UNIVERSITY OF CALIFORNIA

Los Angeles

Machine Learning and Asset Pricing Models

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Management

by

Rafael Amaral Porsani

2018

ABSTRACT OF THE DISSERTATION

Machine Learning and Asset Pricing Models

by

Rafael Amaral Porsani

Doctor of Philosophy in Management

University of California, Los Angeles, 2018

Professor Richard W. Roll, Co-Chair

Professor Avanidhar Subrahmanyam, Co-Chair

Even though statistical-learning techniques have become increasingly popular in many scientific areas, few studies in the field of cross-sectional asset pricing have incorporated these in their essence. In the first chapter of this dissertation, we suggest a framework for testing the empirical performance of linear asset-pricing factor models, and for investigating anomalies, which employs an array of such techniques, bringing artificial intelligence and asset-pricing a step closer. The methodology utilized in our work combines a range of supervised learning algorithms with the model testing strategies of Avramov and Chordia (2006) and Brennan et al. (1998).

Chapter 2 presents results generated by applying our framework to multiple asset pricing models. While simple in nature, the estimation procedure we use can have implications for risk management, the study of anomalies, the creation of optimal investment policies, and the general study of expected returns. Some of the concepts explored herein may take an added role in future studies which investigate these subjects, helping to reshape the way we think about asset prices and financial-market anomalies.

The title of this dissertation is given after its first two chapters. Chapter 3 is titled "The Building Blocks of Employment: A Signal Processing Analysis". As argued by Hawking (2016), artificial intelligence and growing automation have decimated jobs in traditional manufacturing, and may engender further job destruction into the middle

classes, promoting a widening of wealth inequality in their wake. In this chapter, we contribute to the general study of employment, a theme of critical importance today, by utilizing signal processing techniques to decompose into a myriad of building blocks the employment-to-population time series during the years 1975 to 2000 - a prolonged period where strong job gains were registered and recoveries from recessions were quick. An analysis of the main resulting components is presented. The components of employment produced by our signal-processing modeling approach are made available to researchers interested in better comprehending the employment rate during this period, and forces tied to job gains then.

Chapter 4 is co-authored with Mahyar Kargar, and it is titled "The Evolution of Global Financial Integration: A Multivariate Analysis of Currencies and Equities". In this study, we rely on principal component regressions and canonical correlation analyses to show that not only currencies became more integrated with each other from the mid-nineties through the early years of the twenty-first century, but also different assets classes – currencies and equities – became more closely associated throughout the same period. Our framework suggests that a common set of latent factors was, during these years, ever more capable of explaining returns from disparate assets; although such buoyant trend in integration subsequently faced strong headwinds, no longer being present in recent times.

The dissertation of Rafael Amaral Porsani is approved.

Ying N. Wu

Francis A. Longstaff

Avanidhar Subrahmanyam, Committee Co-Chair

Richard W. Roll, Committee Co-Chair

University of California, Los Angeles

2018

To my dearly loved

parents, sister and niece

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the UCLA Anderson School of Management for providing me with an incredible learning experience and for their generous funding. I also thank the UCLA Graduate Division for their generous support.

This dissertation wouldn't have been possible without the encouragement and love of my parents, Milton and Dalva, and of my sister, Juliana. I would like to give a special thank you to all of you. I also thank my other family members and friends, whose support has inspired me throughout the years: the Jarkvist family, the Swart family, the Amaral family, the Ramos family, Laura Kim Tao, Fiona Yeung, Marcio Osako, Rodrigo Kim, Emily Steinbarth, André Parra, Davi Takeshita, Eduardo Izumi, Anthony Vo, Mirella de Bittencourt Cardoso, Jimmy Yeung, Mahima Garg, Vanessa Leoni and Marcio Kimura, among so many others. I also thank my former team members and mentors at Goldman Sachs, Crédit Agricole and Citigroup.

I am extremely thankful to the members of my dissertation committee for their guidance and support, and for sharing with me their knowledge and wisdom: Avanidhar Subrahmanyam (Co-Chair), Richard W. Roll (Co-Chair), Francis A. Longstaff and Ying Nian Wu. Additionally, I thank Eduardo Schwartz, Terence Tao, Frederic Paik Schoenberg, Hanno Lustig, Peter Carr, Ivo Welch, Robert Geske, Teri Geske, Mikhail Chernov, Jack Hughes and John Riley. I would also like to thank Mahyar Kargar, co-author of Chapter 4 of this dissertation. I also thank my other doctoral colleagues at the UCLA Anderson Graduate School of Management.

Lastly, I thank Elisa Dunn for her support while I was in the Master of Financial Engineering Program at UCLA. Moreover, I thank Lydia Heyman, Craig Jessen, Leanna Cortez and Delores Rhaburn for their support.

VITA

2005	Visiting Student, Brown University.
2006	Bachelor of Administration, Universidade de São Paulo (USP).
2007–2008	Investment Analyst, The Citigroup Private Bank. Fulfilled non-discretionary client trades related to fixed-income securities, equities, futures contracts and options, including exotic options.
2009	Investment Analyst, Crédit Agricole. Acted as equity research analyst, covering twenty real estate firms.
2009–2012	Associate/Senior Analyst, Goldman Sachs & Co. Conducted trades on a daily basis, identified and pursued business opportunities, working closely with different product groups (Alternative Investments, Structured Products, Equity Research).
2011	CFA® Charterholder.
2013	Master of Financial Engineering, Anderson School of Management, University of California, Los Angeles.
2013	Intern, Wilshire Associates. Conducted empirical study on the incorporation of statistical factors into firm's proprietary risk-analysis framework.
2015–2017	Teaching Assistant, Anderson School of Management, University of California, Los Angeles. Taught sections of Fixed-Income Markets, Management 232B and Management 237F.
2017	M.S. in Statistics, University of California, Los Angeles.

CHAPTER 1

Machine Learning, Asset Pricing Models and Financial Market Anomalies - Introductory Comments and Methodology

1.1 Introduction

Correctly specified multi-factor models allow risk managers and investors to properly decompose returns into systematic influences and firm-specific effects, contributing therefore to facilitate the investment process and to elucidate the manners in which risk can be managed. Such models are not only intuitive and popular in finance studies; they are key ingredients to many risk products offered by leading analytics firms around the globe[1]. And while the publication of Ross' Arbitrage Pricing Theory (Ross (1973) and Ross (1976)), and of Merton's Intertemporal Capital Asset Pricing Model (Merton (1973)) helped sprawl a burgeoning interest among the finance community on multi-factor linear models, the debate over the empirical performance of such models has never been more relevant. In the wake of arguably the worst financial disasters since the Great Depression, the European debt crisis of 2009-2012 and the 2007-2009 market collapse, investors and portfolio managers have become increasingly aware of the need to utilize models that more accurately take into consideration the impact of systematic shocks on firm value. But how can we empirically evaluate different factor models? In this article, we propose a machine

[1]Wilshire and Barra, for example, are both firms that present its clients with risk models based on factors (see Steven J. Foresti and Minassian (2015) and Inc. (2009), respectively); and so are numerous others.

learning framework for assessing asset pricing models, which complements Brennan et al. (1998), henceforth BCS, and is inspired by the work of Avramov and Chordia (2006).

The BCS framework for testing asset pricing models involves assessing whether equity characteristics such as, for example, firm size or lagged returns, have explanatory power relative to arbitrage pricing theory benchmarks. It entails evaluating conjectured risk-based asset pricing models through regressions of risk-adjusted returns - computed for individual stocks - on company-specific characteristics, via Fama-Macbeth-type procedures. This approach is capable of providing a myriad of insights onto the co-movement of asset prices, as the methodology proposed by BCS allows one to empirically test models and to simultaneously investigate numerous anomalies. With over nintey characteristics having been reported as being able to explain cross-sectional stock returns (see McLean and Pontiff (2015)), and considering that more than three hundred factors have been proposed in asset-pricing studies (Harvey et al. (2015)), the cross-sectional testing procedure developed by BCS is arguably now more relevant than ever.

A well-known application of the concepts described in BCS, we note, can be found in Avramov and Chordia (2006), hereafter AC, who modify the testing procedures shown in BCS by incorporating into the approach suggested by the latter models that condition risk loadings on different variables, merging therefore econometric specifications proposed by Ferson and Harvey (1999) with the cross-sectional analysis found in BCS. AC, in this manner, allow risk and expected return to vary with conditioning information - which may be company-specific.

In this chapter, we propose an approach that extends the work of AC. Our framework introduces an array of learning machines into asset pricing tests that, similarly to those conducted by AC, admit loadings to change in time conditionally upon firm-specific and/or business-cycle-related information.

The framework we suggest focuses on incorporating various machine-learning-based techniques into the estimation of time series specifications of returns, which are then used as "inputs" in asset pricing tests. Our framework goes beyond the notion of using a single regularization approach in estimation procedures, and instead combines a range of supervised learning algorithms with asset-pricing tests, further blending artificial intelligence with the study of linear multi-factor asset-pricing models. The methodology we use is discussed in detail in Section 1.2, and is corroborated by findings displayed in Chapter 2 (see Section 2.1.2 and Appendix A of such chapter).

In the empirical analysis section of our work, shown in Chapter 2, we present results from two related studies. In the first of these studies, we assess how often the least squares estimator, and machine-learning-based ones, are selected as optimal estimators across different time series specifications of returns when our estimation framework that relies on cross-validation to choose learning algorithms for different problems is applied.

This study suggests, as we'll see, that estimators other than the least squares one may be of help in training/estimating time series specifications, especially when betas are allowed to vary with multiple conditioning variables. Along these lines, our results indicate that common algorithms used in machine learning, such as lasso and ridge regression, and variations of these, may aid in that task.

As shown in Chapter 2, results from the aforementioned study, furthermore, also imply that the least squares algorithm should, preferably, be avoided altogether when many conditioning variables are used to model betas. In two of the models we examine, which each let risk loadings depend on five conditioning variables, for example, OLS estimation was never deemed optimal.

Over the course of this dissertation, additionally, we present evidence suggesting that the multifaceted machine-learning-based estimation approach we use is justified, as no single learning algorithm emerges, for any of the sixteen models examined in our work, as being chosen in an overwhelmingly frequent manner as an ideal one. The frequencies with which different learning algorithms are selected as optimal are rather "dispersed" across multiple algorithms. This, in turn, helps to mitigate the notion that a single supervised learning algorithm should be used to estimate time series specifications of returns across all stocks.

The second of the two abovementioned studies is naturally composed of asset pricing tests that use as "inputs" time series specifications estimated using our machine learning framework. Here, we find that even though AC indicate that conditional betas may offer a possible risk-based explanation for value and size anomalies, none of the models we study, including those that allow betas to vary with conditioning information, is able to capture the impact of either size or value - book-to-market - in the cross-section of stock returns. Put simply, we cannot find a risk-based explanation for these anomalies when our machine-learning-driven estimation framework is employed, regardless of the model being considered. This result is particularly relevant given the debate in Daniel and Titman (1997), who argue that characteristics affect the cross-section of returns, as opposed to coefficients tied to the Fama-French Small-Minus-Big and High-Minus-Low factors.

On a related note, our asset pricing tests, furthermore, also show that while past return anomalies are indeed very difficult to explain from a risk-based perspective, turnover effects appear to be easier to capture, and may be explicated by a number of different models.

Our work provides a number of contributions to the asset-pricing literature. Firstly, we lay out a framework that furthers the integration between statistical learning and asset-pricing tests. We go beyond the notion of using a single algorithm that relies on regularization, and use instead a multi-algorithmic system that is itself corroborated by findings presented in this dissertation. The methodology we use brings artificial intelligence and asset-pricing a step closer, yielding new insights into the pricing capabilities of different models. The former is a somewhat new and vibrant area of research, which may help us better comprehend already existing models in finance, and reshape the way we think about asset prices. While simple in nature, the estimation procedure we have developed may ultimately have implications for risk management, the study of anomalies, the creation of optimal investment policies, and the study of risk premia. Secondly, our results imply that common machine learning algorithms, and variations of them, may aid in estimating time series specifications of returns. Our findings, furthermore, help to substantiate the notion that using multiple algorithms to estimate time series specifications, as opposed to a single one, is warranted.

We also note that, in providing an alternative framework that focuses on estimating time series specifications using individual securities, while making use of multiple machine learning techniques, we help contribute to the future development of studies that place securities, as opposed to portfolios, front and center. As stated by Lo and MacKinlay (1990), financial statisticians may end up rejecting the null hypothesis of exact pricing too frequently to the extent that they utilize in their analysis portfolios formed based on firm characteristics that were earlier found to be related to cross-sectional returns. And as BCS point out, Roll (1977) makes nearly an opposite argument, explaining that portfolios may hide important risk or return-related elements associated with individual securities, leading researchers to fail to reject the null when they otherwise should not. From these arguments, it should be clear that using portfolios in asset pricing tests is problematic.

The asset pricing tests we conduct are guided by those shown in AC, who themselves use, as we have mentioned above, Ferson and Harvey (1999) and BCS as inspiration for their own tests. While AC utilize the least squares algorithm to estimate time series specifications of returns, we instead estimate these through a multifaceted machine learning approach. In this manner, in assessing anomalies, and considering the tests conducted herein, our paper is mostly related to BCS and AC, while also being related to Ferson and Harvey (1999). Furthermore, given our framework also permits analyzing the pricing ability of asset-pricing models which contain different factors, our approach may also be, albeit to a lesser extent, associated with the literature that deals more closely with factor identification - see, for example, Bryzgalova (2015) or Feng et al. (2017). Bryzgalova (2015) proposes an estimator that relies on regularization to identify factors, and Feng et al. (2017) suggest using the two-stage lasso method of Belloni et al. (2014) to select a control model from a collection of factors. Employing a framework that uses numerous learning machines simultaneously, while allowing for company-specific and economy-wide information to affect risk-loadings, which these papers do not do, may help elucidate the pricing ability tied to different factor-model specifications and complement their studies, while engendering a novel way to understand asset-pricing models altogether.

In the next section, we present in detail our machine learning framework for estimating time series specifications and testing asset-pricing models. The general econometric specification used in our work is presented, and a discussion of the statistical learning system utilized in our tests is showcased. The estimation procedure we suggest uses simultaneously twelve supervised learning algorithms, all of which can be interpreted from a Bayesian perspective. We present an overview of these, and then proceed to discuss how, by combining them, we create a system that uses multiple learning algorithms to test asset pricing models. Our data is then described, and additional information on our estimation procedure is ultimately presented at the Appendix of this chapter, followed by a discussion of the connection between regularization and Bayesian estimation. In Chapter 2, our empirical analysis is shown. Concluding remarks and supplementary studies are outlined at the end of that chapter.

1.2 A Machine Learning Framework

The estimation approach we suggest makes usage of an array of concepts and algorithms employed in statistical learning studies, allowing for an assessment of the performance of complex models from a novel perspective. In Chapter 2, Section 2.1.2, we present evidence corroborating the usage of such approach, and further findings substantiating its usage are also shown in Appendix A of that same chapter.

While we focus our work on assessing the empirical power of models that allow loadings to vary with conditioning variables, simpler models that utilize constant betas are also evaluated in this dissertation. Over the course of our investigation, we re-visit models inspired by those used in AC, drawing new insights from market prices. Unscaled and various conditional versions of the Fama and French (1993) three factor model, the FF 3 factor model plus a long-term reversals factor, the FF 3 factor model plus a momentum factor, and the Fama and French (2015) five factor model are studied herein. Our framework is described in detail on the following pages.

1.2.1 The Econometric Specification

The models utilized in our work can be collectively described through the econometric specification presented next, which encompasses cases where betas are presumed to be either constant or time-varying[2]. Our general specification is of the form

$$R_{j,t} = E_{t-1}[R_{j,t}] + \beta^{'}_{j,t-1} \left\{ F_t - \left[E_{t-1} \left(F_t \right) \right] \right\} + \epsilon_{j,t}, \tag{1.1}$$

where $R_{t,j}$ is the return of stock j in excess of the one-month treasury rate at time t; $E_{t-1}[\cdot\,]$ is an expectation term, taken conditionally upon a sigma-algebra $\mathcal{F}_{t-1}$ that represents the information investors have at time $t-1$. F_t is a $K \times 1$ vector of excess returns on the risk factor-mimicking portfolios at t. Throughout this dissertation, we'll refer to the elements of F_t simply as risk factors. $\beta_{j,t-1}$ is a time dependent, $K \times 1$ vector, holding exposures of stock j to these factors; and $\epsilon_{j,t}$ represents a factor model disturbance satisfying $E_{t-1}\left[\epsilon_{j,t}|F_t\right] = 0$. Expected returns, moreover, are given by:

$$E_{t-1}[R_{j,t}] = \alpha_j + \beta^{'}_{j,t-1} E_{t-1}\left[F_t\right] \tag{1.2}$$

[2]It should not go without saying that our specification is largely based on and resembles closely that shown in Ferson and Harvey (1999), while being presented in a fashion similar to that showcased in such study as well. It is also inspired by AC, who conduct asset pricing tests where betas are allowed to vary with company-specific information.

Each of the K risk exposures is further presumed to vary in a linear manner with different conditioning variables. Our models, consequently, assume that betas are governed by

$$\beta_{j,t-1} = b_{0,j} + b_{1,j}\Theta_{j,t-1}, \tag{1.3}$$

where $b_{0,j}$ is a $K \times 1$ vector, $b_{1,j}$ is a $K \times S$ matrix, and $\Theta_{j,t-1}$ is an $S \times 1$ vector of scaling variables, which contain conditioning information. $\Theta_{j,t-1}$, may include, for instance, firm-specific characteristics or macroeconomic-related information.

Some comments are in order. Firstly, note that although our econometric specification is primarily motivated by Ferson and Harvey (1999) and AC, models that allow betas to vary linearly with diverse variables are interesting in their own right, as they are more flexible than their fixed-beta counterparts, and in this manner have a greater potential than the latter for providing risk-based explanations to financial-market anomalies. Furthermore, notice that it is also possible that betas may reasonably change according to company-specific characteristics. Gomes et al. (2003), for example, provide a general equilibrium model which implies that firm size and book-to-market ratios may impact risk loadings - such model helped motivate the work of AC. To see in an intuitive manner how characteristics and betas may be connected, moreover, consider for example a small firm which grows in size by acquiring other firms, or by entering into new lines of business. Such firm may, in doing so, become more similar itself to a portfolio, and its betas might in that regard change to reflect its newly "combined portfolio" risk-exposures. In essence, if the lines of business of a company change, it would be plausible for one to assume that the way in which such firm responds to risk might do so as well. While our specification permits one to take considerations such as this one explicitly into account, it also allows for accounting for a potential impact of economy-wide information on risk exposures - and note that Ferson and Harvey (1999) find evidence suggesting that risk loadings may indeed depend on business-cycle-related information.

