol{\Sigma}^{-1}_{i\cdot} (\bar{\boldsymbol{d}}_{i\cdot} - \boldsymbol{\mu}^0_{i\cdot}), \tag{2.17}$$

where $\boldsymbol{\Sigma}_{i\cdot} = \dfrac{1}{B} (\boldsymbol{\zeta}^*_{b,i} - \boldsymbol{\zeta}^*_{b,\cdot})(\boldsymbol{\zeta}^*_{b,i} - \boldsymbol{\zeta}^*_{b,\cdot})'$.

[2] For the asymptotic theory associated multivariate block bootstraps and block length selection criteria, see Jentsch et al. (2015).

(d) Compute the test statistic $T_{\max} = \max_i S_{i\cdot}$, where $S_{i\cdot}$ is either statistic computed in step (c).

(e) Compute either sup t or the Hotelling T^2 bootstrap statistics. For the Hotelling T^2 statistic:

$$T^2_{b,i\cdot} = n(\boldsymbol{\zeta}^*_{b,i} - \boldsymbol{\zeta}^*_{b,\cdot})'\boldsymbol{\Sigma}^{-1}_{i\cdot}(\boldsymbol{\zeta}^*_{b,i} - \boldsymbol{\zeta}^*_{b,\cdot}), \tag{2.18}$$

(f) Compute the bootstrap statistic $T^*_{b,\max} = \max_i S_{b,i\cdot}$, where $S_{b,i\cdot}$ is either statistic computed in step (e).

(g) Compute the p-value for the hypothesis $H_{0,\mathcal{M}}$:

$$P_{H_{0,\mathcal{M}}} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}_{\{T_{\max} > T^*_{b,\max}\}}. \tag{2.19}$$

(h) Reject $H_{\mathcal{M},0}$ if $P_{H_{0,\mathcal{M}}} < \alpha$ and remove $e_{\mathcal{M}} = \arg\max_i S_{i\cdot}$ from $\mathcal{M}$.

(i) Repeat steps (4b) to (4h) until $H_{\mathcal{M},0}$ is not rejected. The $(1-\alpha)$ model confidence set, denoted by $\widehat{\mathcal{M}}^*_{1-\alpha}$, consists of the remaining models.

In the next section, the moving block bootstrap procedure is implemented in simulations for a design with dependent losses.

3 Simulation Results

I consider a simulation design similar to Design I.B of Hansen et al. (2011), with the adaption to multivariate models that the vector of losses within a model can admit some degree of dependence. I consider different parameter values for cross correlation of a covariance matrix with Kronecker structure. The block bootstrap length is set to $l = 2$ and the number of bootstrap iterations to $B =$ 1,000. The number of simulation replications is set to $C =$ 2,500 for each sample size $n \in \{25, 50, 75, 100, 200, 300, 400, 500, 600, 800, 1000, 1500, 2000, 3000, 4000, 5000\}$.

3.1 Simulation Design with Dependent Losses

Let m_0 denote the number of true models from a set of m candidates models. This design uses a $(n \times mP)$ matrix of losses $\boldsymbol{L} = [\boldsymbol{L}_{i_1}, \ldots, \boldsymbol{L}_{i_m}]$ drawn from a multivariate normal distribution $N_{mP}(\theta, \Sigma)$. Each dependent variable in a model admits a loss with mean 0 if the model belongs to $\mathcal{M}^*$, and mean $1/(m - m_0)$ otherwise. The covariance matrix is set as $\Sigma = \Sigma_\phi \otimes \Sigma_\rho$. Σ is parameterized so that the covariance matrix between the losses of models i and j, $\boldsymbol{L}_i$ and $\boldsymbol{L}_j$, is Σ_ϕ; and that for a given model i, the covariance between each $(n \times 1)$ loss vector l^s_i for any given dependent variable s, is Σ_ρ. The $(s, q)^{th}$ and $(i, j)^{th}$ elements of Σ_ϕ and Σ_ρ are defined as

$\Sigma_\phi(s,q) = \phi^{|s-q|}$ and $\Sigma_\rho(i,j) = \rho^{|i-j|}$ for $s,q = 1,\ldots,P$ and $i,j = 1,\ldots,m$, respectively. Σ_ϕ and Σ_ρ are of dimension $(P \times P)$ and $(m \times m)$, respectively. The results of the simulation for $m_0 =$ 1, 2 and 5 true models, $m = 10$ candidate models, and $P = 5$ dependent variables are presented in Figures 1 to 3 for the supremum t statistic, and in Figure 4 to 6 for the Hotelling T^2 statistic. Additional simulation results using the supremum t statistic for $P = 10$ dependent variables are available in Appendix A.3. The top panel of each figure plots the frequency at which the true model is selected by the MCS procedure. This frequency reflects the ability of the procedure to include the true model(s), and is interpreted as the size property of the procedure. The bottom panel of each figure plots the average cardinality (the number of elements in the set) of the MCS. This property illustrates the ability of the procedure to eliminate the inferior models.

Overall, the procedure behaves well and delivers the expected coverage probability and number of selected models. The top panels in Figures 1 and 4 verify Corollary 1 of Hansen et al. (2011) stated in Appendix A.1 for both considered statistics, which implies that if the cardinality of the true MCS $\mathcal{M}^*$ is 1, then the coverage probability $P(\mathcal{M}^* = \widehat{\mathcal{M}}^*_{1-\alpha})$ of the MCS is 1 in the limit. This result is attained for sample sizes greater than 600 for the sup t statistic, and greater than 1,000 for the T^2 statistic. In conjunction with their bottom panels, the top panels in Figures 1 and 4 show that not only the procedure includes the true model in the MCS asymptotically, but only the true model, with probability 1. For parameterizations where there exists more than one true model (Figures 2, 3, 5 and 6), the frequency at which the true models are included in the MCS reaches the 95% coverage probability rapidly, and even exceeds that threshold for small sample sizes when $m_0 = 2$ (Figures 2 and 5). When there are 5 true models, frequency with the sup t statistic reaches 95% coverage after 1,000 observations - faster than with the T^2 statistic. This happens at the expense of power, especially for low values of the within-model correlation parameter ρ. For $\rho = 0$, the T^2 statistic selects close to 5 models for sample sizes as low as 200, where the sup t statistic requires at least $n = 500$.

The bottom panels of each figure captures the power properties of the MCS procedure. All else equal, the average number of selected models decreases for greater values of the between-model correlation parameter ϕ. This additional power reflects the information captured by ϕ, making it easier for the procedure to reject incorrect models. However, greater values of the within-model correlation parameter ρ increase the average number of selected models, making it harder to reject incorrect models, in contrast with the results of Hansen et al. (2011). This pattern remains consistent for $m_0 = 1$, 2 and 5 true models and holds true for both statistics. For the size property, there is no consistent pattern with respect to the different values of the correlation parameters. In the next section, I propose to test a large number of asset pricing factors models that have received support in the literature using the MCS procedure.

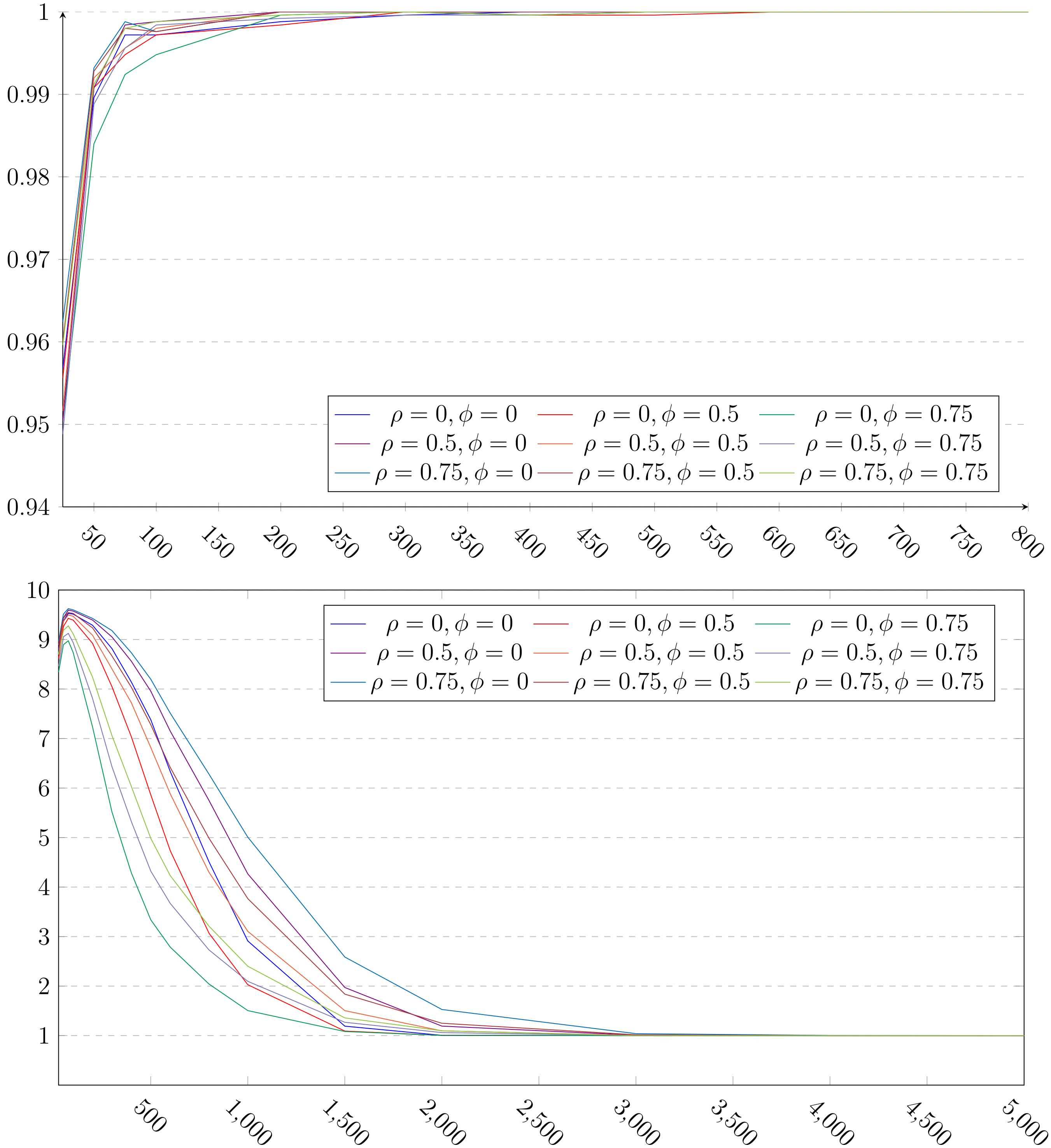


Figure 1: Simulation design for the supremum t statistic with dependent losses, $m = 10$ candidate models, $m_0 = 1$ true model, $P = 5$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true model is included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size. In the top panel, the frequency curve remains the same for sample sizes larger than $n = 600$ and is truncated for clarity.

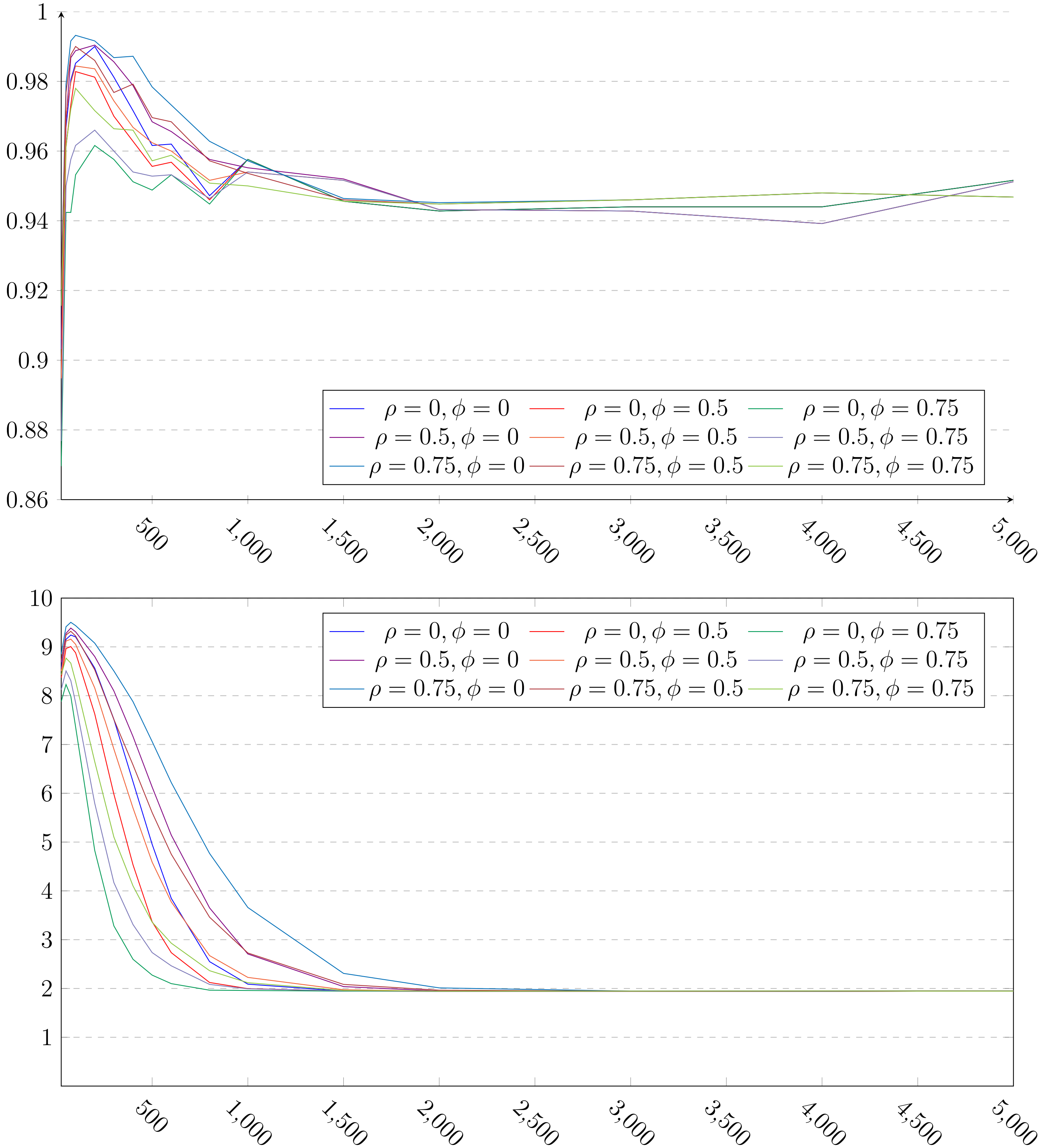


Figure 2: Simulation design for the supremum t statistic with dependent losses, $m = 10$ candidate models, $m_0 = 2$ true models, $P = 5$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true models are included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size.

