Department of Economics- FEA/USP

Risk Factors' CPDAG Roots and the Cross-Section of Expected Returns

FERNANDO MORAES

RODRIGO DE-LOSSO

WORKING PAPER SERIES Nº 2020-18

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2020-18

Risk Factors' CPDAG Roots and the Cross-Section of Expected Returns

Fernando Moraes (fernandotm@al.insper.edu.br)

Rodrigo De-Losso (delosso@usp.br)

Research Group: NEFIN

Abstract: The Asset pricing literature has produced hundreds of risk factor candidates aimed at explaining the cross-section of expected excess returns, although risk factors which are in fact capable of providing independent information remains an open question. Appling a sparse model, Kozak, Nagel, and Santosh (2020) achieve satisfactory results on explaining cross-sectional returns only with PCs (*principal components*). In this paper, we propose a new methodology that seeks to reduce risk factor predictor dimensions by estimating the joint risk factor distribution with CPDAG (*complete partial directed acyclic graph*), in addition to selecting the CPDAG root as the only new risk factor candidate set. Our approach yields a significant shrinkage in the original set of risk factors, whereas our findings lead to sparse models that pose better results than those attained with the standard models and with alternative methods proposed by PCs factor zoo related research papers.

Keywords: Risk factors, factor zoo, DAG, CPDAG.

JEL Codes: G12, C55, D85.

Risk factors' CPDAG roots and the cross-section of expected returns

Fernando Moraes[1] Rodrigo De-Losso [2]

September 11, 2020

Abstract

The Asset pricing literature has produced hundreds of risk factor candidates aimed at explaining the cross-section of expected excess returns, although risk factors which are in fact capable of providing independent information remains an open question. Appling a sparse model, Kozak, Nagel, and Santosh (2020) achieve satisfactory results on explaining cross-sectional returns only with PCs (*principal components*). In this paper, we propose a new methodology that seeks to reduce risk factor predictor dimensions by estimating the joint risk factor distribution with CPDAG (*complete partial directed acyclic graph*), in addition to selecting the CPDAG root as the only new risk factor candidate set. Our approach yields a significant shrinkage in the original set of risk factors, whereas our findings lead to sparse models that pose better results than those attained with the standard models and with alternative methods proposed by PCs factor zoo-related research papers.

Keywords: Risk factors, factor zoo, DAG, CPDAG.

JEL Codes: G12, C55, D85.

.

[1] Email: fernandotm@al.insper.edu.br.

[2] Email: delosso@usp.br.

1. Introduction

Cochrane (2011) states the issue of the factor zoo as the need to answer which risk factors are in fact capable of providing independent information on the cross-section of expected excess returns. Considering that asset pricing literature has produced over two hundred different potential risk factors in the last decades (Harvey, Liu, and Zhu (2016), the factor zoo phenomenon has become a high-dimensional econometric problem, which additionally brings forth new methodological challenges for empirical research (i.e. overfitting, data mining, and design matrix dimension reduction)[3]. Fortunately, though, increased alternative statistical methods geared towards these high-dimensional challenges[4] have enabled researchers to start addressing the issue of the factor zoo[5].

In this paper, we propose a new methodology aimed at diminishing risk factor predictor dimension by estimating the joint risk factor distribution with a *complete partial directed acyclic graph* (CPDAG) and by selecting the CPDAG root as the new candidate set to explain cross-sectional returns. The main driving force behind the decision to apply the CPDAG to the joint risk factor distribution entails its ability to address high-dimensional problems as well as track causal relations. Given the vast set of potential risk factor candidates available to explain cross-sectional returns, these CPDAG properties enable us to identify a sparse set of risk factors that potentially spans all the remaining risk factors. Based on the reasoning of Ross's (1976) APT model, this new sparse set of risk factors relates to natural candidates capable of explaining cross-sectional returns. In this methodology, the CPDAG root is used to solve high-dimensional problems since this set of results poses only a couple of risk factors compared to the original set of risk factors. We then apply the Fama-MacBeth procedure to verify whether risk factors selected using our methodology provide better results as opposed to alternative models.

Our findings achieve a sparse risk factor model that poses better results than standard models documented in the asset pricing literature, as well as certain principal component-related methods proposed by a bunch of factor zoo papers. In addition to significant factor risk premia parameters, our model yields the highest in-sample average

[3] Hastie, Tibshirani, and Friedman (2009); and Abu-Mostafa, Magdon-Ismail, and Lin (2012).
[4] Some of these methods are dubbed 'Machine Learning' techniques.' To learn more, see Hastie, Tibshirani, and Friedman (2009).
[5] Harvey, Liu, and Zhu (2016); Green, Hand, and Zhang (2017); Yan and Zheng (2017); Feng, Giglio and Xiu (2020); Freyberger, Neuhierl and Weber (2020); and Kozak, Nagel and Santosh (2020).

adjusted R^2, and regarding out-of-sample results, the lowest root-mean-square error across the models.

This paper adds a chapter to such a new literature about high-dimensional cross-sectional asset pricing models. This research field applies a wide range of statistical methods, such as bootstrapping[6], lasso[7], multiple-test corrections[8], and principal component analysis[9] to achieve robust estimators in high-dimensional environments, in addition to evaluating which risk factors are in fact capable of explaining the cross-section of expected returns.

Feng, Giglio, and Xiu (2020); and Freyberger, Neuhierl, and Weber (2020) attain non-sparse results whenever the issue at stake concerns how to explain the cross-section of expected returns. Kozak, Nagel, and Santosh (2020), in turn, find that sparsity models are only capable of achieving satisfactory results to explain cross-sectional returns when they use the principal components of portfolio returns as risk factors. Applying the *instrumented principal component*, Kelly, Pruitt, and Su (2019) conclude that only five latent factors manage to provide satisfactory results when it comes to explaining average cross-sectional returns. We therefore witness the emergence of the well-known fact that the cross-section of expected returns can only be adequately described by PCs in a sparse representation. Whenever we apply risk factors, it becomes increasingly challenging to explain the cross-section of expected returns satisfactorily with a sparse model. Considering this stylized fact, we define the principal component model as our benchmark to evaluate our methodology. Our methodology presents an advantage over the principal component analysis since PCs do not present directed economic interpretation and our risk factor selected set does.

In summary, this paper seeks to add a new method to the existing factor zoo-related literature, thereby enabling a significant shrinkage in the original set of risk factors, as well as enabling investigations on joint risk factor distribution. To the best of our knowledge, this is also the first paper that uses a CPDAG model to describe joint risk factor distribution, in addition to using CPDAG properties to select risk factor candidates.

[6] See Harvey and Liu (2019); and Yan and Zheng (2017).
[7] See Kozak, Nagel, and Santosh (2020); Feng, Giglio and Xiu (2020); and Freyberger, Neuhierl and Weber (2020).
[8] See Harvey, Liu, and Zhu (2016); and Green, Hand, and Zhang (2017).
[9] See Kelly, Pruitt, and Su (2019); and Gu, Kelly, and Xiu (2019).

2. Methodology

We describe our research method in two main steps. First, we estimate a *complete partial directed acyclic graph* (CPDAG) to represent the joint risk factor distribution. We employ an analytical framework to make a comparison between the CPDAG *root* and the Ross's (1976) APT model. We conclude our first step by selecting the CPDAG *root* risk factor subset as our candidate to explain the cross-section of expected returns. The risk factor dimension is reduced since the CPDAG *root* set is bounded by the original set. Second, we apply the Fama-MacBeth (FM) procedure[10] to verify whether risk factors selected using our methodology provide better results compared to certain well-known sparse "standard models" from the asset pricing literature, as well as other methodologies proposed by papers addressing the factor zoo that we describe ahead. Finally, we perform an out-of-sample evaluation on all models.

2.1 Risk factor selection using the CPDAG root

2.1.1 DAG joint distribution function representation

Before we discuss the representation of the joint risk factor distribution using a *direct acyclic graph* (DAG), in addition to its implication on the asset pricing theory, first it is important to introduce the graph terminology used herein.

We define *graph* $\mathcal{G} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}})$ as a set of *vertices* $\boldsymbol{\mathcal{V}} = \{1, \dots, P\}$ and a set of *edges* $\boldsymbol{\mathcal{E}} \subseteq \boldsymbol{\mathcal{V}} \times \boldsymbol{\mathcal{V}}$, where these vertices are a random variable vector $\boldsymbol{X} \in \mathbb{R}^P$, while the edges describe the relationship between two variables (X_i, X_j and $i, j \in \boldsymbol{\mathcal{V}}$). We define a *directed edge* $(i \to j)$ as $(i,j) \in \boldsymbol{\mathcal{E}}$ although $(j,i) \notin \boldsymbol{\mathcal{E}}$, whereas an *undirected edge* $(i - j)$ is defined as $(i,j) \in \boldsymbol{\mathcal{E}}$, and $(j,i) \in \boldsymbol{\mathcal{E}}$. A *skeleton* of $\mathcal{G}$ is defined as the $\boldsymbol{\mathcal{E}}$ set, without taking into account edge directions. Vertices i and j are said to be *adjacent* if $(i - j) \in \boldsymbol{\mathcal{E}}$; $(j \to i) \in \boldsymbol{\mathcal{E}}$; or if $(j \leftarrow i) \in \boldsymbol{\mathcal{E}}$. We define a_i as the set of all adjacent vertices of vertex i. A *v-structure* is defined by any three vertices (i, j, k) such that $(i \to j) \in \boldsymbol{\mathcal{E}}$, $(k \to j) \in \boldsymbol{\mathcal{E}}$, and i and k are not adjacent. A vertex i is called a *parent* of j if $\exists i \to j$, and we define p_j as the set of all parent vertices[11] of j. A vertex i is a *root* of $\mathcal{G}$ if $p_i = \emptyset$, and $P_{\mathcal{G}} = \{i : p_i = \emptyset\}$ is the set of all graph roots. A direct acyclic graph ($\mathcal{G}_{DAG} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}})$) is a $\mathcal{G} =$

[10] Fama and MacBeth (1973).
[11] Note that if $\exists i \to j \Leftrightarrow i \in p_j$.

$(\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}})$ in which all edges are directed, feature at least one root and do not contain any cycles[12].

The $P(\boldsymbol{X})$ joint distribution of $\boldsymbol{X}$ is said to be factorized by its Markovian parents if:

$$P(\boldsymbol{X}) = \prod_{i=1}^{P} P(X_i|\{\boldsymbol{X}\} \setminus X_i) = \prod_{i=1}^{P} P(X_i|m_i) \tag{1}$$

where m_i is the Markovian parent set for X_i, which is defined as the minimal subset of $\{\{\boldsymbol{X}\} \setminus X_i\}$ such that $P(X_i|\{\boldsymbol{X}\} \setminus X_i) = P(X_i|m_i)$. If equation (1) stands, then $P(\boldsymbol{X})$ can be factorized according to $\mathcal{G}_{DAG} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}})$, by setting $m_i = p_i$ for every $i \in \boldsymbol{\mathcal{V}}$. In other words, the conditional independence relationship for $P(\boldsymbol{X})$ can be inferred by the $\mathcal{G}_{DAG}$ edge structure since $i \to j \in \mathcal{E} \Leftrightarrow X_i \in m_j$. Following the seminal work by Pearl (2009), the causal relationship for a DAG can be interpreted as follows: X_i is a direct cause of X_j if and only if $\exists i \to j$. It is worth noting that each directed edge indicates a conditional dependence relationship that cannot be attributed to any other variable. This paper explores this DAG property with the aim of finding the original set of risk factors responsible for spanning of all excess return space.

2.1.2 Implications of DAG risk factor representation for the asset pricing theory

Ross's (1976) APT model allows for the following beta representations:

$$r_{i,t} = \sum_{p=1}^{P} \beta_{i,p} f_{p,t} + \varepsilon_{i,t} \tag{2}$$

$$E(r_{i,t}) = \sum_{p=1}^{P} \beta_{i,p} \lambda_p \tag{3}$$

where $r_{i,t}$ represents the excess returns of asset i for period t; $f_{p,t}$ represents the $p-th$ risk factor for period t; $\varepsilon_{i,t}$ is an error term; $\beta_{i,p}$ represents the loadings of asset i on the $p-th$ risk factor; and λ_p represents the p factor risk premia[13]. This asset pricing model allows us to state the well-known theoretical hypothesis that all excess return spaces are

[12] A cycle is a sequence of more than one edges in which the first and last vertices are the same, according to set $c = \{(i \to j), (j \to k), (k \to i)\}$ as an example. For a formal definition, see Diestel (2012).

[13] The APT theory assumes that $E(\varepsilon_i) = 0$; $cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$; and $cov(\varepsilon_i, f_p) = 0$ for each i and p.

spanned the risk factors. As highlighted before, the issue of the factor zoo consists of the fact that the observable set of risk factors is large when compared to the available asset's excess return time series length. Thus, with the purpose of selecting a sparse set of risk factors, we define the observable set of risk factors as $\boldsymbol{f} = (f_1, \dots, f_P)^T$ where P is usually large, and further assume the existence of a "true" risk factor set such that $\boldsymbol{f}_{TRUE} \subseteq \boldsymbol{f}$ and $\boldsymbol{f}_{TRUE}$ spans the entire excess return space, including $\{\boldsymbol{f}\} \setminus \{\boldsymbol{f}_{TRUE}\}$. In other words, we assume that risk factors comprising the stochastic discount factor ($\boldsymbol{f}_{TRUE}$) are included in the original $\boldsymbol{f}$ set.