As a last step, we combine Equations[3] 1.1, 1.2 and 1.3, obtaining:

$$R_{j,t} = \alpha_j + \left(b_{0,j} + b_{1,j}\Theta_{j,t-1}\right)' F_t + \epsilon_{j,t} \tag{1.4}$$

A common approach to estimating the econometric model described by Equation 1.4 would be to run time-series regressions of firm excess returns on factor realizations and interactions between the latter and conditioning variables. The estimation approach we propose, instead, utilizes the data itself to try and determine which estimator may better approximate the relationship described by Equation 1.4.

1.2.2 Evidence-Based Allocation of Estimators

The estimation procedure we adopt is markedly different from what has been proposed in the asset-pricing literature so far. It "digs" into the data and uses ideas drawn from statistical learning, combined with multiple supervised learning algorithms, in an attempt to find what is the "true" time series relationship between returns and factors.

Instead of using any single learning algorithm to estimate the econometric specification of interest, our procedure inspects the data, uses it to determine what may be a prospective ideal way to estimate it, and then entrusts the now-verified method with that task. In taking these steps, we seek to make full use of information contained in the available data, using "checks-and-balances" to ensure that Equation 1.4 is being estimated to the best of our ability. In implementing this approach on a stock-by-stock basis, we provide a framework that is rigorous from a statistical learning perspective.

The estimation strategy underscoring our asset pricing tests can be described succinctly through the steps outlined below. To implement it, we suggest that one should

Step 1: Take an asset pricing model described by Equations 1.1, 1.2 and 1.3, and note

[3] A similar implementation of this step is also presented in Ferson and Harvey (1999).

that these expressions can be used to derive the specification shown in Equation 1.4. This is the econometric model our framework will estimate/address.

Step 2: Select a given stock j.

Step 3: Divide the data sample for stock j into three segments or folds ($l_j \in \{1, 2, 3\}$) containing each an equal number of observations. The observations are first ordered in time; the oldest one-third of the observations are placed in the first fold, $l_j = 1$; the next occurring third in placed in the second fold, $l_j = 2$; and the remaining ones are then included in the third fold.

Step 4: Use the combined observations from folds $l_j = 1$ and $l_j = 2$ to estimate/train the model for the selected stock, using least squares and a variety of other estimators/learning algorithms (hereafter we'll refer to observations used to train/estimate a model as a training set).

A total of twelve supervised learning algorithms (denoted by $\mathcal{M} \in \{1...12\}$) are used in our work (these are described in detail further below)).

Step 5: Pick observations in fold $l_j = 3$ as a test set. Compute out-of-sample mean squared errors (MSE) for the $\mathcal{M}$ algorithms used in Step 4, using observations from this fold/test set.

Step 6: Repeat Steps 4 and 5, using folds $l_j = 1$ and $l_j = 2$ as test sets, one at a time, in Step 5, and the remaining folds, in each case, as training sets. As a result, three out-of-sample MSE estimates (one for each test set), for each algorithm, are generated after completion of this step.

Step 7: Evaluate what is the average out-of-sample MSE for each learning algorithm over the three test sets used. These serve as indications of the abilities of the different algorithms in estimating the specification in question, for the selected stock and given the available data.

Step 8: Select the learning algorithm that yielded the lowest average out-of-sample

MSE, and estimate the time series specification for stock j using it, now with a training set that contains all available observations. This step yields our final estimated version of the econometric model/specification of interest for stock j.

Step 9: Iterate over Steps 2 through 8, using all stocks.

By ordering observations according to the times when they occur, in Step 3, and dividing them into folds that respect their natural ordering, we attempt to preserve the structure of our data to the best of our ability, while at the same time being able to conduct multiple trials with each learning algorithm - here, three trials are conducted with each algorithm, for each stock. This approach also allows us to better conform with the machine learning literature, which typically employs K-fold cross-validation, normally with $K \geq 3$, to validate methods or tuning constants[4].

Put simply, our work relies on cross-validation as a mechanism to provide evidence on which estimator should be used for any given problem. Three-fold cross validated mean squared errors - computed using out-of-sample data - are utilized to determine model fit. In other words, we use out-of-sample errors, across three separate trials (an average over three trials is computed) to assess model fit for each stock; and the algorithm that yields the best fit for a given problem (stock and model) is deemed as an optimal estimation vehicle for it.

The approach we use has, in a sense, an artificial intelligence aspect to it. A machine imbued with artificial intelligence, when confronted with a problem, processes the available data pertaining to such problem and reacts in a manner that is compatible with it, having the ability to react differently depending on the problem it faces. In a similar manner, we "process" the data pertaining to a given stock and take an action - where the action here is the utilization of a given estimator - depending on such data.

[4]The popular function cv.glmnet, in R, commonly used in machine learning studies to perform cross-validation for lasso and ridge regression, for example, does not admit values for K less than 3.

Following the completion of all nine steps described above, asset pricing tests - explained in detail further below, under Section 1.2.4 - are conducted while utilizing the estimated time series models/specifications. The spirit of undergoing these steps lies in conducting asset pricing tests with time series specifications that have been estimated "as best as possible", as opposed to trusting any single learning algorithm to estimate them - additional information on our estimation approach is presented in Appendix A of this chapter. In Chapter 2, we show that when cross-validation is applied to determine which learning algorithms should be optimally used, algorithms other than least squares are chosen for 85% of stocks or more, depending on the model being examined - suggesting there is room for learning algorithms other than OLS to help in the estimation of time series specifications. When many conditioning variables are utilized to model risk loadings, OLS's optimality falls to nearly zero, and in two of our models, it is exactly zero, indicating that the least squares algorithm should be ideally avoided in these high dimensional cases. Furthermore, the frequencies with which algorithms are chosen are dispersed across various learning algorithms, irrespective of the model being examined, substantiating the usage of a multi-algorithmic framework.

1.2.3 Supervised Learning Algorithms

In the discussion that follows, we first present a generic optimization problem, which encapsulates all twelve supervised learning algorithms used in our study. After providing details on its algebraic formulation, we proceed to describe which variations of it comprise our estimation methods. A Bayesian interpretation of the twelve training procedures we use is then presented.

1.2.3.1 The General Formulation

The estimators we use in steps 4 and 8 - and on any iterations over these - can be interpreted as minimizers $\left(\hat{\alpha}_j, \hat{b}_j\right)$ of variants of the function below[5]

$$\mathcal{L}(\alpha_j, b_j) = \frac{1}{n_j}\sum_{t=1}^{n_j} \ell(R_{j,t} - \alpha_j - b_j^{'} X_t) + \lambda P(b_j), \tag{1.5}$$

where $\ell(\cdot)$ is a generic loss function; n_j denotes the number of available observations for stock j in the training set being used, α_j is a scalar and b_j is a vector of coefficients containing $(S+1)K$ elements; P is the elastic net (Zou and Hastie (2005) and Friedman et al. (2010)) penalty function, with tuning parameter $\lambda \geq 0$:

$$P(b_j) \equiv P_e(b_j) = e||b_j||_1 + (1-e)\frac{1}{2}||b_j||_2^2,\ 0 \leq e \leq 1, \tag{1.6}$$

where $|| \cdot ||_1$ and $|| \cdot ||_2$ stand for the L_1 and L_2 norms, and e is the elastic net tuning parameter.

A remaining task is to define X_t in Equation 1.5. In unpenalized versions of 1.5, i.e., when λ is taken as zero, this variable represents a $(S+1)K \times 1$ feature vector that contains realizations of F_t and interactions between the latter and conditioning variables found in $\Theta_{j,t-1}$[6]. In such cases, $\hat{\alpha}_j$ and $\hat{b}_j$ can be seen, respectively, as direct estimates for α_j, the intercept term in Equation 1.4, and the coefficients that collectively define time-varying betas for stock j, shown in Equation 1.3.

Alternatively, when regularization is employed in 1.5, i.e., λ is presumed to be greater than zero, the elements in X_t are as before, with the exception that they are standardized

[5]Our general formulation is from Yi and Huang (2015).

[6]For example, when default spread (denote it here as Def_{t-1}) is used as a conditioning variable, and F_t contains the excess return on the market, HML and SMB (MKT_t, HML_t, SMB_t, respectively), the vector X_t has, as its elements, $x_{1,t} = MKT_t$, $x_{2,t} = HML_t$, $x_{3,t} = SMB_t$, $x_{4,t} = Def_{t-1}MKT_t$, $x_{5,t} = Def_{t-1}HML_t$ and $x_{6,t} = Def_{t-1}SMB_t$.

to have mean zero and variance one. As is customary in the penalized regression literature, $\hat{\alpha}_j$ and $\hat{b}_j$ are then converted back to estimates of the original coefficients of interest.

1.2.3.2 Twelve Variants

Twelve variants of the general $\mathcal{L}(\cdot)$ loss function shown in Equation 1.5 are used in our framework. Put simply, we allow $\ell(\cdot)$ to denote either squared loss, absolute loss - which is more robust to vertical outliers than squared loss - or Huber loss - a robust loss function that can handle leverage points better than the previous two functions. These three loss functions are described in Yi and Huang (2015). The squared loss function here can be written as $l(\delta) = \frac{\delta^2}{2}$; absolute loss is $l(\delta) = \delta\left(\frac{1}{2} - I(\delta < 0)\right)$, with $I(\cdot)$ denoting an indicator function that is equal to one if the condition inside parentheses is satisfied, and zero otherwise; and Huber loss is a function that can take one of two functional forms, depending on its argument. It can be written as: $l(\delta) = \frac{\delta^2}{2k}$ if $|\delta| \leq k$, and if $|\delta| > k$, $l(\delta) = |\delta| - \frac{k}{2}$. As Yi and Huang (2015) point out, the Huber loss function is quadratic for small values of δ, i.e., if $|\delta| \leq k$, and linear if δ is larger than k in absolute terms.

These three loss functions, moreover, are then combined with four regularization-based estimation assumptions, yielding twelve different learning algorithms. The regularization assumptions are: (i) λ equal to zero, i.e., an unpenalized procedure; $\lambda > 0$ and $e = 0$ (a ridge or L_2 penalty); $\lambda > 0$ and $e = 1$ (a lasso or L_1 penalty); and $\lambda > 0$, with $e = 0.5$ (an added elastic net penalty).

1.2.3.3 Multifaceted Bayesian Estimates

The optimization problems that define our estimators can be thought of from a Bayesian perspective. Put simply, the twelve learning algorithms used in our machine intelligence approach can be interpreted as seeking out Bayes posterior modes for model coefficients, (α_j, b_j). Tibshirani (1996), for example, points out that utilizing independent double-exponential priors centered at zero for each element in b_j ($p(b_{j,p}) = \frac{1}{2\tau} exp\left(-\frac{|b_{j,p}|}{\tau}\right)$, $\forall p \in \{1, ..., (S+1)\,K\}$), in conjunction with the assumption that $\epsilon_{j,t}$ is Gaussian, is equivalent to adding a lasso constraint to the residual sum of squares[7].

In essence, using uninformative, Gaussian or double-exponential independent priors for coefficients in b_j - when computing a Bayes posterior mode for model coefficients - is equivalent to setting $\lambda = 0$; $\lambda > 0$ and $e = 0$; and $\lambda > 0$ and $e = 1$ in $\mathcal{L}(\cdot)$, respectively. Also, setting $\lambda > 0$ and $e = 0.5$ in $\mathcal{L}(\cdot)$ is equivalent to using an elastic net prior for b_j[8].

Furthermore, assuming $\epsilon_{j,t}$ follows Gaussian, double-exponential or least-favorable distributions[9] (which resemble Gaussian distributions in their centers, and double-exponential distributions in their tails), correspondingly, is tantamount to utilizing squared loss, absolute loss or Huber loss functions in $\mathcal{L}(\cdot)$[10].

These three error structures (Gaussian, double-exponential/Laplacian, and least-favorable), combined with three different specifications of priors for elements in b_j (uninformative, double-exponential and Gaussian) and with an elastic net prior for b_j[11], yield twelve variants of Bayesian estimation problems that have, as counterparts, the precise regression-type problems we discussed above.

[7]See Murphy (2012) or Trevor Hastie and Friedman (2017) for an introduction to regularization-based techniques. In the discussion in this section, we assume uninformative priors are used for intercept coefficients, and priors for elements of b_j, i.e., slope coefficients, are centered at zero.

[8]The elastic net prior was first introduced by Zou and Hastie (2005). For our version of $\mathcal{L}(\cdot)$ which uses $\lambda > 0$ and $e = 0.5$, the elastic net prior can be written as $p(b_j) \propto exp\left(\gamma_1||b_j||_1 + \gamma_2||b_j||_2^2\right)$, with γ_1 and γ_2 being real-valued constants satisfying $\gamma_1 > 0$ and $\gamma_2 > 0$. $||\cdot||_1$ and $||\cdot||_2$ denote the L_1 and L_2 norms, respectively.

[9]Vide Huber (1964).

[10]An additional discussion on the analogy between regularization and Bayesian estimation is presented in the appendix of this chapter. We refer the reader to it for additional details on this analogy.

[11]See footnote above for its functional form.

1.2.3.4 Optimization

Solutions to optimization problems that use squared loss and different types of regularization (L_1 penalty, L_2 penalty or our elastic net penalty - a mixture of L_1 and L_2 penalties) were obtained using cyclical coordinate descent (Friedman et al. (2010)).

Cyclical coordinate descent, the approach we use in squared loss penalized procedures, is an efficient method for dealing with convex problems that use L_1 or L_2 penalties, or mixtures of these. However, as Yi and Huang (2015) argue, the Huber loss function only exhibits first-order differentiability, and the absolute loss function is not at all differentiable, making cyclical coordinate descent be ill-suited for problems that use regularization and Huber loss or absolute loss. In problems that use regularization and loss functions different than squared loss (i.e. those which involve Huber loss or absolute loss functions), we utilize instead a semismooth Newton coordinate descent approach (Yi (2017)).

The simpler problems, which don't use penalty terms (i.e., $\lambda = 0$), but which still require an optimization procedure ($\ell(\cdot)$ is Huber loss or $\ell(\cdot)$ is absolute loss), were solved using either Koenker (2017) (used when $\ell(\cdot)$ is absolute loss and $\lambda = 0$ - Koenker (2017) provides an implementation of the algorithm for quantile regression found in Koenker and d'Orey (1987)); or iteratively re-weighted least squares (see Venables and Ripley (2002)) - the iteratively re-weighted least squares method was used in procedures that employed Huber loss ($\ell(\cdot)$ is Huber loss) and $\lambda = 0$.

In all cases where a penalized approach was used, we followed Tibshirani (1996) in using five-fold cross-validation to select the penalty tuning parameter utilized in the estimation. Similarly as done when validating the learning algorithm to be used with each stock, observations here were also divided into folds while respecting their natural ordering in time.

As is typical in empirical studies that employ robust regression procedures with the Huber loss function (see e.g., Seo et al. (2016)), when using Huber loss ($\ell(\cdot)$ is Huber loss) and no penalty term ($\lambda = 0$), we select the Huber tuning constant, k, by letting it be equal to $1.345s$, where s represents an estimate for the standard deviation of the error term of the selected stock.

Both s and k were determined simultaneously via the iteratively re-weighted least squares procedure. This iterative procedure essentially consists of picking a value for s, computing $k = 1.345s$, and then training the model (i.e., finding a solution for the coefficients being sought - $\hat{\alpha}_j$, $\hat{\beta}_j$) using such value for k. After the coefficients have been found, a robust estimate for the standard deviation of the error term is computed using residuals implied by these coefficients. This estimate, the new s, is then used to compute a new k. The model is then trained again using the newly computed value for k. This procedure is iterated until the coefficients being sought ($\hat{\alpha}_j$, $\hat{\beta}_j$ here) converge (Fox and Weisberg (2013) contains a detailed explanation of this popular procedure). Robust estimates for standard deviations of error terms were obtained via the rlm function in R, which was also used to implement this iterative procedure[12].

Furthermore, note that when regularization is used ($\lambda > 0$) in conjunction with a Huber loss function in Equation 1.5, two tuning constants need to be selected: the penalty constant λ, and the Huber tuning constant k. We break down the problem of choosing these two tuning parameters into two parts. Firstly, we tackle the problem of choosing the Huber tuning constant k. To choose it, we first find the minimizers $\left(\hat{\alpha}_j, \hat{b}_j\right)$ of the function described by Equation 1.5, $\mathcal{L}(\cdot)$, while setting λ to 0 and letting $\ell(\cdot)$ be Huber loss. In other words, we first use iteratively re-weighted least squares to find standard

[12] The precise R functions used to implement all of the aforementioned optimizations are as follows: cv.glmnet - used to perform five-fold cross validation using cyclical coordinate descent when training models while using regularization and squared loss; predict.cv.glmnet - fits models using cyclical coordinate descent, using the optimal value for λ returned by cv.glmnet, and returns desired coefficients; cv.hqreg - we use it to perform five-fold cross validation using semismooth Newton coordinate descent, when training models with regularization and absolute loss or Huber loss; predict.cv.hqreg - fits models with semismooth Newton coordinate descent, using the optimal value for λ returned by cv.hqreg, and returns desired coefficients; rq - used for training models when no regularization is used and absolute loss is employed; and rlm - we utilize it with the option scale.est = "Huber" to fit models with iteratively re-weighted least squares when no regularization is employed and a Huber loss function is used.

Huber estimates for our parameters. This approach, after complete, yields a value for s. Next, we simply let k be equal to $1.345s$. After k has been chosen, our attention then turns to the problem of selecting λ. While having k fixed at $1.345s$, the penalty tuning parameter λ is afterwards simply chosen via five-fold cross validation, as described above.

In concluding this section, we note that our analysis involves solving several million optimization problems for which there is no analytic solution. To implement it with celerity, we utilize multiple multi-core computers, rented online via Amazon's high performance elastic cloud, and exploit various parallel computing techniques throughout our fitting procedures.

1.2.4 Asset Pricing Tests

After estimating the time series econometric specification (Equation 1.4) using the strategy/steps outlined above, we are ready to implement our asset pricing test. Hereafter, we'll refer to the next steps as the "second-stage" of our model testing procedure. This second-stage involves: (i) using coefficients from the estimated time-series specifications to compute risk-adjusted returns; (ii) regressing computed risk-adjusted returns cross-sectionally on firm characteristics, every month; (iii) and then averaging the loadings on each characteristic across all months[13]. These steps are presented in mathematical terms next.

Following the estimation of (1.4), risk-adjusted returns are estimated as:

$$\tilde{R}_{j,t} = R_{j,t} - \hat{\beta}'_{j,t-1} F_t, \tag{1.7}$$

where $\tilde{R}_{j,t}$ is the risk-adjusted return of stock j at month t, $\hat{\beta}_{j,t-1}$ is the $K \times 1$ vector of factor exposures of stock j, at time $t-1$, estimated for stock j via our machine learning

[13]The steps that compose our second-stage procedure, which are described here, were originally described in BCS. We also average the intercept terms from our cross-sectional regressions across all months.

approach - i.e., we obtain $\hat{\beta}_{j,t-1}$ from the estimated time series specification (1.4); and $R_{j,t}$ and F_t are, as before, the excess return of stock j at month t and a $K \times 1$ vector of risk factor realizations.