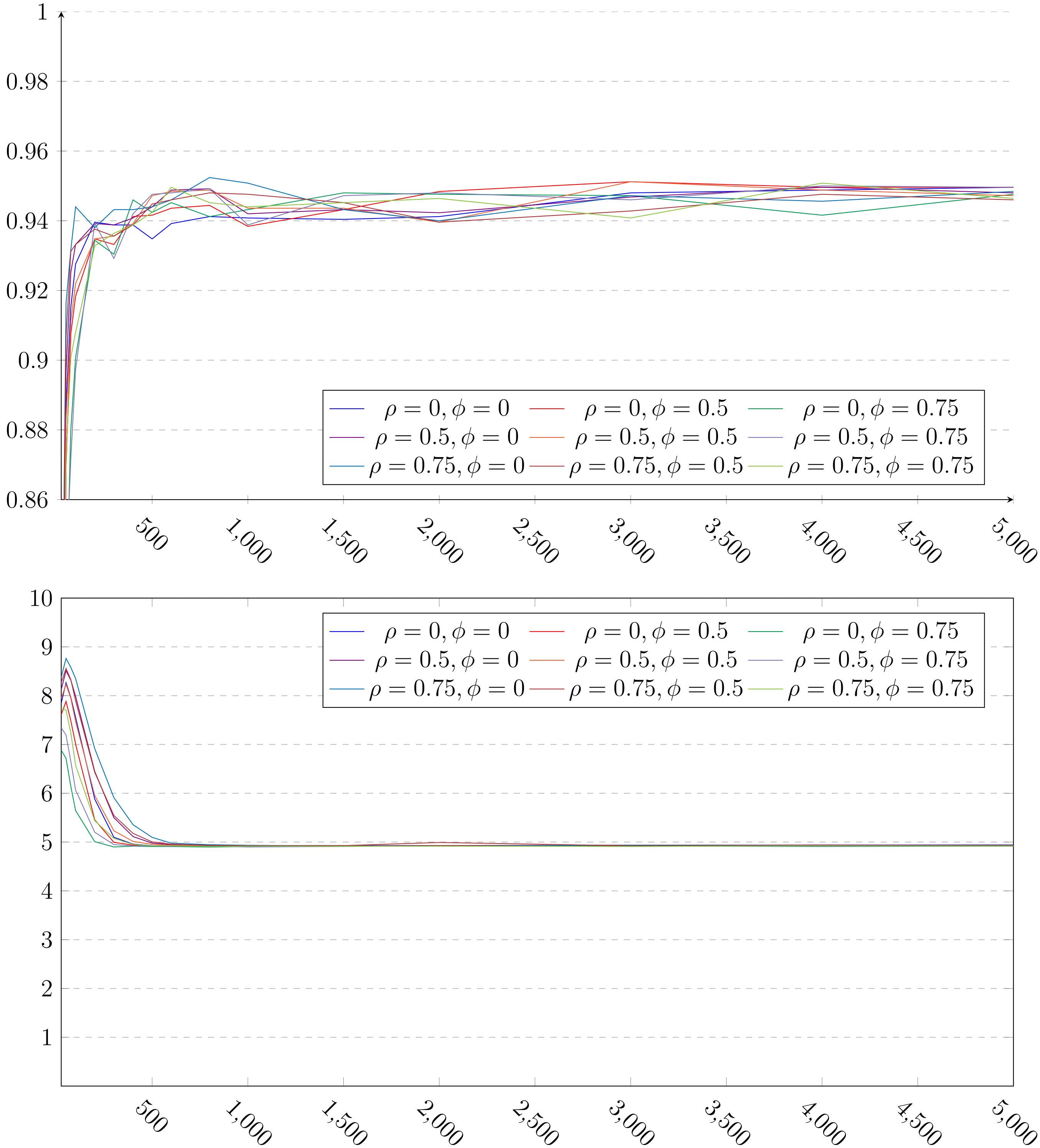


Figure 3: Simulation design for the supremum t statistic with dependent losses, $m = 10$ candidate models, $m_0 = 5$ true models, $P = 5$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true models are included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size.

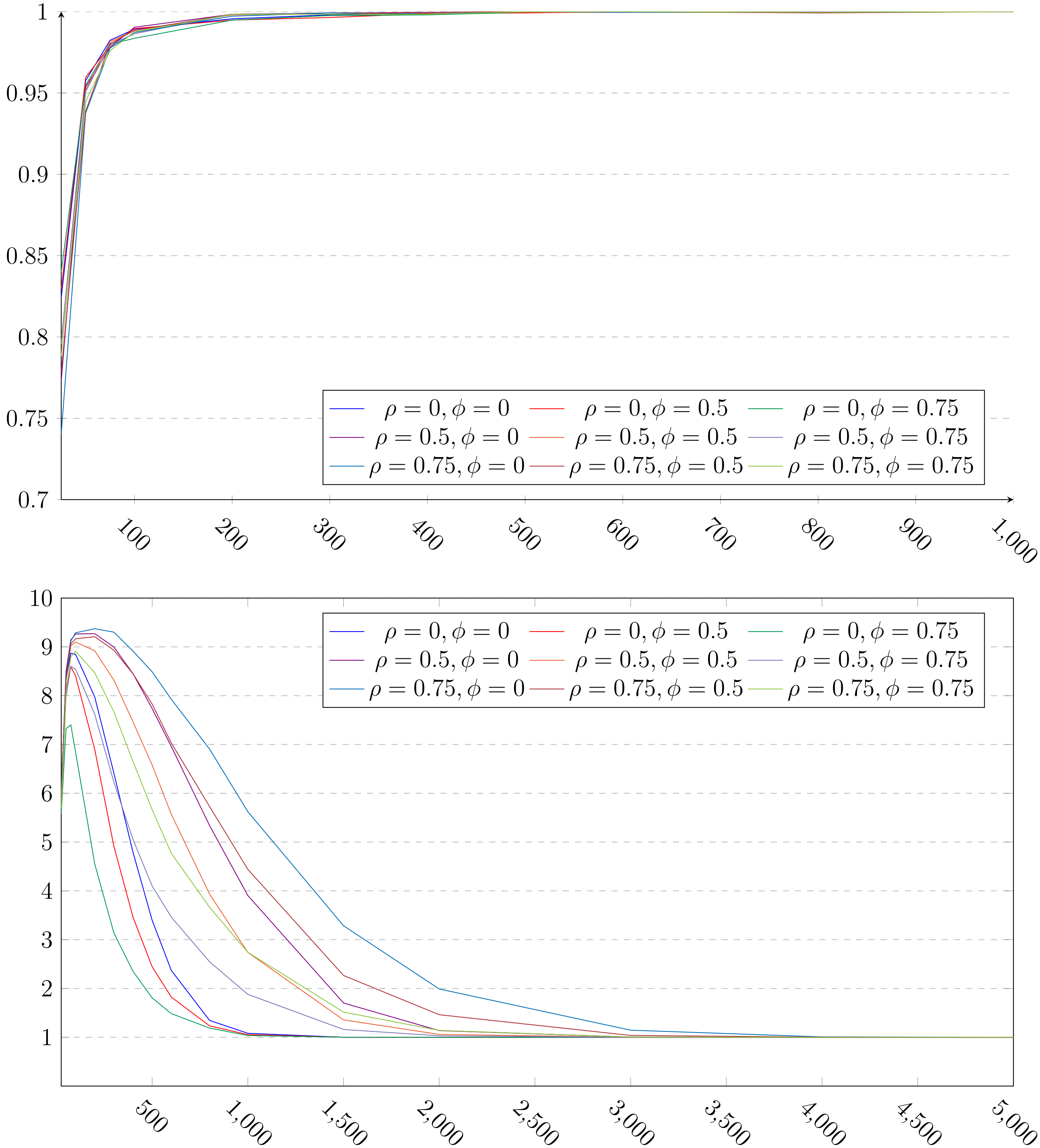


Figure 4: Simulation design for the Hotelling T^2 statistic with dependent losses, $m = 10$ candidate models, $m_0 = 1$ true model, $P = 5$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true model is included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size. In the top panel, the frequency curve remains the same for sample sizes larger than $n = 1000$ and is truncated for clarity.

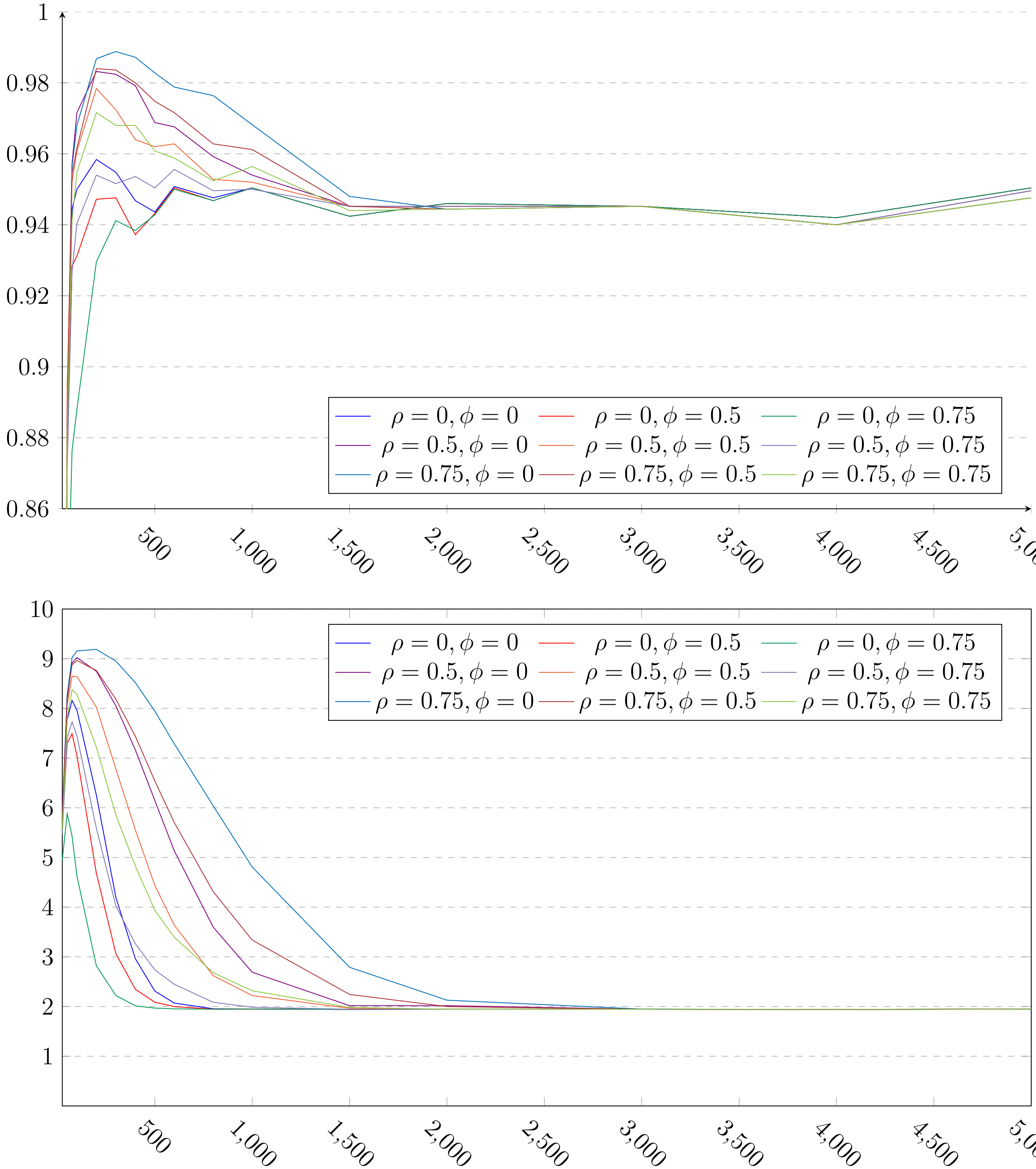


Figure 5: Simulation design for the Hotelling T^2 statistic with dependent losses, $m = 10$ candidate models, $m_0 = 2$ true models, $P = 5$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true models are included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size.

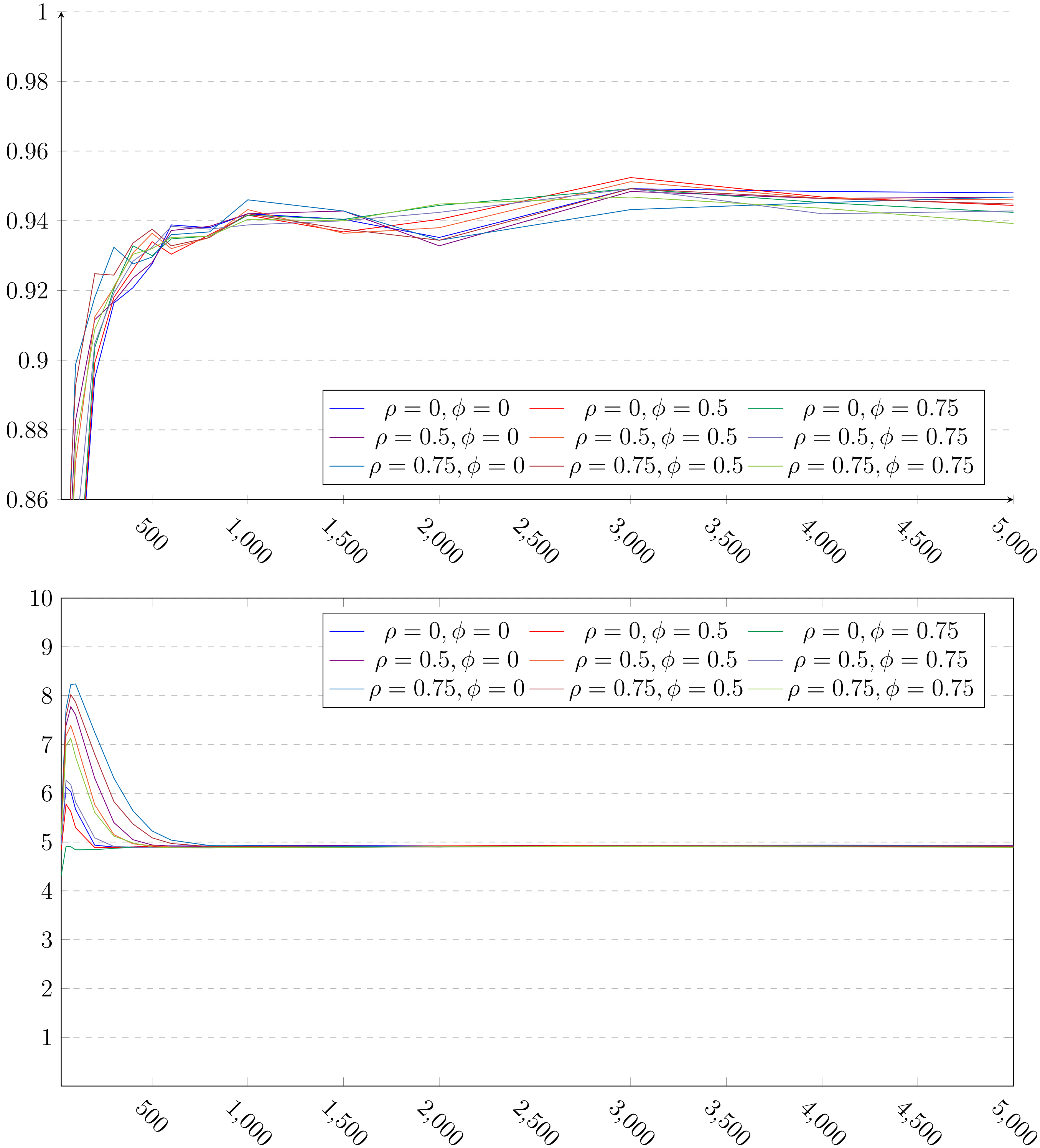


Figure 6: Simulation design for the Hotelling T^2 statistic with dependent losses, $m = 10$ candidate models, $m_0 = 5$ true models, $P = 5$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true models are included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size.