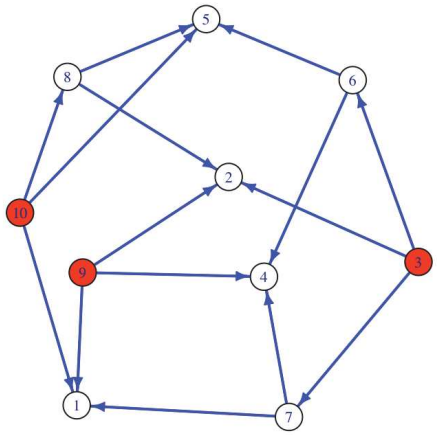
First, we define $P(\boldsymbol{f})$ as the joint risk factor distribution function. By stating a restricted version of the APT model, where we set $cov(f_i, f_j) = 0$ for $i \neq j$ $(i, j \in (1, \dots, P))$ as an additional assumption, we can establish a relationship between $\boldsymbol{f}_{TRUE}$ and the $\mathcal{G}_{DAG} = (\boldsymbol{f}, \boldsymbol{\mathcal{E}})$ root. Note that if $cov(f_i, f_j) = 0$ when $i, j \in \boldsymbol{f}_{TRUE}$, then the joint distribution of $\boldsymbol{f}_{TRUE}$ can be expressed as:

$$P(\boldsymbol{f}_{TRUE}) = \prod\nolimits_{i \in \{\boldsymbol{f}_{TRUE}\}} P(f_i) \tag{4}$$

Equation (4) allows us to observe that $m_i = \emptyset$ for all $i \in \boldsymbol{f}_{TRUE}$. Since every excess return is spanned by the $\boldsymbol{f}_{TRUE}$ set, if we add another set of risk factors, defined as $\boldsymbol{f}_{OTHERS}$, such that $\boldsymbol{f} = \boldsymbol{f}_{TRUE} \cup \boldsymbol{f}_{OTHERS}$, and hold the assumption that $P(\boldsymbol{f})$ can be expressed as (1), we therefore have $m_i \neq \emptyset$ for all $i \in \boldsymbol{f}_{OTHERS}$. Thus, $P(\boldsymbol{f})$ can be represented by $\mathcal{G}_{DAG} = (\boldsymbol{f}, \boldsymbol{\mathcal{E}})$ where $\boldsymbol{f}_{TRUE}$ is given by $\boldsymbol{f}_S = \{i : m_i = \emptyset\}$, which is identical to $P_{\mathcal{G}_{DAG}}$. In other words, the true set of risk factors corresponds to the set of all $\mathcal{G}_{DAG}$ roots. By selecting only the $\mathcal{G}_{DAG} = (\boldsymbol{f}, \boldsymbol{\mathcal{E}})$ roots as the risk factor candidate to explain cross-sectional asset returns, we are able to proceed with our first shrinkage step since the number of elements in $\boldsymbol{f}_{TRUE}$ is bounded by the number of elements in $\boldsymbol{f}$.

Figure 1 plots a DAG example for a $P(\boldsymbol{f})$ distribution where $P = 10$, $\boldsymbol{f}_{TRUE} = \{3,9,10\}$, which are vertices highlighted in red, while $\boldsymbol{f}_{OTHERS} = \{1,2,4,5,6,7,8\}$. As we can see, $p_i \neq \emptyset$ applies to all $i \in \boldsymbol{f}_{OTHERS}$ since these risk factors are generated by $\boldsymbol{f}_{TRUE}$ and its Markovian parent set is not empty by definition ($m_i = p_i$). On the other hand, when it comes to the red vertices ($i \in \boldsymbol{f}_{TRUE}$), $p_i = \emptyset$ since these risk factors are independent among them and are not generated by any other variable.

Figure 1: DAG example



Note: The figure shows a DAG example for a $P(\boldsymbol{f})$ distribution where $\boldsymbol{f} = \boldsymbol{f}_{TRUE} \cup \boldsymbol{f}_{OTHERS}$, $\boldsymbol{f}_{TRUE} = \{3,9,10\}$, $\boldsymbol{f}_{OTHERS} = \{1,2,4,5,6,7,8\}$, $p_1 = \{7,9,10\}, p_2 = \{3,8,9\}$, $p_3 = \emptyset$, $p_4 = \{6,7,9\}$, $p_5 = \{6,8,10\}, p_6 = \{3\}$; $p_7 = \{3\}$, $p_8 = \{10\}$, $p_9 = \emptyset$, and $p_{10} = \emptyset$. Each node represents a risk factor and the directed edge $i \rightarrow j$ indicates that i belongs to the Markovian parent set of i. The red nodes are the DAG roots.

2.1.3 DAG representation using a CPDAG

Before we proceed with our risk factor selection methodology, we must introduce some additional concepts regarding DAG structures. Two or more different DAGs can describe the same independence information for a joint distribution function[14]. Therefore, it is possible for two different DAGs, which describe the same conditional independence information from the joint risk factor distribution, to pose different root sets. In this case, any inferences on the $\boldsymbol{f}_S$ set explaining cross-sectional returns are insufficient since $\boldsymbol{f}_S$ can be obtained using a single DAG estimation. Moreover, every DAG describing the same conditional independence information for $P(\boldsymbol{f})$ comprises a set $\boldsymbol{\mathcal{C}}$, which is defined as a *Markov equivalence class*[15] for $P(\boldsymbol{f})$. As demonstrated by Andersson et al. (1997), set $\boldsymbol{\mathcal{C}}$ is uniquely represented by a *complete partial directed acyclic graph* ($\mathcal{G}_{CPDAG}$)[16], which states that:

$\mathcal{G}_{CPDAG}$ is said to *represent* a $\mathcal{G}_{DAG}$ if $\mathcal{G}_{DAG} \in \boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{C}}$ is described by $\mathcal{G}_{CPDAG}$ (5)

In setting $\boldsymbol{\Lambda} = \left\{\mathcal{G}_{DAG(i)} = \left(\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}}_{DAG(i)}\right)\right\}_{i=1}^{G}$ as the set that contains all DAGs comprising $\boldsymbol{\mathcal{C}}$, the $\mathcal{G}_{CPDAG} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}}_{CPDAG})$ is defined by: i) if $(i \rightarrow j) \in \boldsymbol{\mathcal{E}}_{CPDAG}$, then

[14] See Verma and Pearl (1991).

[15] $\mathcal{G}_{DAG(i)} \in \boldsymbol{\mathcal{C}}$ and $\mathcal{G}_{DAG(j)} \in \boldsymbol{\mathcal{C}}$ if and only if $\mathcal{G}_{DAG(i)}$ and $\mathcal{G}_{DAG(j)}$ presents the same *skeleton* and *v-structures*.

[16] See Andersson et al. (1997) for formal evidence.

$(i \rightarrow j) \in \boldsymbol{\mathcal{E}}_{DAG(i)}$ for at least one $\mathcal{G}_{DAG(i)}$; and ii) if $(i - j) \in \boldsymbol{\mathcal{E}}_{CPDAG}$[17], then $(i \rightarrow j) \in \boldsymbol{\mathcal{E}}_{DAG(i)}$ for at least one $\mathcal{G}_{DAG(i)}$ and $(i \leftarrow j) \in \boldsymbol{\mathcal{E}}_{DAG(j)}$ for at least one other $\mathcal{G}_{DAG(j)}$. These two properties mean that the $\mathcal{G}_{CPDAG}$ root presents the following property[18]:

$$P_{\mathcal{G}_{CPDAG}} \subseteq \bigcap_{i=1}^{G} P_{\mathcal{G}_{DAG(i)}} \tag{6}$$

where $P_{\mathcal{G}_{DAG(i)}}$ is the root set of $\mathcal{G}_{DAG(i)}$. According to (5), $\boldsymbol{f}$ can be expressed as $\mathcal{G}_{CPDAG} = (\boldsymbol{f}, \boldsymbol{\mathcal{E}}_{CPDAG})$ since $\mathcal{G}_{CPDAG}$ uniquely represents $\boldsymbol{\mathcal{C}}$ from $P(\boldsymbol{f})$, where $\boldsymbol{f}_{TRUE}$ is given by $P_{\mathcal{G}_{CPDAG}}$ following the same reasoning from the previous section. Relation (6) allows us to observe how our first shrinkage step is sparser when we select the $\mathcal{G}_{CPDAG}$ root instead of the $\mathcal{G}_{DAG}$ root since $P_{\mathcal{G}_{CPDAG}}$ is bounded by $\bigcap_{i=1}^{G} P_{\mathcal{G}_{DAG(i)}}$.

2.1.4 Estimation procedure for the high-dimensional CPDAG

The PC-algorithm proposed by Spirtes et al. (2000) has been widely applied to estimate a $\mathcal{G}_{CPDAG}$ representation for high-dimensional problems[19]. Furthermore, it is consistent in generating sparse graphs for the conditional dependence information of high-dimensional joint distribution functions, as evidenced by Harris and Drton (2013). This PC-algorithm property is desirable since we are interested in carrying out a sparse selection from the original risk factor dataset, with the $\mathcal{G}_{CPDAG}$ expected to have fewer roots as the number of estimated edges decreases[20]. In order to estimate $\mathcal{G}_{CPDAG}$, we apply a modified version of the PC-algorithm proposed by Colombo and Maathuis (2014) (PC-CM). As highlighted by these authors, the original PC-algorithm is order-dependent, meaning that the final estimation depends on the order upon which input variables are given. According to the authors, the order-dependent property can be very problematic in a high-dimensional environment. The PC-CM consists of three procedures: i) determination of the skeleton; ii) establishment of the v-structures; and iii) orientation of the remaining undirected edges.

---

[17] Unlike $\mathcal{G}_{DAG}$, $\mathcal{G}_{CPDAG}$ allows for undirected edges $(i - j)$, and according to Pearl (2009), the $(i - j)$ relation can be interpreted as a correlation effect between X_i and X_j, which cannot be attributed to any other variable. To learn more, see Verma and Pearl (1991).

[18] See Chickering (2002) for formal proof.

[19] For CPDAG-applied empirical papers on high-dimensional sets, see Kalisch et al. (2010); Nagarajan et al. (2010); Stekhoven et al. (2012); and Zhang et al. (2012).

[20] See Kalisch and Bühlmann (2007).

The first procedure starts from an undirected graph $\mathcal{G}_{U(0)} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}}_{(o)})$, where $\boldsymbol{\mathcal{E}}_{(o)}$ contemplates all possible edges across $\boldsymbol{\mathcal{V}}$. For the first step, for every $i, j \in \boldsymbol{\mathcal{V}}$, we test whether $X_i \perp X_j$, and if it is significant, we then remove $(i - j)$ from $\boldsymbol{\mathcal{E}}_{(o)}$; however, if it is not significant, we keep $(i - j)$ on $\boldsymbol{\mathcal{E}}_{(o)}$. The first step leads to a new $\mathcal{G}_{U(1)} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}}_{(1)})$. In the second step, for every $i, j \in \boldsymbol{\mathcal{V}}$ such that $(i - j) \in \boldsymbol{\mathcal{E}}_{(1)}$, we test $X_i \perp X_j | \boldsymbol{S}_1$ for different sets of $\boldsymbol{S}_1$, where $\#(\boldsymbol{S}_1) = 1$ and $\boldsymbol{S}_1 \subseteq a_i(\boldsymbol{\mathcal{E}}_{(1)}) \setminus \{j\}$, until we confirm whether $X_i \perp X_j | \boldsymbol{S}_1$ is significant, removing $(i - j)$ from $\boldsymbol{\mathcal{E}}_{(1)}$ and setting $\widehat{\boldsymbol{S}}_{i,j} = \boldsymbol{S}_1$, where $\boldsymbol{S}_{i,j}$ is the separation set of i and j; or, on the other hand, we exhaust all possible sets for S_1, maintaining $(i - j)$ from $\boldsymbol{\mathcal{E}}_{(1)}$ and setting $\widehat{\boldsymbol{S}}_{i,j} = \emptyset$. The second step yields a new $\mathcal{G}_{U(2)} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}}_{(2)})$, and we repeat the second step exactly as the first step, though with $\#(\boldsymbol{S}_2) = 2$ instead of 1. For every $i, j \in \boldsymbol{\mathcal{V}}$, we stop the algorithm whenever $\#(a_i(\boldsymbol{\mathcal{E}}_{(l-1)}) \setminus \{j\}) < (l - 1)$, where l is the number of steps.