After risk-adjusted returns have been computed, a series of cross-sectional regressions are then implemented - one for every month in our sample period[14]. The following cross-sectional model (t is kept fixed) is used:

$$\tilde{R}_{j,t} = z_{0,t} + z_{1,t}^{'} C_{t-1} + \tilde{\epsilon}_t \tag{1.8}$$

The variable $z_{0,t}$ is a scalar; C_{t-1} is a $M \times 1$ vector which contains information on firm characteristics and $z_{1,t}$ is a $M \times 1$ vector of loadings on these characteristics. Estimates for $z_{0,t}$ and $z_{1,t}$ are calculated following BCS and AC via least squares[15], i.e., by regressing $\tilde{R}_{j,t}$ on C_t and an intercept term.

The cross-sectional regression described above is repeated for every month in our sample. For each of the M characteristics, an aggregate value/coefficient is then computed, over all months. Let m be an index tied to a given characteristic included in the cross-sectional regression, and i denote a $T \times 1$ unit vector, where T is the total number of months included in our study. We denote $z_{1,m}$ as the $T \times 1$ vector containing coefficient estimates for characteristic m obtained from all T cross-sectional regressions. The aggregate coefficient for characteristic m is computed as:

$$\hat{z}_{1,m} = \left(i^{'} i\right)^{-1} i^{'} z_{1,m}, \tag{1.9}$$

[14] As we'll see in section 1.3, our sample period goes from Jan. 1965 to June 2017, implying a total of 630 cross-sectional regressions are conducted as part of our model testing strategy.

[15] The extent to which machine learning may aid in this "second-stage" cross-sectional procedure could be the subject of future studies. There are many ways in which machine learning can be applied to asset-pricing tests. In our work, we focus on the estimation of time series specifications/models, which are notoriously difficult to estimate when individual stocks are used.

implying the coefficients $\hat{z}_{1,m}$, $(m = 1, ..., M)$, which we are after are simply computed as their average estimated value over all months. A similar procedure is implemented for the intercept terms, $z_{0,t}$ - i.e., we obtain an aggregate value for the intercept ($\hat{z}_0$) by averaging our intercept estimates over all months. Standard errors for each aggregate coefficient are then obtained using their individual monthly estimates.

As BCS point out, under the null of exact pricing, these aggregate coefficients should not be statistically significant. This approach allows us to determine not only if a model fails, but to also understand which anomalies it can explain, and thus which ones are harder overall to account for.

The asset pricing tests that we implement and the models we test (described in Section 2.1.1) are inspired by those shown in AC, who themselves offer in their paper an adaptation of the work of BCS. Both AC and BCS use individual securities in their tests, as we do. BCS, however, differently from us and from AC, estimate time series specifications using rolling regressions. Betas from these estimated specifications are then used by them to compute risk-adjusted returns (Equation 1.7), and these risk-adjusted returns are then regressed cross-sectionally on firm characteristics (Equation 1.8).

As we mentioned in our introductory comments (Section 1.1), AC have in turn modified the work of BCS by modeling betas as functions of conditioning variables, incorporating time variation in betas into their model specification, instead of using rolling regressions to estimate them. In their work, they estimate Equation 1.4 using a single standard OLS procedure for each stock. Differently from AC, we use an array of machine learning algorithms to estimate Equation 1.4. Our approach, as we have argued in Section 1.2, is supported both by findings displayed in Chapter 2, Section 2.1.2, and by a supplementary study showcased in Appendix A of that chapter.

1.2.5 Contribution

Prior to concluding this section, we note that our work and methodology contributes to the extant literature in a number of ways: (i) we propose a machine-learning-based framework for estimating time series specifications of returns using individual stocks, and for testing asset asset pricing models. Our approach goes beyond the simple use of regularization, incorporating an array of learning algorithms into asset pricing tests, and is corroborated by results shown in separate studies, which are described in this dissertation (see Section 2.1.2 and Appendix A of Chapter 2); (ii) our analysis yields insights into the capabilities of different conditional beta-pricing models (see Section 2.1.3), and into the capabilities of models that use constant betas - we have also included in Chapter 2, for comparison purposes, results attained from OLS-based-tests; (iii) we present evidence that substantiates the idea that utilizing multiple learning algorithms to estimate time series models of stock returns, instead of using a single estimator, is justified; (iv) our findings indicate that machine-learning-based estimators may help in estimating such models; (v) our methodology helps us to confirm that the potential for estimators other than the least squares one to aid in the estimation of time series models of stock returns is larger when many conditioning variables are used to model betas; and (vi) our framework brings machine learning and asset-pricing in general a step closer, while also aiding in shifting the paradigm in asset-pricing studies from focusing on portfolios, to relying on individual securities.

1.3 Description of Data

The data utilized in our analyses is essentially composed, as in AC and BCS, of time series of monthly returns and of characteristics associated with ordinary stocks listed either on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX) or the Nasdaq stock market; and, additionally, of time series of both factor realizations and returns on the one-month Treasury bill. Share returns and volume data were obtained from the CRSP monthly stock returns file, while accounting data stems from Compustat annual tapes; factor realizations and returns on the one-month Treasury are, as customary, from French (n.d.).

Firstly, in line with AC and with various additional studies (e.g. Fama and French (1992) and Kothari et al. (1995)), in an attempt to control for a potential Compustat survival bias, we remove from our data set the initial two years of accounting data for every firm. Subsequently, our sample is constructed using a simple screening procedure.

To be included in our sample, we require each stock to be associated with at least one hundred and eighty valid observations from January 1965 to June 2017, our sample period. Valid observations[16] - those employed in our study - are considered such to the extent that (i) the return on the month when such observation is registered - the current month - and on the preceding twelve months, for the stock associated with it, are available in the CRSP database; (ii) data availability pertaining to the firm tied to such observation is such that we can compute its book-to-market ratio as of December of the year preceding that when the observation is recorded; and (iii) there is enough data in the aforementioned databases to calculate share turnover, volume as a percentage of total shares, and firm size, as measured by market capitalization, for its related stock, two months prior to the present month. Utilization of this screening procedure yielded a total of 3577 stocks, all of which are included in our analyses.

Table 1.1 presents descriptive statistics for all stocks included in our study. Size denotes market capitalization, in billions of dollars, from two months prior to the current

[16] Valid observations are simply those which are "complete", i.e., those for which there is no missing data. As we'll see next, our study requires information on a number of characteristics to be available for each stock.

month[17]. Share turnover is monthly volume, as a percentage of total shares, from two months prior to the current month[18]. Turnover is segmented between that of shares traded on the New York Stock Exchange and American Stock Exchange (NYSE-AMEX Turnover), and that of stocks traded on the Nasdaq stock market (Nasdaq Turnover). BMK represents the book-to-market ratio. Book-to-market is simply computed as the ratio of book value of equity (proxied through shareholder's equity plus deferred taxes, from two calendar years prior to the current year), to last December's market capitalization, except that, as in BCS, Fama and French (1992) and AC, book-to-market ratios lower than the 0.005 or in excess of the 0.995 fractiles were respectively replaced with the 0.005 and 0.995 fractile values[19]. Lastly, the variables RET 2-3, RET 4-6, and RET 7-12 are, as in AC and BCS, correspondingly, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to the present month[20].

Logarithmic transforms of all of the anomaly-related variables cited above - with the exception of those based on past returns - are employed in our study, and all of the aforementioned characteristics are included in our empirical analyses as deviations from

[17]When firm size is used as a conditioning variable, the value of market capitalization that is used is that from two months prior to the current month. Similarly, when used as a characteristic in cross-sectional regressions (Equation 1.8), the value of market capitalization that is used is that from two months prior to the present month. The current/present month, in the context of our estimation procedures, is the one when the excess return (in the case of time series models) is realized, or when risk-adjusted returns (in the case of cross-sectional regressions) are realized.

[18]Similarly as done with firm size, when turnover is used as a characteristic in cross-sectional regressions (Equation 1.8), the turnover value that is used is that from two months prior to the present month. In other words, turnover is lagged two periods relative to response variables.

[19]When book-to-market is used as a conditioning variable, the accounting data that is used is from two calendar years prior to the current year (the year when excess return is realized), and market capitalization is from December of the year prior to the current one. The same is true for cross-sectional regressions (Equation 1.8). When book-to-market is used as a characteristic in cross-sectional regressions, the accounting data that is used is from two calendar years prior to the current year (here the year when risk-adjusted returns are realized), and market capitalization is from December of the year prior to the current one.

[20]In our cross-sectional regressions, Equation 1.8, assuming the response variables - risk-adjusted returns - are realized at month t, we have: RET 2-3 is computed using returns from months $t-2$ and $t-3$; RET 4-6 is computed using returns from months $t-4$, $t-5$ and $t-6$; and, in the same manner, RET 7-12 is calculated using returns from months $t-7$, $t-8$, $t-9$, $t-10$, $t-11$ and $t-12$.

their cross-sectional means. When referring to characteristics as explanatory variables in cross-sectional regressions, or as conditioning variables, throughout the study, we are alluding to their transformed versions.

In addition to the characteristics described above (firm size, book-to-market, the turnover of AMEX/NYSE and of Nasdaq stocks, RET 2-3, RET 4-6 and RET 7-12), the vector of characteristics used in our cross-sectional regressions (1.8) also has a dummy (Nasdaq Dummy) that takes the value of one when a given stock trades on the Nasdaq market, and zero otherwise[21].

Lastly, as we'll see further below, a key conditioning variables used in our work is the business-cycle-related variable default spread - or default premium. This variable measures the yield differential between BAA and AAA rated bonds, and was computed using yields retrieved from the website of the Board of Governors of the Federal Reserve[22].

[21]The Nasdaq Dummy variable is included in our cross-sectional regressions with a two-month lag relative to risk-adjusted returns, the dependent variable in these.

[22]Default spread enters the study with a two-month lag relative to excess returns, the response variable in our time series models (Equation 1.4). To clarify, if we let $R_{j,t}$ denote the excess return of stock j in month t, and use for example default spread as a single conditioning variable - i.e., $\theta_{j,t-1} \equiv Def_{t-1}$, the variable Def_{t-1} will contain in effect the value of default spread registered at month $t-2$.

1.4 Appendix A - Additional Information on Framework

Consider the problem of finding the true set of parameters for a data generating process using an available sample. To approach this problem, a financial statistician would apply a given estimator/learning algorithm to the data sample at hand, trying to learn what those parameters are. This would give him or her a set of parameter estimates - for simplicity purposes, let's denote that set here simply as $\hat{\beta}$.

As financial statisticians, we have to be able to imagine that we could have observed things differently. In that case, a different realization of the data would have occurred, and a different sample would be available instead. Utilization of the same learning algorithm on this alternative sample would likely yield a different set of parameter estimates. One may thus imagine a variety of ways in which the data could realize itself, and the distribution of parameter estimates that would be produced through such approach.

Consider next using the least squares estimator to tackle the current problem, and note that the available data sample is an intrinsic part of that problem. This process yields a set of parameter estimates, $\hat{\beta}$, which may, for the given data sample - the present realization of the data - be far from the true set of parameters being sought.

Let's contemplate a different estimator whose distribution may have higher variance than that of the least squares one. This estimator may be less likely on average - where the average is taken across possible realizations of the data - to find the true parameters being sought, relative to the least squares one. However, for the specific problem being tackled, which again depends on the actual realization of the data, it may produce a set of parameter estimates that is closer to the true set of parameters, relative to the least squares estimator. Similarly, a biased estimator, which on average yields a set of estimates that are different from the true set of parameters being sought, for this particular data sample, i.e. realization of the data, can yield a set of estimates closer to the truth.

The approach we use in this dissertation seeks the algorithm that can yield the best set of parameter estimates for a problem - and once more, note that the current/actual realization of the data is an important part of that problem; the other ones being in our case the conjectured model and a given stock[23]. This is accomplished using trials which resemble similar versions of the original problem - an approach known as cross-validation. This process is repeated in our framework across all stocks, for a given model.

This approach is inspired at heart on what machine learning in general tries to accomplish, which is to find the true set of parameters given a unique realization of the data at one's disposal. Our framework combines this notion with multiple learning algorithms, integrating in a comprehensive manner machine learning into the estimation of time series specifications.

As we alluded to in Section 1.2.2, our approach has an artificial intelligence aspect to it. A machine imbued with artificial intelligence determines what action it should take when facing a given problem, and then takes an action in line with that assessment. Similarly, in our approach, a machine - the computer - determines which action to take when facing a given problem - the action here being the usage of a certain learning algorithm - and then takes an action in line with that assessment.

The Gauss-Markov theorem relates the least squares estimator with other linear estimators. A linear estimator, in the context of the Gauss-Markov theorem, is one that can be expressed as a linear function of the vector of responses (see Flinn (2004)). Non-linear estimators are a different class of estimators. These can have greater potential than the least squares one for getting closer to true model parameters, and should thus be subject to further studies. The cross-validation-based estimation procedure we use functions as a non-linear estimator. Its non-linear feature comes the fact that coefficient estimates produced by it stem from an algorithm that requires using individual candi-

[23]To simplify our discussion, in Section 1.2.2 and in other portions of our work we related an estimation problem with a conjectured model and a given stock, although as we note here the "present realization of the data" is an intrinsic part of it.

date estimators that are non-linear; in addition to the fact that such algorithm employs a decision rule (cross-validation) concerning these; and to the fact that this rule involves MSE computations, which have in them quadratic terms that depend on the data.

Aside from the least squares estimator, the individual candidate estimators we use are non-linear. The ridge estimator[24] in which the penalty tuning parameter, λ, depends on the data (such is our case, as we use cross-validation to determine λ), is non-linear. Similarly, our other estimators that stem from using penalties other than the L_2 penalty in conjunction with squared loss are also non-linear - this assertion holds even if these are in closed form (such as soft thresholding in Lasso with orthogonal design matrix). The remaining candidate estimators employed in our approach, which stem from using absolute loss or Huber loss functions, also cannot be expressed as a linear function of the vector of responses, and are thus similarly non-linear.

Some of the drawbacks of using estimation approaches such as ours, it should not go without saying, are that they are not analytically tractable, can be computationally intensive, and, although we have attempted to provide intuition into our approach here, ours and similar approaches will still likely be less intuitive than approaches that make usage of a single linear estimator.

[24]Which results from the usage of an L_2 penalty and squared loss.

1.5 Appendix B - Analogy Between Regularization and Bayesian Estimation

We show the equivalence between regularization in linear regression and Bayesian estimation, when errors are assumed to follow a Gaussian distribution, and priors for slope coefficients are taken to be normal, centered at zero and independent from each other[25].

Proofs for the additional error densities and priors used in our study are analogous.

Consider the following version of a Gaussian linear model:

$$R_{j,t} = \alpha_j + b_j^{'} X_t + \epsilon_{j,t}$$

$$\epsilon_{j,t} \sim N\left(0, \sigma_j^2\right),$$

with each variable being as described in Section 1.2.

Let K denote here the number of elements in X_t, i.e., K in this proof represents the total number of model features and thus slope coefficients. Next, assume that we believe individual slope coefficients, $b_{j,k}$, are normally and identically distributed, with mean zero; and independent of each other. In order words, assume $p\left(b_{j,k}\right) = N\left(0, \tau_j\right)$, $b_{j,k} \perp b_{j,k'}$, for $k \neq k'$, $k \in \{1...K\}$, $k' \in \{1...K\}$. Take an uninformative prior for the intercept coefficient α_j and define $\theta_j \equiv (\alpha_j, b_j)$.

We want to the find the solution to the following problem:

$$\begin{aligned} max_{\theta_j} p\left(\theta_j | D_j\right) \quad &\iff \quad max_{\theta_j} \frac{p\left(D_j | \theta_j\right) p\left(\theta_j\right)}{p(D_j)} \\ &\iff \quad max_{\theta_j} p\left(D_j | \theta_j\right) p\left(\theta_j\right) \end{aligned}$$

Substituting in our prior for b_j and using knowledge about the distribution of the data,

[25] This proof is presented by Hazlett (2016) and Wu (2017) and is included here for clarifying purposes.

this becomes:

$$
\begin{aligned}
max_{\theta_j} p\left(\theta_j|D_j\right) &\iff max_{\theta_j} \prod_{t=1}^{n_j} N(\alpha_j + b_j^{'}X_t, \sigma_j^2) \prod_{k=1}^{K} N(0, \tau_j) \\
&\iff max_{\theta_j} \sum_{t=1}^{n_j} log\left(N(\alpha_j + b_j^{'}X_t, \sigma_j^2)\right) + \sum_{k=1}^{K} log\left(N(0, \tau_j)\right) \\
&\iff max_{\theta_j} \sum_{t=1}^{n_j} -\frac{\left(R_{j,t} - \alpha_j - b_j^{'}X_t\right)^2}{2\sigma_j^2} + \sum_{k=1}^{K} \frac{-b_{j,k}^2}{2\tau_j^2} \\
&\iff min_{\theta_j} \sum_{t=1}^{n_j} \left(R_{j,t} - \alpha_j - b_j^{'}X_t\right)^2 + \sum_{k=1}^{K} b_{j,k}^2 \left(\frac{\sigma_j^2}{\tau_j^2}\right) \\
&\iff min_{\alpha_j, b_j} \frac{1}{n_j} \sum_{t=1}^{n_j} \left(R_{j,t} - \alpha_j - b_j^{'}X_t\right)^2 + \lambda ||b_j||_2^2,
\end{aligned}
$$

Which is the same problem as the one described by Equations 1.5 and 1.6, when one uses a squared loss function, $\ell(\delta) = \frac{\delta^2}{2}$ and takes $e = 0$, while letting $\lambda = \frac{\sigma_j^2}{n_j \tau_j^2}$. ■

TABLE 1.1: Descriptive Statistics.

	Mean	Median	Std. Dev.
Book-to-Market Ratio (BMK)	0.94	0.74	0.84
Size (US $ Billion)	2.81	0.35	10.57
NYSE-AMEX Turnover (% per month)	8.13	5.98	9.61
Nasdaq Turnover (% per month)	11.17	6.34	18.94
RET 2-3 (%)	2.64	1.44	16.06
RET 4-6 (%)	3.96	2.22	19.88
RET 7-12 (%)	8.03	4.65	29.85

This table displays descriptive statistics for all stocks included in our study. Time series averages - over the 630 months that go from January 1965 to June 2017 - of cross-sectional means, medians and standard deviations are shown. BMK represents the ratio of book value of equity (proxied through shareholder's equity plus deferred taxes, from two calendar years prior to the current year), to last December's market capitalization, except that, as in BCS, Fama and French (1992) and AC, book-to-market ratios lower than the 0.005 or in excess of the 0.995 fractiles were respectively replaced with the 0.005 and 0.995 fractile values. Size denotes the market value of each firm, in billions of dollars, from two months prior to the current month. Share turnover is monthly volume, as a percentage of total shares, from two months prior to the current month. Turnover is segmented between that of NYSE and AMEX shares, and that of Nasdaq-traded stocks. The variables RET 2-3, RET 4-6, and RET 7-12 are, as in AC and BCS, correspondingly, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to the present month.