4 Empirical Results

In my empirical analysis, I apply the MCS procedure to a large set of multivariate asset pricing factor models. Smaller subsets of models determined by category of factors are available in Appendix A.4. The multivariate linear factor models presented in this section take the following form:

$$E[R_s - r^f] = \beta_s E[f], \tag{4.1}$$

where R_s represents the returns of portfolio s, r^f is the risk-free rate, β_s is a vector of factor loadings for portfolio s defined as $\beta_s = [\beta_{1,s}\, \beta_{2,s} \ldots \beta_{K,s}]$, and f is a vector of factors defined as $f = [f_1\, f_2 \ldots f_K]'$. The expected returns model states that expected excess returns on the test portfolios are proportional to the expected returns on the factors. The loadings or sensitivities on the factors can be obtained by estimating the following time series regression:

$$R_{s,t} - r^f_t = \alpha_s + f_{1,t}\beta_{1,s} + \ldots + f_{K,t}\beta_{K,s} + \epsilon_{s,t}, \qquad \text{for} \quad t = 1, \ldots, n, \tag{4.2}$$

where $R_{s,t}$, r^f_t, and $f_{k,t}$ are the time-t counterpart of the variables in equation (4.1), and $\epsilon_{s,t}$ is the error term associated with portfolio s at time t. When the factors are themselves tradable, the loadings are interpreted as portfolio weights. If the test portfolio returns $R_{s,t}$ are in excess of a benchmark rate (often the risk-free rate), the well-known mean-variance efficiency conditions imply that the regression intercepts must equal zero, i.e. $\alpha_s = 0$ for $s = 1, \ldots, P$. I impose this additional restriction in my empirical results for robustness. In equation (4.2), the factors are identical across the P equations and I allow for cross-correlations across portfolios. The multivariate regression is equivalent to a system of seemingly unrelated equations (SUR), which can be estimated via ordinary least squares (OLS). I use the Fama and French research portfolios available on Professor French's website as test portfolios.[3] The return series are value-weighted monthly portfolio returns of U.S. stocks on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), and the NASDAQ Stock Market. Portfolios are rebalanced each June and are sorted by characteristics. The characteristics are the market equity (size) ($P = 18$), the double-sorted size and book-to-market ($P = 25$), and the industry ($P = 5$, $P = 12$, and $P = 49$), as suggested by Lewellen et al. (2010). The portfolios formed on size are sorted according to the firm's market equity value (ME), and the portfolios formed on size and book-to-market are the intersection 5 portfolios formed on size and 5 portfolios formed on book-to-market (BE/ME). For both the size and the book-to-market portfolios, we use the lowest 30%, the middle 40%, and the top 40% portfolios returns, along with the quintile and the decile portfolios returns. The industry portfolios are sorted according to the industry that the issuing firm falls under using the Compustat Standard Industrial Classification (SIC) codes for the previous fiscal year, or the Center for Research in Security Prices (CRSP) SIC code if the latter

[3] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research

is unavailable. Summary statistics and industry classifications for the test portfolios are presented in Appendix A.2. I consider the time-t quadratic loss:

$$\boldsymbol{L}_{i,t} = [(R_{s,t} - \widehat{R}^{i}_{s,t})^2, \ldots, (R_{P,t} - \widehat{R}^{i}_{P,t})^2]. \tag{4.3}$$

To compute each statistic, I use a moving-block bootstrap with block length $l = 12$ for in-sample tests and $l = 3$ for out-of-sample test.

4.1 Candidate factors models

I consider 14 popular asset pricing models over a time period from July 1972 to June 2013, totaling 492 observations. As seen in the simulation results, this sample size is often sufficient to achieve $(1 - \alpha)$ coverage. I estimate the model parameters using the Fama and French test portfolios. The candidate models are as follow:

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + e_{st} \tag{CAPM}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t + e_{st} \tag{FF3}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t + \beta_{4s} UMD_t + e_{st} \tag{CAR}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t + \beta_{4s} LIQ_t + e_{st} \tag{PS}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t + \beta_{4s} RMW_t + \beta_{5s} CMA_t + e_{st} \tag{FF5}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} ICRF_t + e_{st} \tag{HKM}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} DHML_t + e_{st} \tag{AF}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} HML^*_t + \beta_{3s} UMD^*_t + \beta_{4s} PMU^*_t + e_{st} \tag{NM}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} MOM_t + \beta_{3s} TREND_t + e_{st} \tag{HZZ}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} MGMT_t + \beta_{4s} PERF_t + e_{st} \tag{SY}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} LIQZ_t + e_{st} \tag{LIU}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} PEAD_t + \beta_{3s} FIN_t + e_{st} \tag{DHS}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} MOMEV_t + \beta_{3s} VALEV_t + e_{st} \tag{AMP}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t + \beta_{4s} UMDQ_t + \beta_{5s} QMJ_t + e_{st} \tag{AFP}$$

For the capital asset pricing model (CAPM), the MKT factor is defined as the return on the market portfolio net of the risk-free rate, which varies over time. The Small Minus Big (SMB) and High Minus Low (HML) factors of the Fama and French (1993) (FF3) are defined as follows:

$$\begin{aligned}
\text{SMB} &= (\text{Small Value} + \text{Small Neutral} + \text{Small Growth})/3, \\
&\quad - (\text{Big Value} + \text{Big Neutral} + \text{Big Growth})/3, \\
\text{HML} &= (\text{Small Value} + \text{Big Value})/2 - (\text{Small Growth} + \text{Big Growth})/2.
\end{aligned}$$

The SMB factor represents the spread between the mean return of three small portfolios and three big portfolios, and the HML factor represents the spread between the mean return of two value portfolios and the mean return of two growth portfolios. Value stocks are stocks considered underpriced by the market, while growth stocks are stocks expected to grow significantly in the future. The Up Minus Down (UMD) factor in the Carhart (1997) (CAR) model is computed as the spread between the mean return of winning stocks and losing stocks. It represents the momentum in a stock, i.e., the tendency of the return to be positive if the last period return was also positive, and vice-versa for negative returns. Pástor and Stambaugh (2003) (PS) uses the Fama and French (1993) factors and a liquidity factor, LIQ. This liquidity factor proxies for aggregate market liquidity. Fama and French (2015) (FF5) defines the Robust Minus Weak (RMW) and Conservative Minus Aggressive (CMA) factors as follows:

$$\text{RMW} = (\text{Small Robust} + \text{Big Robust})/2 - (\text{Small Weak} + \text{the})/2,$$
$$\text{CMA} = (\text{Small Conservative} + \text{Big Conservative})/2 - (\text{Small Aggressive} + \text{Big Aggressive})/2.$$

The RMW is the spread between the mean return of a robust portfolio and a weak portfolio, in terms of operating profitability. The CMA is the spread between the mean return of a conservative portfolio and an aggressive portfolio. The $ICRF$ factor of He et al. (2017) (HKM) is the innovations of the autoregressive involving the intermediary capital ratio and its lagged value. The intermediary capital ratio (ICR) is defined as

$$ICR = \text{Market capitalization}/(\text{Market capitalization} + \text{book asset - book equity}).$$

The devil's HML factor of Asness and Frazzini (2013) (AF) is the HML factor with the modification that the portfolios are sorted based on the book-to-price ratio instead of the traditional price-to-book ratio. Novy-Marx (2013) (NM) uses an industry-adjusted version of the HML factor, the Up Minus Down (UMD^*) factor, which represents the momentum in a stock, i.e., the tendency of the return to be positive if the last period return was also positive, and vice-versa for negative returns, as well as an industry-adjusted profitability factor, PMU^*, defined as the spread between gross profits-to-assets ratios for profitable and unprofitable firms. The Han et al. (2016) (HZZ) model use the MOM momentum factor, and the $TREND$ trend factor, which reflects the short, medium, and long term moving average prices at different time horizons. The Stambaugh and Yuan (2017) (SY) uses two mispricing factors, $MGMT$ and $PERF$, which summarizes information in mispricing related to firm management and firm performance, respectively. The $LIQZ$ factor of Liu (2006) (LIU) acts as a proxy for the proportion of zero daily volume for a given number of days. Daniel et al. (2020) (DHS) factors exploit mispricing in both the short and long term: the $PEAD$ factor captures post earnings announcement drift anomalies, i.e., the subdued reaction of market to those earnings surprises, and the FIN factor captures the long term mispricing. Asness et al. (2013) (AMP) uses value (the ratio of book value to market value) and momentum factors, categorized by asset class. The QMJ factor, or "quality minus junk", of Asness et al. (2019) (AFP) reflects the quality premium earned by high quality stocks over low quality stocks.

4.2 Discussion

The empirical analysis highlights two key facts. First, there are stark differences between the in-sample and out-of-sample results. The in-sample MCS consistently contains more models than the out-of-sample MCS. Second, as the number of test portfolios grows, the MCS includes fewer equally predictable models. Notably, the capital asset pricing model containing only the market premium factor is never included in MCS for out-of-sample predictions. Tables 1 to 6 present the 95% and 75% model confidence sets for the 14 candidate models for the two considered statistics. The left panels display the results of the in-sample tests, and the right panels that of the out-of-sample tests, for the 12-, 24- and 60-month horizons. In each case, I also perform the MCS procedure imposing the mean-variance efficiency condition by suppressing the regression intercept. Tables 1 and 4 show the results for the 5- and 12-industry portfolios, Tables 2 and 5 show the results for the 18 size-sorted and 25 size- and book-to-market-sorted portfolios, and Tables 3 and 6 show the results for the 49-industry portfolios.

Table 1 presents the results for the 5 and 12 industry-sorted portfolios as test assets using the sup t statistic. For in-sample predictions at the 95% confidence level, the candidate models are indistinguishable in their capacity to explain variation in five-industry portfolio returns. For the 75% MCS, the set consists only of the Fama and French (2015) model in all 4 cases (for the 5- and 12-industry portfolios, with and without mean-variance conditions); however the confidence level is lower. For out-of-sample tests with a 12-month horizon, the MCS contains fewer models than for in-sample predictions. Only the Fama and French (1993), the Pástor and Stambaugh (2003), and the Stambaugh and Yuan (2017) models survive the procedure. When imposing $\alpha_i = 0$, the Fama and French (2015) model is selected in addition to these 3 models. For a 24-month horizon, 4 and 3 models are included in the MCS with and without mean-variance efficiency, respectively. For $h = 60$, the Asness and Frazzini (2013), Stambaugh and Yuan (2017), and Liu (2006) models are included for both restricted and unrestricted regressions. When using the 12-industry portfolios, 12 of the 14 candidate models are selected by the in-sample MCS. The picture becomes clearer in out-of-sample results: for short time horizon (12 and 24 months), the Fama and French (2015) model is the only included model in the MCS. The out-of-sample results underscore the short term stability of the RMW and the CMA factors, as the three-factor model (Fama and French (1993)) is never selected alongside the five-factor model. This finding does not hold for the 60-month horizon, however. Table 2 displays the results for 18 size-sorted, and 25 size- and book-to-market-sorted portfolios. The short term out-of-sample results show that the procedure only retains at most 2 models for the size-sorted test portfolios, and only the Fama and French (2015) model for the 25 size- and book-to-market-sorted test portfolios. At 5-year horizon, the 6 models are selected for 18

Table 1: MCS p-values for the candidate factor models, using the supremum t statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from July 1972 to June 2013. For out-of-sample tests, the horizon h is 12, 24, and 60 months. R_{5IND} and R_{12IND} denote the portfolio returns for 5 industry-sorted, and 12 industry-sorted portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 10{,}000$.

		In-sample		Out-of-sample					
		July 1972 to June 2013		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{5IND}	CAPM	0.0769*	0.1423*	0.0037	0.0000	0.0027	0.0009	0.0118	0.0115
	FF3	0.0987*	0.1423*	0.3053**	0.1111*	0.0251	0.0182	0.0118	0.0115
	CAR	0.0987*	0.1423*	0.0037	0.0000	0.0039	0.0079	0.0118	0.0115
	PS	0.0987*	0.1423*	0.3053**	0.1111*	0.0251	0.0182	0.0118	0.0115
	FF5	1.0000**	1.0000**	0.0037	0.4607**	1.0000**	1.0000**	0.0045	0.0015
	HKM	0.0769*	0.1423*	0.0037	0.0000	0.0065	0.0079	0.0118	0.0115
	AF	0.0769*	0.1423*	0.0037	0.0000	0.0039	0.0079	0.1168*	0.1361*
	NM	0.0769*	0.1423*	0.0037	0.0000	0.0027	0.0009	0.0118	0.0115
	HZZ	0.0769*	0.1423*	0.0037	0.0000	0.0752*	0.0079	0.0118	0.0115
	SY	0.0987*	0.1423*	1.0000**	1.0000**	0.1041*	0.3829**	1.0000**	1.0000**
	LIU	0.0769*	0.1423*	0.0037	0.0000	0.0251	0.3829**	0.5254**	0.6050*
	DHS	0.0769*	0.1423*	0.0037	0.0000	0.6179**	0.0182	0.0045	0.0015
	AMP	0.0769*	0.1423*	0.0037	0.0000	0.0005	0.0003	0.0045	0.0015
	AFP	0.0987*	0.1423*	0.0037	0.0000	0.0065	0.0079	0.0118	0.0115
R_{12IND}	CAPM	0.0437	0.0542*	0.0000	0.0000	0.0000	0.0000	0.0108	0.0022
	FF3	0.1578*	0.1268*	0.0318	0.0009	0.0004	0.0000	0.0217	0.0165
	CAR	0.1578*	0.1268*	0.0008	0.0152	0.0006	0.0001	0.1387*	0.0639*
	PS	0.1578*	0.1268*	0.0008	0.0152	0.0004	0.0001	0.0223	0.0165
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	0.0223	0.0165
	HKM	0.0437	0.0542*	0.0000	0.0009	0.0004	0.0000	0.0217	0.0165
	AF	0.1578*	0.1268*	0.0000	0.0009	0.0004	0.0000	1.0000**	1.0000**
	NM	0.0437	0.0542*	0.0000	0.0000	0.0000	0.0000	0.0108	0.0022
	HZZ	0.0602*	0.0542*	0.0000	0.0000	0.0004	0.0000	0.0217	0.0165
	SY	0.1578*	0.1268*	0.0008	0.0009	0.0004	0.0000	0.0338	0.0165
	LIU	0.0602*	0.1032*	0.0008	0.0009	0.0004	0.0001	0.0223	0.0165
	DHS	0.0602*	0.1032*	0.0008	0.0000	0.0004	0.0001	0.0217	0.0165
	AMP	0.0602*	0.1032*	0.0000	0.0000	0.0001	0.0000	0.0108	0.0022
	AFP	0.1578*	0.1268*	0.0008	0.0152	0.0004	0.0001	0.0217	0.0165

Table 2: MCS p-values for the candidate factor models, using the supremum t statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from July 1972 to June 2013. For out-of-sample tests, the horizon h is 12, 24, and 60 months. R_{ME} and $R_{ME/BEME}$ denote the portfolio returns for 18 size-sorted, and 25 size- and book-to-market-sorted portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 10{,}000$ for the size-sorted portfolios, and to $B = 1{,}000$ for the size- and book-to-market-sorted portfolios.