To evaluate $X_i \perp X_j | \boldsymbol{S}$ empirically, we perform a partial correlation test by setting α as the significance level. The null-hypothesis can be stated as $H_0: \rho_{i,j|\boldsymbol{S}} = 0$, as opposed to the alternative hypothesis of $H_1: \rho_{i,j|\boldsymbol{S}} \neq 0$. We follow the methodology set forth by Kalisch and Bühlmann (2007) and apply Fisher's z-transformation:

$$Z(i, j|\boldsymbol{S}) = \frac{1}{2} log\left(\frac{1 + \hat{\rho}_{i,j|\boldsymbol{S}}}{1 - \hat{\rho}_{i,j|\boldsymbol{S}}}\right) \tag{7}$$

where $\hat{\rho}_{i,j|\boldsymbol{S}}$ is the partial correlation estimator calculated from a linear regression. Thus, we reject the null-hypothesis if:

$$|Z(i, j|\boldsymbol{S})|(N - \#(\boldsymbol{S}) - 3)^{\frac{1}{2}} > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{8}$$

where N is the number of observations and Φ is a cdf $N(0,1)$. It is important to point out that the role of α on (9) reaches well beyond the test's significance level. It is a shrinkage degree parameter that controls the sparsity level of $\widehat{\boldsymbol{\mathcal{E}}}_{CPDAG}$ and, consequently, the final CPDAG root set. As in Colombo and Maathuis (2014), we apply the Bayesian Information Criterion (BIC) to set the significance level starting from (8); thus, we select α, which minimizes:

$$-2L\left(\hat{\Sigma}_{\hat{\mathcal{G}}_{CPDAG}(\alpha)}, \hat{\mu}\right) + \left(\sum_{i>j} 1\left(\hat{\Sigma}_{\hat{\mathcal{G}}_{CPDAG}(\alpha)_{i,j}} \neq 0\right) + P\right) log(N) \tag{9}$$

where $L(\cdot)$ is a P-dimensional log-likelihood of a multinormal distribution; $1(\cdot)$ is an indicator function equal to 1 if $\hat{\Sigma}_{\hat{\mathcal{G}}_{CPDAG}(\alpha)_{i,j}} \neq 0$ and, otherwise, is zero; $\hat{\Sigma}_{\hat{\mathcal{G}}_{CPDAG}(\alpha)}$ is the covariance matrix estimator[21] based on $\hat{\mathcal{G}}_{CPDAG}(\alpha)$; and $\hat{\mu}$ is the $\boldsymbol{\mathcal{V}}$ vector mean**.** From the first part, we obtain an estimated skeleton defined to be $\hat{\mathcal{G}}_{Skel} = \left(\boldsymbol{\mathcal{V}}, \widehat{\boldsymbol{\mathcal{E}}}_{Skel}\right)$ where $\widehat{\boldsymbol{\mathcal{E}}}_{Skel} = \boldsymbol{\mathcal{E}}_{(l-1)}$ and the separation set $\widehat{\boldsymbol{S}}_{i,j}$ for every $i, j \in \boldsymbol{\mathcal{V}}$.

The second procedure, in turn, consists of assigning directions to the skeleton's edges to create the graph's v-structures. It is worth noting that for every $k, i, j \in \boldsymbol{\mathcal{V}}$[22]:

$$(i - j) \notin \widehat{\boldsymbol{\mathcal{E}}}_{Skel}, k \in a_i \text{ and } k \in a_j \Rightarrow \exists (i - k - j) \tag{10}$$

Consequently, when we look at (11) if $k \notin \widehat{\boldsymbol{S}}_{i,j}$, we orient $(i - k - j)$ as $(i \rightarrow k \leftarrow j)$. This orientation provides us with a new set of edges defined as $\widehat{\boldsymbol{\mathcal{E}}}_{PDAG}$**.**

Finally, we undertake the third and final procedure (orientation of the remaining undirected edges) pursuant to three rules:

1. Orient $(i - j)$ as $(i \rightarrow j)$ if $(k \rightarrow i) \in \widehat{\boldsymbol{\mathcal{E}}}_{PDAG}$**,** and $j \notin a_k$; otherwise, a new v-structure is created.
2. Orient $(i - j)$ as $(i \rightarrow j)$ if $(i \rightarrow k) \in \widehat{\boldsymbol{\mathcal{E}}}_{PDAG}$**,** and $(k \rightarrow j) \in \widehat{\boldsymbol{\mathcal{E}}}_{PDAG}$; otherwise, a directed cycle is created.
3. Orient $(i - j)$ as $(i \rightarrow j)$ if $(i - k) \in \widehat{\boldsymbol{\mathcal{E}}}_{PDAG}$**,** $(k \rightarrow j) \in \widehat{\boldsymbol{\mathcal{E}}}_{PDAG}$, $(i - l) \in \widehat{\boldsymbol{\mathcal{E}}}_{PDAG}$ and $(l \rightarrow j) \in \widehat{\boldsymbol{\mathcal{E}}}_{PDAG}$; otherwise, either a new v-structure or a directed cycle is created.

The end goal of the PC-CM algorithm is a $\hat{\mathcal{G}}_{CPDAG} = \left(\boldsymbol{\mathcal{V}}, \widehat{\boldsymbol{\mathcal{E}}}_{CPDAG}\right)$, where we obtain $\widehat{\boldsymbol{\mathcal{E}}}_{CPDAG}$ after applying the aforementioned rules to $\widehat{\boldsymbol{\mathcal{E}}}_{PDAG}$. Given $\hat{\mathcal{G}}_{CPDAG}$, the $\hat{P}_{\mathcal{G}_{CPDAG}}$ set is formally defined as $\hat{P}_{\mathcal{G}_{CPDAG}} = \left\{i : p_{CPDAG,i} = \emptyset\right\}$. Finally, we set $\boldsymbol{f}_{TRUE} = \hat{P}_{\mathcal{G}_{CPDAG}}$.

2.2 The Fama-MacBeth procedure

We apply the FM procedure to estimate the APT model composed of (2) and (3). We focus our analyses on $\boldsymbol{\lambda}$ factor risk premia parameters, with the aim of evaluating whether our methodology is effective in selecting risk factors fully capable of explaining

[21] See Norton and Richardson (2002) for detailed information on $\hat{\Sigma}_{\hat{\mathcal{G}}_{CPDAG}(\alpha)}$ estimators.

[22] By definition, $k \in a_i \Rightarrow (i - k) \in \widehat{\boldsymbol{\mathcal{E}}}_{Skel}$**.**

the cross-section of expected returns $\boldsymbol{E}(r_{i,t})$. Since our empirical findings for the $\hat{P}_{\mathcal{G}_{CPDAG}}$ set pose a low dimension, the APT model can be estimated with very well-known econometric methods[23], such as, for instance, the *generalized method of moments* (GMM) and the *iterated nonlinear seemingly unrelated regression* (ITNLSURE). However, we apply the FM procedure since our asset dataset for excess returns is an unbalanced panel. Our econometric model can be defined as:

$$r_{i,t} = \sum_{p=1}^{P} \beta_{i,p} f_{p,t} + \varepsilon_{i,t} \tag{11}$$

$$r_{i,t} = \lambda_0 + \sum_{p=1}^{P} \hat{\beta}_{i,p,t} \lambda_p + e_t \tag{12}$$

where (12) is identical to equation (2), which is estimated using a time-rolling window procedure whose length[24] equals t^*. Thus, this first-pass time series regression results in a sequence of estimated betas like $\{\widehat{\boldsymbol{\beta}}_{i,t}\}_{t=t^*}^{T}$. These betas are inputs for the second-pass cross-sectional regression described by (12), which leads to a sequence of estimated factor risk premia such as $\{\hat{\boldsymbol{\lambda}}_t\}_{t=t^*}^{T}$. The final factor risk premia estimator can be expressed as:

$$\hat{\boldsymbol{\lambda}} = \frac{\sum_{t=t^*}^{T} \hat{\boldsymbol{\lambda}}_t}{(T - t^*)} \tag{13}$$

We follow Shanken (1992) to compute the $\hat{\boldsymbol{\lambda}}$ covariance matrix with

$$\widehat{\boldsymbol{\Sigma}}_{\lambda shanken} = \left(1 + \hat{\boldsymbol{\lambda}}^T \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}^{-1} \hat{\boldsymbol{\lambda}}\right) \left(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\lambda}} - \frac{\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}}{(T - t^*)}\right) + \frac{\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}}{(T - t^*)} \tag{14}$$

where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}$ is the estimated risk factor covariance matrix[25], while $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\lambda}}$ is the regular estimated factor risk premia covariance matrix[26].

---

[23] See Campbell et al. (1997), Cochrane (2009), and Goyal (2012) for methods used to estimate low-dimensional factor risk premia models.

[24] We set $t^* = 60$ in our research paper, using the same value employed by Fama-MacBeth (1973).

[25] $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}} = \frac{\sum_{t=t^*}^{T} (f_t - \bar{f})(f_t - \bar{f})^T}{(T - t^*)^2}$.

[26] $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\lambda}} = \frac{\sum_{t=t^*}^{T} (\hat{\lambda}_t - \hat{\lambda})(\hat{\lambda}_t - \hat{\lambda})^T}{(T - t^*)^2}$.

2.3 Standard models and alternative methodologies

We select four well-known "standard sparse models" from the asset pricing literature to compare the FM results to our own findings. We choose to employ the Fama-French three-factor model (FF3)[27]; the Novy-Marx four-factor model (NM4)[28]; the Carhart four-factor model (C4)[29]; and a fourth model (P5) contemplating every risk factor from FF3 NM4 and C4. Since the first step of our methodology relies on a large set of potential risk factor candidates, we feel it would not be fair to make a direct comparison with standard models. Thus, aimed at seeking a fair comparison for purposes of evaluating our research, we select a number of other approaches documented by the factor zoo literature.

Kozak, Nagel, and Santosh's (2020) paper on the factor zoo suggests that a small group of principal components[30] (*PC*s) of risk factors are able to achieve highly satisfactory results in explaining cross-sectional returns. Bearing this in mind, we also proceed to test an FM procedure using two alternative risk factor sets. The first set relates to the *PC*s from the original high-dimensional risk factor set. As previously mentioned, the principal component analysis poses a disadvantage since it lacks a directed economic interpretation. The second set concerns the risk factors that yield the highest factor loadings on the first principal component of the original risk factor distribution ($PC1$)[31]**.** Since the first principal component is a latent factor, which individually explains the largest proportion of the risk factor covariance matrix, each of these selected risk factors is a preeminent candidate to explain the cross-section of expected returns.

2.4 Out-of-sample evaluation

Finally, we compute the one-step-ahead forecast for each model, with the aim of evaluating and comparing out-of-sample (OOS) results among them. Since we are able to calculate the estimator $\{\widehat{\boldsymbol{\beta}}_{i,t}\}_{t=t^*}^{T}$ as well as $\{\hat{\boldsymbol{\lambda}}_t\}_{t=t^*}^{T}$ for each methodology we propose

[27] Risk factors are Mkt, SMB, and HML. See Fama and French (1993).

[28] Risk factors are Mkt, SMB, HML, and GP. See Novy-Marx (2013).

[29] Risk factors are Mkt, SMB, HML, and MOM. See Carhart (1997). In this paper, we apply the six-month MOM risk factor (see Jegadeesh and Titman (1993).

[30] See Kelly, Pruitt and Su (2019); and Gu, Kelly, and Xiu (2019) for principal component analysis applications regarding the factor zoo.

[31] Factor loading is the correlation coefficient between the principal component and the original random variable.

herein, and our interest lies in explaining the cross-section of expected returns, we estimate the one-step-ahead forecast as follows:

$$\hat{r}_{i,t+1} = \hat{\lambda}_{0,t} + \widehat{\boldsymbol{\beta}}_{i,t}\hat{\boldsymbol{\lambda}}_{f,t} \tag{15}$$

It is worth noting that the forecast computed with (15) is entirely out-of-sample. For each cross-section, we calculate the root-mean-square-error ($RMSE_t$)[32] for $t = t^*, \dots, T-1$, and then compare root-mean-square-error averages ($AV.RMSE$)[33] and root-mean-square-error medians ($M.RMSE$)[34] across all models. It also worth mentioning that inputs for (15) are obtained using the model described by (11) and (12), which is focused on explaining – and not forecasting –, the cross-section of expected returns. Thus, our interest lies only on the comparative performance among OOS models.

3. Dataset

We apply our research to the factor zoo dataset compiled by Kozak, Nagel, and Santosh (2020), with monthly data ranging from January 1981 to December 2016[35]. It consists of 51 risk factors, the first being the *Excess Market Return* gathered from the French Library[36], while the remaining 50 are zero-investment, long-short portfolios composed of well-known traits described in the asset pricing literature. Table A1 of the Appendix summarizes risk factor descriptions and statistics.

In regards to cross-sectional returns, although portfolios do not produce missing data by construction (balance panel), they do have a tendency of showing a bias towards traits used to build them, as highlighted by Harvey and Liu (2019). Consequently, and due to the fact that the FM procedure supports a large unbalanced panel, we choose to focus on individual assets from the CRSP stock return dataset. To compose excess asset returns, we set one-month maturity USD LIBOR interest rates as risk-free. Since we adopt a 60-month time window for the first-pass of our FM procedure, we accordingly disregard assets with less than 60 observations. Additionally, we remove stocks from the financial

[32] $RMSE_t = \sqrt{\frac{\sum_{i=1}^{I}(\hat{r}_{i,t} - r_{i,t})^2}{I}}$, where I is the number of assets presents on the cross-section in the period t.

[33] $AV.RMSE = \frac{\sum_{t=t^*}^{T-1} RMSE_t}{T - t^* - 1}$.

[34] $M.RMSE = median(RMSE_t), t = t^*, \dots, T-1$.

[35] Data can be downloaded from: https://www.serhiykozak.com/data.

[36] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

sector. Using these procedures, our dataset accounts for 14,317 individual stocks (CSRP). Considering that this first dataset also comprises small-caps, which may yield significantly illiquid stocks, we create a second dataset which exclude stocks whose prices are lower than USD 5.00, thereby leading to a dataset encompassing 10,221 individual assets (CRSP without small-caps).