CHAPTER 2

Machine Learning, Asset Pricing Models and Financial Market Anomalies - Empirical Analysis

2.1 Empirical Analysis

2.1.1 The Sixteen Models being Examined

In this section, we present results engendered by applying our framework to different constant and time-varying beta models. We test a total of sixteen different models in this dissertation using our machine learning approach.

We examine: (i) four models that have the three Fama and French (1993) factors as risk factors; (ii) four models that have the three Fama and French (1993) factors plus a long-term reversals factor (LTREV) as risk factors; (iii) four models that employ the three Fama and French (1993) factors plus momentum (MOM - a winners-minus-losers factor) as risk factors; and (iv) four models that use the five Fama and French (2015) as risk factors.

The four models in (i) differ in how betas are modeled. The same can be said about the models in (ii), (iii) and (iv). For each of the sets of factors we mentioned above (the factors used in (i), (ii), (iii) and (iv)), different specifications for risk loadings are used. In other words, the vector of conditioning variables, $\Theta_{j,t-1}$ can take on four different specifications, for every set of factors.

In particular, we: (a) let betas be constant in time. This means $\Theta_{j,t-1}$ is empty for all stocks[1]. This is arguably the most common formulation used in asset-pricing studies, which normally assume that betas do not vary in time; (b) allow betas to vary linearly with size and book-to-market - in other words, $\Theta_{j,t-1}$, in this case, is taken to be a 2×1 vector containing the characteristics firm size ($Size_{t-1}$) and book-to-market (BMK_{t-1}). Note that the subscript j here is relevant: $\Theta_{j,t-1}$ is different for each stock j here; (c) model betas as linear functions of default spread - this simply says that $\Theta_{j,t-1}$ is taken to be a scalar, containing only information about the conditioning variable default spread (Def_{t-1}); and (d) employ a flexible specification for risk loadings, which scales betas using size, book-to-market, default spread, and interactions between size and default spread and book-to-market and default spread. Put differently, the vector $\Theta_{j,t-1}$ in this case is a 5×1 vector that contains the characteristics firm size ($Size_{t-1}$) and book-to-market (BMK_{t-1}), in addition to information on default spread (Def_{t-1}) and two other conditioning variables formed by interacting the previous two characteristics with default spread ($Size_{t-1}Def_{t-1}$ and $BMK_{t-1}Def_{t-1}$).

The four sets of factors we use, combined with these four different specifications for the vector $\Theta_{j,t-1}$, which describe how betas change in time, yield a total of sixteen models. In the following sections, we present two separate analyses involving these models. Firstly, we show how frequently each of the twelve estimators contemplated in our study are selected as optimal, for the time series specifications (Equation 1.4) of every one of the sixteen models we test. This analysis, as we'll argue below, suggests that the multifaceted estimation approach we use is warranted. No single algorithms emerges as being superior across all stocks and models, and the evidence we find suggests that different learning algorithms may all be able to contribute to estimating time series specifications. Such analysis also sheds light on the performance of least squares relative to the other estimators we use. After presenting it, we turn our attention to the second analysis, of the two we alluded to above: our asset pricing tests.

[1] Alternatively, we can think about $\Theta_{j,t-1}$ here as being simply a scalar equal to zero.

2.1.2 Least Squares Versus Competing Learning Algorithms

Here, we display the frequency with which each of the twelve candidate learning algorithms discussed in Chapter 1 was selected as being optimal, for the time series specifications (Equation 1.4) of all of the sixteen models we investigate in this dissertation.

2.1.2.1 Summary of Findings - Least Squares Versus Competing Learning Algorithms

Our results show some interesting empirical regularities. We'll discuss them in the context of tables which display our findings next, but prior to that, let's first summarize/introduce them here. We find that: (1) the least squares algorithm is chosen as an optimal estimation vehicle for only 14.17% of the stocks in our sample, when the three FF factors are used as risk factors and betas are assumed to be constant. This is the simplest time series econometric specification we use in our tests. It has only three explanatory variables: the three FF factors. This finding implies that our other estimators, which include common estimators used in machine learning (e.g. lasso, ridge estimators) seem to be optimal for over 85% of the stocks in our universe, for this simple model; (2) The percentage of times in which algorithms other than OLS are chosen as optimal is higher than 85% for all other models we use. Findings (1) and (2) suggest that there may be room for learning algorithms other than the least squares one to aid in the estimation of time series models (Equation 1.4) of returns; (3) moreover, the frequency with which OLS is chosen as optimal falls rapidly as we increase the number of conditioning variables being used, regardless of which risk factors are being employed. In other words, the more we let the vector of conditioning variables $\Theta_{j,t-1}$ depend on a larger number of variables, the more algorithms different from OLS emerge as being optimal. This implies, in essence, that there may be greater potential for machine-learning-based estimators to help when many conditioning variables are used to model betas; (4) the evidence suggests that OLS estimation should, preferably, be avoided altogether when a large number of conditioning variables are used in $\Theta_{j,t-1}$. When the five Fama and French (2015) factors are used as risk factors, and betas are modeled as functions of five conditioning variables[2], for example, OLS was never selected as the ideal algorithm; and (5) no single learning algorithm is selected with a very large probability/frequency, regardless of the model being investigated, and the distribution of frequencies tends to be quite dispersed across multiple algorithms. This in turn helps to dispel the notion that using a single estimator for all stocks might be optimal, and suggests instead that a multifaceted approach, such as the one we advocate for, is justified.

2.1.2.2 Results for the Sixteen Models being Examined - Least Squares Versus Competing Learning Algorithms

Let's turn our focus to the specific tables that contain the abovementioned results, and look first at models that use the standard three Fama and French (1993) factors as risk factors. Table 2.1 shows the frequencies with which the twelve learning algorithms we use were chosen as optimal, for all of our four models that use the three Fama and French (1993) factors as risk factors. The three-fold cross-validated mean squared error, computed for the time series model - Equation 1.4 - was used to determine quality of fit. This simply means that for each stock, we evaluated each algorithm using three separate trials, conducted by assigning observations to folds. The algorithm that produced the best fit for the time series specification was chosen as optimal (see Section 1.2.2 for additional details).

All frequencies shown in Table 2.1 are in percentage terms. This table contains four

[2] Size, book-to-market, default spread, and interactions between size and default spread and book-to-market and default spread. In other words, the vector $\Theta_{j,t-1}$ in this case is, as mentioned before, a 5×1 vector that contains the characteristics firm size ($Size_{t-1}$) and book-to-market (BMK_{t-1}), in addition to default spread (Def_{t-1}) and two other conditioning variables formed by interacting the previous two characteristics with default spread ($Size_{t-1}Def_{t-1}$ and $BMK_{t-1}Def_{t-1}$).

columns, each of which corresponds to a different model that uses the three FF factors as risk factors. The models differ according to how betas are modeled. The first column of this table corresponds to a model where betas were taken as constant, and the three FF factors were taken as risk factors.

As we indicated in our preview of the results above, we see from the findings reported in this column that the least squares learning algorithm was chosen as optimal for only 14.17% of the stocks in our sample, when a basic model that has constant betas and the three Fama & French factors was used to model returns. Naturally, this implies that other algorithms seem to be optimal for almost 86% of our stocks, when this model is used.

In the second column of Table 2.1, the conditioning variables used to model betas were firm size and book-to-market. This column therefore corresponds to the specification for betas described in item (b), of Section 2.1.1. In the third column, a single conditioning variable was used to model risk loadings: default spread. This column, in this manner, corresponds to the specification for betas described in item (c), of Section 2.1.1. In the fourth column, betas were modeled according to specification (d) from Section 2.1.1. In other words, they were taken as linear functions of five conditioning variables: size, book-to-market, default spread, and two interaction terms - an interaction term between size and default spread and a second interaction term between book-to-market and default spread.

The second, third and fourth columns of Table 2.1 show that least squares fares worse, in terms of algorithm selection frequency, relative to the case where betas are taken as constant (column 1), when conditioning variables are used to model betas. When a single conditioning variable was used (column 3 - that variable being default spread), for example, least squares was chosen as optimal for only 4.72 % of the stocks in our sample. When five conditioning variables were used (column 4), OLS was selected for a meager fraction - 0.11% - of stocks. We see from this table, furthermore, that the frequency with which OLS is chosen as optimal falls rapidly as the number of number of conditioning variables - and thereby explanatory variables - increases.

Put simply, the findings shown on this table suggest that there may be room for learning algorithms different from OLS to help in estimating time series models of stock returns. They further indicate that the potential for these algorithms - i.e., those different from OLS - to collectively aid in the estimation of such models is greater when a large number of conditioning variables are used.

Lastly, we also see here that no individual learning algorithm was overwhelmingly chosen as optimal, irrespective of the model being examined. The distribution of frequencies is rather dispersed across different algorithms. This finding buttresses the notion that using a multifaceted approach - with multiple algorithms - is warranted, and helps to dispel the idea that one should use a single algorithm to estimate time series models for all stocks.

Tables 2.2, 2.3 and 2.4 are similar to Table 2.1, except that different sets of factors were used when constructing each of these tables. In Table 2.2, the risk factors are the three Fama and French (1993) factors, plus a long-term reversals factor; in Table 2.3, the risk factors are the three Fama and French (1993) factors, plus a momentum factor; and in Table 2.4, the risk factors are the five Fama and French (2015) factors.

The results depicted in these tables are in line with those reported in Table 2.1, and therefore the conclusions that can be drawn from Tables 2.2, 2.3 and 2.4 are analogous to those we drew from Table 2.1. In other words, the conclusions we drew from Table 2.1 are robust to using the other sets of factors we cited above.

Take, for example, the four models investigated in Table 2.4. When the five Fama and French (2015) factors are used as risk factors, and betas are presumed to be constant in time (column 1 of Table 2.4), for instance, least squares was selected for only 10.37% of all stocks. When five conditioning variables were used to model betas, and the same factors were used, least squares was never chosen as optimal; this table also shows that the frequency with which OLS is selected as an estimation vehicle decreases with celerity as the number of conditioning variables rises; and the distribution of frequencies is dispersed across different algorithms.

2.1.3 Pricing Abilities of Conditional Beta Models and of Models with Constant Betas

What does our framework say about the pricing abilities of models that admit time-varying betas, and those of models which assume that risk loadings are constant in time? Our results for asset pricing tests involving the sixteen models discussed in the introductory paragraphs of this chapter are presented here.

As explained in Section 1.2.2, to conduct our asset pricing tests, we initially conjecture that stock returns can be explained by a linear multi-factor asset pricing model, described by Equations 1.1, 1.2 and 1.3. These three equations are then combined, yielding the time series model/specification shown in Equation 1.4. We estimate the derived time series model for different stocks via our machine learning approach. Subsequently, betas from the estimated time series models are used as inputs in the computation of risk-adjusted returns - see Equation 1.7. These risk-adjusted returns are then regressed cross-sectionally on a number of characteristics, every month from January 1965 to June 2017, our sample period. Our tests, we note, are described in detail in Section 1.2.4.

The characteristics used as explanatory variables in our cross-sectional regressions (Equation 1.8)) are firm size; the book-to-market ratio; a turnover variable that measures the turnover of NYSE-AMEX stocks; a turnover variable for Nasdaq stocks; a Nasdaq dummy equal to one if the stock in question is traded on the Nasdaq exchange, and zero otherwise; and our three past return variables, RET 2-3, RET 4-6 and RET 7-12. A detailed description of how each of these variables is computed is shown in Section 1.3.

The results from our tests have been summarized in four tables, which display the average coefficients on the characteristics cited above, produced by our cross-sectional regressions, over all 630 months from January 1965 to June 2017 - absolute values of t statistics are shown in parenthesis. These tables (Table 2.5, 2.6, 2.7 and 2.8) all have a similar structure. Each of them corresponds to one of the four sets of factors used in this study. Table 2.5, for example, contains results for tests involving models that have the three Fama and French (1993) factors as risk-factors.

The tables are divided into four columns, and each column refers to a different model being tested. Models shown in each table differ according to how betas are modeled. In the first column, betas were taken as being constant - this specification thus corresponds to item (a) in Section 2.1.1. In the second column, betas were presumed to vary with size and book-to-market - specification (b) in Section 2.1.1; in the third and forth columns, risk loadings were modeled following specifications (c) and (d), respectively, from Section 2.1.1 - in other words, in the third column default spread is used as a conditioning variable, and in the forth column betas depend on many conditioning variables: size, book-to-market, default spread, an interaction term between size and default spread and another one between book-to-market and default spread.

2.1.3.1 Summary of Findings - Asset Pricing Tests

As we did when discussing the findings from our first analysis, which touches on how least squares fares relative to other learning algorithms, we'll first present a brief summary of our findings, and then subsequently proceed to discuss them in the context of tables which present all specific results. These can be summarized as follows: (1) AC's work imply that conditional betas may offer a potential explanation for size and value anomalies. We don't find in our work evidence corroborating this assertion. In particular, we see that size and value anomalies seem to be very difficult to explain. None of the sixteen models we use is able to capture the effect of size or book-to-market (the value anomaly) in the cross-section of stock returns. Put differently, none of the models employed throughout our tests yielded non-significant coefficients on size or book-to-market variables - these were all statistically significant (absolute values of t-statistics above 1.96); (2) Past return anomalies are also exceedingly difficult to explain, and the models examined in this dissertation were also unable to account for the impact of past return variables in the cross-section of returns; (3) a model that uses the three FF factors plus a long-term reversals factor, and employs a flexible formulation for betas, allowing them to vary conditionally upon multiple variables (specification (d) for betas in Section 2.1.1) can capture turnover effects, when a five percent level of significance is used (it produces t-statistics that in absolute terms lie below the threshold value of approximately 1.96, for both turnover variables). Other models that use the same factors (HML, SMB, the excess return on the market, and LTREV) are not able to accomplish this same feat; (4) all four models that have the three FF factors, plus momentum, as risk factors can also capture turnover effects at that level of significance - i.e., coefficients for both turnover variables are not significant at that level; (5) Despite the fact that a number of models we have examined were able to explicate turnover anomalies, only two models yielded t-statistics for turnover variables that were below 1.65 in absolute terms, and were thus able to capture turnover effects "more convincingly". These were both models that used the five Fama and French (2015) as risk factors, and had conditional betas. One of these is the model that uses such factors, and permits betas to change with firm size and book-to-market (specification (b) in Section 2.1.1); and the other is the model that uses these factors while allowing betas to change with these same characteristics, default spread, and interactions between these characteristics and default spread (specification (d) from Section 2.1.1). Notice that, for the purpose of capturing effects from characteristics in the cross section, lower absolute values for t-statistics are desirable.

Next, we turn our attention to the specific tables that report these findings.

2.1.3.2 Four Models that have the Three Fama and French (1993) Factors as Risk Factors

Table 2.5 presents average coefficients from cross-sectional regressions (Equation 1.8) of risk-adjusted returns - the dependent variables - on firm characteristics. As we have mentioned above, risk-adjusted returns were computed using Equation 1.7, with betas obtained from time series specifications estimated through our machine-learning framework, and each of the columns shown here describes results for a different model. All of the four models used to construct this table use the three Fama and French (1993) factors as risk factors, and models differ according to how betas are modeled.

We see from the first column of Table 2.5 that a simple model that uses the three Fama & French factors as risk factors, and assumes betas are constant, is not able to explain size and value anomalies - the coefficients on size and book-to-market variables are both statistically significant. Allowing betas to vary conditionally upon size and book-to-market, the very same variables tied to size and value anomalies in the cross-section of returns, is also not helpful in capturing these effects in the cross-section (see column 2). In column 3, betas were allowed to change with default spread - they were modeled according to specification (c) from Section 2.1.1. This table shows that using default spread as a conditioning variable is also unhelpful in explaining size and value anomalies. Lastly, on the last column of Table 2.5, betas were modeled using our most flexible specification (see specification (d) from Section 2.1.1) - they were permitted to change with size, book-to-market, default spread, and with interactions between the two former characteristics and the latter macro-related variable. When our machine-learning framework is used to estimate time-series models/specifications, this flexible formulation is also unable to capture the impact of size and value in the cross-section of returns.

In other words, our framework suggests that the three FF factors are unable to explain these well-known effects, even when loadings are allowed to vary with different conditioning variables. As we'll see below, these findings are robust to replacing the 3 FF factors with the 3 FF factors and either a long-term reversals factor or a momentum factor. And they are also robust to utilizing the five Fama and French (2015) factors in lieu of the 3 FF factors.

Our framework also indicates that past return variables are, indeed, exceedingly difficult to explain. None of the four models investigated in table 2.5 produced a non-significant coefficient for past return variables employed in our study. The additional twelve models we examine - better discussed below - were also not capable of producing non-significant coefficients for these variables.

On an remaining note, moreover, we see from Table 2.5 that coefficients on both turnover variables are significant for three of the four models that use the three FF factors as risk factors (see columns 1, 2 and 3 of Table 2.5). Although one of the models that uses the three FF factors as risk factors (see column 4 of Table 2.5), furthermore, delivers a t-statistic with absolute value below 1.96 - the approximate threshold for a five percent significance level - for the Nasdaq turnover variable, that model still produces a significant coefficient for the NYSE/AMEX turnover variable. These results point towards the inability of models built around the three FF factors to also explicate turnover effects in the cross-section of returns.

2.1.3.3 Four Models that have the Three Fama and French (1993) Factors plus a Long-Term Reversal Factor (LTREV) as Risk Factors

Carhart (1997) is well-known for adding a factor based on the momentum effect - documented in Jegadeesh and Titman (1993) - to the popular set of factors composed by SMB, HML and the excess return on the market. Indeed, his work invariably lead to a plethora of papers which made use of similar four-factor models. As Swedroe (2017) argues, interestingly, another effect associated with past returns - and thereby with momentum - is also evidenced in Jegadeesh and Titman (1993): long-term reversals in stock prices.

Here, we assess whether a factor based on this effect is able to improve the pricing power of the Fama & French three factor model, when betas are presumed to be either constant or time-varying, and time series models of stock returns are estimated through our machine learning approach. Table 2.6 displays our findings.

Interestingly, Table 2.6 indicates that a conditional model that utilizes the three FF factors and a long-term reversals factor as risk factors is capable of capturing turnover effects, insofar as a five percent level of significance is employed - t-statistic approximately equal to 1.96 - see column 4. When the aforementioned factors are used in conjunction with the assumption that betas change with multiple variables - specification (d) for betas, from Section 2.1.1 - the coefficients on both turnover variables lie below the 1.96 threshold. Columns 1 to three, which display results for models that use the same factors, but employ simpler specifications for betas - indicate that the other models which we investigate, and that use excess market return, SMB, HML and LTREV as risk factors , are unable to deliver similar pricing results. As stated earlier, moreover, one can see here that none of the models which employ these factors can account for either size, value, or past return anomalies in the cross-section.