		In-sample		Out-of-sample					
		July 1972 to June 2013		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{ME}	CAPM	0.0000	0.0000	0.0000	0.0001	0.0002	0.0034	0.0004	0.0014
	FF3	0.9729**	0.9683*	0.0017	0.0103	0.0002	0.0707*	0.1448*	0.1451*
	CAR	0.9729**	0.9683*	0.0003	0.0103	0.0002	1.0000**	0.1448*	0.1451*
	PS	0.9729**	0.9683*	0.0003	0.0103	0.0002	0.0092	1.0000**	1.0000**
	FF5	1.0000**	1.0000*	1.0000**	1.0000**	1.0000**	0.0092	0.1448*	0.1451*
	HKM	0.0000	0.0000	0.0000	0.0001	0.0002	0.0034	0.0010	0.0025
	AF	0.7248**	0.7182*	0.0003	0.0103	0.0002	0.0092	0.1448*	0.1451*
	NM	0.0000	0.0000	0.0003	0.0020	0.0002	0.0034	0.0115	0.0067
	HZZ	0.0000	0.0000	0.0000	0.0000	0.0002	0.0055	0.0004	0.0014
	SY	0.0002	0.0006	0.0003	0.0041	0.0002	0.0092	0.0355	0.0339
	LIU	0.0000	0.0000	0.0000	0.0000	0.0002	0.0034	0.0010	0.0025
	DHS	0.0000	0.0000	0.0000	0.0000	0.0002	0.0055	0.0010	0.0021
	AMP	0.0000	0.0000	0.0000	0.0020	0.0002	0.0092	0.0004	0.0014
	AFP	0.9729**	0.9683*	0.4865**	0.0103	0.0002	0.0092	0.1448*	0.1451*
$R_{ME/BEME}$	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FF3	0.3940**	0.3970**	0.0000	0.0000	0.0000	0.0000	0.0580*	0.0220
	CAR	0.3940**	0.3970**	0.0090	0.0000	0.0000	0.0040	0.0580*	0.0250
	PS	0.3940**	0.3970**	0.0000	0.0000	0.0000	0.0000	0.0580*	0.0220
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	HKM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AF	0.0080	0.0070	0.0000	0.0000	0.0000	0.0000	0.0580*	0.0220
	NM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010
	HZZ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	SY	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0580*	0.0220
	LIU	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	DHS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AMP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010
	AFP	0.3940**	0.3970**	0.0000	0.0000	0.0000	0.0040	0.0580*	0.0250

Table 3: MCS p-values for the candidate factor models, using the supremum t statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from July 1972 to June 2013. For out-of-sample tests, the horizon h is 12, 24, and 60 months. R_{49IND} denote the portfolio returns for 49 industry-sorted portfolios. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 1{,}000$.

		In-sample		Out-of-sample					
		July 1972 to June 2013		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{49IND}	CAPM	0.0020	0.0040	0.0000	0.0000	0.0000	0.0010	0.0330	0.0060
	FF3	0.2710**	0.3240**	0.0080	0.0000	0.0310	0.0100	0.0330	0.2850**
	CAR	0.2710**	0.3240**	0.0000	0.0000	0.0310	0.0100	0.0330	0.0120
	PS	0.2710**	0.3240**	1.0000**	0.0000	1.0000**	1.0000**	0.0330	0.0120
	FF5	1.0000**	1.0000**	0.0080	1.0000**	0.0000	0.0010	0.0330	0.0120
	HKM	0.0020	0.0040	0.0000	0.0000	0.0000	0.0010	0.0330	0.0120
	AF	0.2620**	0.3240**	0.0000	0.0000	0.0310	0.0090	1.0000**	0.0120
	NM	0.0020	0.0040	0.0000	0.0000	0.0000	0.0010	0.0330	0.0120
	HZZ	0.0020	0.0040	0.0000	0.0000	0.0000	0.0010	0.0330	0.0120
	SY	0.2710**	0.3240**	0.0080	0.0000	0.0050	0.0010	0.0330	0.0120
	LIU	0.0020	0.0040	0.0000	0.0000	0.0000	0.0010	0.0330	0.0120
	DHS	0.0020	0.0040	0.0080	0.0000	0.0310	0.0100	0.0330	0.0120
	AMP	0.0020	0.0040	0.0000	0.0000	0.0000	0.0010	0.0330	0.0120
	AFP	0.2710**	0.3240**	0.0020	0.0000	0.0310	0.0100	0.0330	1.0000**

Table 4: MCS p-values for the candidate factor models, using the Hotelling T^2 statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from July 1972 to June 2013. For out-of-sample tests, the horizon h is 12, 24, and 60 months. R_{5IND} and R_{12IND} denote the portfolio returns for 5 industry-sorted, and 12 industry-sorted portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by the and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 1{,}000$.

		In-sample		Out-of-sample					
		July 1972 to June 2013		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{5IND}	CAPM	0.0030	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FF3	0.0280	0.0360	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000
	CAR	0.0930*	0.0930*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	PS	0.0280	0.0360	0.0000	0.0080	0.0000	0.0000	0.0000	0.0000
	FF5	0.0930*	1.0000**	0.0000	1.0000**	1.0000**	1.0000**	0.0000	0.0000
	HKM	0.0030	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AF	0.0030	0.0010	0.0000	0.0000	0.0000	0.0000	0.0020	0.0000
	NM	0.0030	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	HZZ	0.0030	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	SY	0.0280	0.0360	1.0000**	0.0040	0.0070	0.1670*	1.0000**	1.0000**
	LIU	0.0030	0.0010	0.0000	0.0000	0.0000	0.0000	0.6320**	0.7340**
	DHS	0.0030	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AMP	0.0040	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AFP	1.0000**	0.0930*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R_{12IND}	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FF3	0.3960**	0.0410	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	CAR	0.0490	0.0410	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	PS	0.3960**	0.2340*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	0.0000	0.0000
	HKM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AF	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000**	1.0000**
	NM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	HZZ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	SY	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	LIU	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	DHS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AMP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AFP	0.0490	0.0410	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5: MCS p-values for the candidate factor models, using the Hotelling T^2 statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from July 1972 to June 2013. For out-of-sample tests, the horizon h is 12, 24, and 60 months. R_{ME} and $R_{ME/BEME}$ denote the portfolio returns for 18 size-sorted, and 25 size- and book-to-market-sorted portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 1{,}000$.

		In-sample		Out-of-sample					
		July 1972 to June 2013		$h=12$		$h=24$		$h=60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{ME}	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FF3	0.3430**	0.3120**	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000
	CAR	0.3430**	0.3120**	0.0000	0.0000	0.0000	1.0000**	0.0000	0.0000
	PS	0.3430**	0.3120**	0.0000	0.0000	0.0000	0.0000	1.0000**	1.0000**
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	0.0000	0.0000	0.0000
	HKM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AF	0.3190**	0.2900**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	NM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	HZZ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	SY	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	LIU	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	DHS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AMP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AFP	0.3430**	0.3120**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$R_{ME/BEME}$	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FF3	0.0110	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	CAR	0.0110	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	PS	0.0110	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	HKM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AF	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	NM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	HZZ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	SY	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	LIU	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	DHS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AMP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AFP	0.0110	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 6: MCS p-values for the candidate factor models, using the Hotelling T^2 statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from July 1972 to June 2013. For out-of-sample tests, the horizon h is 12, 24, and 60 months. R_{49IND} denote the portfolio returns for 49 industry-sorted portfolios. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 1{,}000$.

		In-sample		Out-of-sample					
		July 1972 to June 2013		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{49IND}	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FF3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	CAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	PS	0.0000	0.0000	1.0000**	0.0000	1.0000**	1.0000**	0.0000	0.0000
	FF5	1.0000**	1.0000**	0.0000	1.0000**	0.0000	0.0000	0.0000	0.0000
	HKM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AF	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000**	0.0000
	NM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	HZZ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	SY	0.0000	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	LIU	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	DHS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AMP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	AFP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000**

portfolios. Table 3 shows the results for the 49-industry portfolios. For in-sample predictions, 7 of the 14 candidate models are selected by the procedure. For the out-of-sample predictions, the liquidity model of Pástor and Stambaugh (2003) is the only model selected at the 12- and 24-month horizons for unrestricted regressions. For longer horizons ($h = 60$), the Asness and Frazzini (2013) model is the only selected model. For restricted regressions, the Fama and French (2015) model is selected at the 12-month horizon, while the Pástor and Stambaugh (2003) model remains the best model for 24-month horizons. At the 60-month horizon, the Fama and French (1993) and the Asness et al. (2019) models survive the procedure.

Tables 4, 5, and 6 show the results of the MCS procedure for the Hotelling T^2 statistic. For the industry portfolios, the T^2 statistic eliminates significantly more models in-sample than the sup t statistic. This pattern also holds in simulations when the within-model correlation parameter ρ is close to 0. For out-of-sample predictions with 5 test portfolios, the results are mixed. In the short term ($h = 12$), the Stambaugh and Yuan (2017) mispricing model is the only included candidate for unrestricted regressions, but the Fama and French (2015) is declared winner for restricted regressions. These two models are also included in the MCS for $h = 24$ under mean-variance efficiency conditions. At the $h = 60$ horizon, only the Stambaugh and Yuan (2017) and Liu (2006) models remain. For the 12-industry test portfolios, the Fama and French (2015) model is again the only included model for out-of-sample test at the 1- and 2-year horizons. For a 5-year horizon, the Asness and Frazzini (2013) model is the only selected model. The results for the in-sample predictions based on the 18 size-sorted portfolios are robust to using the T^2 statistic, as displayed in Table 5. For 25 portfolios however, the MCS based on the T^2 statistic eliminates all models but the Fama and French (2015) model, for in-sample and out-of-sample predictions. For 49-industry portfolios, the Fama and French (2015) model is again declared winner for in-sample tests. The results of out-of-sample predictions for the Hotelling statistic are consistent with that of the supremum statistic, with the exception that only the Asness et al. (2019) model is selected for unrestricted regressions at the 60-month horizon.

5 Conclusion

This paper provides a multivariate extension of the model confidence set procedure originally proposed by Hansen et al. (2011) for univariate models, and proposes two statistics to test equal predictive ability: a supremum-type t statistic and a Hotelling-type T^2 statistic. A confidence set approach provides valuable insights and has significant strengths. First, the candidate models do not need to follow a certain structure, e.g. with respect to nesting; second, a baseline model is not required, and the model confidence set procedure allows models to be viewed as statistically equal. The model confidence set procedure also allows us to establish the significance of models as opposed to the marginal contributions of new factors.

The extensive simulation study showcases the asymptotic size and power properties of procedure.

The procedure is adequately sized, even in small samples. In many cases, the desired coverage probability is achieved in samples as small as $n = 500$, and the procedure can eliminate all inferior models around $n = 800$ when there are a large number of good models. Simulations also show that one of the key properties of the MCS procedure, namely that the estimated MCS converges to the true MCS in probability when the latter is a singleton, holds true in the multivariate case.

The empirical analysis answers several outstanding questions with regards to the factor proliferation problems encountered in the asset pricing. Namely, how do models featuring recently discovered factors compare, and does a particular model stand out? I apply the MCS procedure to a large set of candidate models and I find that the prominent Fama and French (2015) five-factor model is declared winner for out-of-sample tests with a large number of dependent variables. For in-sample predictions, the candidate models are often indistinguishable in their capacity to explain expected returns.

A Appendix

A.1 Assumptions and Theorems in Hansen et al. (2011)

The following assumptions and theorems do not depend on the dimension of $\boldsymbol{L}$, $\boldsymbol{d}$, or $\boldsymbol{\mu}$, and can be applied to multivariate loss functions.

Assumption 3. *For any $\mathcal{M} \subset \mathcal{M}^0$, (a) $\limsup_{n\to\infty} P(\delta_{\mathcal{M}} = 1 | H_{0,\mathcal{M}}) \leq \alpha$, and (b) $\lim_{n\to\infty} P(\delta_{\mathcal{M}} = 1 | H_{A,\mathcal{M}}) = 1$.*

Additionally, a given model i must not by eliminated by $e_{\mathcal{M}}$ asymptotically if it belongs to the set of superior objects:

Assumption 4. *(Hansen et al. (2011)) $\lim_{n\to\infty} P(e_{\mathcal{M}} \in \mathcal{M}^* | H_{A,\mathcal{M}}) = 0$.*

Given Assumptions 3 and 4, Theorem 1 and Corollary 1 of Hansen et al. (2011) also apply to the multivariate case and are stated without proof.

Theorem A.1. *Under Assumptions 3 and 4, (i) $\liminf_{n\to\infty} P(\mathcal{M}^* \subset \widehat{\mathcal{M}}^*_{1-\alpha}) \geq 1 - \alpha$ and (ii) $\lim_{n\to\infty} P(i \in \widehat{\mathcal{M}}^*_{1-\alpha}) = 0$ for all $i \notin \mathcal{M}^*$.*

Corollary A.1.1. *Under Assumptions 3 and 4, and that $\mathcal{M}^*$ is a singleton, we have $\lim_{n\to\infty} P(\mathcal{M}^* = \widehat{\mathcal{M}}^*_{1-\alpha}) = 1$.*

Theorem A.2. *Suppose that $P(\delta_{\mathcal{M}} = 1, e_{\mathcal{M}} \in \mathcal{M}^*)$. Then,*

$$P(\mathcal{M}^* \subset \widehat{\mathcal{M}}^*_{1-\alpha}) \geq 1 - \alpha. \tag{A.1}$$

A.2 Summary Statistics

Table 7: Summary statistics for the monthly factors, from July 1972 to June 2013: monthly average, standard deviation, minimum, and maximum.