4. Results

4.1 CPDAG roots and selected risk factors

Figure 2 shows the graphic representation of $P(\boldsymbol{f})$ as $\hat{\mathcal{G}}_{CPDAG} = \left(\boldsymbol{f}, \hat{\boldsymbol{\mathcal{E}}}_{CPDAG}\right)$. In this illustration, each node represents a risk factor; blue lines denote direct edges; and red lines denote undirected edges. Since $\hat{\mathcal{G}}_{CPDAG}$ describes the factorization of $P(\boldsymbol{f})$ with its Markovian parents, Figure 2 allows us to infer the conditional dependence relationship among all risk factors. A blue directed edge $(i \rightarrow j)$ indicates that $i \in p_j$, meaning that $f_i \in m_j$. In other words, if Figure 2 presents $(i \rightarrow j)$, this entails that f_i belongs to the Markovian parent set of f_j given for every other risk factor and, consequently, f_j is conditionally dependent on f_i. A red undirected edge $(i - j)$ signals that $f_i \in m_j$, and $f_j \in m_i$, meaning that we are unable to infer a direct conditional relationship between f_i and f_j since both risk factors belong to each other's Markovian parent sets. Still concerning this illustration, if the i node is red, it therefore indicates that $i \in \hat{P}_{\mathcal{G}_{CPDAG}}$, meaning that $m_i = \emptyset$. In other words, the Markovian parent set of f_i is empty, signaling that f_i does not present any dependency on any other risk factor, thereby making it one of our risk factor candidates to explain cross-sectional returns. Thus, as we mentioned before, our final risk factor selection is given by $\boldsymbol{f}_{TRUE} = \hat{P}_{\mathcal{G}_{CPDAG}}$.

Results attained with $\hat{\mathcal{G}}_{CPDAG}$ lead to some interesting features for $P(\boldsymbol{f})$. First, our final risk factor set poses only six elements from an original set of 51, and succeeds in performing the significant shrinkage we are pursuing herein. This selected set, as described in Table 1, consists of the following risk factors: *Accruals*, *Short Interest*, *Value Profitability*, *Gross Margins*, *Debt Issuance,* and *Return on Equity*. Figure 2 shows us that none of the risk factors are conditional dependents on either *Accruals* or *Short Interest,* which might lead us to conclude that these two risk factors pose a very unique type of information. Additionally, according to $\hat{\mathcal{G}}_{CPDAG}$: i), *Return on Equity* partially

spans *Return on Book Equity* and *Return on Assets*; ii) *Value Profitability* partially spans *Sales to Price*, *Asset Turnover* and *Value Momentum Profitability*; iii) *Gross Margins* partially span *Sales to Price;* and iv) *Debt Issuance* partially spans *Firm Age* and *Share Repurchases*. Besides the $\hat{\mathcal{G}}_{CPDAG}$ root set, all the remaining risk factors pose, on average, 1.6 conditional dependence edges and span an additional 1.4 risk factors, given all other factors.

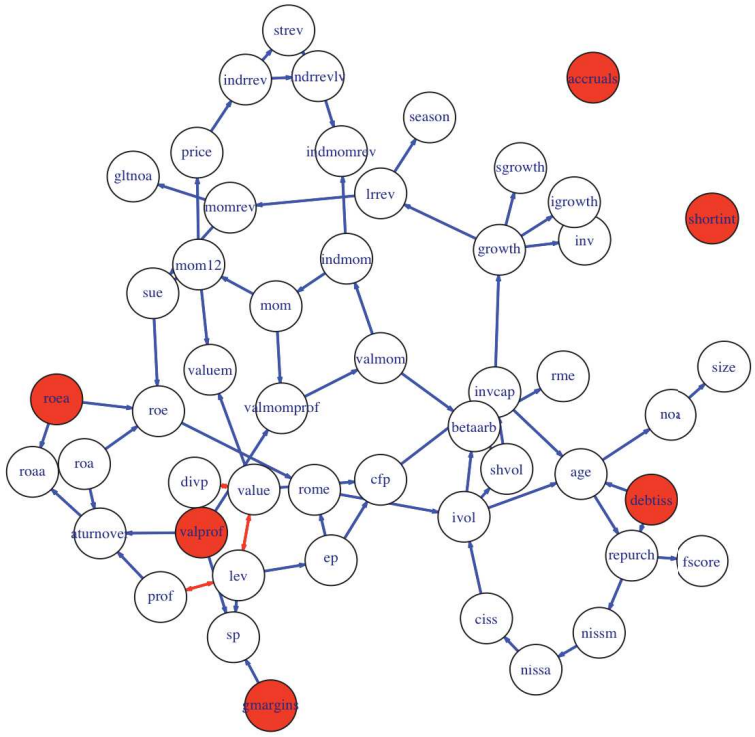
The optimal regularization parameter (α) is 0.001 from a grid of $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$. These five values for hyperparameter α are defined in accordance with the fact that α is also the significance level for the hypothesis test described in (9). Figure A1 of the Appendix displays the $\hat{\mathcal{G}}_{CPDAG}$ object for the entire α grid, and as we can see – with the exception of $\alpha = 0.0001$ –, the theoretical results described by Kalisch and Bühlmann (2007) are verified since the number of roots decreases insofar as the value of α increases. Table A2 of the Appendix describes the selected risk factor pursuant to each value of α, as well as its respective BIC values. As we can see, $\alpha = 0.001$ reaches the lowest BIC value, which therefore makes it our selected model. It is interesting to note that certain risk factors, such as *Accruals*, *Short Interest*, *Value-Profitability* and *Gross Profitability*, are selected using more than one model, suggesting strong evidence in favor of their conditional independence in regards to other risk factors.

Table 1: Risk factor selected using the $\widehat{\mathcal{G}}_{CPDAG}$ *root*

Description	Ret.	S.R.	Reference	Code
Accruals	-0.029	-0.188	Sloan (1996)	accruals
Short Interest	-0.006	-0.048	Dechow, Kothari, and Watts (1998)	shortint
Value-Profitability	0.131	0.845	Novy-Marx (2013)	valprof
Gross Margins	-0.011	-0.068	Novy-Marx (2013)	gmargins
Debt Issuance	0.016	0.098	Spiess and Affleck-Graves (1999)	debtiss
Return on Equity (annual)	0.038	0.235	Haugen and Baker (1996)	roea

Note: The table above displays the selected risk factors based on the estimated $\widehat{\mathcal{G}}_{CPDAG}$ root set ($\hat{P}_{\mathcal{G}_{CPDAG}}$) for the optimal alpha, according to the BIC criteria ($\alpha = 0.001$). For each selected risk factor, the table includes annualized average excess returns, annualized Sharpe ratios, literature references, and code names.

Figure 2: Graphic representation of joint risk factor distribution with $\widehat{\mathcal{G}}_{CPDAG}$



Note: The figure above displays the graphic representation of the joint risk factor distribution using $\widehat{\mathcal{G}}_{CPDAG}$. As described in section 2.1.4, we estimate the joint risk factor distribution with $\widehat{\mathcal{G}}_{CPDAG} = \mathcal{G}_{CPDAG}\left(\boldsymbol{\nu} = \boldsymbol{f}, \widehat{\boldsymbol{\varepsilon}}_{CPDAG}(\alpha)\right)$. To select the tuning parameter alpha, we set an alpha grid as $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$, and then proceed to evaluate the BIC criteria (described by (10)). We find an optimal regularization parameter $\alpha = 0.001$. In this illustration, each node represents a risk factor; blue lines denote direct edges; and red lines denote undirected edges. Nodes in red are graph roots ($\hat{P}_{\mathcal{G}_{CPDAG}}$).

4.2 FM Results

After estimating $\hat{\mathcal{G}}_{CPDAG}$ and selecting the risk factor with $\hat{P}_{\mathcal{G}_{CPDAG}}$, we are able to verify whether this methodology is in fact capable of explaining the cross-section of expected returns. Table 2 displays the estimated factor risk premia attained with the FM approach used for our selected risk factors. When we examine these results, we verify that the intercept is significant for CRSP samples both with and without small-caps, therefore signaling that certain idiosyncratic risks remain in cross-sectional returns even after they are controlled by our risk factors. Nevertheless, both regressions pose the highest average adjusted R^2 comparing to all others alternatives models. (around 15.8% and 17.8% for CRSP samples both with and without small-caps, respectively), as well as significant factor risk premia for *Short Interest* and *Debt Issuance*. Following Dechow, Kothari, and Watts (1998), *Short Interest* is the ratio between sorted and outstanding

shares. As described by Spiess and Affleck-Graves (1999), *Debt Issuance* is an indicator function equal to 1 if the firm's cash flow statement issues long-term debt or, otherwise, zero. The high average adjusted R^2 and the significant *Short Interest* and *Debt Issuance* factor risk premia for CRSP samples both with and without small-caps favor our selection methodology. Table A3 of the Appendix displays the FM procedure's results for models with risk factors selected using alternative alphas. Furthermore, it shows us that our optimal value for alpha provides the best average adjusted R^2 for CRSP samples both with and without small-caps. This result favors the BIC criteria alpha selection methodology.

Table 2: FM results for risk factor model selected with the $\widehat{\mathcal{G}}_{CPDAG}$ *root*

Coefficient		**CRSP**	**CRSP without Small-Caps**
Intercept	Estimate	0.0033**	0.0046**
	S. Error	0.0014	0.0011
	p-value	0.0157	0.0000
Accruals	Estimate	0.0017	0.0030
	S. Error	0.0026	0.0024
	p-value	0.5176	0.2210
Short Interest	Estimate	0.0034***	0.0070***
	S. Error	0.0012	0.0021
	p-value	0.0076	0.0011
Value-Profitability	Estimate	0.0009	0.0009
	S. Error	0.0025	0.0023
	p-value	0.7269	0.6791
Gross Margins	Estimate	0.0003	0.0023
	S. Error	0.0027	0.0026
	p-value	0.9060	0.3813
Debt Issuance	Estimate	0.0039*	0.0048*
	S. Error	0.0022	0.0027
	p-value	0.0875	0.0838
Return on Equity	Estimate	(0.0019)	(0.0027)
(annual)	S. Error	0.0028	0.0026
	p-value	0.5010	0.3010
Av. adjusted R^2		0.158	0.178

Note: The table displays the Fama-MacBeth (FM) results for models whose risk factors were selected based on the estimated $\widehat{\mathcal{G}}_{CPDAG}$ root set ($\hat{P}_{\mathcal{G}_{CPDAG}}$). The full FM procedure is described in Section 2.2. To select the tuning parameter alpha, we set an alpha grid as $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$, and then evaluate the BIC criteria (described by AS (10). We find an optimal regularization parameter $\alpha = 0.001$. Results are reported for CRSP datasets both with and without small-caps. The average number of securities in each cross-sectional regression is 4,041 and 2,885, respectively, for CRSP datasets both with and without small-caps. For each model and dataset, the table includes the estimated risk premia, standard error and the p-value for null hypothesis $H_0: \lambda = 0$, as opposed to the alternative hypothesis $H_1: \lambda \neq 0$. Subscriptions $*,**$ and $***$ entail that the null hypothesis is rejected at 10%, 5% and 1% levels of significance, respectively.

Table 3 brings the standard models. All of these pose lower average adjusted R^2 when compared to our methodology. We notice that the intercept is also significant across all models and samples. When we consider the full CRSP sample, we find that standard models do not show any significant estimated factor risk premia. In turn, when we remove small-caps from the CRSP dataset, the risk premium for the *Excess Market Return* becomes significant for FF3 and NM4 models, although it remains insignificant in the C4 model. *Gross Profitability* is significant in the NH4 model, whereas *Momentum* is insignificant in the C4 model.

Table 3: FM Results for standard models

Base		**CRSP**				**CRSP without Small-Caps**			
Coefficient		**FF3**	**NM4**	**C4**	**P5**	**FF3**	**NM4**	**C4**	**P5**
Intercept	Estimate	0.0032**	0.0035***	0.0029**	0.0031**	0.0039***	0.0038***	0.0037**	0.0036**
	S. Error	0.0013	0.0013	0.0012	0.0012	0.0009	0.0009	0.0008	0.0008
	p-value	0.0182	0.0070	0.0197	0.0106	0.0000	0.0000	0.0000	0.0000
Excess Market Return	Estimate	0.0026	0.0025	0.0028	0.0026	0.0046*	0.0046*	0.0045	0.0044
	S. Error	0.0027	0.0027	0.0027	0.0026	0.0025	0.0025	0.0025	0.0025
	p-value	0.3372	0.3610	0.2963	0.3213	0.0715	0.0697	0.0690	0.0717
Size	Estimate	(0.0025)	(0.0029)	(0.0031)	(0.003)	-0.0059**	-0.0065**	-0.006**	(0.006)
	S. Error	0.0028	0.0028	0.0027	0.003	0.0026	0.0026	0.0026	0.003
	p-value	0.3823	0.3002	0.2571	0.217	0.0253	0.0124	0.0199	0.012
Value	Estimate	(0.0024)	(0.0025)	(0.0021)	(0.0024)	(0.0037)	(0.0037)	(0.0038)	(0.0039)
	S. Error	0.0028	0.0028	0.0027	0.0027	0.0027	0.0027	0.0026	0.0026
	p-value	0.3986	0.3700	0.4321	0.3886	0.1640	0.1609	0.1475	0.1441
Gross Profitability	Estimate		0.0034		0.0035		0.0047*		0.0047
	S. Error		0.0028		0.0027		0.0026		0.0026
	p-value		0.2205		0.1981		0.0707		0.0703
Momentum (6m)	Estimate			(0.0002)	0.0000			0.0003	0.0005
	S. Error			0.0030	0.0029			0.0028	0.0028
	p-value			0.9343	0.9987			0.9101	0.8476
Av. adjusted R^2		0.048	0.060	0.060	0.071	0.058	0.071	0.066	0.082

Note: The table above displays the Fama-MacBeth (FM) results for standard models. Section 2.2 describes the full FM procedure. Columns FF3, NM4, C4 and P5 refer to standard models described in Section 2.3. Results are reported for CRSP datasets both with and without small-caps. The average number of securities in each cross-sectional regression is 4,041 and 2,885, respectively, for CRSP datasets both with and without small-caps. For each model and dataset, the table includes the estimated risk premia, standard error and the p-value for null hypothesis $H_0: \lambda = 0$, as opposed to the alternative hypothesis $H_1: \lambda \neq 0$. Subscriptions $*, **$ and $***$ entail that the null hypothesis is rejected at 10%, 5% and 1% levels of significance, respectively.