2.1.3.4 Four Models that have the Three Fama and French (1993) Factors plus Momentum as Risk Factors

To the extent that our estimation framework is utilized, can adding a momentum factor to the Fama & French three factor model - and to variants of it which admit time-varying risk-loadings - help explain the impact of characteristics on the cross-section of returns? The results in Table 2.7 suggest that utilizing a momentum factor together with the three Fama & French factors may be of help in explaining turnover anomalies. Shown in this table are average coefficients from cross-sectional regressions that have as response variables risk-adjusted returns computed when the three FF factors plus a momentum - winners-minus-losers - factor are used to model returns - and various specifications for betas are used.

This table shows that the coefficient on the NYSE-AMEX Turnover variable is not statistically significant, at a five percent level of significance, for all four models that use SMB, HML, excess market return and momentum as factors. The coefficient on the Nasdaq Turnover variable, furthermore, is also not statistically significant at such level, for all models that make usage of a momentum factor in conjunction with the three FF factors. In short, all four models that use the three FF factors plus momentum as risk factors, which we examine, can account for turnover effects in the cross-section at a five percent level. It goes without saying, results for the models examined in this table also indicate that they cannot explicate value, size, and past return anomalies in the cross-section of individual stock returns.

Additional comments on turnover effects are presented next, in our discussion of models that use the 5 Fama and French (2015) factors as risk factors.

2.1.3.5 Four Models that have the Five Fama and French (2015) Factors as Risk Factors

Fama and French (2015) introduce profitability-related and investment-related factors to their three factor model. Table 2.8 presents results obtained when models that use the 5 Fama and French (2015) factors, while admitting different specifications for betas (see Section 2.1.1), are tested via our statistical-learning framework. It is important to note that two of the conditional models that use the 5 Fama and French (2015) factors as risk factors appear to be better able to capture turnover effects in the cross-section of returns, relative to other models investigated in this dissertation.

Although many of the models examined in our work are apt at explicating turnover anomalies at a five percent significance level (i.e., they can deliver t-statistics for turnover variables whose absolute values are below the threshold value of 1.96), they are unsuccessful in that regard to the extent that one uses a lower critical value of 1.65 for t-statistics for determining significance. Two of the models we examine, however, are able to explain turnover effects in a more "convincing manner", delivering t-statistics for turnover coefficients whose absolute values are below such threshold. These two models both use the five Fama and French (2015) factors as risk-factors, and have betas that depend on conditioning information.

In the second column of Table 2.8, betas were allowed to change linearly with firm-size and book-to-market (specification (b) from Section 2.1.1). In the fourth column of Table 2.8, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables scaling betas (specification (d) from Section 2.1.1)). In both of these columns, we see that t-statistics tied to NYSE-AMEX Turnover and Nasdaq Turnover coefficients are such that we cannot conclude that turnover effects are present in the data when a critical t-statistic of 1.65 is used.

2.2 Final Remarks

To this date, over ninety characteristics have been reported as having the capacity to predict cross-sectional stock returns (McLean and Pontiff (2015)), and more than three hundred factors have been suggested and found to be significant (Harvey et al. (2015)) in financial studies. In this dissertation, we have argued that the cross-sectional testing framework used by AC - which is itself inspired by the work of BCS - can be a powerful tool for investigating these findings, as it permits one to test models and acquire a granular understanding of multiple anomalies, all while admitting the possibility that loadings may change with conditioning information. The methodology we have proposed is motivated by and expands the work of AC and BCS, integrating into asset pricing tests a machine-learning-based estimation procedure that can help shed new light onto financial market anomalies.

In essence, one of the most well-known concepts in the empirical asset-pricing literature - advocated by Roll (1977), and later ratified by numerous studies, such as Pukthuanthong et al. (2017) - is arguably the notion that asset-pricing tests should, preferably, focus on making usage of securities, as opposed to utilizing portfolios. The framework proposed in this dissertation incorporates, in a simultaneous fashion, a spectrum of statistical learning techniques into tests that employ individual stocks. Our framework contributes, in this regard, to further foster analyses that take into consideration comments found in Roll (1977), which indicate that usage of portfolios in asset pricing tests may be problematic, and arguments of similar caliber, made following the publication of the latter study (see e.g. Lo and MacKinlay (1990)).

In summary, the machine-learning-driven methodology for estimating time series specifications of returns, and for testing asset pricing models, which we have proposed, yielded two separate analyses, presented throughout this chapter. In the first of these, we compared least squares with other learning algorithms, assessing how frequently OLS was chosen as an optimal estimation vehicle, relative to other estimators, across different time series specifications. There, we showed that least squares was selected as optimal for a relatively small percentage of stocks in our sample (14.17%), when the three Fama and French (1993) factors were used to model returns and betas were presumed to be constant. The percentage of stocks for which OLS is chosen is even smaller for other models, reaching zero or values close to zero for models in which betas depend on many conditioning variables. Collectively, these results suggest that learning algorithms different from the least squares one may offer help in estimating time series models of returns (Equation 1.4).

Our findings, on an additional note, also contribute to dispelling the notion that a single learning algorithm (such as lasso or elastic net, for example) should be used to estimate time series specifications of returns: simply put, the distribution of frequencies with which estimators are selected as optimal appears to be quite dispersed across multiple algorithms, regardless of the model being investigated.

Lastly, we verify that the frequency with which least squares is selected as optimal falls very rapidly as the number of conditioning variables used to model betas, and thus explanatory variables, increases. This suggests that OLS overfitting may be a concrete point of concern in the estimation of time series specifications which use conditional betas. In Appendix C, we present a separate study that corroborates this conclusion.

In our second analysis, we showed results from asset pricing tests that make usage of our estimation framework, for sixteen different models. While AC's work suggests that conditional betas could help to explain the well-known value and size anomalies, we have not found evidence in support of that notion. None of the conditional models - and also none of the non-conditional ones - we examined could explain either value or size effects in the cross-section of returns. Essentially, a number of effects tied to anomaly-related variables (firm-level characteristics) are still left unaccounted for, even when risk loadings are allowed to vary with both characteristics and default spread. Such finding ultimately lends a degree of plausibility to the hypothesis that some anomaly-related effects may potentially be associated with mispricing, as opposed to risk. Indeed, Fleckenstein et al. (2014) uncovers strong evidence suggesting that arbitrage opportunities - and thus mispricing - may be ubiquitous in financial markets.

Along these lines, we find that past return anomalies seem to indeed be hard to explain from a risk-based perspective - none of the sixteen models we assessed was either able to account for these. And while different models were capable of capturing turnover/liquidity effects in the cross section of returns, only two models did so in a “convincing” fashion. Both of these had the five Fama and French (2015) as risk factors. In one of them, risk loadings were presumed to vary with firm size and book-to-market; and in the second one, loadings were allowed to change according to a more complex formulation - they varied with these same characteristics, default spread, and interaction terms between the latter variable and the former characteristics.

In this dissertation, in conclusion, we have proposed a multi-algorithmic framework that imbues machines with an ample array of intelligences. We find evidence supporting the usage of such approach, displaying it in Section 2.1.2, and additional evidence further corroborating its usage is also presented in a separate analysis at the end of this chapter - see Appendix A, below.

The estimation strategy presented here further integrates statistical learning concepts into asset-pricing tests involving individual stocks, having an artificial intelligence "flavor" to it, being particularly useful for investigating models that condition betas on a large number of variables. In this regard, we hope that some of the concepts explored herein may take an added role in future studies which focus on risk management, the study of anomalies, the creation of optimal investment policies, and the general study of expected returns.

2.3 Appendix A - Average Ranks of Learning Algorithms

Here, we present an analysis that complements the first of our studies, shown in Section 2.1.2, where we assessed how often OLS and machine-learning-based algorithms are chosen as optimal - across sixteen alternative time series specifications of returns - when the estimation framework we suggest is utilized.

Table 2.9 displays average ranks for the twelve learning algorithms we use throughout our empirical analyses. The columns in this table each depicts average ranks calculated for a different model. Average ranks can take values going from one to twelve, and they represent the average quality of fit attained by a given learning algorithm, for the time series specification of the model in question. For each problem (model and stock), we rank all twelve learning algorithms from one to twelve, according to quality of fit.The algorithm that achieves the best fit for a given problem is assigned to it a value of one, and the one with the worst fit is assigned to it a value of twelve. An average rank was then computed for each learning algorithm utilizing the ranks they received across all stocks, for a given model. Similarly as done in the study presented in Section 2.1.2, the three-fold cross-validated mean squared error associated with the time series specification (Equation 1.4) for the model being considered was used to determine quality of fit.

The four models being examined in Table 2.9 all use the three Fama and French (1993) factors as risk factors. They differ according to how betas are modeled. In the first column, betas were taken as constant (i.e., they were modeled according to specification (a) from Section 2.1.1). In the second column, betas vary linearly with firm-size and book-to-market (i.e., they were modeled following specification (b) from Section 2.1.1). In the third column, default spread was used as the only conditioning variable (betas were here thus modeled using specification (c) from Section 2.1.1). In the fourth column, betas were allowed to change with five conditioning variables: book-to-market, size, default spread, and interactions between the former characteristics and the latter variable (put differently, they were modeled using specification (d) from Section 2.1.1).

Interestingly, the least squares algorithm achieves a meager 6.13 average rank when betas are taken as constant and the three Fama and French (1993) factors are used as risk factors (see column 1 of Table 2.9). When a single conditioning variable is used to model betas (column 3), OLS fares even worse and its average rank falls abruptly to 9.27. Moreover, the performance of the OLS estimator falls even further as more conditioning variables are used to model betas. When five conditioning variables are employed (column 4), OLS attains an abysmal performance, displaying an average rank of 11.10.

Tables 2.10, 2.11 and 2.12 are analogous to Table 2.9, except that different risk factors were used in these. In Table 2.10, the three Fama and French (1993) factors plus a long-term reversals factor were used as risk factors; in Table 2.11, the three Fama and French (1993) factors plus a momentum factor were used as risk factors; and in Table 2.12, the five Fama and French (2015) factors were utilized as risk factors. Similar findings to the ones we alluded to above are presented in these tables. We see that least squares received an average rank even worse than 6.13 - the one it attained for our simplest time series specification - for all models examined in these tables. In all models where five conditioning variables were used, the OLS estimator exhibits a terrible performance, achieving average ranks worse than 11 in all such models. This finding further corroborates the notion that OLS estimation should be avoided when many conditioning variables are used to model betas - a conclusion we drew in Section 2.1.2, in light of results shown in our frequency tables.

Importantly, furthermore, none of the learning algorithms is able to attain average ranks very close to one for the models examined in this dissertation. This finding helps to further dispel the notion that a single learning algorithm should be used to estimate time series specifications of returns, and again - in line with findings from our frequency analyses - buttresses the idea that a multi-algorithmic estimation framework, such as the one we suggest, is warranted.

2.4 Appendix B - OLS Benchmarks for Asset Pricing Tests

For comparison purposes, we display in this section results from asset pricing tests that are obtained when least squares is used to estimate time series specifications of returns (Equation 1.4), instead of our machine-learning-based framework.

Tables 2.13, 2.14, 2.15 and 2.16 showcase results from these benchmark tests. These four tables are similar to Tables 2.5, 2.6, 2.7 and 2.8, respectively, with the exception that to construct them, time series specifications were estimated using the least squares algorithm, and risk-adjusted returns (Equation 1.7) were computed using betas from these OLS-estimated specifications. In Table 2.13, the risk factors are the three Fama and French (1993) factors; in Table 2.14, they are the three Fama and French (1993) factors, plus a long-term reversals factor; Table 2.15 uses the three Fama and French (1993) factors plus a momentum factor as risk factors; in Table 2.16, the risk factors are the five Fama and French (2015) factors. Each of these tables examines four models, and these differ in how betas are modeled. Betas were themselves modeled according to four different specifications, described in Section 2.1.1. These tables are included here for reference purposes.

There are varied differences between results from asset pricing tests obtained when our framework is used vis-à-vis least squares. Some of the coefficients obtained in our asset pricing tests are not significant when our framework is used to estimate time series specifications, and significant when least squares is used instead. For example, turnover coefficients for the model that uses the three Fama and French (1993) factors plus a long-term reversals factor as risk factors, and which allows betas to vary with five conditioning variables, are not significant when our framework is used, and significant when least squares is used instead. Conversely, there are also coefficients produced through our tests that are significant when our framework is used, and not significant when the least squares algorithm replaces it. For instance, coefficients that assess the value effect across the four flexible models which use five conditioning variables are significant when our estimation framework is used, and not significant when least squares is used to estimate time series specifications.

We note that varied differences in results from asset pricing tests that use these two estimation approaches are to be expected. In essence, if OLS is doing a poor job in estimating time series specifications, as Appendix A in this chapter suggests, and our estimation framework attains its intent of obtaining better approximations for time series specifications, then coefficient estimates from individual cross-sectional regressions (Equation 1.8) can be different across these two approaches (as these cross-sectional regressions use as "inputs" the estimated time series specifications), leading to differences in reported average coefficients and inferences.

2.5 Appendix C - Least Squares Overfitting - Supplementary Analysis

In this section, we present a supplementary analysis which buttresses the notion that OLS overfitting is a concrete point of concern when one estimates time series specifications of returns which use conditional betas.

For every month from January 1965 to June 2017 (630 months in total), we implement a cross-sectional regression of risk-adjusted returns on firm characteristics. Table 2.17 displays the average R^2 values, over all 630 months, from these cross-sectional regressions. This table possesses four columns, and each of these columns refers to a separate asset pricing model. While all of these four models employ the three Fama and French (1993) factors as risk factors, the four models differ depending on how betas are being specified. In the first column, betas were presumed to be constant in time (specification (a) for betas from Section 2.1.1). In the second column, betas were modeled as linear functions of two variables: size and book-to-market (specification (b) from Section 2.1.1). In the third column, betas were modeled using one conditioning variable: default spread (specification (c) from Section 2.1.1). In the fourth column, betas were modeled using five conditioning variables: size, book-to-market, default spread, an interaction term between size and default spread and a second interaction term between book-to-market and default spread (specification (d) from Section 2.1.1). In the first row of Table 2.17, least squares was utilized in the estimation of time series specifications (Equation 1.4); and in the second row of the same table, our machine-learning-based approach, which controls for overfitting, was used to estimate time series specifications instead. Betas from the estimated time series specifications were subsequently used to obtain risk adjusted returns, following the formula described by Equation 1.7, and these were then regressed every month on the same firm characteristics used in our asset pricing tests (see Section 1.3).

Table 2.17 shows that when least squares is used to estimate time series specifications, there is a decreasing relationship between average R^2 values from cross-sectional regressions, and the number of conditioning variables used to model betas. A possible explanation for this phenomenon is that least squares may increasingly overfit time series specifications as the number of conditioning variables, and thus explanatory variables, increases. Importantly, we find that the decreasing pattern is no longer present when our machine learning approach, which controls for overfitting, is used in the estimation.

These results suggest that the decreasing pattern observed in the first row of this table may indeed have been caused by the least squares algorithm overfitting the time series specifications at hand. They imply, consequently, that OLS may "truly" overfit time series specifications that use conditional betas: OLS overfitting in such models is, in this manner, not simply "a theoretical matter".

Another reason why this analysis is relevant lies in the fact that AC use, in their work, average R^2 values from cross-sectional regressions of risk-adjusted returns, which result from OLS estimation of time series specifications, on firm characteristics, to aid - as an informal measure - in assessing the pricing ability of different models. Our findings indicate, as we have mentioned, that this measure may be affected by the least squares algorithm overfitting time series specifications.

Table 2.1: Least squares versus competing learning algorithms, with excess market return, SMB and HML used as risk factors.

	Constant	BMK + Size	Def	(BMK + Size) ×Def
Squared Loss (OLS)	14.29	3.19	4.70	0.11
Squared Loss & Ridge Penalty	10.15	13.22	16.16	15.32
Squared Loss & Lasso Penalty	6.18	10.65	6.93	9.78
Squared Loss & Elastic Net Penalty	4.89	5.93	4.86	8.64
Absolute Loss	12.55	3.35	4.50	0.08
Absolute Loss & Ridge Penalty	9.25	10.20	12.33	10.74
Absolute Loss & Lasso Penalty	5.12	8.67	7.30	10.46
Absolute Loss & Elastic Net Penalty	3.49	6.26	5.81	7.58
Huber Loss	13.87	3.58	5.37	0.22
Huber Loss & Ridge Penalty	11.38	15.74	15.49	16.47
Huber Loss & Lasso Penalty	4.39	11.27	8.92	11.77
Huber Loss & Elastic Net Penalty	4.45	7.94	7.63	8.83

This table shows the frequencies with which the twelve learning algorithms we employ in this dissertation (described in Section 1.2.3) were chosen as optimal, when our estimation framework is used (see Section 1.2.2). Each of the four columns shown depicts frequencies computed for a different model. Frequencies in each column add up to 100. They are in percentage terms and correspond to the proportion of stocks in our sample for which algorithms were deemed optimal for the time series specification of the model in question (Equation 1.4). The four models being examined here all use the three Fama and French (1993) factors as risk factors. They differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. For each problem (model and stock), we evaluate each algorithm using three separate trials, conducted by assigning observations to folds. The algorithm that produced the best fit for the time series specification (Equation 1.4) was chosen as optimal. The three-fold cross-validated mean squared error was used to determine quality of fit. Section 1.2.2 contains additional details on this procedure.

Table 2.2: Least squares versus competing learning algorithms, with excess market return, SMB, HML and LTREV used as risk factors.

	Constant	BMK + Size	Def	(BMK + Size) ×Def
Squared Loss (OLS)	11.24	1.73	2.94	0.03
Squared Loss & Ridge Penalty	8.53	13.08	13.89	14.23
Squared Loss & Lasso Penalty	8.16	11.21	8.58	11.10
Squared Loss & Elastic Net Penalty	5.59	5.62	5.45	8.83
Absolute Loss	11.43	1.54	2.96	0.03
Absolute Loss & Ridge Penalty	9.14	9.62	12.08	9.31
Absolute Loss & Lasso Penalty	6.35	9.34	7.94	10.04
Absolute Loss & Elastic Net Penalty	4.78	7.55	7.02	8.05
Huber Loss	12.52	2.77	4.11	0.06
Huber Loss & Ridge Penalty	10.48	15.94	17.11	15.10
Huber Loss & Lasso Penalty	6.49	12.78	9.81	12.94
Huber Loss & Elastic Net Penalty	5.28	8.83	8.11	10.29

This table shows the frequencies with which the twelve learning algorithms we employ in this dissertation (described in Section 1.2.3) were chosen as optimal, when our estimation framework is used (see Section 1.2.2). Each of the four columns shown depicts frequencies computed for a different model. Frequencies in each column add up to 100. They are in percentage terms and correspond to the proportion of stocks in our sample for which algorithms were deemed optimal for the time series specification of the model in question (Equation 1.4). The four models being examined here all use the three Fama and French (1993) factors plus a long-term reversals factor as risk factors. They differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. For each problem (model and stock), we evaluate each algorithm using three separate trials, conducted by assigning observations to folds. The algorithm that produced the best fit for the time series specification (Equation 1.4) was chosen as optimal. The three-fold cross-validated mean squared error was used to determine quality of fit. Section 1.2.2 contains additional details on this procedure.