	MKT	SMB	HML	UMD	LIQ	RMW	CMA	ICRF	DHML	HML*	UMD*	PMU*
Mean	0.0049	0.0017	0.0043	0.0069	0.0045	0.0029	0.0039	-0.0003	0.0043	0.0043	0.0062	0.0027
S.d.	0.0465	0.0314	0.0297	0.0451	0.0352	0.0236	0.0201	0.0683	0.0364	0.0149	0.0289	0.0118
Min	-0.2324	-0.1687	-0.1110	-0.3439	-0.1278	-0.1833	-0.0688	-0.2795	-0.1798	-0.0502	-0.2338	-0.0462
Max	0.1610	0.2171	0.1290	0.1836	0.1119	0.1331	0.0958	0.3965	0.2700	0.0656	0.1218	0.0679

	MOM	TREND	MGMT	PERF	LIQZ	PEAD	FIN	MOMEV	VALEV	UMDQ	QMJ
Mean	0.0069	0.0099	0.0069	0.0065	0.0059	0.6535	0.8208	0.0052	0.0039	0.0067	0.0038
S.d.	0.0451	0.0347	0.0288	0.0392	0.0371	1.8576	3.9457	0.0284	0.0275	0.0433	0.0236
Min	-0.3439	-0.1667	-0.0893	-0.2145	-0.1321	-9.0283	-24.5554	-0.1427	-0.1887	-0.3456	-0.0910
Max	0.1836	0.1716	0.1458	0.1852	0.1417	11.9816	20.4176	0.1647	0.1658	0.1707	0.1239

Table 8: Correlation between monthly factors, from July 1972 to June 2013.

	MKT	SMB	HML	UMD	LIQ	RMW	CMA	ICRF	DHML	HML*	UMD*	PMU*
MKT	1.0000	0.2613	-0.2896	-0.1419	0.0210	-0.2625	-0.3984	0.7622	-0.1226	-0.0558	0.0320	0.0338
SMB		1.0000	-0.2197	0.0008	0.0020	-0.4305	-0.1320	0.0913	-0.1076	-0.0212	0.0166	0.0029
HML			1.0000	-0.1680	0.0508	0.1333	0.6922	-0.0098	0.7727	-0.0103	-0.0376	0.0277
UMD				1.0000	-0.0346	0.1095	0.0261	-0.2651	-0.6504	0.0934	0.0329	-0.1457
LIQ					1.0000	-0.0054	0.0143	-0.0005	0.0975	0.0152	-0.1309	-0.0460
RMW						1.0000	0.0314	-0.1679	-0.0249	0.0772	-0.0429	0.0165
CMA							1.0000	-0.1994	0.4850	-0.0140	-0.0654	0.0318
ICRF								1.0000	0.1245	-0.0765	0.0261	0.0727
DHML									1.0000	-0.0519	-0.0415	0.0894
HML*										1.0000	-0.1817	-0.2241
UMD*											1.0000	0.2804
PMU*												1.0000

	MOM	TREND	MGMT	PERF	LIQZ	PEAD	FIN	MOMEV	VALEV	UMDQ	QMJ
MKT	-0.1421	0.1024	-0.5396	-0.2603	-0.6594	-0.0998	-0.5130	-0.0653	-0.0722	0.0093	-0.5242
SMB	0.0007	0.0436	-0.3904	-0.0943	-0.1968	0.0204	-0.4871	0.0043	-0.0555	-0.0131	-0.4652
HML	-0.1676	-0.0292	0.7204	-0.3034	0.4776	-0.1532	0.6529	-0.1857	0.3761	-0.0513	-0.0448
UMD	1.0000	-0.0972	0.0572	0.7188	0.1717	0.4615	0.0960	0.5142	-0.3150	0.0974	0.2902
LIQ	-0.0346	0.0250	-0.0162	0.0200	0.0050	-0.0176	0.0133	-0.1018	0.0960	-0.0817	-0.0243
RMW	0.1096	-0.0367	0.2684	0.4411	0.3477	-0.0920	0.5629	0.0285	0.0791	-0.0078	0.7613
CMA	0.0264	-0.0029	0.7710	-0.0454	0.5077	-0.0028	0.5961	-0.0788	0.2552	-0.0581	0.0591
ICRF	-0.2651	0.1622	-0.2717	-0.4179	-0.4324	-0.1850	-0.2767	-0.1473	0.0584	-0.0026	-0.4532
DHML	-0.6502	0.0406	0.4858	-0.6371	0.2373	-0.4072	0.4087	-0.4046	0.4578	-0.0620	-0.2640
HML*	0.0935	-0.0512	0.0741	0.0891	0.0808	0.0073	0.0598	0.0957	-0.0568	0.0263	0.1074
UMD*	0.0330	-0.0682	-0.0415	0.0134	-0.0729	0.0736	-0.0464	0.0634	-0.0314	0.0933	-0.0375
PMU*	-0.1458	0.1289	-0.0286	-0.0914	0.0168	-0.0326	0.0055	-0.0422	0.0098	0.0437	-0.0306
MOM	1.0000	-0.0970	0.0573	0.7186	0.1718	0.4614	0.0960	0.5144	-0.3150	0.0972	0.2901
TREND		1.0000	-0.0041	-0.0879	-0.0327	-0.1077	-0.0202	-0.0621	0.0123	-0.0211	-0.0154
MGMT			1.0000	0.0096	0.6265	-0.0027	0.7993	-0.0365	0.2581	0.0040	0.3458
PERF				1.0000	0.1781	0.3841	0.1509	0.3721	-0.2286	0.0256	0.6570
LIQZ					1.0000	0.0181	0.6376	0.0308	0.1414	-0.0239	0.4273
PEAD						1.0000	-0.0473	0.2883	-0.2300	0.1320	0.1401
FIN							1.0000	-0.0340	0.2697	-0.0308	0.5285
MOMEV								1.0000	-0.5884	0.1221	0.1267
VALEV									1.0000	-0.0843	-0.0221
UMDQ										1.0000	-0.0062
QMJ											1.0000

Table 9: Summary statistics for the monthly research portfolios, for 5, 12 industry-sorted portfolios, 12 size-sorted portfolios, and 25 size- and size/book-to-market-sorted portfolios, from July 1972 to June 2013: monthly average return, standard deviation, minimum, and maximum.

5 industry-sorted portfolios	1	2	3	4	5								
Monthly average return	0.58%	0.58%	0.51%	0.58%	0.50%								
Standard deviation	4.70%	4.51%	5.82%	5.05%	5.43%								
Minimum	-25.62%	-21.42%	-23.00%	-21.06%	-24.18%								
Maximum	21.16%	16.77%	19.84%	29.01%	19.71%								
12 industry-sorted portfolios	1	2	3	4	5	6	7	8	9	10	11	12	
Mean	0.69%	0.45%	0.58%	0.74%	0.55%	0.49%	0.59%	0.54%	0.58%	0.58%	0.56%	0.37%	
Standard deviation	4.43%	6.67%	5.53%	5.67%	4.85%	6.87%	4.86%	4.12%	5.40%	5.05%	5.62%	5.54%	
Minimum	-21.63%	-32.71%	-29.18%	-19.01%	-25.19%	-26.41%	-16.44%	-12.94%	-28.83%	-21.06%	-22.53%	-29.81%	
Maximum	18.3%	42.62%	21.07%	23.60%	19.71%	20.32%	21.22%	18.26%	25.28%	29.01%	20.59%	18.85%	
18 size-sorted portfolios	1	2	3	4	5	6	7	8	9	10	11	12	13
Mean	0.68%	0.69%	0.48%	0.67%	0.70%	0.70%	0.66%	0.46%	0.68%	0.68%	0.72%	0.68%	0.73%
Standard deviation	6.20%	5.49%	4.51%	6.32%	5.99%	5.53%	5.27%	4.47%	6.29%	6.46%	6.13%	5.92%	5.76%
Minimum	-30.01%	-27.74%	-21.37%	-30.23%	-29.83%	-27.66%	-25.72%	-20.91%	-29.52%	-31.08%	-29.53%	-30.05%	-28.72%
Maximum	26.36%	22.18%	17.3%	27.23%	24.42%	22.01%	19.71%	17.61%	29.11%	27.82%	25.16%	23.75%	24.22%
18 size-sorted portfolios	14	15	16	17	18								
Mean	0.67%	0.68%	0.64%	0.59%	0.44%								
Standard deviation	5.40%	5.37%	5.25%	4.82%	4.45%								
Minimum	-26.75%	-26.83%	-24.92%	-22.87%	-20.32%								
Maximum	20.32%	21.83%	18.53%	17.63%	17.61%								
25 size- and size/book-to-market-sorted portfolios	1	2	3	4	5	6	7	8	9	10	11	12	13
Mean	0.10%	0.73%	0.77%	0.98%	1.09%	0.38%	0.71%	0.83%	0.91%	0.95%	0.42%	0.77%	0.74%
Standard deviation	8.10%	6.99%	5.98%	5.69%	6.09%	7.38%	6.12%	5.49%	5.28%	6.19%	6.84%	5.63%	5.07%
Minimum	-34.82%	-31.55%	-29.37%	-29.49%	-29.47%	-33.33%	-32.26%	-29.00%	-25.64%	-29.44%	-30.33%	-29.65%	-24.94%
Maximum	38.48%	40.62%	27.58%	27.26%	33.3%	27.75%	25.54%	25.76%	27.01%	29.13%	24.18%	24.39%	21.36%
25 size- and size/book-to-market-sorted portfolios	14	15	16	17	18	19	20	21	22	23	24	25	
Mean	0.86%	1.07%	0.55%	0.63%	0.69%	0.83%	0.83%	0.40%	0.60%	0.59%	0.50%	0.70%	
Standard deviation	4.99%	5.74%	6.23%	5.36%	5.21%	4.93%	5.77%	4.87%	4.68%	4.50%	4.90%	5.57%	
Minimum	-23.63%	-26.72%	-26.54%	-29.43%	-25.56%	-22.48%	-24.44%	-22.24%	-23.02%	-22.31%	-27.10%	-19.59%	
Maximum	22.82%	28.62%	25.79%	19.86%	23.43%	23.78%	27.32%	21.83%	16.11%	18.27%	19.18%	23.62%	

Table 10: Summary statistics for the monthly research portfolios, for 49 industry-sorted portfolios, from July 1972 to June 2013.

49 industry-sorted portfolios	1	2	3	4	5	6	7	8	9	10	11	12	13
Mean	0.65%	0.74%	0.68%	0.69%	0.99%	0.35%	0.80%	0.53%	0.41%	0.70%	0.69%	0.50%	0.64%
Standard deviation	6.51%	4.62%	6.91%	5.48%	6.41%	7.23%	7.90%	5.87%	4.86%	6.87%	7.86%	5.42%	5.20%
Minimum	-29.64%	-18.46%	-27.07%	-20.19%	-25.32%	-35.01%	-32.48%	-25.27%	-22.25%	-31.45%	-31.91%	-21.02%	-19.71%
Maximum	28.45%	18.99%	37.95%	25.51%	32.38%	26.42%	39.3%	30.73%	18.22%	31.79%	35.89%	20.52%	31.29%
49 industry-sorted portfolios	14	15	16	17	18	19	20	21	22	23	24	25	26
Mean	0.61%	0.63%	0.59%	0.61%	0.56%	0.37%	0.40%	0.56%	0.73%	0.45%	0.83%	0.65%	0.83%
Standard deviation	5.87%	6.16%	7.58%	6.44%	7.36%	7.81%	7.03%	6.59%	6.55%	7.28%	6.73%	7.63%	6.57%
Minimum	-28.60%	-31.15%	-33.11%	-31.89%	-32.12%	-32.99%	-26.93%	-31.91%	-32.80%	-36.50%	-30.83%	-32.87%	-30.47%
Maximum	21.68%	31.94%	58.92%	34.4%	23.61%	30.3%	30.37%	23.02%	22.87%	49.56%	24.53%	29.15%	31.88%
49 industry-sorted portfolios	27	28	29	30	31	32	33	34	35	36	37	38	39
Mean	0.51%	0.68%	0.81%	0.75%	0.54%	0.59%	0.37%	0.53%	0.47%	0.77%	0.61%	0.62%	0.58%
Standard deviation	10.96%	7.76%	10.69%	5.64%	4.11%	4.86%	6.60%	5.71%	7.54%	9.93%	7.85%	7.22%	5.77%
Minimum	-33.61%	-34.83%	-38.11%	-18.97%	-12.94%	-16.30%	-28.85%	-28.24%	-33.88%	-30.31%	-32.62%	-30.75%	-27.08%
Maximum	79.63%	26.95%	45.55%	23.70%	18.26%	21.20%	24.06%	24.8%	24.94%	40.42%	26.85%	21.08%	24.19%
49 industry-sorted portfolios	40	41	42	43	44	45	46	47	48	49			
Mean	0.59%	0.53%	0.55%	0.61%	0.60%	0.54%	0.62%	0.21%	0.69%	0.18%			
Standard deviation	5.86%	5.82%	5.47%	5.71%	6.25%	6.25%	5.66%	7.74%	6.46%	6.78%			
Minimum	-28.82%	-28.52%	-29.28%	-29.72%	-32.17%	-27.23%	-26.86%	-37.59%	-26.57%	-26.94%			
Maximum	20.19%	18.51%	17.47%	26.52%	28.23%	24.55%	26.31%	66.01%	19.51%	21.00%			

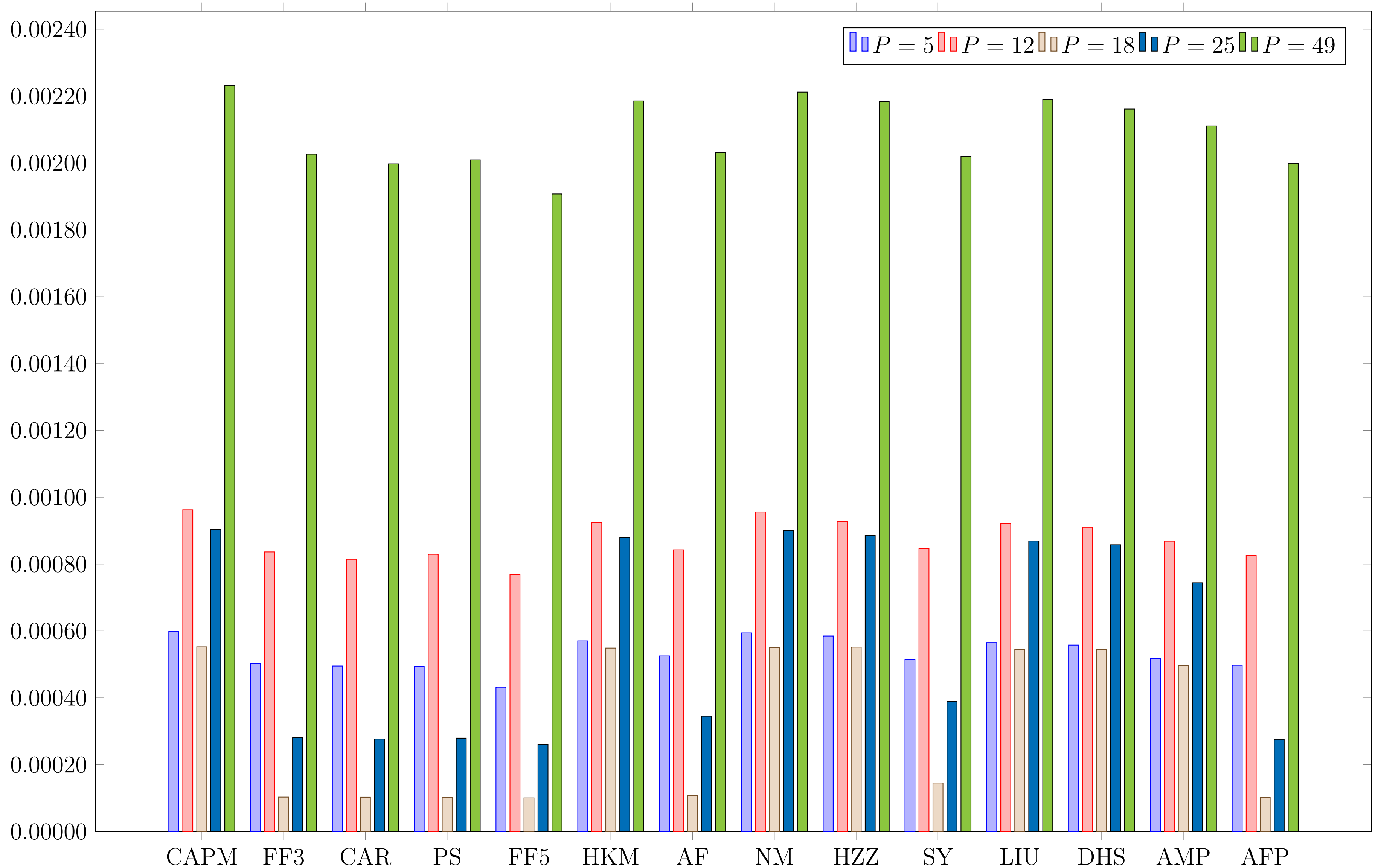


Figure 7: Average root mean squared error (RMSE) for 5 and 12 industry-sorted portfolios, 18 size-sorted, 25 size- and book-to-market-sorted portfolios, and 49 industry-sorted portfolios, for monthly data from July 1972 to June 2013.