For our next step, we compute the FM procedure for the first four PCs of $\boldsymbol{f}$, whose results are described in Table 4. The first four PCs account for approximately 80% of the total cumulative variance of $\boldsymbol{f}$, as we can see in Figure A2 of the Appendix, resulting in promising risk factor candidates to explain cross-sectional returns. Again, the intercept is significant across all models and samples. Considering the average adjusted R^2, our

methodology still poses a better performance when compared to PC models. Table 4 below evidences how none of the PCs are significant for full CRSP samples. In regards to the dataset sample without small-caps, only the first principal component is significant across all models, while the fourth principal component is significant in the PC4 model.

Table 4: FM Results for PC models

Base		**CRSP**				**CRSP without Small-Caps**			
Coefficient		**PC1**	**PC2**	**PC3**	**PC4**	**PC1**	**PC2**	**PC3**	**PC4**
Intercept	Estimate	0.0042**	0.0035**	0.0021	0.0024*	0.0057***	0.0046***	0.0035***	0.0035***
	S. Error	0.0019	0.0016	0.0015	0.0014	0.0017	0.0013	0.0013	0.0011
	p-value	0.0286	0.0296	0.1735	0.0912	0.0007	0.0007	0.0053	0.0011
PC1	Estimate	0.0191	0.0167	0.0148	0.0148	0.0263**	0.0263**	0.0243**	0.0237**
	S. Error	0.0129	0.0129	0.0126	0.0125	0.0119	0.0122	0.0121	0.0120
	p-value	0.1397	0.1948	0.2410	0.2372	0.0278	0.0319	0.0455	0.0499
PC2	Estimate		(0.0047)	(0.0018)	(0.000)		(0.0056)	(0.0027)	0.000
	S. Error		0.0080	0.0076	0.007		0.0073	0.0072	0.007
	p-value		0.5582	0.8148	0.986		0.4390	0.7055	0.962
PC3	Estimate			0.0045	0.0035			0.0050	0.0044
	S. Error			0.0066	0.0065			0.0058	0.0057
	p-value			0.4985	0.5916			0.3866	0.4441
PC4	Estimate				(0.0063)				-0.0098**
	S. Error				0.0050				0.0044
	p-value				0.2081				0.0273
Av. adjusted R^2		0.042	0.074	0.100	0.124	0.055	0.088	0.116	0.143

Note: The table above displays the Fama-MacBeth (FM) results for PC models. Section 2.2 describes the full FM procedure. Columns PC1, PC2, PC3 and PC4 refer to principal component models described in Section 2.3. Results are reported for CRSP datasets both with and without small-caps. The average number of securities in each cross-sectional regression is 4,041 and 2,885, respectively, for CRSP datasets both with and without small-caps. For each model and dataset, the table includes the estimated risk premia, the standard error and the p-value for null hypothesis $H_0: \lambda = 0$, as opposed to the alternative hypothesis $H_1: \lambda \neq 0$. Subscriptions $*,**$ and $***$ indicate that the null hypothesis is rejected at 10%, 5% and 1% levels of significance, respectively.

In regards to the PC1 (first principal component) risk premium significance in the CRSP sample without small-caps, we also test models with the four-risk factors that yield the highest factor loadings for the first principal component for $\boldsymbol{f}$, leading to models that we dub "PC1 loadings"**.** Figure A3 of the Appendix shows factor loadings for the first principal component of $\boldsymbol{f}$**.** It is worth pointing out that none of the four factors with the highest loadings (*Firm Age*, *Investment to Capital*, and *Share Issuance* [*Monthly* and *Annual*]) belong to $\hat{P}_{\mathcal{G}_{CPDAG}}$. Table 5 summarizes estimated results for PC1 loadings models, allowing us to conclude that the regression from Model 3 in the full CRSP dataset is the only situation in which an insignificant intercept is achieved, in spite of its average adjusted R^2 (near 8.9%) being considerably lower than the 15.8% percentage verified in our methodology for the same samples. As we can see, unlike previous alternative models

- and with the only exception being Model 1 -, PC1 loadings models yield significant factor risk premia estimators for CRSP samples both with and without small-caps. Nevertheless, the average adjusted R^2 from all PC1 models is lower than the one submitted by our methodology.

Table 5: FM Results for PC1 loadings models

Base		**CRSP**				**CRSP without Small-Caps**			
Coefficient		**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Intercept	Estimate	0.0032*	0.0031*	0.0027	0.0033**	0.0049***	0.005***	0.0047***	0.0052***
	S. Error	0.0018	0.0017	0.0017	0.0016	0.0015	0.0014	0.0014	0.0014
	p-value	0.0771	0.0728	0.1086	0.0454	0.0014	0.0007	0.0011	0.0002
Firm Age	Estimate	(0.0053)	(0.0051)*	(0.0046)	(0.0047)	(0.006)***	(0.0071)**	(0.0068)**	(0.0071)**
	S. Error	0.0032	0.0031	0.0030	0.0030	0.0029	0.0028	0.0027	0.0028
	p-value	0.1011	0.0942	0.1272	0.1155	0.0187	0.0119	0.0138	0.0108
Investment-to-Capital	Estimate		0.007**	0.0065**	0.0064**		0.0081***	0.0077***	0.008***
	S. Error		0.0031	0.0031	0.003		0.0029	0.0028	0.003
	p-value		0.0258	0.0350	0.037		0.0048	0.0057	0.005
Share Issuance (monthly)	Estimate			0.0051*	0.0051*			0.0057**	0.006**
	S. Error			0.0030	0.0031			0.0028	0.0028
	p-value			0.0967	0.0965			0.0437	0.0347
Share Issuance (annual)	Estimate				0.0050				0.0058**
	S. Error				0.0031				0.0029
	p-value				0.1130				0.0459
Av. adjusted R^2		0.043	0.065	0.089	0.107	0.057	0.081	0.106	0.125

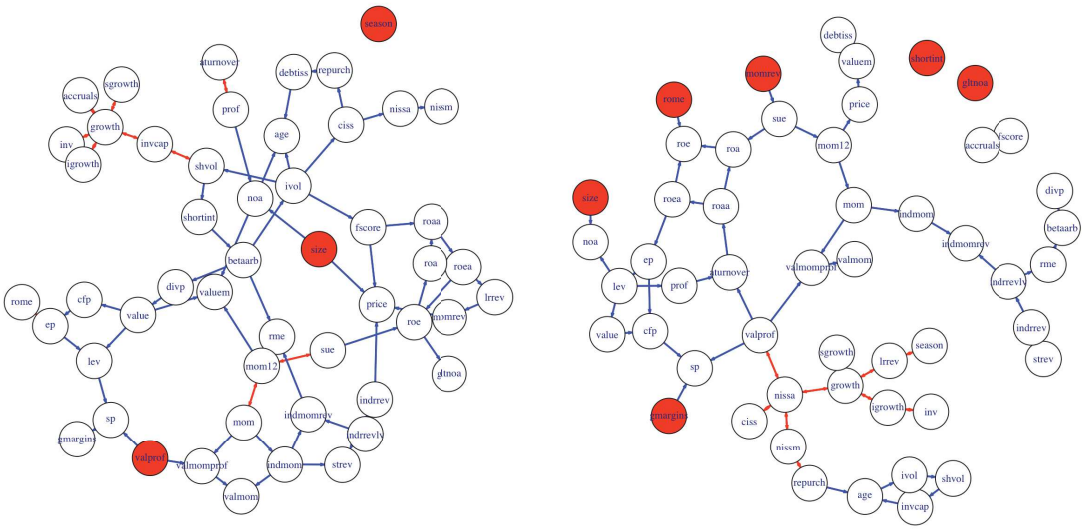
Note: The table above displays the Fama-MacBeth (FM) results for the PC1 loadings models. Section 2.2 describes the full FM procedure. Columns Model1, Model2, Model3 and Model4 refer to PC1 loadings models described in Section 2.3. Results are reported for CRSP datasets both with and without small-caps. The average number of securities in each cross-sectional regression is 4,041 and 2,885, respectively, for CRSP datasets both with and without small-caps. For each model and dataset, the table includes the estimated risk premia, the standard error and the p-value for null hypothesis $H_0: \lambda = 0$, *as opposed to the alternative hypothesis* $H_1: \lambda \neq 0$. *Subscriptions* $*, **$ *and* $***$ *indicate that the null hypothesis is rejected at 10%, 5% and 1% levels of significance, respectively.*

Generally speaking, we find that our methodology is the only approach that poses significant estimated factor risk premia parameters for the full CRSP sample, with the exception of PC1 Model 1. After comparing our methodology to all other models, we verify that our results generate a higher average adjusted R^2 in both samples. Thus, FM results support our model when we take into consideration each sparse model tested so far. In regards to in-sample average adjusted R^2 numbers, our final result amounts, on one hand, to approximately 15.8% for the full CRSP sample; and, on the other, to 17.8% for the CRSP sample with small-caps, a satisfactory outcome according to Campbell and Thompson (2008).

4.3 Time-Varying Robustness

Variations on the explanatory power of individual factors in the cross-section of expected returns over time are something commonly described in research papers addressing the factor zoo[37]. Taking such a fact into consideration, and seeking to examine this well-known issue, we proceed to split our samples in two so that the first part ranges from January 1981 to December 1998, while the second encompasses the time frame from January 1999 to December 2016. For these samples we apply the same methodology described in Section 2.1. Figure 3 displays the graphic representation of $P(\boldsymbol{f})$ as $\hat{\mathcal{G}}_{CPDAG}$ for the first and second half samples.

Figure 3: Graphic representation of joint risk factor distribution using $\widehat{\mathcal{G}}_{CPDAG}$ for different sample periods



Note: The figure above displays the graphic representation of joint risk factor distribution using $\widehat{\mathcal{G}}_{CPDAG}$. As described in section 2.1.4, we estimate the joint risk factor distribution with $\widehat{\mathcal{G}}_{CPDAG} = \mathcal{G}_{CPDAG}\left(\boldsymbol{\nu} = \boldsymbol{f}, \widehat{\boldsymbol{\varepsilon}}_{CPDAG}(\alpha)\right)$. To select the tuning parameter alpha, we set an alpha grid as $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$, and then evaluate the BIC criteria (described by (10)). Results are presented for the first and second half sample periods. The first half ranges from January 1981 to December 1998, while the second half ranges from January 1999 to December 2016. We find an optimal regularization parameter $\alpha = 0.05$ for the first half, and $\alpha = 0.01$ for the second half. In each figure, each node represents a risk factor; blue lines denote direct edges; and red lines denote undirected edges. The nodes in red refer to graph roots ($\hat{P}_{\mathcal{G}_{CPDAG}}$).

[37] See Freyberger, Neuhierl, and Weber (2020); and Green, Hand, and Zhang (2017).

For the first half, the optimal alpha is 0.05, leading to a less sparse graph and, consequently, to fewer roots when compared to the full-size sample results. Although the selected alpha for the second half is 0.01, or 0.001 higher than the same full-size sample alpha, the number of selected risk factors for the second half samples is the same as in the full-size sample. Table 6 compares the risk factor selected using our methodology to these three different sample periods. This allows us to observe a very low intersectional among these selected risk factor sets. None of the risk factors are selected by samples at the same time. We only have two common selected risk factors that are common to both the second half and the full sample (*Short Interest* and *Gross Margins*), and one element common to both the first and second half samples (*Size*), as well as between the first half and the full-size sample (*Value Profitability*).

Table 6: Risk factor selected using the $\widehat{\mathcal{G}}_{CPDAG}$ *root* for different sample periods

Description	Ret.	S.R.	Sample Period All	First Half	Second Half
Accruals	-0.029	-0.188	X		
Short Interest	-0.006	-0.048	X		X
Value-Profitability	0.131	0.845	X	X	
Gross Margins	-0.011	-0.068	X		X
Debt Issuance	0.016	0.098	X		
Return on Equity (annual)	0.038	0.235	X		
Size	-0.027	-0.170		X	X
Seasonality	0.076	0.482		X	
Return on Market Equity	0.073	0.441			X
Momentum-Reversal	-0.074	-0.451			X
Growth in LTNOA	-0.017	-0.132			X
Selected Alpha			**0.001**	**0.05**	**0.01**

Note: The table above displays the selected risk factors based on the estimated $\widehat{\mathcal{G}}_{CPDAG}$ root set ($\hat{P}_{\mathcal{G}_{CPDAG}}$) and the alpha selected using BIC criteria. To select the tuning parameter alpha, we set an alpha grid as $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$, and then evaluate the BIC criteria (described by (10)). Results are shown for all three different sample periods. The first half ranges from January 1981 to December 1998, while the second half ranges from January 1999 to December 2016. For each selected risk factor, the table includes annualized average excess returns and annualized Sharpe ratios.