TABLE 2.3: Least squares versus competing learning algorithms, with excess market return, SMB, HML and MOM used as risk factors.

	Constant	BMK + Size	Def	(BMK + Size) ×Def
Squared Loss (OLS)	11.35	1.57	2.12	0.00
Squared Loss & Ridge Penalty	9.20	11.77	13.89	14.71
Squared Loss & Lasso Penalty	7.27	11.13	8.67	11.27
Squared Loss & Elastic Net Penalty	4.56	6.54	6.12	9.20
Absolute Loss	10.29	1.85	2.35	0.06
Absolute Loss & Ridge Penalty	9.84	10.51	11.83	8.55
Absolute Loss & Lasso Penalty	6.32	8.97	7.80	8.97
Absolute Loss & Elastic Net Penalty	4.58	7.44	7.58	9.25
Huber Loss	12.86	2.52	3.97	0.03
Huber Loss & Ridge Penalty	11.49	15.46	16.10	13.67
Huber Loss & Lasso Penalty	7.30	12.27	10.51	13.47
Huber Loss & Elastic Net Penalty	4.95	9.98	9.06	10.82

This table shows the frequencies with which the twelve learning algorithms we employ in this dissertation (described in Section 1.2.3) were chosen as optimal, when our estimation framework is used (see Section 1.2.2). Each of the four columns shown depicts frequencies computed for a different model. Frequencies in each column add up to 100. They are in percentage terms and correspond to the proportion of stocks in our sample for which algorithms were deemed optimal for the time series the model in question (Equation 1.4). The four models being examined here all use the three Fama and French (1993) factors plus a momentum factor as risk factors. They differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. For each problem (model and stock), we evaluate each algorithm using three separate trials, conducted by assigning observations to folds. The algorithm that produced the best fit for the time series specification (Equation 1.4) was chosen as optimal. The three-fold cross-validated mean squared error was used to determine quality of fit. Section 1.2.2 contains additional details on this procedure.

Table 2.4: Least squares versus competing learning algorithms, with excess market return, SMB, HML, CMA and RMW used as risk factors.

	Constant	BMK + Size	Def	(BMK + Size) $\times$Def
Squared Loss (OLS)	10.40	0.78	2.24	0.00
Squared Loss & Ridge Penalty	8.75	11.80	14.12	14.40
Squared Loss & Lasso Penalty	8.53	11.99	8.72	11.74
Squared Loss & Elastic Net Penalty	5.70	6.96	6.96	10.01
Absolute Loss	9.87	0.89	2.52	0.00
Absolute Loss & Ridge Penalty	7.94	9.45	12.13	7.24
Absolute Loss & Lasso Penalty	7.41	10.34	7.88	9.70
Absolute Loss & Elastic Net Penalty	5.26	7.80	6.88	9.48
Huber Loss	11.29	1.59	3.05	0.03
Huber Loss & Ridge Penalty	10.76	15.01	16.69	14.57
Huber Loss & Lasso Penalty	8.75	13.98	10.12	12.69
Huber Loss & Elastic Net Penalty	5.34	9.39	8.69	10.15

This table shows the frequencies with which the twelve learning algorithms we employ in this dissertation (described in Section 1.2.3) were chosen as optimal, when our estimation framework is used (see Section 1.2.2). Each of the four columns shown depicts frequencies computed for a different model. Frequencies in each column add up to 100. They are in percentage terms and correspond to the proportion of stocks in our sample for which algorithms were deemed optimal for the time series specification of the model in question (Equation 1.4). The four models being examined here all use the five Fama and French (2015) factors as risk factors. They differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. For each problem (model and stock), we evaluate each algorithm using three separate trials, conducted by assigning observations to folds. The algorithm that produced the best fit for the time series specification (Equation 1.4) was chosen as optimal. The three-fold cross-validated mean squared error was used to determine quality of fit. Section 1.2.2 contains additional details on this procedure.

Table 2.5: Average coefficients from cross-sectional regressions, with excess market return, SMB and HML used as risk factors, and time series econometric specifications estimated through a multi-algorithmic machine learning framework.

	Constant	BMK + Size	Def	(BMK + Size)×Def
Intercept	0.02	0.09	0.04	0.11
	(0.34)	(1.72)	(0.80)	(1.85)
Nasdaq Dummy	0.10	0.11	0.12	0.12
	(0.74)	(0.86)	(1.07)	(1.04)
Size	-0.09	-0.09	-0.09	-0.09
	(4.19)	(4.23)	(4.08)	(3.95)
BMK	0.14	0.09	0.14	0.09
	(3.78)	(2.77)	(3.67)	(2.81)
RET 2-3	0.69	0.68	0.63	0.64
	(3.09)	(3.14)	(2.84)	(2.99)
RET 4-6	0.57	0.59	0.59	0.63
	(2.95)	(3.24)	(3.16)	(3.54)
RET 7-12	0.69	0.69	0.70	0.71
	(5.58)	(5.84)	(5.71)	(5.98)
NYSE/AMEX Turnover	-0.09	-0.08	-0.08	-0.08
	(2.44)	(2.29)	(2.33)	(2.13)
Nasdaq Turnover	-0.09	-0.09	-0.09	-0.09
	(2.14)	(2.03)	(2.12)	(1.95)

For every month from January 1965 to June 2017 (630 months in total), we conduct a cross-sectional regression of risk-adjusted returns on anomaly-related variables - firm characteristics (see Equation 1.8). This table depicts the average loadings, over all 630 months, on such characteristics. All coefficients were multiplied by 100 and absolute values of t-statistics are shown in parenthesis. Risk adjusted returns, the dependent variables in each cross-sectional regression, were computed using Equation 1.7, with betas obtained through our machine learning estimation framework. This table has four columns, each of which corresponds to a different model being tested. All of these four models employ the three Fama and French (1993) factors as risk factors. Models differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. Nasdaq Dummy is a dummy variable that takes the value of 1 if the share in question is negotiated in the Nasdaq Stock Exchange, and zero otherwise. The variables Size and BMK measure firm size and the book-to-market ratio, respectively. RET 2-3, RET 4-6 and RET 7-12 are our past return variables. They measure, respectively, as in AC, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to that when returns are recorded. NYSE/AMEX Turnover and Nasdaq Turnover represent the monthly turnover of shares traded on the AMEX/NYSE and Nasdaq exchanges, correspondingly. A detailed description of how each of these variables is computed is presented in Section 1.3.

TABLE 2.6: Average coefficients from cross-sectional regressions, with excess market return, SMB, HML and LTREV used as risk factors, and time series econometric specifications estimated through a multi-algorithmic machine learning framework.

	Constant	BMK + Size	Def	(BMK + Size)×Def
Intercept	0.04	0.11	0.06	0.13
	(0.67)	(2.06)	(1.12)	(2.19)
Nasdaq Dummy	0.09	0.11	0.10	0.10
	(0.71)	(0.85)	(0.76)	(0.80)
Size	-0.09	-0.09	-0.09	-0.09
	(4.27)	(4.14)	(4.26)	(4.00)
BMK	0.14	0.09	0.14	0.09
	(3.66)	(2.77)	(3.52)	(2.79)
RET 2-3	0.68	0.65	0.64	0.66
	(3.09)	(3.07)	(2.91)	(3.11)
RET 4-6	0.54	0.57	0.55	0.60
	(2.82)	(3.17)	(3.00)	(3.39)
RET 7-12	0.69	0.69	0.69	0.69
	(5.56)	(5.89)	(5.67)	(5.87)
NYSE/AMEX Turnover	-0.09	-0.08	-0.08	-0.07
	(2.43)	(2.20)	(2.29)	(1.93)
Nasdaq Turnover	-0.09	-0.09	-0.10	-0.09
	(2.17)	(1.98)	(2.13)	(1.95)

For every month from January 1965 to June 2017 (630 months in total), we conduct a cross-sectional regression of risk-adjusted returns on anomaly-related variables - firm characteristics (see Equation 1.8). This table depicts the average loadings, over all 630 months, on such characteristics. All coefficients were multiplied by 100 and absolute values of t-statistics are shown in parenthesis. Risk adjusted returns, the dependent variables in each cross-sectional regression, were computed using Equation 1.7, with betas obtained through our machine learning estimation framework. This table has four columns, each of which corresponds to a different model being tested. All of these four models employ the three Fama and French (1993) factors plus a long-term reversals factor as risk factors. Models differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. Nasdaq Dummy is a dummy variable that takes the value of 1 if the share in question is negotiated in the Nasdaq Stock Exchange, and zero otherwise. The variables Size and BMK measure firm size and the book-to-market ratio, respectively. RET 2-3, RET 4-6 and RET 7-12 are our past return variables. They measure, respectively, as in AC, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to that when returns are recorded. NYSE/AMEX Turnover and Nasdaq Turnover represent the monthly turnover of shares traded on the AMEX/NYSE and Nasdaq exchanges, correspondingly. A detailed description of how each of these variables is computed is presented in Section 1.3.

Table 2.7: Average coefficients from cross-sectional regressions, with excess market return, SMB, HML and MOM used as risk factors, and time series econometric specifications estimated through a multi-algorithmic machine learning framework.

	Constant	BMK + Size	Def	(BMK + Size)×Def
Intercept	0.11	0.18	0.13	0.18
	(2.24)	(3.39)	(2.51)	(3.14)
Nasdaq Dummy	0.10	0.12	0.10	0.11
	(0.75)	(0.94)	(0.78)	(0.86)
Size	-0.10	-0.10	-0.09	-0.09
	(4.57)	(4.41)	(4.23)	(4.15)
BMK	0.14	0.12	0.15	0.12
	(3.76)	(3.67)	(3.86)	(3.59)
RET 2-3	0.67	0.65	0.65	0.66
	(3.09)	(3.15)	(3.04)	(3.15)
RET 4-6	0.57	0.55	0.58	0.60
	(3.07)	(3.17)	(3.26)	(3.51)
RET 7-12	0.69	0.66	0.71	0.70
	(5.71)	(5.84)	(6.03)	(6.17)
NYSE/AMEX Turnover	-0.06	-0.07	-0.07	-0.06
	(1.76)	(1.89)	(1.81)	(1.76)
Nasdaq Turnover	-0.06	-0.07	-0.07	-0.07
	(1.48)	(1.51)	(1.57)	(1.52)

For every month from January 1965 to June 2017 (630 months in total), we conduct a cross-sectional regression of risk-adjusted returns on anomaly-related variables - firm characteristics (see Equation 1.8). This table depicts the average loadings, over all 630 months, on such characteristics. All coefficients were multiplied by 100 and absolute values of t-statistics are shown in parenthesis. Risk adjusted returns, the dependent variables in each cross-sectional regression, were computed using Equation 1.7, with betas obtained through our machine learning estimation framework. This table has four columns, each of which corresponds to a different model being tested. All of these four models employ the three Fama and French (1993) factors plus a momentum factor as risk factors. Models differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. Nasdaq Dummy is a dummy variable that takes the value of 1 if the share in question is negotiated in the Nasdaq Stock Exchange, and zero otherwise. The variables Size and BMK measure firm size and the book-to-market ratio, respectively. RET 2-3, RET 4-6 and RET 7-12 are our past return variables. They measure, respectively, as in AC, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to that when returns are recorded. NYSE/AMEX Turnover and Nasdaq Turnover represent the monthly turnover of shares traded on the AMEX/NYSE and Nasdaq exchanges, correspondingly. A detailed description of how each of these variables is computed is presented in Section 1.3.

TABLE 2.8: Average coefficients from cross-sectional regressions, with excess market return, SMB, HML, CMA and RMW used as risk factors, and time series econometric specifications estimated through a multi-algorithmic machine learning framework.

	Constant	BMK + Size	Def	(BMK + Size)×Def
Intercept	-0.01	0.11	0.05	0.13
	(0.27)	(2.02)	(0.94)	(2.25)
Nasdaq Dummy	0.19	0.17	0.17	0.16
	(1.44)	(1.28)	(1.30)	(1.22)
Size	-0.12	-0.11	-0.12	-0.11
	(5.39)	(5.31)	(5.42)	(5.07)
BMK	0.14	0.09	0.13	0.09
	(3.66)	(2.75)	(3.56)	(2.78)
RET 2-3	0.64	0.61	0.58	0.59
	(2.93)	(2.89)	(2.66)	(2.81)
RET 4-6	0.51	0.53	0.52	0.56
	(2.68)	(2.94)	(2.84)	(3.18)
RET 7-12	0.65	0.66	0.66	0.67
	(5.39)	(5.61)	(5.51)	(5.78)
NYSE/AMEX Turnover	-0.07	-0.05	-0.06	-0.05
	(1.85)	(1.48)	(1.75)	(1.40)
Nasdaq Turnover	-0.05	-0.04	-0.04	-0.04
	(1.19)	(0.92)	(0.89)	(0.87)

For every month from January 1965 to June 2017 (630 months in total), we conduct a cross-sectional regression of risk-adjusted returns on anomaly-related variables - firm characteristics (see Equation 1.8). This table depicts the average loadings, over all 630 months, on such characteristics. All coefficients were multiplied by 100 and absolute values of t-statistics are shown in parenthesis. Risk adjusted returns, the dependent variables in each cross-sectional regression, were computed using Equation 1.7, with betas obtained through our machine learning estimation framework. This table has four columns, each of which corresponds to a different model being tested. All of these four models employ the five Fama and French (2015) factors as risk factors. Models differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. Nasdaq Dummy is a dummy variable that takes the value of 1 if the share in question is negotiated in the Nasdaq Stock Exchange, and zero otherwise. The variables Size and BMK measure firm size and the book-to-market ratio, respectively. RET 2-3, RET 4-6 and RET 7-12 are our past return variables. They measure, respectively, as in AC, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to that when returns are recorded. NYSE/AMEX Turnover and Nasdaq Turnover represent the monthly turnover of shares traded on the AMEX/NYSE and Nasdaq exchanges, correspondingly. A detailed description of how each of these variables is computed is presented in Section 1.3.

Table 2.9: Average ranks of learning algorithms, with excess market return, SMB and HML used as risk factors.

	Constant	BMK + Size	Def	(BMK + Size) $\times$Def
Squared Loss (OLS)	6.13	9.92	9.27	11.10
Squared Loss & Ridge Penalty	6.33	5.73	5.34	5.14
Squared Loss & Lasso Penalty	6.86	5.89	6.60	5.41
Squared Loss & Elastic Net Penalty	6.70	5.88	6.22	5.14
Absolute Loss	6.77	9.98	9.35	11.13
Absolute Loss & Ridge Penalty	7.11	6.06	5.89	5.92
Absolute Loss & Lasso Penalty	7.97	5.92	6.38	5.26
Absolute Loss & Elastic Net Penalty	7.95	5.89	6.28	5.19
Huber Loss	4.91	8.89	8.27	10.56
Huber Loss & Ridge Penalty	5.26	4.82	4.59	4.82
Huber Loss & Lasso Penalty	6.02	4.54	4.98	4.23
Huber Loss & Elastic Net Penalty	5.99	4.47	4.81	4.09

This table depicts average ranks for the twelve learning algorithms used in our study (described in Section 1.2.3). Each of the four columns shown presents average ranks computed for a different model. Average ranks may range from one to twelve, and represent the average quality of fit attained by an algorithm, for the time series specification of the model in question. For each problem (model and stock), the twelve learning algorithms we use were ranked from one - best quality of fit - to twelve - worst quality of fit. An average rank was then computed for each learning algorithm using the ranks they received across all stocks, for a given model. The three-fold cross-validated mean squared error associated with the time series specification was used to determine quality of fit. The four models being examined here all use the three Fama and French (1993) factors as risk factors. They differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables.

Table 2.10: Average ranks of learning algorithms, with excess market return, SMB, HML and LTREV used as risk factors.

	Constant	BMK + Size	Def	(BMK + Size) ×Def
Squared Loss (OLS)	6.74	10.42	9.79	11.14
Squared Loss & Ridge Penalty	6.60	5.74	5.50	5.26
Squared Loss & Lasso Penalty	6.76	5.65	6.30	5.20
Squared Loss & Elastic Net Penalty	6.71	5.61	6.02	5.01
Absolute Loss	7.03	10.38	9.79	11.21
Absolute Loss & Ridge Penalty	7.03	6.13	5.75	6.12
Absolute Loss & Lasso Penalty	7.58	5.59	6.12	5.18
Absolute Loss & Elastic Net Penalty	7.60	5.56	6.00	5.10
Huber Loss	5.25	9.37	8.79	10.59
Huber Loss & Ridge Penalty	5.29	4.91	4.57	5.10
Huber Loss & Lasso Penalty	5.70	4.34	4.75	4.10
Huber Loss & Elastic Net Penalty	5.69	4.27	4.62	3.97

This table depicts average ranks for the twelve learning algorithms used in our study (described in Section 1.2.3). Each of the four columns shown presents average ranks computed for a different model. Average ranks may range from one to twelve, and represent the average quality of fit attained by an algorithm, for the time series specification of the model in question. For each problem (model and stock), the twelve learning algorithms we use were ranked from one - best quality of fit - to twelve - worst quality of fit. An average rank was then computed for each learning algorithm using the ranks they received across all stocks, for a given model. The three-fold cross-validated mean squared error associated with the time series specification was used to determine quality of fit. The four models being examined here all use the three Fama and French (1993) factors plus a long-term reversals factor as risk factors. They differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables.

Table 2.11: Average ranks of learning algorithms, with excess market return, SMB, HML and MOM used as risk factors.

	Constant	BMK + Size	Def	(BMK + Size) ×Def
Squared Loss (OLS)	6.75	10.45	10.13	11.14
Squared Loss & Ridge Penalty	6.59	5.73	5.41	5.33
Squared Loss & Lasso Penalty	6.86	5.63	6.16	5.10
Squared Loss & Elastic Net Penalty	6.75	5.56	5.89	4.88
Absolute Loss	7.09	10.43	10.09	11.17
Absolute Loss & Ridge Penalty	6.90	6.10	5.77	6.27
Absolute Loss & Lasso Penalty	7.57	5.58	5.89	5.13
Absolute Loss & Elastic Net Penalty	7.57	5.52	5.75	5.03
Huber Loss	5.30	9.48	9.18	10.63
Huber Loss & Ridge Penalty	5.19	4.94	4.70	5.34
Huber Loss & Lasso Penalty	5.73	4.32	4.58	4.03
Huber Loss & Elastic Net Penalty	5.68	4.25	4.43	3.93

This table depicts average ranks for the twelve learning algorithms used in our study (described in Section 1.2.3). Each of the four columns shown presents average ranks computed for a different model. Average ranks may range from one to twelve, and represent the average quality of fit attained by an algorithm, for the time series specification of the model in question. For each problem (model and stock), the twelve learning algorithms we use were ranked from one - best quality of fit - to twelve - worst quality of fit. An average rank was then computed for each learning algorithm using the ranks they received across all stocks, for a given model. The three-fold cross-validated mean squared error associated with the time series specification was used to determine quality of fit. The four models being examined here all use the three Fama and French (1993) factors plus a momentum factor as risk factors. They differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables.