Table 11: Industry classifications for the 5, 12, and 49 industry-sorted test portfolios.

Portfolios	Industry classification
$P = 5$	Consumer goods, manufacturing, business equipment, healthcare, others.
$P = 12$	Consumer nondurables, consumer durables, manufacturing, energy, chemicals, business equipment, telecommunications, utilities, wholesale and retail, healthcare, financial services, others.
$P = 49$	Agriculture, food products, candy & soda, beer & liquor, tobacco products, recreation, entertainment, printing and publishing, consumer goods, apparel, healthcare, medical equipment, pharmaceutical products, chemicals, rubber and plastic products, textiles, construction materials, construction, steel works, fabricated products, machinery, electrical equipment, automobiles and trucks, aircraft, shipbuilding, railroad equipment, defense, precious metals, non-metallic and industrial metal mining, coal, petroleum and natural gas, utilities, communication, personal services, business services, computers, computer software, electronic equipment, measuring and control equipment, business supplies, shipping containers, transportation, wholesale, retail, restaurants, hotels, motels, banking, insurance, real estate, trading, others.

A.3 Additional Simulation Results

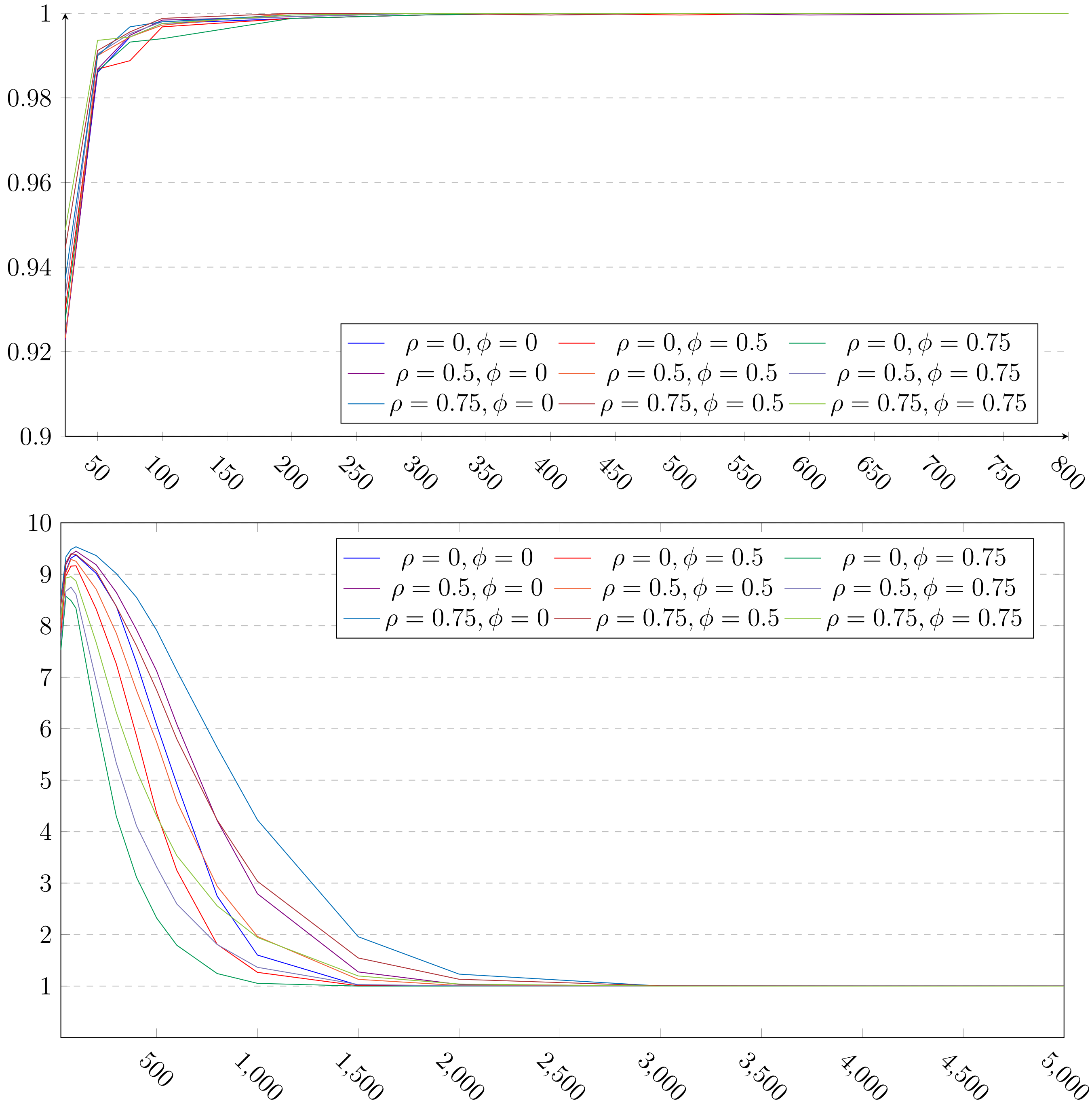
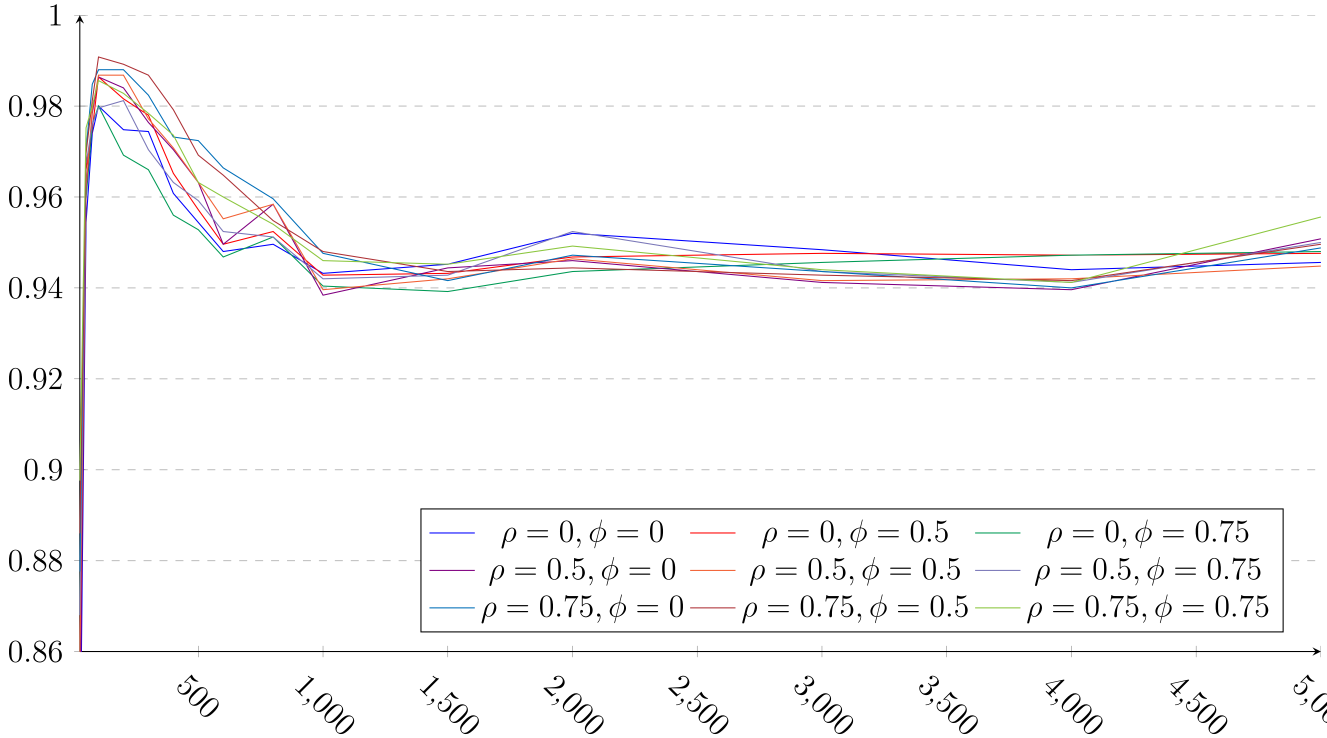


Figure 8: Simulation design for the supremum t statistic with dependent losses, $m = 10$ candidate models, $m_0 = 1$ true model, $P = 10$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true model is included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size. In the top panel, the frequency curve remains the same for sample sizes larger than $n = 800$ and is truncated for clarity.



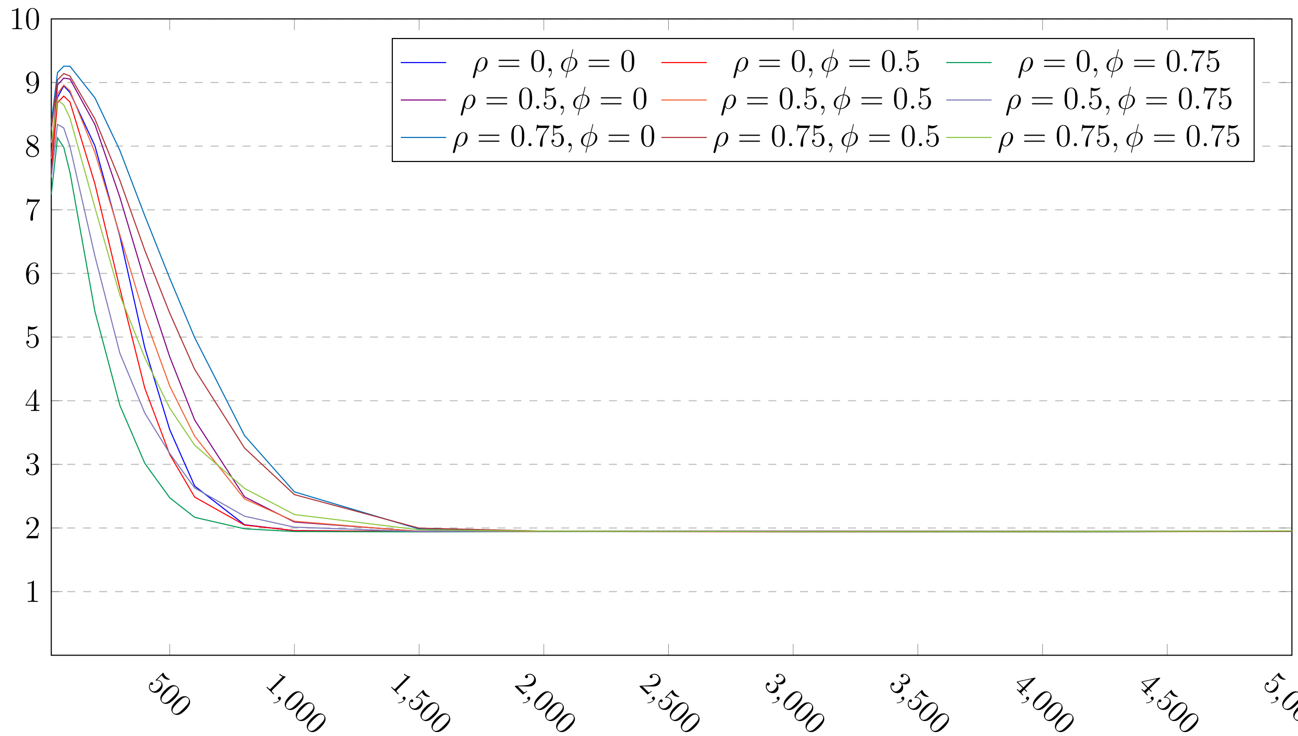


Figure 9: Simulation design for the supremum t statistic with dependent losses, $m = 10$ candidate models, $m_0 = 2$ true models, $P = 10$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true models are included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size.