In regards to the FM procedure, and given that we keep using the 60-month time window for the first-pass, results from the second-pass regression apply to cross-sections estimated from January 1986 to December 1998 in the first half, and from January 2004 to December 2016 for the second half. Table 7 brings the FM procedure's results for the first part of the sample, where we can see that regressions for the CRSP dataset both with and without small-caps do not pose any significant risk factor premia. For the second part

of the sample (Table 8), only regressions for the CRSP dataset without small-caps provide significant factor risk premia (*Size* and *Return on Market Equity*).

Splitting the dataset in two enables us to verify the factor zoo evidence on the time-varying explanatory power of individual factors in the cross-section of expected returns. Nevertheless, results are considerably poorer when compared to those obtained using the full sampling period, thereby suggesting that 13-year monthly data time frames for FM second-pass regressions may be too short a sample to reach satisfactory results[38].

Table 7: FM results for risk factor models selected using the $\widehat{\mathcal{G}}_{CPDAG}$ *root* for the first half sample

Coefficient		**CRSP**	**CRSP without Small-Caps**
Intercept	Estimate	0.0022	0.0057**
	S. Error	0.0031	0.0023
	p-value	0.4751	0.0169
Value-Profitability	Estimate	0.0019	-0.0004
	S. Error	0.0030	0.0025
	p-value	0.5273	0.8817
Size	Estimate	-0.0011	-0.0023
	S. Error	0.0035	0.0030
	p-value	0.7535	0.4585
Seasonality	Estimate	0.0049	0.0053
	S. Error	0.0040	0.0038
	p-value	0.2232	0.1643
Av. adjusted R^2		0.0738	0.0850

Note: The table above displays the Fama-MacBeth (FM) results for models whose risk factors were selected based on the estimated $\widehat{\mathcal{G}}_{CPDAG}$ root set ($\hat{P}_{\mathcal{G}_{CPDAG}}$) for the first part of our sample. Section 2.2. describes the full FM procedure. To select the tuning parameter alpha, we set an alpha grid as $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$, and then evaluate the BIC criteria (described by AS (10)). We find an optimal regularization parameter $\alpha = 0.05$. Results are reported for CRSP datasets both with and without small-caps. The average number of securities in each cross-sectional regression is 3,825 and 2,527, respectively, for CRSP datasets both with and without small-caps. For each model and dataset, the table includes the estimated risk premia, the standard error and the p-value for null hypothesis $H_0: \lambda = 0$, as opposed to the alternative hypothesis $H_1: \lambda \neq 0$. Subscriptions $*,**$ and $***$ indicate that the null hypothesis is rejected at 10%, 5% and 1% levels of significance, respectively.

[38] See Cavalcante Filho et al. (2020) for interesting observations on sample sizes required to obtain robust risk premium estimators.

Table 8: FM results for risk factor models selected using the $\widehat{\mathcal{G}}_{CPDAG}$ *root* for the second half sample

Coefficient		CRSP	CRSP without Small-Caps
Intercept	Estimate	0.0033*	0.0032**
	S. Error	0.0017	0.0014
	p-value	0.0567	0.0252
Short Interest	Estimate	-0.0003	0.0045
	S. Error	0.0033	0.0034
	p-value	0.9173	0.1906
Gross Margins	Estimate	-0.0030	-0.0017
	S. Error	0.0030	0.0029
	p-value	0.3284	0.5462
Size	Estimate	-0.0008	-0.0046*
	S. Error	0.0028	0.0028
	p-value	0.7801	0.0995
Return on Market Equity	Estimate	-0.0015	-0.0033*
	S. Error	0.0020	0.0017
	p-value	0.4418	0.0501
Momentum-Reversal	Estimate	-0.0007	-0.0014
	S. Error	0.0030	0.0027
	p-value	0.8246	0.6060
Growth in LTNOA	Estimate	-0.0004	-0.0017
	S. Error	0.0025	0.0025
	p-value	0.8899	0.5081
Av. adjusted R^2		0.1412	0.1631

Note: The table above displays the Fama-MacBeth (FM) results for models whose risk factors were selected based on the estimated $\widehat{\mathcal{G}}_{CPDAG}$ root set ($\widehat{P}_{\mathcal{G}_{CPDAG}}$) for the second part of our sample. Section 2.2 describes the full FM procedure. To select the tuning parameter alpha, we set an alpha grid as $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$, and then evaluate the BIC criteria (described by AS (10)). We find an optimal regularization parameter $\alpha = 0.01$. Results are reported for CRSP datasets both with and without small-caps. The average number of securities in each cross-sectional regression is 4,128 and 3,103, respectively, for CRSP datasets both with and without small-caps. For each model and dataset, the table includes the estimated risk premia, the standard error and the p-value for null hypothesis $H_0: \lambda = 0$, as opposed to the alternative hypothesis $H_1: \lambda \neq 0$. Subscriptions $*, **$ and $***$ indicate that the null hypothesis is rejected at 10%, 5% and 1% levels of significance, respectively.

4.4 Out-of-Sample Results

It is widely known in the high-dimensional statistics literature that, as the model's complexity increases, the in-sample performance (IN) follows suit and permanently increases. On the other hand, out-of-sample performances (OOS) will start to decrease after the training model first stops fitting the data structure, then starts fitting the data noise[39]. We can therefore state that a desirable result is a model posing both IN and OOS

[39] See Abu-Mostafa, Magdon-Ismail, and Lin (2012) for more information.

satisfactory results. Thus, we examine out-of-sample (OOS) results from the computed one-step-ahead forecast, as described in section 2.4.

Table 9 displays statistics for the $AV.RMSE$ (root-mean-square-error averages) and $M.RMSE$ (root-mean-square-error medians) of each model tested in the previous sections. As already pointed out, our methodology poses the lowest statistics for the OOS results across all metrics and datasets. This outperform is statistically significant in both samples as we observe the Diebold-Mariano test results in Table A4 and A5 on the Appendix. Table A6 of the Appendix summarizes the OOS results for models with alternative values for the alpha shrinkage parameter with, as noted before, the model with the optimal alpha presenting the best results. Thus, the OOS analysis also favors our methodology when compared to every other approach.

Table 9: All out-of-sample model results

Base		**CRSP**		**CRSP without Small-Caps**	
Model/Metric		***AV.RMSE***	***M.RMSE***	***AV.RMSE***	***M.RMSE***
$\widehat{\mathcal{G}}_{CPDAG}$ *roots*		0.1715	0.1542	0.1258	0.1118
Standard	FF3	0.1763	0.1620	0.1298	0.1173
	NM4	0.1775	0.1627	0.1306	0.1185
	C4	0.1776	0.1632	0.1307	0.1183
	P5	0.1787	0.1632	0.1314	0.1192
PCs	PC1	0.1726	0.1600	0.1278	0.1171
	PC2	0.1747	0.1611	0.1290	0.1174
	PC3	0.1763	0.1626	0.1299	0.1177
	PC4	0.1776	0.1629	0.1307	0.1182
PC1 Loadings	Model1	0.1728	0.1596	0.1278	0.1169
	Model2	0.1740	0.1610	0.1286	0.1174
	Model3	0.1755	0.1618	0.1296	0.1185
	Model4	0.1767	0.1628	0.1303	0.1188

Note: The table above displays the root mean square error averages ($AV.RMSE$) and medians ($M.RMSE$) across all models for the out-of-sample one-step-ahead forecast described in Section 2.4. In regards to our $\widehat{\mathcal{G}}_{CPDAG}$ root, we set an alpha grid as $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$ to select the tuning parameter alpha, after which we evaluate the BIC criteria (described by (10)). We find an optimal regularization parameter $\alpha = 0.001$.

5. Conclusion

The risk factor literature has produced hundreds of potential candidates to explain the cross-section of expected returns, which has led to the issue of the factor zoo. Fortunately, the combination of advanced computational power, enhanced datasets and novel econometric methods has allowed economists to start addressing this matter.

With this paper, we propose a new methodology aimed at reducing the original risk factor candidate set dimension by applying a CPDAG to estimate the joint risk factor distribution, as well as selecting the CPDAG root as the new candidate set to explain cross-sectional returns. We therefore achieve sparsity since the CPDAG root set poses a much lower dimension than its original risk factor set. As we pointed out before - in accordance with Ross's (1976) APT model structure -, the CPDAG root is a natural candidate to span the remaining risk factors and, consequently, the cross-section of expected returns. After the shrinkage selection, we then apply the FM procedure to evaluate our methodology based on the CRSP monthly stock return dataset ranging from January 1981 to December 2016, in addition to Kozak, Nagel, and Santosh (2020) risk factor datasets.

Our findings indicate that our methodology yields better both in and out-of-sample results when compared to standard models or related principal component analysis methods documented in the factor zoo literature. This satisfactory result shows that CPDAG analysis may prove to be a useful tool in helping to understand joint risk factor distribution, in addition to how its properties can be applied to select risk factors, with the purpose of reaching a sparse risk factor model aimed at explaining the cross-section of expected returns.