Table 2.12: Average ranks of learning algorithms, with excess market return, SMB, HML, CMA and RMW used as risk factors.

	Constant	BMK + Size	Def	(BMK + Size) $\times$Def
Squared Loss (OLS)	7.23	10.79	10.12	11.12
Squared Loss & Ridge Penalty	6.52	5.74	5.41	5.29
Squared Loss & Lasso Penalty	6.64	5.36	6.05	5.06
Squared Loss & Elastic Net Penalty	6.57	5.30	5.77	4.81
Absolute Loss	7.46	10.72	10.10	11.23
Absolute Loss & Ridge Penalty	7.06	6.26	5.86	6.42
Absolute Loss & Lasso Penalty	7.31	5.35	5.91	5.12
Absolute Loss & Elastic Net Penalty	7.30	5.33	5.81	4.99
Huber Loss	5.74	9.88	9.17	10.62
Huber Loss & Ridge Penalty	5.23	5.03	4.71	5.36
Huber Loss & Lasso Penalty	5.45	4.12	4.63	4.04
Huber Loss & Elastic Net Penalty	5.48	4.10	4.45	3.93

This table depicts average ranks for the twelve learning algorithms used in our study (described in Section 1.2.3). Each of the four columns shown presents average ranks computed for a different model. Average ranks may range from one to twelve, and represent the average quality of fit attained by an algorithm, for the time series specification of the model in question. For each problem (model and stock), the twelve learning algorithms we use were ranked from one - best quality of fit - to twelve - worst quality of fit. An average rank was then computed for each learning algorithm using the ranks they received across all stocks, for a given model. The three-fold cross-validated mean squared error associated with the time series specification was used to determine quality of fit. The four models being examined here all use the five Fama and French (2015) factors as risk factors. They differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables.

TABLE 2.13: Average coefficients from cross-sectional regressions, with excess market return, SMB and HML used as risk factors, and time series econometric specifications estimated through OLS.

	Constant	BMK + Size	Def	(BMK + Size)×Def
Intercept	-0.06	-0.02	-0.05	0.01
	(1.16)	(0.37)	(1.10)	(0.19)
Nasdaq Dummy	0.09	0.14	0.15	0.21
	(0.66)	(1.29)	(1.33)	(3.15)
Size	-0.08	-0.07	-0.08	-0.07
	(3.89)	(3.39)	(3.71)	(3.59)
BMK	0.14	0.06	0.13	0.03
	(3.85)	(1.94)	(3.45)	(0.86)
RET 2-3	0.69	0.65	0.58	0.54
	(3.12)	(3.11)	(2.66)	(2.55)
RET 4-6	0.57	0.62	0.58	0.62
	(3.02)	(3.56)	(3.25)	(3.72)
RET 7-12	0.69	0.65	0.67	0.66
	(5.57)	(5.70)	(5.50)	(6.13)
NYSE/AMEX Turnover	-0.09	-0.10	-0.09	-0.09
	(2.65)	(2.97)	(2.66)	(3.02)
Nasdaq Turnover	-0.09	-0.10	-0.09	-0.09
	(2.19)	(2.57)	(2.02)	(2.51)

For every month from January 1965 to June 2017 (630 months in total), we conduct a cross-sectional regression of risk-adjusted returns on anomaly-related variables - firm characteristics (see Equation 1.8). This table depicts the average loadings, over all 630 months, on such characteristics. All coefficients were multiplied by 100 and absolute values of t-statistics are shown in parenthesis. Risk adjusted returns, the dependent variables in each cross-sectional regression, were computed using Equation 1.7, with betas obtained through OLS estimation. This table has four columns, each of which corresponds to a different model being tested. All of these four models employ the three Fama and French (1993) factors as risk factors. Models differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. Nasdaq Dummy is a dummy variable that takes the value of 1 if the share in question is negotiated in the Nasdaq Stock Exchange, and zero otherwise. The variables Size and BMK measure firm size and the book-to-market ratio, respectively. RET 2-3, RET 4-6 and RET 7-12 are our past return variables. They measure, respectively, as in AC, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to that when returns are recorded. NYSE/AMEX Turnover and Nasdaq Turnover represent the monthly turnover of shares traded on the AMEX/NYSE and Nasdaq exchanges, correspondingly. A detailed description of how each of these variables is computed is presented in Section 1.3.

TABLE 2.14: Average coefficients from cross-sectional regressions, with excess market return, SMB, HML and LTREV used as risk factors, and time series econometric specifications estimated through OLS.

	Constant	BMK + Size	Def	(BMK + Size)×Def
Intercept	-0.06	-0.01	-0.05	0.03
	(1.09)	(0.18)	(1.08)	(0.62)
Nasdaq Dummy	0.08	0.13	0.13	0.20
	(0.61)	(1.21)	(1.13)	(3.00)
Size	-0.08	-0.06	-0.08	-0.06
	(3.91)	(3.31)	(3.82)	(3.57)
BMK	0.14	0.05	0.11	0.01
	(3.69)	(1.60)	(3.11)	(0.21)
RET 2-3	0.67	0.61	0.59	0.52
	(3.07)	(2.99)	(2.76)	(2.55)
RET 4-6	0.55	0.56	0.54	0.48
	(2.95)	(3.33)	(3.04)	(2.99)
RET 7-12	0.68	0.63	0.64	0.64
	(5.53)	(5.65)	(5.47)	(6.16)
NYSE/AMEX Turnover	-0.09	-0.10	-0.09	-0.10
	(2.71)	(3.01)	(2.71)	(3.27)
Nasdaq Turnover	-0.09	-0.10	-0.09	-0.09
	(2.23)	(2.52)	(2.08)	(2.46)

For every month from January 1965 to June 2017 (630 months in total), we conduct a cross-sectional regression of risk-adjusted returns on anomaly-related variables - firm characteristics (see Equation 1.8). This table depicts the average loadings, over all 630 months, on such characteristics. All coefficients were multiplied by 100 and absolute values of t-statistics are shown in parenthesis. Risk adjusted returns, the dependent variables in each cross-sectional regression, were computed using Equation 1.7, with betas obtained through OLS estimation. This table has four columns, each of which corresponds to a different model being tested. All of these four models employ the three Fama and French (1993) factors plus a long-term reversals factor as risk factors. Models differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. Nasdaq Dummy is a dummy variable that takes the value of 1 if the share in question is negotiated in the Nasdaq Stock Exchange, and zero otherwise. The variables Size and BMK measure firm size and the book-to-market ratio, respectively. RET 2-3, RET 4-6 and RET 7-12 are our past return variables. They measure, respectively, as in AC, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to that when returns are recorded. NYSE/AMEX Turnover and Nasdaq Turnover represent the monthly turnover of shares traded on the AMEX/NYSE and Nasdaq exchanges, correspondingly. A detailed description of how each of these variables is computed is presented in Section 1.3.

Table 2.15: Average coefficients from cross-sectional regressions, with excess market return, SMB, HML and MOM used as risk factors, and time series econometric specifications estimated through OLS.

	Constant	BMK + Size	Def	(BMK + Size)×Def
Intercept	0.05	0.08	0.05	0.08
	(0.92)	(1.90)	(1.01)	(2.13)
Nasdaq Dummy	0.09	0.16	0.16	0.24
	(0.65)	(1.48)	(1.38)	(3.76)
Size	-0.09	-0.06	-0.08	-0.06
	(4.21)	(3.30)	(3.88)	(3.12)
BMK	0.14	0.08	0.12	0.04
	(3.81)	(2.80)	(3.24)	(1.55)
RET 2-3	0.67	0.63	0.57	0.52
	(3.11)	(3.15)	(2.72)	(2.65)
RET 4-6	0.57	0.60	0.57	0.55
	(3.17)	(3.75)	(3.40)	(3.67)
RET 7-12	0.68	0.63	0.66	0.61
	(5.70)	(6.00)	(5.81)	(6.36)
NYSE/AMEX Turnover	-0.06	-0.09	-0.07	-0.08
	(1.80)	(2.70)	(2.17)	(2.76)
Nasdaq Turnover	-0.05	-0.06	-0.05	-0.07
	(1.29)	(1.70)	(1.29)	(1.93)

For every month from January 1965 to June 2017 (630 months in total), we conduct a cross-sectional regression of risk-adjusted returns on anomaly-related variables - firm characteristics (see Equation 1.8). This table depicts the average loadings, over all 630 months, on such characteristics. All coefficients were multiplied by 100 and absolute values of t-statistics are shown in parenthesis. Risk adjusted returns, the dependent variables in each cross-sectional regression, were computed using Equation 1.7, with betas obtained through OLS estimation. This table has four columns, each of which corresponds to a different model being tested. All of these four models employ the three Fama and French (1993) factors plus a momentum factor as risk factors. Models differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. Nasdaq Dummy is a dummy variable that takes the value of 1 if the share in question is negotiated in the Nasdaq Stock Exchange, and zero otherwise. The variables Size and BMK measure firm size and the book-to-market ratio, respectively. RET 2-3, RET 4-6 and RET 7-12 are our past return variables. They measure, respectively, as in AC, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to that when returns are recorded. NYSE/AMEX Turnover and Nasdaq Turnover represent the monthly turnover of shares traded on the AMEX/NYSE and Nasdaq exchanges, correspondingly. A detailed description of how each of these variables is computed is presented in Section 1.3.

TABLE 2.16: Average coefficients from cross-sectional regressions, with excess market return, SMB, HML, CMA and RMW used as risk factors, and time series econometric specifications estimated through OLS.

	Constant	BMK + Size	Def	(BMK + Size)×Def
Intercept	-0.13	-0.04	-0.08	0.00
	(2.69)	(0.93)	(1.65)	(0.09)
Nasdaq Dummy	0.19	0.21	0.23	0.25
	(1.49)	(2.05)	(2.29)	(4.15)
Size	-0.10	-0.08	-0.10	-0.08
	(5.08)	(4.50)	(4.99)	(4.53)
BMK	0.13	0.04	0.12	0.01
	(3.69)	(1.36)	(3.29)	(0.26)
RET 2-3	0.62	0.58	0.47	0.40
	(2.88)	(2.88)	(2.17)	(1.98)
RET 4-6	0.50	0.53	0.47	0.45
	(2.75)	(3.25)	(2.72)	(2.86)
RET 7-12	0.63	0.59	0.58	0.56
	(5.25)	(5.36)	(5.05)	(5.57)
NYSE/AMEX Turnover	-0.08	-0.07	-0.07	-0.07
	(2.21)	(2.38)	(2.09)	(2.31)
Nasdaq Turnover	-0.06	-0.04	-0.03	-0.02
	(1.38)	(1.16)	(0.85)	(0.64)

For every month from January 1965 to June 2017 (630 months in total), we conduct a cross-sectional regression of risk-adjusted returns on anomaly-related variables - firm characteristics (see Equation 1.8). This table depicts the average loadings, over all 630 months, on such characteristics. All coefficients were multiplied by 100 and absolute values of t-statistics are shown in parenthesis. Risk adjusted returns, the dependent variables in each cross-sectional regression, were computed using Equation 1.7, with betas obtained through OLS estimation. This table has four columns, each of which corresponds to a different model being tested. All of these four models employ the five Fama and French (2015) factors as risk factors. Models differ according to how betas are modeled. In the first column, risk loadings are assumed to be constant. In the second column, risk loadings vary linearly with firm-size and book-to-market. In the third column, beta coefficients were conditioned on default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables. Nasdaq Dummy is a dummy variable that takes the value of 1 if the share in question is negotiated in the Nasdaq Stock Exchange, and zero otherwise. The variables Size and BMK measure firm size and the book-to-market ratio, respectively. RET 2-3, RET 4-6 and RET 7-12 are our past return variables. They measure, respectively, as in AC, the cumulative returns over the second through third, fourth through sixth, and seventh through twelfth months prior to that when returns are recorded. NYSE/AMEX Turnover and Nasdaq Turnover represent the monthly turnover of shares traded on the AMEX/NYSE and Nasdaq exchanges, correspondingly. A detailed description of how each of these variables is computed is presented in Section 1.3.

Table 2.17: Least squares overfitting - supplementary analysis.

	Constant	BMK + Size	Def	(BMK + Size)$\times$Def
$\bar{R}^2$ (Least Squares)	3.00	2.59	2.94	2.51
$\bar{R}^2$ (Machine Learning)	3.11	2.85	3.16	2.95

For every month from January 1965 to June 2017 (630 months in total), we conduct a cross-sectional regression of risk-adjusted returns on anomaly-related variables - firm characteristics. This table depicts the average R^2 values, over all 630 months, from such cross-sectional regressions. This table has four columns, each of which corresponds to a different asset pricing model. All of these four models employ the three Fama and French (1993) factors as risk factors. Models differ according to how betas are modeled. In the first column, betas are assumed to be constant. In the second column, risk loadings vary linearly with two-variables: firm-size and book-to-market. In the third column, beta coefficients were conditioned on a single variable: default spread (the yield differential between BAA and AAA corporate bonds). In the fourth column, default spread, size, book-to-market, and interactions between size and default spread and book-to-market and default spread were used as conditioning variables (five conditioning variables in total). In the first row, least squares was used to estimate time series specifications (Equation 1.4). In the second row, our multi-algorithmic machine learning approach, which controls for overfitting, was used to estimate time series specifications (Equation 1.4) instead. Parameters from the estimated time series specifications were then used to obtain risk adjusted returns, following the formula described by Equation 1.7, and these were subsequently regressed every month on the same firm characteristics used in our asset pricing tests (see Section 1.3). We see here that when least squares is used to estimate time series specifications, there is a decreasing relationship between average R^2 values from cross-sectional regressions and the number of conditioning variables used to model betas. A possible explanation for this phenomenon is that least squares may increasingly overfit time series specifications as the number of conditioning variables, and thus explanatory variables, increases. Importantly, the decreasing pattern is no longer present when our machine learning approach, which controls for overfitting, is used in the estimation (second row). These findings suggests that overfitting resulting from OLS estimation of conditional time series specifications may be very much real, and not merely a theoretical concern - see Appendix C of Chapter 2.

CHAPTER 3

The Building Blocks of Employment: A Signal Processing Analysis

3.1 Introduction

The last US recession ended in the second-quarter of 2009, and the unemployment rate (BLS (2018a)) has since fallen considerably (see Figure 3.1). Indeed, unemployment is close to the lowest level it's been in the last 40 years, suggesting at first sight the predominance of very healthy labor-market conditions in the country. One could conjecture, however, that such precipitous drop may potentially be, in part, attributed to the fact that, more and more, individuals are dropping out of the labor-force. In truth, the unemployment rate - according to the Bureau of Labor Statistics (BLS) - is measured as the estimated number of unemployed individuals (people who are simultaneously jobless, looking for a job, and available for work) divided by the estimated labor force (which is comprised of unemployed plus employed individuals[1]). A decrease in the number of individuals looking for a job leads to a decrease in both the number of unemployed individuals (the numerator of the unemployment rate), and to a similar decrease in the labor force (the denominator of such rate), ultimately producing a decline in the rate itself[2].

As stated by Leamer (2012), a more indicative measure of the overall health of the labor

[1]In section 3.2 we provide a comprehensive summary of which individuals are considered to be employed by the BLS, and on how employment-related data is collected.

[2]Note that this follows from a basic property of fractions. If a, b and x are non-zero positive integers, with $a < b$, then $\frac{a-x}{b-x} < \frac{a}{b}$.

market is the employment-to-population ratio (illustrated in Figure 3.2). As highlighted by Leamer (2012), we see here that US workers have been, relatively speaking, having a hard time finding jobs - the employment-to-population ratio is at a lower level than it was in the late 1990's. The 4% decline seen over the last recession, he argues, was very large vis-à-vis declines registered in previous recessions, and the employment-to-population ratio is still depressed relative to the levels achieved prior to the two most recent recessions.

What may have caused the recent decline in the employment-to-population ratio? Abraham and Kearney (2018) investigate this subject. Abraham and Kearney (2018) argue that although population aging contributed to the recent decline in the ratio, the effects of within-group employment rate declines among young and prime age adults on the overall employment-to-population ratio have been even larger than the effects of population aging. The authors conclude that these declines were driven by labor demand factors, the most important of them being robotization of work and trade.

In addition to the decline in employment we alluded to above, the future scenario for labor and employment may also be one marked by challenges. Put simply, there is a potential for job losses due to automation going forward. As Hawking (2016) has argued, artificial intelligence and automation have not only destroyed jobs in traditional manufacturing[3], but may also promote additional job destruction into the middle classes, generating a widening of inequality while posing a marked challenge for society. Along these lines, Frey and Osborne (2013) investigate the susceptibility of jobs to computerization, and conclude that 47% of total US employment is at high risk of being automated, and could be so over the next decade or two.

In this study, we model and interpret the time series of employment-to-population (the employment rate) from April 1975 to December 2000, a period when employment exhib-

[3]In Figure 3.3, we show the estimated total number of employees in the Manufacturing sector Manufacturing jobs took a hard hit in the last recession, falling by about 30% then.

ited strong gains. By understanding it better, one could possibly develop recipes/policies for improving the current and future labor environments in the United States. In our view, identifying and analyzing relevant components of the series during this period is a step towards promoting such understanding. The model shown here explains well the data at hand, and may contribute to fostering research that can help further improve employment levels and offset impacts to employment from the challenges that lay ahead. The employment-to-population components singled out via our signal processing modeling approach are made available to researchers in an online appendix to this article.

3.2 Description of Data

Our data comes from the U.S. Bureau of Labor Statistics, and was retrieved through the Federal Reserve Bank of Saint Louis (FRED)[4]. FRED provides about 469,000 US and international time series, from numerous sources, in its website[5]. The data we use is monthly, and the observations originally come from the Current Population Survey (CPS) conducted early each month by the BLS. This survey, which measures employment and unemployment in the country, has been conducted in the US every month since 1940 (though the data that is available on FRED starts in 1948). The BLS reports that there are roughly 60,0000 eligible households in the sample for this survey. The sample is selected so as to be representative of the US population. Every month, government employees contact these 60,000 households and ask individuals living in them (aged 16 and above) questions about their labor activities. Individuals are considered employed if they: (i) “did any work for pay or profit during the week when the survey is conducted”;

[4]BLS (2018b).