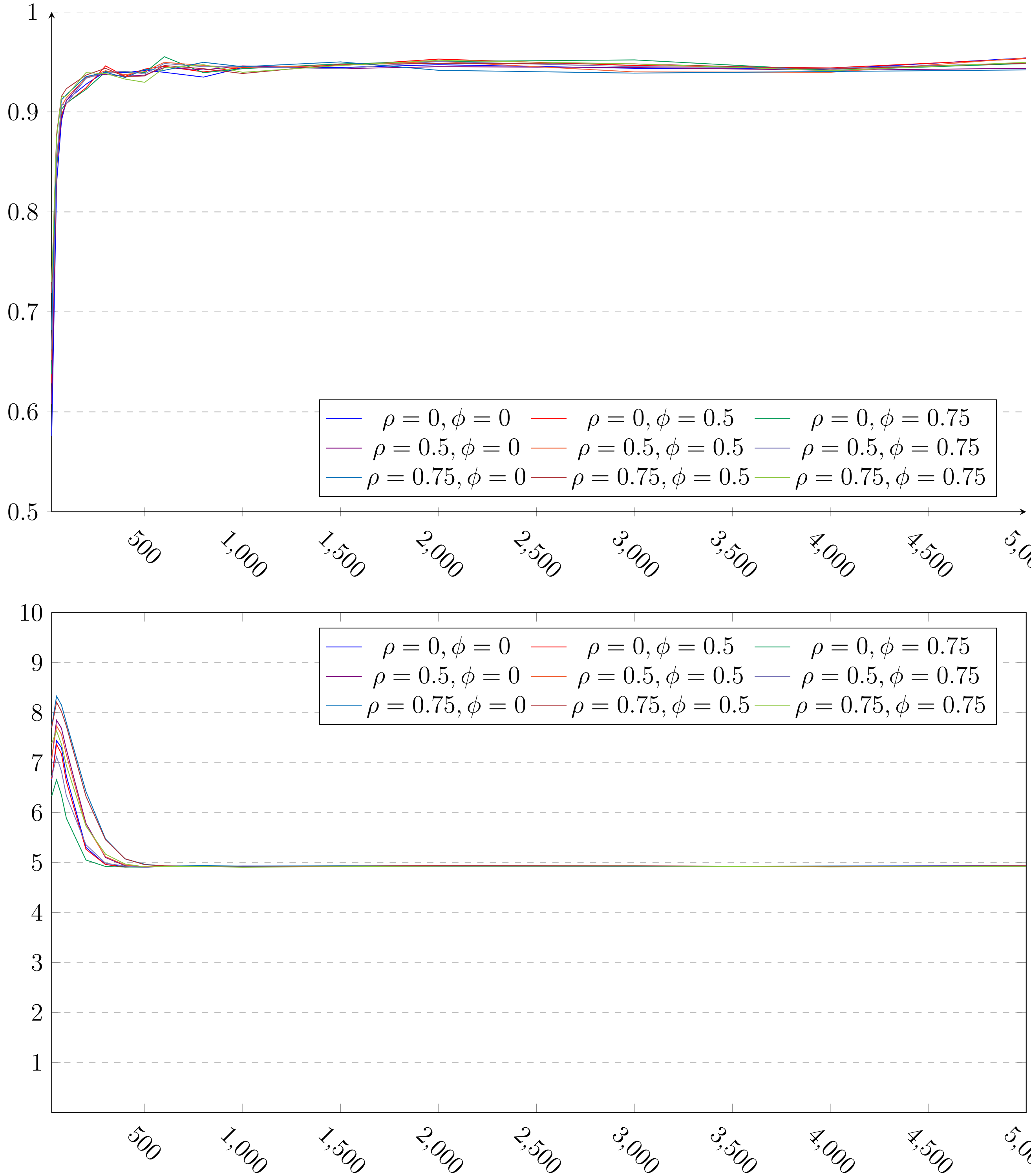


Figure 10: Simulation design for the supremum t statistic with dependent losses, $m = 10$ candidate models, $m_0 = 5$ true models, $P = 10$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true models are included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size.

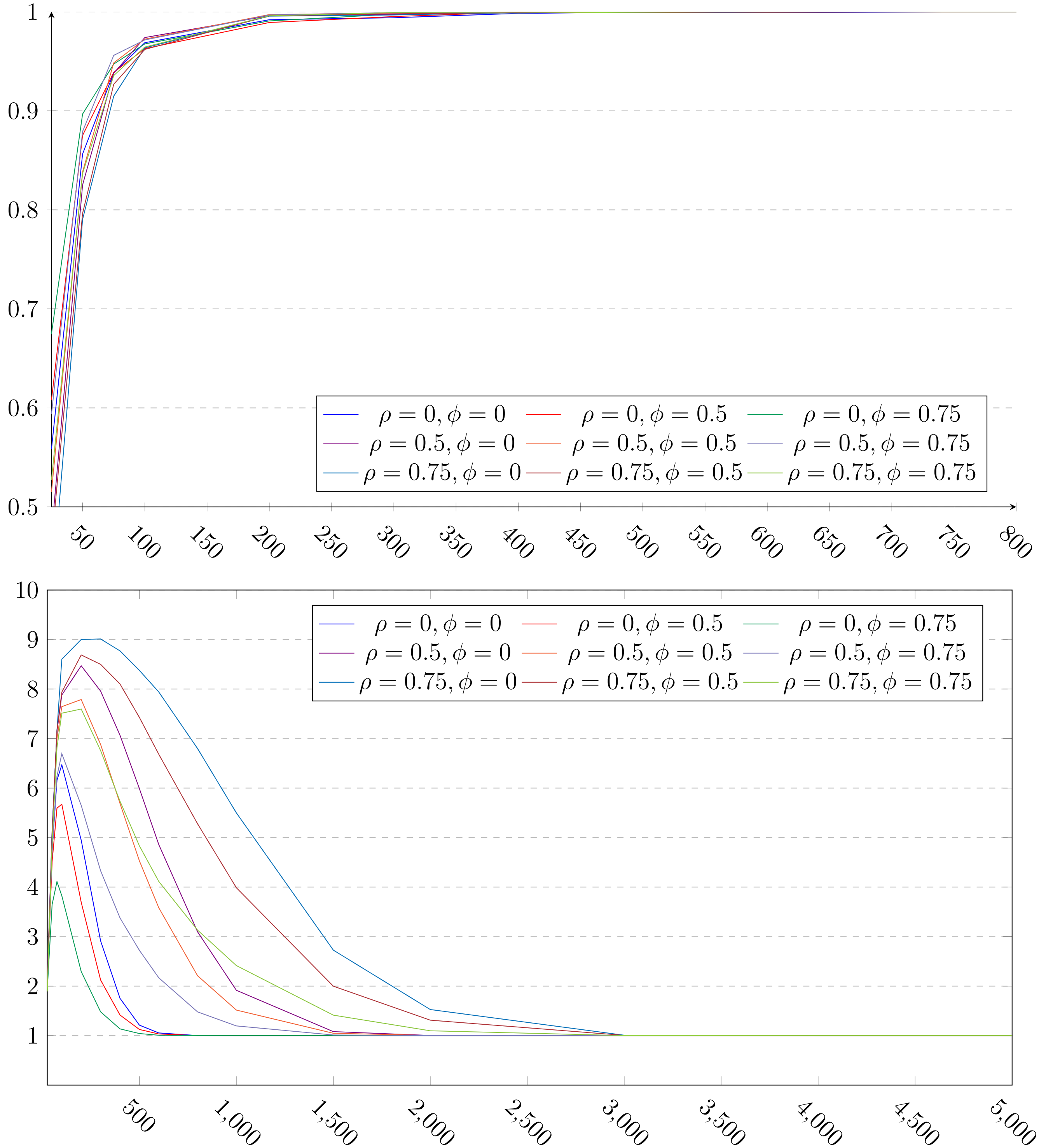


Figure 11: Simulation design for the Hotelling T^2 statistic with dependent losses, $m = 10$ candidate models, $m_0 = 1$ true model, $P = 10$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true model is included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. In both panels, the horizontal axis shows the sample size. In the top panel, the frequency curve remains the same for sample sizes larger than $n = 800$ and is truncated for clarity.

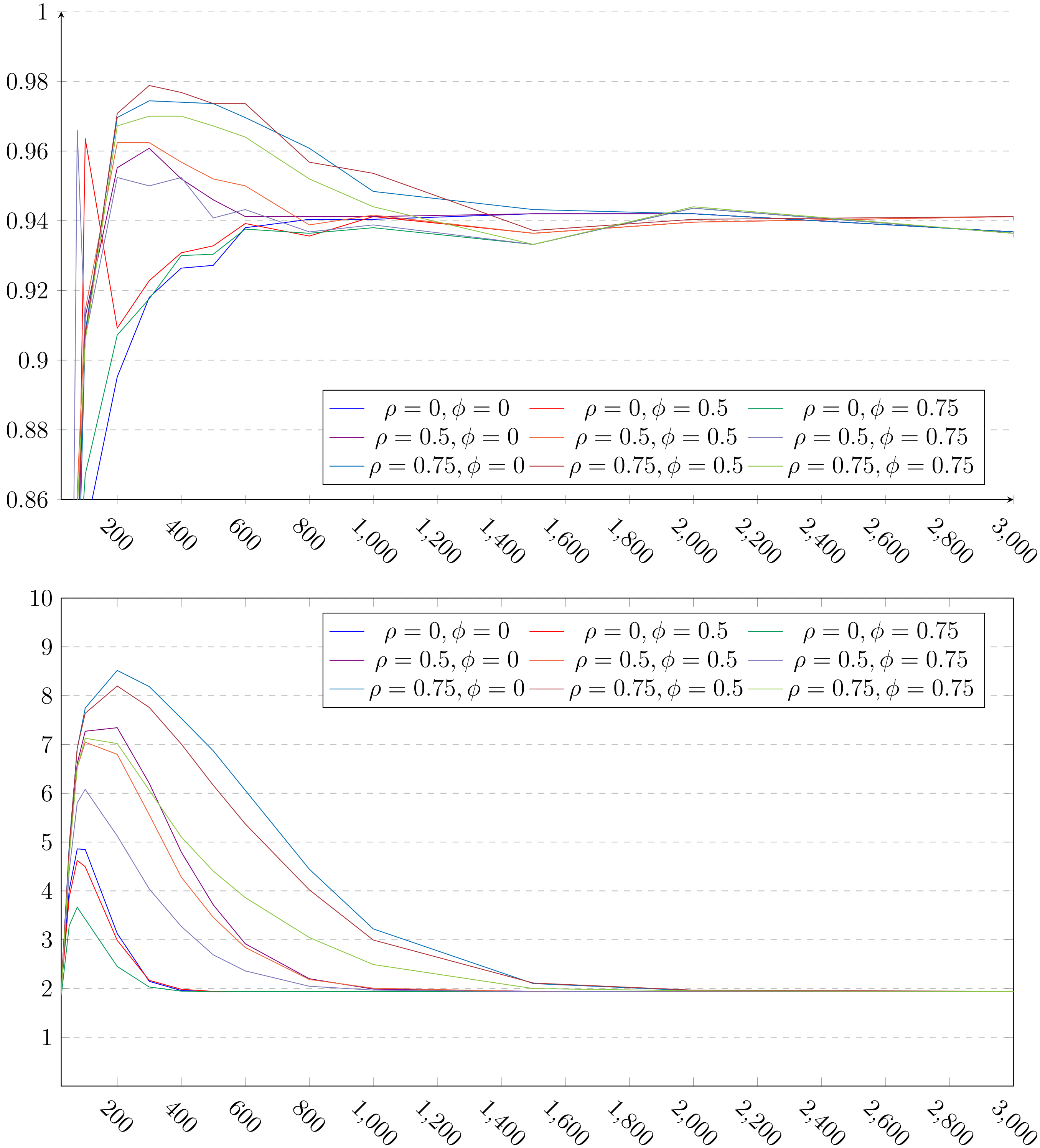


Figure 12: Simulation design for the Hotelling T^2 statistic with dependent losses, $m = 10$ candidate models, $m_0 = 2$ true models, $P = 10$ dependent variables, and $\alpha = 0.05$. In the top panel, the vertical axis shows the frequency at which the true model is included in the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$, and in the bottom panel, the vertical axis shows the average cardinality of the estimated model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$.

A.4 Additional Empirical Results

A.4.1 Augmented Fama and French (1993) models

$$R_{st} = \alpha_s + \beta_{1s}MKT_t + e_{st} \tag{CAPM}$$

$$R_{st} = \alpha_s + \beta_{1s}MKT_t + \beta_{2s}SMB_t + \beta_{3s}HML_t + e_{st} \tag{FF3}$$

$$R_{st} = \alpha_s + \beta_{1s}MKT_t + \beta_{2s}SMB_t + \beta_{3s}HML_t + \beta_{4s}RMW_t + \beta_{5s}CMA_t + e_{st} \tag{FF5}$$

$$R_{st} = \alpha_s + \beta_{1s}MKT_t + \beta_{2s}SMB_t + \beta_{3s}HML_t + \beta_{4s}UMDQ_t + \beta_{5s}QMJ_t + e_{st} \tag{AFP}$$

Table 12: MCS p-values for augmented Fama and French (1993) models, using the supremum t statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from January 1964 to October 2014. For out-of-sample tests, the horizon h is 12, 24, and 60. R_{5IND}, R_{12IND}, R_{ME}, $R_{ME/BEME}$, and R_{49IND} denote the portfolio returns for the 5-industry, 12-industry, size-sorted, size- and book-to-market-sorted, and 49-industry portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 10{,}000$.

		In-sample		Out-of-sample					
		January 1964 - October 2014		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{5IND}	CAPM	0.0640*	0.0560*	0.0020	0.0040	0.0060	0.0030	0.0140	0.0050
	FF3	0.0640*	0.0560*	1.0000**	0.0430	0.0060	0.0040	0.1880*	0.1420*
	FF5	1.0000**	1.0000**	0.3030**	1.0000**	1.0000**	1.0000**	0.0140	0.0230
	AFP	0.0640*	0.0560*	0.0020	0.0040	0.0060	0.0030	1.0000**	1.0000**
R_{12IND}	CAPM	0.0150	0.0340	0.0090	0.0070	0.0000	0.0000	0.0060	0.0020
	FF3	0.0530*	0.0360	0.0310	0.0070	0.0000	0.0000	0.5490**	0.1180*
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	0.0060	0.0020
	AFP	0.0530*	0.0360	0.0090	0.0070	0.0000	0.0000	1.0000**	1.0000**
R_{ME}	CAPM	0.0000	0.0000	0.0080	0.0220	0.0010	0.0040	0.0000	0.0000
	FF3	0.8470**	0.8160**	0.0080	0.5330**	0.0010	0.0140	0.1040*	0.1680*
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	AFP	0.8470**	0.8160**	0.5270**	0.6810**	0.0010	0.0140	0.0930*	0.1680*
$R_{ME/BEME}$	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0020
	FF3	0.2050*	0.2060*	0.0000	0.0000	0.0000	0.0000	0.0240	0.0130
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	AFP	0.2050*	0.2060*	0.0000	0.0000	0.0000	0.0000	0.0240	0.0130
R_{49IND}	CAPM	0.0010	0.0020	0.0020	0.0000	0.0000	0.0010	0.0080	0.0030
	FF3	0.1360*	0.1370*	0.0130	0.0000	1.0000**	1.0000**	0.3900**	0.2850**
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	0.0000	0.0000	0.0080	0.0030
	AFP	0.1360*	0.1370*	0.0090	0.0000	0.3880**	0.1170*	1.0000**	1.0000**

A.4.2 Modified Fama and French (1993) factor models

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + e_{st} \tag{CAPM}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t + e_{st} \tag{FF3}$$
$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} DHML_t + e_{st} \tag{AF}$$

Table 13: MCS p-values for modified Fama and French (1993) models, using the supremum t statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from January 1964 to October 2014. For out-of-sample tests, the horizon h is 12, 24, and 60. R_{5IND}, R_{12IND}, R_{ME}, $R_{ME/BEME}$, and R_{49IND} denote the portfolio returns for the 5-industry, 12-industry, size-sorted, size- and book-to-market-sorted, and 49-industry portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 10{,}000$.