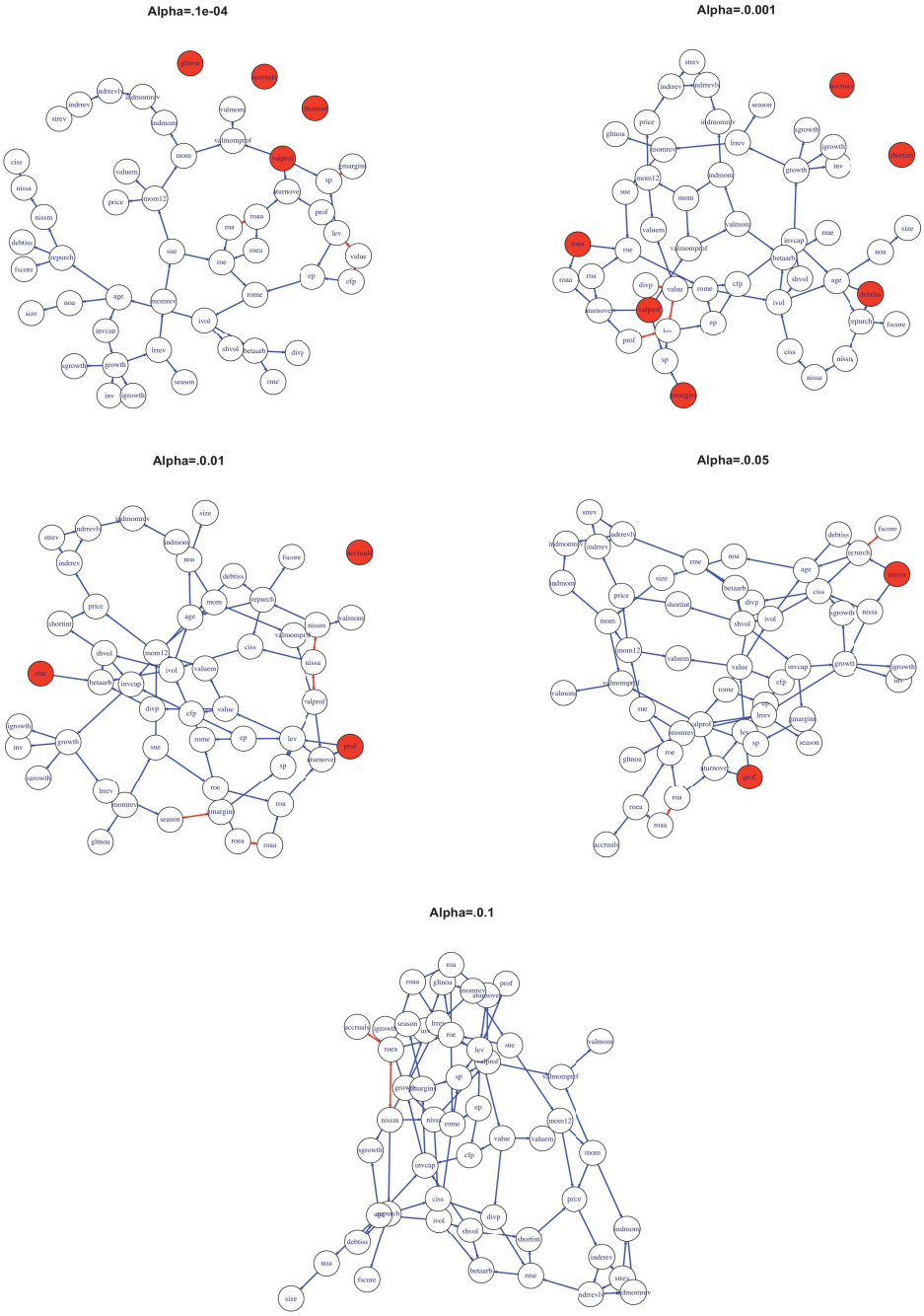
Appendix

Table A1: Descriptive statistics of risk factors

Description	Ret.	S.R.	Reference	Code
Excess Market Return	0.064	0.418	(Sharpe, 1964)	Rme
Size	-0.027	-0.170	(Fama and French,1993)	Size
Value	0.039	0.242	(Fama and French, 1993)	value
Gross Profitability	0.024	0.151	(Novy-Marx, 2013)	prof
Value-Profitability	0.131	0.845	(Novy-Marx, 2013)	valprof
Piotroski's F-score	0.038	0.221	(Piotroski, 2000)	fscore
Debt Issuance	0.016	0.098	(Spiess and Affleck-Graves, 1999)	debtiss
Share Repurchases	0.029	0.172	(Ikenberry, Lakonishok, and Vermaelen, 1995)	repurch
Share Issuance (annual)	-0.100	-0.563	(Pontiff and Woodgate, 2008)	nissa
Accruals	-0.029	-0.188	(Sloan, 1996)	accruals
Asset Growth	-0.090	-0.556	(Cooper, Gulen, and Schill, 2008)	growth
Asset Turnover	0.036	0.246	(Soliman, 2008)	aturnover
Gross Margins	-0.011	-0.068	(Novy-Marx, 2013)	gmargins
Dividend Yield	0.022	0.155	(Naranjo, Nimalendran, and Ryngaert, 1998)	divp
Earnings/Price	0.052	0.313	(Basu, 1977)	Ep
Cash Flow/Market Value of Equity	0.048	0.295	(Lakonishok, Shleifer, and Vishny, 1994)	Cfp
Net Operating Assets	0.004	0.026	(Hirshleifer et al., 2004)	Noa
Investment	-0.098	-0.594	(Chen, Novy-Marx, and Zhang, 2011)	Inv
Investment-to-Capital	-0.053	-0.302	(Xing, 2008)	invcap
Investment Growth	-0.107	-0.630	(Xing, 2008)	igrowth
Sales Growth	-0.071	-0.427	(Lakonishok, Shleifer, and Vishny, 1994)	sgrowth
Leverage	0.033	0.195	(Bhandari, 1988)	Lev
Return on Assets (annual)	0.010	0.067	(Chen, Novy-Marx, and Zhang, 2011)	roaa
Return on Equity (annual)	0.038	0.235	(Haugen and Baker, 1996)	roea
Sales-to-Price	0.066	0.394	(Barbee Jr, Mukherji, and Raines, 1996)	Sp
Growth in LTNOA	-0.017	-0.132	(Fairfield, Whisenant, and Yohn, 2003)	gltnoa
Momentum (6m)	0.004	0.023	(Jegadeesh and Titman, 1993)	mom
Industry Momentum	0.042	0.244	(Moskowitz and Grinblatt, 1999)	indmom
Value-Momentum	0.029	0.178	(Novy-Marx, 2013)	valmom
Value-Momentum-Profitability	0.051	0.308	(Novy-Marx, 2013)	Valmomprf
Short Interest	-0.006	-0.048	(Dechow, Kothari, and Watts, 1998)	shortint
Momentum (1y)	0.043	0.259	(Jegadeesh and Titman, 1993)	mom12
Momentum-Reversal	-0.074	-0.451	(Jegadeesh and Titman, 1993)	momrev
Long-term Reversals	-0.058	-0.374	(De Bondt and Thaler, 1985)	lrrev
Value (monthly)	0.027	0.161	(Asness and Frazzini, 2013)	valuem
Share Issuance (monthly)	-0.096	-0.533	(Pontiff and Woodgate, 2008)	nissm
PEAD (SUE)	0.072	0.481	(Foster, Olsen, and Shevlin, 1984)	sue
Return on Book Equity	0.084	0.541	(Chen, Novy-Marx, and Zhang, 2011)	roe
Return on Market Equity	0.073	0.441	(Chen, Novy-Marx, and Zhang, 2011)	rome
Return on Assets	0.047	0.316	(Chen, Novy-Marx, and Zhang, 2011)	roa
Short-term Reversal	-0.069	-0.413	(Jegadeesh, 1990)	strev
Idiosyncratic Volatility	-0.054	-0.323	(Ang, Chen, and Xing, 2006)	ivol
Beta Arbitrage	-0.034	-0.207	(Cooper, Gulen, and Schill, 2008)	betaarb
Seasonality	0.076	0.482	(Heston and Sadka, 2008)	season
Industry Relative Reversals	-0.133	-0.808	(Da, Liu, and Schaumburg, 2014)	indrrev
Industry Rel. Rev. (Low Vol.)	-0.225	-1.542	(Da, Liu, and Schaumburg, 2014)	indrrevlv
Industry Momentum-Reversal	0.143	0.885	(Moskowitz and Grinblatt, 1999)	indmomrev
Composite Issuance	-0.086	-0.543	(Daniel and Titman, 2006)	ciss
Price	-0.015	-0.092	(Blume and Husic, 1973)	price
Firm Age	0.013	0.074	(Barry and Brown, 1984)	age
Share Volume	-0.037	-0.222	(Datar, Naik, and Radcliffe, 1998)	shvol

Note: The table above displays the descriptive statistics of our risk factor dataset compiled by Kozak, Nagel, and Santosh (2020), with monthly data ranging from January 1981 to December 2016. For each risk factor, the table includes annualized average excess returns, annualized Sharpe ratios, literature references, and code names.

Figure A1: Risk factor distribution using $\widehat{\mathcal{G}}_{CPDAG}$, according to the *alpha* parameter



Note: The figure above displays the graphic representation of the joint risk factor distribution as $\widehat{\mathcal{G}}_{CPDAG}$ for different alpha values. As described in section 2.1.4, we estimate the joint risk factor distribution with $\widehat{\mathcal{G}}_{CPDAG} = \mathcal{G}_{CPDAG}\left(\boldsymbol{\nu} = \boldsymbol{f}, \widehat{\boldsymbol{\varepsilon}}_{CPDAG}(\alpha)\right)$. We set an alpha grid as $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$. In each figure, each node represents a risk factor; blue lines denote direct edges; and red lines denote undirected edges. The nodes in red refer to graph roots ($\hat{P}_{\mathcal{G}_{CPDAG}}(\alpha)$).

Table A2: Risk factor selected using the $\widehat{\mathcal{G}}_{CPDAG}$ *root,* according to different alphas

Description	Ret.	S.R.	Alpha				
			0.0001	**0.001**	**0.01**	**0.05**	**0.1**
Accruals	-0.029	-0.188	X	X	X		
Short Interest	-0.006	-0.048	X	X			
Value-Profitability	0.131	0.845	X	X			
Growth in LTNOA	-0.017	-0.132	X				
Gross Margins	-0.011	-0.068		X			none
Debt Issuance	0.016	0.098		X			
Return on Equity (annual)	0.038	0.235		X			
Gross Profitability	0.024	0.151			X	X	
Excess Market Return	0.064	0.418			X		
Share Issuance (monthly)	-0.096	-0.533				X	
BIC			30,747.4	29,549.7	30,864.3	30,838.9	30,754.0

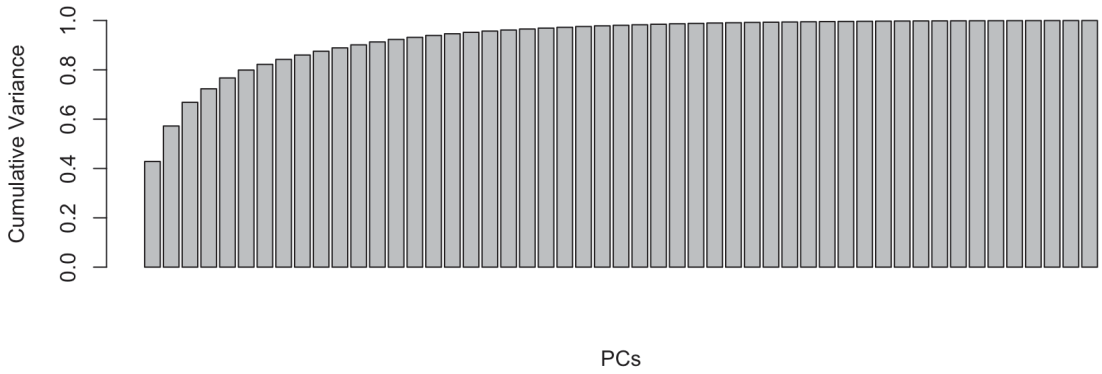
Note: The table above displays the selected risk factors based on the estimated $\widehat{\mathcal{G}}_{CPDAG}$ root set ($\hat{P}_{\mathcal{G}_{CPDAG}}$), as well as the BIC criteria for models with different alpha values. We set an alpha grid as $\alpha = (0.0001, 0.001, 0.01, 0.05, 0.1)$. For each selected risk factor, the table includes annualized average excess returns and annualized Sharpe ratios.

Table A3: FM results for risk factor models selected using the $\widehat{\mathcal{G}}_{CPDAG}$ *root,* according to different alphas

Base		**CRSP**				**CRSP without Small-Caps**			
Coefficient	**Alpha**	**0.0001**	**0.001**	**0.01**	**0.05**	**0.0001**	**0.001**	**0.01**	**0.05**
Intercept	Estimate	0.0037**	0.0033**	0.0036***	0.0038*	0.0043***	0.0046**	0.004***	0.0059*
	S. Error	0.0015	0.0014	0.0013	0.0020	0.0013	0.0011	0.0009	0.0017
	p-value	0.0140	0.0157	0.0057	0.0557	0.0007	0.0000	0.0000	0.0006
Accruals	Estimate	0.0023	0.0017	0.0009		0.0035	0.0030	0.0030	
	S. Error	0.0027	0.0026	0.0029		0.0025	0.0024	0.0025	
	p-value	0.3973	0.5176	0.7596		0.1576	0.2210	0.2328	
Short Interest	Estimate	0.0046**	0.0034***			0.0081***	0.007***		
	S. Error	0.0023	0.0012			0.0022	0.0021		
	p-value	0.0480	0.0076			0.0003	0.0011		
Value-Profitability	Estimate	0.0017	0.0009			0.0011	0.0009		
	S. Error	0.0025	0.0025			0.0024	0.0023		
	p-value	0.4912	0.7269			0.6301	0.6791		
Growth in LTNOA	Estimate	0.0011				(0.0010)			
	S. Error	0.0024				0.0023			
	p-value	0.6544				0.6575			
Gross Margins	Estimate		0.0003				0.0023		
	S. Error		0.0027				0.0026		
	p-value		0.9060				0.3813		
Debt Issuance	Estimate		0.0039*				0.0048*		
	S. Error		0.0022				0.0027		
	p-value		0.0875				0.0838		
Return on Equity (annual)	Estimate		(0.0019)				(0.0027)		
	S. Error		0.0028				0.0026		
	p-value		0.5010				0.3010		
Gross Profitability	Estimate			0.0035	0.0029			0.0051*	0.0051*
	S. Error			0.0028	0.0029			0.0026	0.0026
	p-value			0.2222	0.3161			0.0564	0.0467
Excess Market Return	Estimate			0.0030				0.0055**	
	S. Error			0.0027				0.0025	
	p-value			0.2744				0.0288	
Share Issuance (monthly)	Estimate				0.0040				0.0053*
	S. Error				0.0032				0.0029
	p-value				0.2048				0.0715
Av. adjusted R^2		0.111	0.158	0.093	0.068	0.128	0.178	0.110	0.082

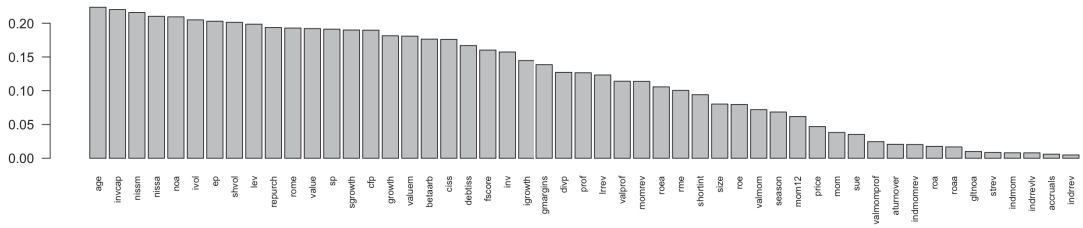
Note: The table above displays the Fama-MacBeth (FM) results for models whose risk factors were selected based on the estimated $\widehat{\mathcal{G}}_{CPDAG}$ root set ($\hat{P}_{\mathcal{G}_{CPDAG}}$). Section 2.2 describes the full FM procedure. Columns refer to models with different alpha values. Results are reported for CRSP datasets both with and without small-caps. The average number of securities in each cross-sectional regression is 4,041 and 2,885, respectively, for CRSP datasets both with and without small-caps. For each model and dataset, the table includes the estimated risk premia, the standard error and the p-value for null hypothesis $H_0: \lambda = 0$, as opposed to the alternative hypothesis $H_1: \lambda \neq 0$. Subscriptions $*,**$ and $***$ indicate that the null hypothesis is rejected at 10%, 5% and 1% levels of significance, respectively.

Figure A2: Principal component cumulative risk factor variance



Note: This figure displays the cumulative risk factor variance explained by its principal components (PCs).

Figure A3: First principal component risk factor loadings' absolute values



Note: This figure displays the absolute values for the first principal component's risk factor loadings, arranged by value.