		In-sample		Out-of-sample					
		January 1964 - October 2014		$h=12$		$h=24$		$h=60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{5IND}	CAPM	0.1100*	0.1360*	0.0180	0.0030	0.0040	0.0020	0.0270	0.0470
	FF3	1.0000**	0.3440**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	AF	0.3070**	1.0000**	0.0330	0.0030	0.0040	0.0020	0.4430**	0.3290**
R_{12IND}	CAPM	0.0250	0.0310	0.0260	0.0140	0.0010	0.0000	0.0190	0.0460
	FF3	1.0000**	0.3950**	1.0000**	1.0000**	0.0030	0.0020	0.3980**	0.5580**
	AF	0.3470**	1.0000**	0.0260	0.0140	1.0000**	1.0000**	1.0000**	1.0000**
R_{ME}	CAPM	0.0000	0.0000	0.0040	0.0060	0.0010	0.0030	0.0000	0.0000
	FF3	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	AF	0.3500**	0.4470**	0.0550*	0.0510*	0.1310*	0.2080*	0.0450	0.0500*
$R_{ME/BEME}$	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0020
	FF3	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	AF	0.0010	0.0040	0.0020	0.0120	0.1140*	0.1210*	0.1000*	0.1210*
R_{49IND}	CAPM	0.0120	0.0200	0.0000	0.0000	0.0050	0.0010	0.0360	0.0240
	FF3	0.6970**	0.7610**	1.0000**	1.0000**	1.0000**	1.0000**	0.0360	0.0280
	AF	1.0000**	1.0000**	0.0000	0.0340	0.0110	0.0040	1.0000**	1.0000**

A.4.3 Modified Fama and French (2015) models

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + e_{st} \tag{CAPM}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t + \beta_{4s} RMW_t + \beta_{5s} CMA_t + e_{st} \tag{FF5}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} HML_t^* + \beta_{3s} UMD_t^* + \beta_{4s} PMU_t^* + e_{st} \tag{NM}$$

Table 14: MCS p-values for modified Fama and French (2015) models, using the supremum t statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from January 1964 to October 2014. For out-of-sample tests, the horizon h is 12, 24, and 60. R_{5IND}, R_{12IND}, R_{ME}, $R_{ME/BEME}$, and R_{49IND} denote the portfolio returns for the 5-industry, 12-industry, size-sorted, size- and book-to-market-sorted, and 49-industry portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 10{,}000$.

		In-sample		Out-of-sample					
		January 1964 - October 2014		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{5IND}	CAPM	0.0610*	0.0610*	0.0180	0.0670*	0.0340	0.0560*	0.1620*	0.1640*
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	NM	0.0610*	0.0610*	0.0320	0.0670*	0.0340	0.0560*	0.1620*	0.1640*
R_{12IND}	CAPM	0.0130	0.0260	0.0590*	0.0090	0.0000	0.0000	0.0350	0.0280
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	NM	0.0130	0.0260	0.0590*	0.0190	0.0000	0.0000	0.0350	0.0280
R_{ME}	CAPM	0.0000	0.0000	0.0070	0.0000	0.0000	0.0030	0.0000	0.0000
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	NM	0.0000	0.0000	0.0080	0.0080	0.0000	0.0030	0.0020	0.0000
$R_{ME/BEME}$	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0030
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	NM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010	0.0030
R_{49IND}	CAPM	0.0020	0.0010	0.0000	0.0000	0.0000	0.0000	0.0720*	0.0300
	FF5	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	NM	0.0020	0.0020	0.0100	0.0000	0.0000	0.0000	0.0720*	0.0300

A.4.4 Liquidity-based models

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + e_{st} \tag{CAPM}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t + \beta_{4s} LIQ_t + e_{st} \tag{PS}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} LIQ_t + e_{st} \tag{LIU}$$

Table 15: MCS p-values for liquidity-based models, using the supremum t statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from January 1964 to October 2014. For out-of-sample tests, the horizon h is 12, 24, and 60. R_{5IND}, R_{12IND}, R_{ME}, $R_{ME/BEME}$, and R_{49IND} denote the portfolio returns for the 5-industry, 12-industry, size-sorted, size- and book-to-market-sorted, and 49-industry portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 10{,}000$.

		In-sample		Out-of-sample					
		January 1964 - October 2014		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{5IND}	CAPM	0.1030*	0.1960*	0.0250	0.0070	0.0010	0.0000	0.0150	0.0200
	PS	1.0000**	1.0000**	1.0000**	1.0000**	0.0190	0.0290	0.0150	0.0200
	LIU	0.2430*	0.3080**	0.0750*	0.1000**	1.0000**	1.0000**	1.0000**	1.0000**
R_{12IND}	CAPM	0.0090	0.0100	0.0020	0.0100	0.0000	0.0000	0.0160	0.0220
	PS	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	0.0160	0.0220
	LIU	0.0130	0.0300	0.0020	0.0100	0.0040	0.0130	1.0000**	1.0000**
R_{ME}	CAPM	0.0000	0.0000	0.0030	0.0050	0.0010	0.0020	0.0000	0.0000
	PS	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	LIU	0.0000	0.0000	0.0000	0.0010	0.0010	0.0020	0.0000	0.0000
$R_{ME/BEME}$	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0030
	PS	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	LIU	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0030
R_{49IND}	CAPM	0.0140	0.0220	0.0010	0.0020	0.0040	0.0010	0.0280	0.0100
	PS	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	LIU	0.0160	0.0240	0.0010	0.0020	0.0340	0.0230	0.0280	0.0130

A.4.5 Momentum-based models

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + e_{st} \tag{CAPM}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} HML_t + \beta_{4s} MOM_t + e_{st} \tag{CAR}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} MOM_t + \beta_{3s} TREND_t + e_{st} \tag{HZZ}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} MOM_t + \beta_{3s} VAL_t + e_{st} \tag{AMP}$$

Table 16: MCS p-values for momentum-based models, using the supremum t statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from January 1972 to December 2017. For out-of-sample tests, the horizon h is 12, 24, and 60. R_{5IND}, R_{12IND}, R_{ME}, $R_{ME/BEME}$, and R_{49IND} denote the portfolio returns for the 5-industry, 12-industry, size-sorted, size- and book-to-market-sorted, and 49-industry portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to $B = 10{,}000$.

		In-sample		Out-of-sample					
		January 1972 - December 2017		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{5IND}	CAPM	0.0630*	0.0820*	0.0090	0.0130	0.2180*	0.0020	0.0140	0.0120
	CAR	1.0000**	1.0000**	1.0000**	1.0000**	0.0000	1.0000**	0.0140	1.0000**
	HZZ	0.0630*	0.0820*	0.0080	0.0000	1.0000**	0.0240	0.7250**	0.0120
	AMP	0.0630*	0.0820*	0.0080	0.0030	0.5790**	0.0000	1.0000**	0.0120
R_{12IND}	CAPM	0.0150	0.0190	0.0010	0.0000	0.0000	0.0000	0.0190	0.0170
	CAR	1.0000**	1.0000**	1.0000**	1.0000*	1.0000**	1.0000**	1.0000**	1.0000**
	HZZ	0.0310	0.0190	0.0210	0.1040*	0.0040	0.0000	0.0230	0.0480
	AMP	0.0310	0.0190	0.0010	0.0000	0.0000	0.0000	0.0190	0.0170
R_{ME}	CAPM	0.0000	0.0000	0.0000	0.0000	0.0010	0.0030	0.0000	0.0000
	CAR	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	HZZ	0.0000	0.0000	0.0000	0.0000	0.0010	0.0030	0.0000	0.0000
	AMP	0.0000	0.0000	0.0000	0.0000	0.0090	0.0070	0.0000	0.0000
$R_{ME/BEME}$	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010
	CAR	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	HZZ	0.0000	0.0000	0.0070	0.0000	0.0000	0.0010	0.0000	0.0010
	AMP	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000	0.0010	0.0040
R_{49IND}	CAPM	0.0220	0.0290	0.0020	0.0000	0.0000	0.0000	0.0470	0.0000
	CAR	1.0000**	1.0000**	0.0020	0.0000	1.0000**	1.0000**	1.0000**	1.0000**
	HZZ	0.0180	0.0210	0.0020	0.0000	0.0000	0.0090	0.0700*	0.0500*
	AMP	0.0220	0.0290	1.0000**	1.0000**	0.0000	0.0000	0.0700*	0.0500*

A.4.6 Market-based models

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + e_{st} \tag{CAPM}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} ICRF_t + e_{st} \tag{HKM}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} SMB_t + \beta_{3s} MGMT_t + \beta_{4s} PERF_t + e_{st} \tag{SY}$$

$$R_{st} = \alpha_s + \beta_{1s} MKT_t + \beta_{2s} PEAD_t + \beta_{3s} FIN_t + e_{st} \tag{DHS}$$

Table 17: MCS p-values for market-based models, using the supremum t statistic, using the quadratic loss function, for in-sample and out-of-sample tests, at the monthly frequency from July 1972 to December 2016. For out-of-sample tests, the horizon h is 12, 24, and 60. R_{5IND}, R_{12IND}, R_{ME}, $R_{ME/BEME}$, and R_{49IND} denote the portfolio returns for the 5-industry, 12-industry, size-sorted, size- and book-to-market-sorted, and 49-industry portfolios, respectively. Inclusion in the 95% and the 75% MCS is denoted by * and **, respectively. $\alpha_i \neq 0$ denotes the presence of regression intercept. The block bootstrap length is $l = 12$ for in-sample tests and $l = 3$ for out-of-sample tests. The number of bootstrap iterations is set to B = 10,000.

		In-sample		Out-of-sample					
		July 1972 to December 2016		$h = 12$		$h = 24$		$h = 60$	
Portfolio type	Model	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$	$\alpha_i \neq 0$	$\alpha_i = 0$
R_{5IND}	CAPM	0.0670*	0.0650*	0.0800*	0.1790*	0.0000	0.0060	0.1500*	0.1590*
	HKM	0.0930*	0.0670*	0.0870*	0.1880*	0.0710*	0.1150*	0.4240**	0.2610**
	SY	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000*	1.0000**	1.0000**
	DHS	0.0930*	0.0810*	0.0870*	0.1880*	0.0710*	0.1150*	0.1500*	0.1590*
R_{12IND}	CAPM	0.0180	0.0260	0.0000	0.0000	0.0000	0.0000	0.1850*	0.0970*
	HKM	0.0500*	0.1050*	0.0000	0.0000	0.0000	0.0010	0.3360**	0.1110*
	SY	1.0000*	1.0000**	1.0000**	1.0000**	1.0000**	1.0000*	1.0000**	1.0000**
	DHS	0.0180	0.0260	0.1790*	0.2510**	0.2150*	0.1990*	0.2360*	0.1110*
R_{ME}	CAPM	0.0000	0.0000	0.0200	0.0000	0.0000	0.0040	0.0000	0.0010
	HKM	0.0000	0.0000	0.0030	0.0000	0.0000	0.0040	0.0000	0.0010
	SY	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	DHS	0.0000	0.0000	0.0200	0.0000	0.0000	0.0040	0.0000	0.0010
$R_{ME/BEME}$	CAPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010
	HKM	0.0000	0.0000	0.0000	0.0010	0.0010	0.0030	0.0000	0.0010
	SY	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	DHS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0010
R_{49IND}	CAPM	0.0080	0.0130	0.0000	0.0000	0.0000	0.0010	0.0380	0.0330
	HKM	0.0080	0.0130	0.0000	0.0000	0.0000	0.0010	0.0380	0.0330
	SY	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**	1.0000**
	DHS	0.0080	0.0130	0.0180	0.0640*	0.0040	0.0010	0.0380	0.0330

References

Aslanidis, N., C. Christiansen, and A. Cipollini (2018): "Predicting bond betas using macro-finance variables," *Finance Research Letters*.

Asness, C. and A. Frazzini (2013): "The devil in HMLs details," *The Journal of Portfolio Management*, 39, 49–68.

Asness, C. S., A. Frazzini, and L. H. Pedersen (2019): "Quality minus junk," *Review of Accounting Studies*, 24, 34–112.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen (2013): "Value and momentum everywhere," *The Journal of Finance*, 68, 929–985.

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

——— (2018): "Comparing asset pricing models," *The Journal of Finance*, 73, 715–754.

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

Daniel, K., D. Hirshleifer, and L. Sun (2020): "Short-and long-horizon behavioral factors," *The review of financial studies*, 33, 1673–1736.

Diebold, F. X. and R. S. Mariano (2002): "Comparing predictive accuracy," *Journal of Business & economic statistics*, 20, 134–144.

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

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

Gu, S., B. T. Kelly, and D. Xiu (2018): "Empirical asset pricing via machine learning," *Chicago Booth Research Paper*, 18-04.

Han, Y., G. Zhou, and Y. Zhu (2016): "A trend factor: Any economic gains from using information over investment horizons?" *Journal of Financial Economics*, 122, 352–375.

Hansen, P. R., A. Lunde, and J. M. Nason (2003): "Choosing the best volatility models: the model confidence set approach," *Oxford Bulletin of Economics and Statistics*, 65, 839–861.

——— (2011): "The model confidence set," *Econometrica*, 79, 453–497.

Harvey, C. R. (2017): "Presidential address: The scientific outlook in financial economics," *The Journal of Finance*, 72, 1399–1440.

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

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

Hou, K., H. Mo, C. Xue, and L. Zhang (2018): "Which factors?" *Review of Finance*, 23, 1–35.

Huang, D., J. Li, and G. Zhou (2018): "Shrinking factor dimension: A reduced-rank approach," *Working Paper*.

Jentsch, C., D. N. Politis, and E. Paparoditis (2015): "Block bootstrap theory for multivariate integrated and cointegrated processes," *Journal of Time Series Analysis*, 36, 416–441.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial economics*, 96, 175–194.

Lintner, J. (1965): "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *The Review of Economics and Statistics*, 47, 13–47.

Liu, W. (2006): "A liquidity-augmented capital asset pricing model," *Journal of Financial Economics*, 82, 631–671.

Mariano, R. S. and D. Preve (2012): "Statistical tests for multiple forecast comparison," *Journal of econometrics*, 169, 123–130.

Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen (2012): "Time series momentum," *Journal of Financial Economics*, 104, 228–250.

Mossin, J. (1966): "Equilibrium in a capital asset market," *Econometrica*, 768–783.

Novy-Marx, R. (2013): "The other side of value: The gross profitability premium," *Journal of Financial Economics*, 108, 1–28.

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

Pukthuanthong, K., R. Roll, and A. Subrahmanyam (2018): "A protocol for factor identification," *The Review of Financial Studies*.

Samuels, J. D. and R. M. Sekkel (2017): "Model confidence sets and forecast combination," *International Journal of Forecasting*, 33, 48–60.

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

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

Treynor, J. L. (1961): "Market value, time, and risk," .

——— (1962): "Toward a theory of market value of risky assets," .