Table A4: Modified Diebold-Mariano test results for the CRSP dataset

	Model 1		$\widehat{\mathcal{G}}_{CPDAG}$	Classic				PCs				PC1 Loadings			
				FF3	NM4	C4	P5	PC1	PC2	PC3	PC4	M1	M2	M3	M4
	$\widehat{\mathcal{G}}_{CPDAG}$			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Model 2	Classic	FF3			0.000	0.000	0.000	0.000	0.000	0.936	0.000	0.000	0.000	0.000	0.675
		NM4				0.000	0.000	0.000	0.000	0.576	0.000	0.000	0.000	0.000	0.000
		C4					0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		P5						0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	PCs	PC1							0.000	0.000	0.000	0.000	0.000	0.000	0.000
		PC2								0.000	0.000	0.000	0.000	0.000	0.000
		PC3									0.000	0.000	0.000	0.000	0.328
		PC4										0.000	0.000	0.000	0.000
	PC1 Loadings	M1											0.000	0.000	0.000
		M2												0.000	0.000
		M3													0.000
		M4													

Note: The table displays the p-value from the modified Diebold-Mariano test for predictive accuracy among two models proposed by D. Harvey, Leybourne, and Newbold (1997)*. The null hypothesis is given by* $H_0: h(e_{1,t}) = h(e_{2,t})$ *against the alternative hypothesis* $H_0: h(e_{1,t}) \neq h(e_{2,t})$*, where* $e_{i,t}$ *is the out-of-sample one-step-ahead forecast error described in Section 2.4 from model* i *and* h *is a quadratic function.*

Table A5: Diebold-Mariano test results for the CRSP dataset without small caps

	Model 1		$\widehat{\mathcal{G}}_{CPDAG}$	Classic				PCs				PC1 Loadings			
				FF3	NM4	C4	P5	PC1	PC2	PC3	PC4	M1	M2	M3	M4
	$\widehat{\mathcal{G}}_{CPDAG}$			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Model 2	Classic	FF3			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		NM4				0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		C4					0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		P5						0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	PCs	PC1							0.000	0.000	0.000	0.786	0.000	0.000	0.000
		PC2								0.000	0.000	0.000	0.000	0.000	0.000
		PC3									0.000	0.000	0.000	0.000	0.000
		PC4										0.000	0.000	0.000	0.000
	PC1 Loadings	M1											0.000	0.000	0.000
		M2												0.000	0.000
		M3													0.000
		M4													

Note: The table displays the p-value from the modified Diebold-Mariano test for predictive accuracy among two models proposed by D. Harvey, Leybourne, and Newbold (1997)*. The null hypothesis is given by* $H_0: h(e_{1,t}) = h(e_{2,t})$ *against the alternative hypothesis* $H_0: h(e_{1,t}) \neq h(e_{2,t})$*, where* $e_{i,t}$ *is the out-of-sample one-step-ahead forecast error described in Section 2.4 from model* i *and* h *is a quadratic function.*

Table A6: Out-of-sample results for risk factor models selected using the $\widehat{\mathcal{G}}_{CPDAG}$ *root,* according to different alphas

Base		**CRSP**		**CRSP without Small-Caps**	
Model/Metric		$AV.RMSE$	$M.RMSE$	$AV.RMSE$	$M.RMSE$
$\widehat{\mathcal{G}}_{CPDAG}$ *roots*	0.0001	0.1769	0.1629	0.1301	0.1183
	0.001	0.1715	0.1542	0.1258	0.1118
	0.01	0.1724	0.1536	0.1270	0.1116
	0.05	0.1744	0.1611	0.1288	0.1178

Note: The table above displays the root mean square error averages ($AV.RMSE$) *and medians* ($M.RMSE$) *across* $\widehat{\mathcal{G}}_{CPDAG}$ *root models concerning the out-of-sample one-step-ahead forecast described in Section 2.4. for different alpha values.*

References

Abu-Mostafa, Y S, M Magdon-Ismail, and H T Lin. 2012. "Learning from Data Vol. 4: AMLBook New York." *NY, USA*.

Andersson, Steen A, David Madigan, Michael D Perlman, and others. 1997. "A Characterization of Markov Equivalence Classes for Acyclic Digraphs." *The Annals of Statistics* 25 (2): 505–41.

Ang, Andrew, Joseph Chen, and Yuhang Xing. 2006. "Downside Risk." *The Review of Financial Studies* 19 (4): 1191–1239.

Asness, Clifford, and Andrea Frazzini. 2013. "The Devil in HML's Details." *The Journal of Portfolio Management* 39 (4): 49–68.

Barbee Jr, William C, Sandip Mukherji, and Gary A Raines. 1996. "Do Sales--Price and Debt--Equity Explain Stock Returns Better than Book--Market and Firm Size?" *Financial Analysts Journal* 52 (2): 56–60.

Barry, Christopher B, and Stephen J Brown. 1984. "Differential Information and the Small Firm Effect." *Journal of Financial Economics* 13 (2): 283–94.

Basu, Sanjoy. 1977. "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis." *The Journal of Finance* 32 (3): 663–82.

Bhandari, Laxmi Chand. 1988. "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence." *The Journal of Finance* 43 (2): 507–28.

Blume, Marshall E, and Frank Husic. 1973. "Price, Beta, and Exchange Listing." *The Journal of Finance* 28 (2): 283–99.

Bondt, Werner F M De, and Richard Thaler. 1985. "Does the Stock Market Overreact?" *The Journal of Finance* 40 (3): 793–805.

Campbell, John Y, John J Champbell, John W Campbell, Andrew W Lo, Andrew W Lo, and A Craig MacKinlay. 1997. *The Econometrics of Financial Markets*. princeton University press.

Campbell, John Y, and Samuel B Thompson. 2008. "Predicting Excess Stock Returns out of Sample: Can Anything Beat the Historical Average?" *The Review of Financial Studies* 21 (4): 1509–31.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." Journal of Finance, 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

Cavalcante Filho, Elias, Fernando Chague, Rodrigo De-Losso, and Bruno Giovannetti. 2020. "US Risk Premia under Emerging Markets Constraints." Available at SSRN 3600947.

Chen, Long, Robert Novy-Marx, and Lu Zhang. 2011. "An Alternative Three-Factor Model." Available at SSRN 1418117.

Chickering, David Maxwell. 2002. "Learning Equivalence Classes of Bayesian-Network Structures." Journal of Machine Learning Research 2 (Feb): 445–98.

Cochrane, John H. 2009. Asset Pricing: Revised Edition. Princeton university press.

———. 2011. "Presidential Address: Discount Rates." The Journal of Finance 66 (4): 1047–1108.

Colombo, Diego, and Marloes H Maathuis. 2014. "Order-Independent Constraint-Based Causal

Structure Learning." The Journal of Machine Learning Research 15 (1): 3741–82.

Cooper, Michael J, Huseyin Gulen, and Michael J Schill. 2008. "Asset Growth and the Cross-Section of Stock Returns." The Journal of Finance 63 (4): 1609–51.

Da, Zhi, Qianqiu Liu, and Ernst Schaumburg. 2014. "A Closer Look at the Short-Term Return Reversal." Management Science 60 (3): 658–74.

Daniel, Kent, and Sheridan Titman. 2006. "Market Reactions to Tangible and Intangible Information." The Journal of Finance 61 (4): 1605–43.

Datar, Vinay T, Narayan Y Naik, and Robert Radcliffe. 1998. "Liquidity and Stock Returns: An Alternative Test." Journal of Financial Markets 1 (2): 203–19.

Dechow, Patricia M, Sagar P Kothari, and Ross L Watts. 1998. "The Relation between Earnings and Cash Flows." Journal of Accounting and Economics 25 (2): 133–68.

Drton, Mathias, and Thomas S Richardson. 2002. "A New Algorithm for Maximum Likelihood Estimation in Gaussian Graphical Models for Marginal Independence." In Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence, 184–91.

Fairfield, Patricia M, J Scott Whisenant, and Teri Lombardi Yohn. 2003. "Accrued Earnings and Growth: Implications for Future Profitability and Market Mispricing." The Accounting Review 78 (1): 353–71.

Fama, Eugene F, and Kenneth R French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." Journal of Financial Economics 33 (1): 3–56. https://doi.org/10.1016/0304-405X(93)90023-5.

Fama, Eugene F, and James D MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." Journal of Political Economy 81 (3): 607–36.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu. 2020. "Taming the Factor Zoo: A Test of New Factors." The Journal of Finance 75 (3): 1327–70.

Foster, George, Chris Olsen, and Terry Shevlin. 1984. "Earnings Releases, Anomalies, and the Behavior of Security Returns." Accounting Review, 574–603.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. 2020. "Dissecting Characteristics Nonparametrically." The Review of Financial Studies 33 (5): 2326–77.

Goyal, Amit. 2012. "Empirical Cross-Sectional Asset Pricing: A Survey." Financial Markets and Portfolio Management 26 (1): 3–38.

Green, Jeremiah, John R M Hand, and X Frank Zhang. 2017. "The Characteristics That Provide Independent Information about Average Us Monthly Stock Returns." The Review of Financial Studies 30 (12): 4389–4436.

Gu, Shihao, Bryan T Kelly, and Dacheng Xiu. 2019. "Autoencoder Asset Pricing Models."

Harris, Naftali, and Mathias Drton. 2013. "PC Algorithm for Nonparanormal Graphical Models." The Journal of Machine Learning Research 14 (1): 3365–83.

Harvey, Campbell R, and Yan Liu. 2019. "Lucky Factors." Available at SSRN 2528780.

Harvey, Campbell R, Yan Liu, and Heqing Zhu. 2016. "… and the Cross-Section of Expected Returns." The Review of Financial Studies 29 (1): 5–68.

Harvey, David, Stephen Leybourne, and Paul Newbold. 1997. "Testing the Equality of Prediction Mean Squared Errors." International Journal of Forecasting 13 (2): 281–91.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. The Elements of Statistical

Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.

Haugen, Robert A, and Nardin L Baker. 1996. “Commonality in the Determinants of Expected Stock Returns.” Journal of Financial Economics 41 (3): 401–39.

Heston, Steven L, and Ronnie Sadka. 2008. “Seasonality in the Cross-Section of Stock Returns.” Journal of Financial Economics 87 (2): 418–45.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang. 2004. “Do Investors Overvalue Firms with Bloated Balance Sheets?” Journal of Accounting and Economics 38: 297–331.

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen. 1995. “Market Underreaction to Open Market Share Repurchases.” Journal of Financial Economics 39 (2–3): 181–208.

Jegadeesh, Narasimhan. 1990. “Evidence of Predictable Behavior of Security Returns.” The Journal of Finance 45 (3): 881–98.

Jegadeesh, Narasimhan, and Sheridan Titman. 1993. “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.” The Journal of Finance 48 (1): 65–91.

Kalisch, Markus, and Peter Bühlmann. 2007. “Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm.” Journal of Machine Learning Research 8 (Mar): 613–36.

Kalisch, Markus, Bernd A G Fellinghauer, Eva Grill, Marloes H Maathuis, Ulrich Mansmann, Peter Bühlmann, and Gerold Stucki. 2010. “Understanding Human Functioning Using Graphical Models.” BMC Medical Research Methodology 10 (1): 14.

Kelly, Bryan T, Seth Pruitt, and Yinan Su. 2019. “Characteristics Are Covariances: A Unified Model of Risk and Return.” Journal of Financial Economics 134 (3): 501–24.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. 2020. “Shrinking the Cross-Section.” Journal of Financial Economics 135 (2): 271–92.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny. 1994. “Contrarian Investment, Extrapolation, and Risk.” The Journal of Finance 49 (5): 1541–78.

Moskowitz, Tobias J, and Mark Grinblatt. 1999. “Do Industries Explain Momentum?” The Journal of Finance 54 (4): 1249–90.

Nagarajan, Radhakrishnan, Sujay Datta, Marco Scutari, Marjorie Beggs, Greg Nolen, and Charlotte Peterson. 2010. “Functional Relationships between Genes Associated with Differentiation Potential of Aged Myogenic Progenitors.” Frontiers in Physiology 1: 160.

Naranjo, Andy, M Nimalendran, and Mike Ryngaert. 1998. “Stock Returns, Dividend Yields, and Taxes.” The Journal of Finance 53 (6): 2029–57.

Novy-Marx, Robert. 2013. “The Other Side of Value: The Gross Profitability Premium.” Journal of Financial Economics 108 (1): 1–28.

Pearl, Judea. 2009. Causality. Cambridge university press.

Piotroski, Joseph D. 2000. “Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers.” Journal of Accounting Research, 1–41.

Pontiff, Jeffrey, and Artemiza Woodgate. 2008. “Share Issuance and Cross-Sectional Returns.” The Journal of Finance 63 (2): 921–45.

Shanken, Jay. 1992. “On the Estimation of Beta-Pricing Models.” The Review of Financial Studies 5 (1): 1–33.

Sharpe, William F. 1964. “Capital Asset Prices: A Theory of Market Equilibrium under

Conditions of Risk." The Journal of Finance 19 (3): 425–42.

Sloan, Richard G. 1996. "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?" Accounting Review, 289–315.

Soliman, Mark T. 2008. "The Use of DuPont Analysis by Market Participants." The Accounting Review 83 (3): 823–53.

Spiess, D Katherine, and John Affleck-Graves. 1999. "The Long-Run Performance of Stock Returns Following Debt Offerings." Journal of Financial Economics 54 (1): 45–73.

Spirtes, Peter, Clark N Glymour, Richard Scheines, and David Heckerman. 2000. Causation, Prediction, and Search. MIT press.

Stekhoven, Daniel J, Izabel Moraes, Gardar Sveinbjörnsson, Lars Hennig, Marloes H Maathuis, and Peter Bühlmann. 2012. "Causal Stability Ranking." Bioinformatics 28 (21): 2819–23.

Verma, Thomas, and Judea Pearl. 1991. Equivalence and Synthesis of Causal Models. UCLA, Computer Science Department.

Xing, Yuhang. 2008. "Interpreting the Value Effect through the Q-Theory: An Empirical Investigation." The Review of Financial Studies 21 (4): 1767–95.

Yan, Xuemin, and Lingling Zheng. 2017. "Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach." The Review of Financial Studies 30 (4): 1382–1423.

Zhang, Xiujun, Xing-Ming Zhao, Kun He, Le Lu, Yongwei Cao, Jingdong Liu, Jin-Kao Hao, Zhi-Ping Liu, and Luonan Chen. 2012. "Inferring Gene Regulatory Networks from Gene Expression Data by Path Consistency Algorithm Based on Conditional Mutual Information." Bioinformatics 28 (1): 98–104.