[5]The series can be downloaded from FRED directly into R using the function “getSymbols” from the “quantmod” library (Ryan (2016)).

or (ii) have a job but couldn't work because of specific circumstances[6]. Note that part-time workers, according to this definition, are considered to be employed. And also note that if a person has more than one job, he or she will simply show up in the survey as being employed - i.e., s/he won't be "counted twice". The employment-to-population ratio is then defined as the number of employed individuals, as a percentage of the total population (in the surveyed households).

A concern that might arise with respect to the employment-to-population series is that it may be influenced by population aging. To ameliorate this concern, we have adjusted the series, creating an employment-to-working-age-population one[7], and subsequently computed the correlation between the original employment-to-population series and our adjusted series during the period in which our analysis takes place. We report that both series exhibit a very high correlation, 0.96, during such period.

3.3 Analysis

To model our data and single out components of the employment-to-population time series, the following steps are taken: **(i)** we detrend/filter the data, extracting a low frequency component from it; **(ii)** we use spectral analysis to identify cycles/seasonal patterns in the detrended data, and then remove these cycles from it; **(iii)** the residuals (original data, minus trend, minus seasonal components) are then analyzed and modeled as an ARMA process; **(iv)** finally, we assess how well we have modeled our time series by evaluating the residuals engendered by our approach.

[6]In particular, according to the BLS, if they have a job but were on vacation; ill; experiencing child care problems; on maternity or paternity leave; prevented from working due to bad weather; involved in a labor dispute; or taking care of personal of family obligation.

[7]Our adjusted series is created by multiplying the original civilian employment-to-population series by the civilian noninstitutional population (BLS (2018c)) - "civilians" are those aged 16 and above - and by then dividing it by working age population - defined by the Organization for Economic Co-operation and Development as the number of individuals aged 15 to 64 - (OECD (2018)). While the civilian population series is presented in thousands of individuals, working age population is not. For this reason, prior to performing the computation above, we first divided working age population by one thousand.

Before implementing these steps, we first assess the autocorrelation (ACF) and partial-autocorrelation functions (PACF) of the original data. On the top portion of Figure 3.4, we show the ACF and PACF of the original data when all possible lags are used; on the bottom of Figure 3.4, the ACF and PACF are depicted for a total of 50 lags. We see that the PACF hints towards the existence of cycles in the data; and so does the ACF plot which includes all possible lags. As we'll show below, the cyclical nature of our data will become more apparent once the series has been detrended.

3.3.1 Detrending

As a first step, we attempt to fit a linear regression model to the data and define the trend component as being of a linear nature. The fitted model takes on the following form: $\hat{Y}_t = -520 + 0.29t$, where $\hat{Y}_t$ denotes the fitted employment-to-population ratio at time t. The $t-statistic$ for our slope coefficient is 35.24, indicating that the trend in this period is statistically significant (even at the 1% level of significance).

Figure 3.5 shows the detrended data. This figure suggests, interestingly, that the resulting detrended series is not stationary: the variance on the left-hand site of Figure 3.5 seems markedly higher than that on the right-hand side. To confirm our visual intuition, we conduct an augmented Dickey-Fuller (ADF) test on the detrended series[8]. The p-value from our test is 0.36 - therefore, we fail to reject the null hypotheses that there exists a unit root in this data. This in turn suggests that the detrended series is indeed non-stationary.

One could attempt to get around this by taking first-differences of the original data; but our main interest here is in the employment-to-population ratio itself, not in first-differences of this quantity. Instead, therefore, we fit a non-linear trend to the data using a penalized smoothing-spline regression procedure (i.e., a spline regression where the knots

[8]The adf.test function in R (Trapletti and Hornik (2017)) selects a lag order of 6 for this test.

are tied to the observations themselves)[9].

The fitted trend (low-frequency) component is of the form $f(x) = \alpha_0 + \sum_{j=1}^{p} \alpha_j max(0, x - k_j)$, where x here is equivalent to t (time), and the $k_j's$ are the times associated with the observations in the employment-to-population series[10]. The fitted $\alpha's$ are found by minimizing $\sum_{t=1}^{n} ||y_t - \alpha_0 - \sum_{j=1}^{p} \alpha_j max(0, x_t - k_j)||^2 + \lambda \sum_{j=1}^{p} \alpha_j^2$, with y_t denoting the employment-to-population ratio at time t, x_t again simply denotes the time t associated with observation y_t; λ is a tuning parameter that penalizes the complexity of our model[11]. In the appendix, we provide a brief derivation of the spline solution.

Figure 3.6 shows our time series and the non-linear trend. The tuning parameter λ was selected so as to yield a detrended time series that looked stationary, while producing a trend component that is reasonably smooth[12]. The detrended/filtered data is shown in Figure 3.7. We can see that after removing the non-linear trend, we obtain a series that exhibits an apparent stationary behavior. An ADF test (Trapletti and Hornik (2017)) was conducted on this time series, producing a very low $p - value$ ($p < 0.01$)[13] - the null hypothesis that there exists a unit root in this series can now be rejected. This is the detrended series that we use in our analysis.

Some comments are in order. Firstly, one could speculate that the overall positive trend seen in this period is associated with an increased integration between the US economy and other economies, which may have led to increasing exports and the creation of local job openings. It is also possible that an increasing participation of women in the

[9] See de Boor (1978) and Wahba (1990) for a comprehensive overview of spline models. Note that one might conjecture that the Hodrick-Prescott filter (Hodrick and Prescott (1997)) could be a suitable filtering alternative. Hamilton (2017), however, presents strong arguments against its usage. Indeed, the very title of their paper, "Why you should never use the Hodrick-Prescott filter" emphatically suggests that one should refrain from employing it altogether.

[10] Note that $p = 309$, the total number of observations.

[11] Also, $n = 309$ and $p = 309$; and k_j denotes, as we mentioned, the time associated with observation j.

[12] A value of 100 was used for λ.

[13] Again a lag order of 6 was used in the test.

labor-force contributed to the steep increase in employment we saw from 1975 to 2000. Lastly, automation and product innovations may destroy jobs in some industries, but can also create jobs in others (and note that these can also make companies more efficient, potentially leading to greater shareholder wealth, which can also spur consumption and jobs) - thus, automation by itself (which one could proxy through labor productivity) may also have helped create jobs in this period. These are simple possible candidate explanations for the low-frequency component extracted for the original series, and should be subject to scrutiny in further studies.

3.3.2 Identifying and Removing Cycles (Spectral Analysis)

Figure 3.8 depicts the ACF and PACF of the detrended data - we can clearly see, from this figure, that the detrended series is influenced by cycles. Figure 3.9 illustrates the raw periodogram of the detrended data. A large portion of the total variation in the data is explained by the 1.013 frequency, corresponding to a period of roughly one year. There are also peaks at $\omega = 1.988 \approx 2$ and $\omega = 3$, harmonics of $\omega = 1$. There is, moreover, considerable power at lower frequencies. In particular, we see a small peak at the frequency of 0.118 (which corresponds to a period of 5.33 years, or 64 months). One could associate this peak with long-term business-cycle fluctuations not captured by our non-linear trend.

In Figure 3.10, we show in detail the yearly seasonal component, which, as we have argued, is responsible for a good portion of the total variation in the detrended/filtered series.The graph suggests that there is a strong employment season from March to May, and a weak one from October to January. Generally speaking, April seems to be the best month to find a job, while October is apparently the worst. The yearly seasonal component declines, in general terms, after April, and starts to increase after October. July seems to actually be a stronger month for employment-population than June, August and September. Also note that the annual seasonal component is negative from September to February, staying positive in other months. Summer vacations of hiring managers may be a possible reason for the slowdown in Summer months; a desire by some CFO's to boost calendar-year-end bottom lines[14], coupled with the holiday season, might help to explain the low year-end figures.

We have also computed approximate approximate confidence intervals for the spectral densities associated with the one-year cycle, the cycles corresponding to $\omega = 2$ and $\omega = 3$ (harmonics), and the 5.33 year cycle. The lower values of these intervals are all higher than most of the other periodogram ordinates, indicating that these cycles are all significant.

Figure 3.11 contains the smoothed periodogram of our detrended series - obtained using a modified Daniell (2,2) kernel. This power spectrum shown here is smoother than the one depicted through the raw periodogram, as expected. The peaks occur at exactly the same frequencies we identified previously ($\omega = 0.118, 1.013, 1.988$ and 3). The smoothed periodogram is specially helpful to confirm the low-frequency peak at $\omega = 0.118$ - given in the raw periodogram we saw reasonable power around this frequency.

Moreover, we have utilized an AR spectral estimator to further substantiate our findings. Figure 3.12 demonstrates that the optimal number of lags to use with such estimator, when using AIC as a criteria, is 39 [15]. Figure 3.13 shows that this procedure produces peaks at nearly the same frequencies we had identified previously. There are clear peaks at $\omega \approx 1, 2$ and 3, as before. The low frequency peak is now located at $\omega = 0.192$, indicating a cycle with period of roughly 5.2 years - which is very close to the 5.33 years period that was uncovered with the raw and smoothed periodograms.

The ACF of the detrended data, after removal of the one-year cycle, is provided in Figure 3.14. We can see from this figure that the series still exhibits a cyclical pattern,

[14]Note that the calendar year is used as the fiscal year by most publicly-traded US firms.

[15]Note that the optimal number of lags, when AIC is used, is not that clear here. There are many lags with AIC values very close to the one obtained with p = 39. We have used other values for p in our analysis, obtaining similar results.

as expected, when only the one-year cycle is removed from it. Figure 3.14 also illustrates the ACF of the detrended data after the removal of both the one-year and the 5.33 years (64 months) cycles. Even when influences from these two cycles are removed, we still see that the resulting series possesses a cyclical behavior. Thus, in the analysis that follows, we have opted to remove the influences of all of the four cycles that are evident in our periodograms, and which have found to be significant: the one-year cycle, the 5.33 years one, and the cycles associated with the second and third harmonics of $\omega = 1$ (i.e., $\omega = 2$ and $\omega = 3$).

3.3.3 Modeling Residuals through an ARMA approach and Model Diagnostics

Next, we turn our attention to our residual data (the series obtained after the trend and all seasonal components have been removed). Figure 3.15 shows the ACF and PACF of the residual data. On the top portion of this figure, we see that the ACF seems to tail off (albeit it alternates between positive and negative territory), while the PACF apparently cuts off after a certain number of lags. Closer inspection (see bottom portion of Figure 3.15) of the PACF suggests that the estimated partial-autocorrelation function cuts off at about lag 14 or 15. When combined, these findings suggest fitting an AR(14) or AR(15) model to the data. We confirm our visual intuition by assessing what would be the best AR model, from an AIC standpoint, to fit to the residual data. Figure 3.16 demonstrates that the lowest AIC (-2.33) is obtained when 14 lags are used, confirming the intuition we developed from evaluating the ACF and PACF of our series.

In Figure 3.17, model diagnostics are presented for the AR(14) model. For the most part, an AR(14) model does a good job in describing our residual data. An inspection of the normal Q-Q plot and the histogram of residuals produced by this model indicates that model residuals are normally distributed. A Shapiro-Wilk normality test conducted on these yields a p-value of 0.57, further supporting this assertion. The Ljung-Box test, however, suggests there may be serial correlation in the residuals from lags 12 to 15.

As a result, we have tried to include MA terms in our model, fitting different ARMA models to the data. Good results were obtained when an ARMA(14,9) is used (see Figure 3.18). Residuals from the ARMA(14,9) model look normal[16]; they resemble a white-noise process (see ACF in Figure 3.18); and Ljung-Box tests provide no evidence of serial correlation in them. Furthermore, the AIC associated with this model was -2.64 - a value that is lower than the one we obtained with the AR(14) model.

3.4 Concluding Remarks

We speculate that different variables may have contributed to the robust gains in employment evidenced in the period modeled in this article. Firstly, the US economy became more integrated with other economies during this period. Growing exports may have helped spur job growth over these years. Secondly, it is possible that an increase in the participation of women (and other groups) in the labor force may have also helped boost jobs then. We have also speculated that automation itself (which could be measured through labor productivity, i.e., output per hours worked) may also have contributed to an increase in the level of employment-to-population in this period: while automation may destroy jobs, it also has the potential to make companies more efficient, contributing to an increase in shareholder wealth, possibly thus having an impact in household consumption (via a wealth effect - households may "feel richer" and consume more), leading then to more employment. Lastly, this was also a period in which a good number of skilled foreign workers came to the US. These may have helped the economy grow further, imparting their knowledge on local workers, and going on to start-up new companies by themselves. In future studies, it would be interesting to investigate how these forces may have driven

[16]The Shapiro-Wilk normality test here yielded a p-value of 0.172 - we thus fail to reject the null that the data comes from a normal distribution, at a 5% level of significance.

employment in this period.

Potentially, having an immigration policy that incentivizes skilled workers to join the work force in the US might be a good idea? Maybe creating incentives for women to become entrepreneurs and/or business leaders should be a priority of our government? Questions such as these warrant scrutiny, and can be investigated under different lights in future research using time series components we have here methodically identified and discussed.

3.5 Appendix - Derivation of Solution to Penalized Spline Regression

Suppose we have training examples of the type (y_i, x_i), $i = 1, ..., n$. Let x be one-dimensional (for example, x can be represent $t = time$). Also assume we have knots k_j, $j = 1, ..., p$. Suppose we try to fit a linear spline of the form $f(x) = \alpha_0 + \sum_{j=1}^{p} \alpha_j max(0, x - k_j)$ to the data, by minimizing $\sum_{i=1}^{n} ||y_i - \alpha_0 - \sum_{j=1}^{p} \alpha_j max(0, x_i - k_j)||^2 + \lambda \sum_{j=1}^{p} \alpha_j^2$.

We look for:

$$\hat{\alpha} = argmin_{\alpha_0, \{\alpha_j\}_{j=1...p}} \sum_{i=1}^{n} ||y_i - \alpha_0 - \sum_{j=1}^{p} \alpha_j max(0, x_i - k_j)||^2 + \lambda \sum_{j=1}^{p} \alpha_j^2$$

Let $\hat{\alpha}$ now denote the $(p+1) \times 1$ vector containing our parameter estimates.

Notice that we can write this problem as:

$$\hat{\alpha} = argmin_{\alpha} ||Y - Z\alpha||^2 + \lambda \alpha^T D \alpha,$$

Where D is a $(p+1) \times (p+1)$ matrix, with 1 in all its diagonal elements aside from the first one, which is 0, and with remaining elements equal to 0. In other words, $D_{rc} = 1$ if both $r = c$ and $r, c \neq 1$; and 0 otherwise (for r,c∈{1...p+1}):

$$D = \begin{bmatrix} 0 & 0 \\ 0 & I_p \end{bmatrix}$$

And Z is a $n \times (p+1)$ matrix, whose entries are:

$$Z = \begin{bmatrix} 1 & max(0, x_1 - k_1) & max(0, x_1 - k_2) & ... & max(0, x_1 - k_j) \\ 1 & max(0, x_2 - k_1) & max(0, x_2 - k_2) & ... & max(0, x_2 - k_j) \\ ... & ... & ... & ... & ... \\ 1 & max(0, x_{n-1} - k_1) & max(0, x_{n-1} - k_2) & ... & max(0, x_{n-1} - k_j) \\ 1 & max(0, x_n - k_1) & max(0, x_n - k_2) & ... & max(0, x_n - k_j) \end{bmatrix}$$

Taking the derivative with respect to α and setting it to zero, we get the FOC:

$$\begin{aligned} -2Z^T(Y - Z\alpha) + 2\lambda D\alpha &= 0 \\ -Z^T Y + Z^T Z\alpha + \lambda D\alpha &= 0 \\ (Z^T Z + \lambda D)\alpha &= Z^T Y \end{aligned}$$

Which implies:

$$\hat{\alpha} = (Z^T Z + \lambda D)^{-1} Z^T Y$$

■

Figure 3.1: Unemployment Rate in the US - Recessions in Blue. This figure depicts the evolution of the unemployment rate in the United States. The rate has decreased substantially since the end of the last recession, suggesting healthy labor market conditions prevail in the country. The unemployment rate, however, is notoriously affected by job destruction and declines in labor force participation.

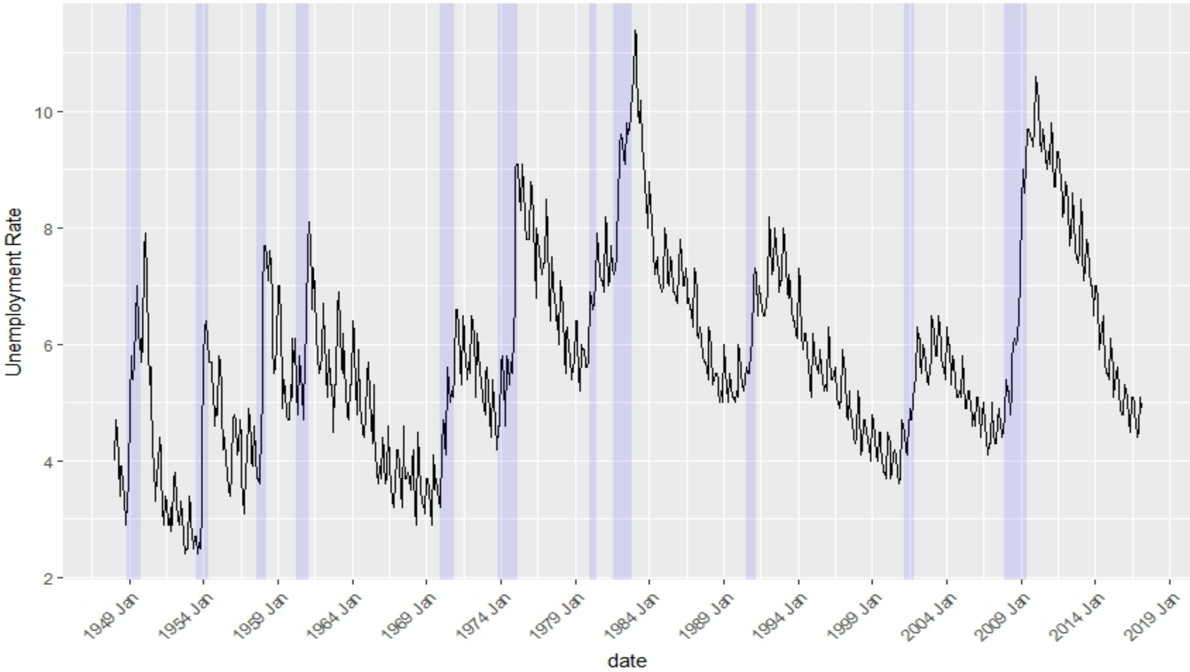


FIGURE 3.2: Employment-to-Population Ratio - Recessions in Blue. This figure illustrates how the employment-to-population ratio has varied across the years. Employment is at a much lower level than it was in 2001. The precipitous 4% decline evidenced over the last recession was very large relative to declines registered in previous recessions.

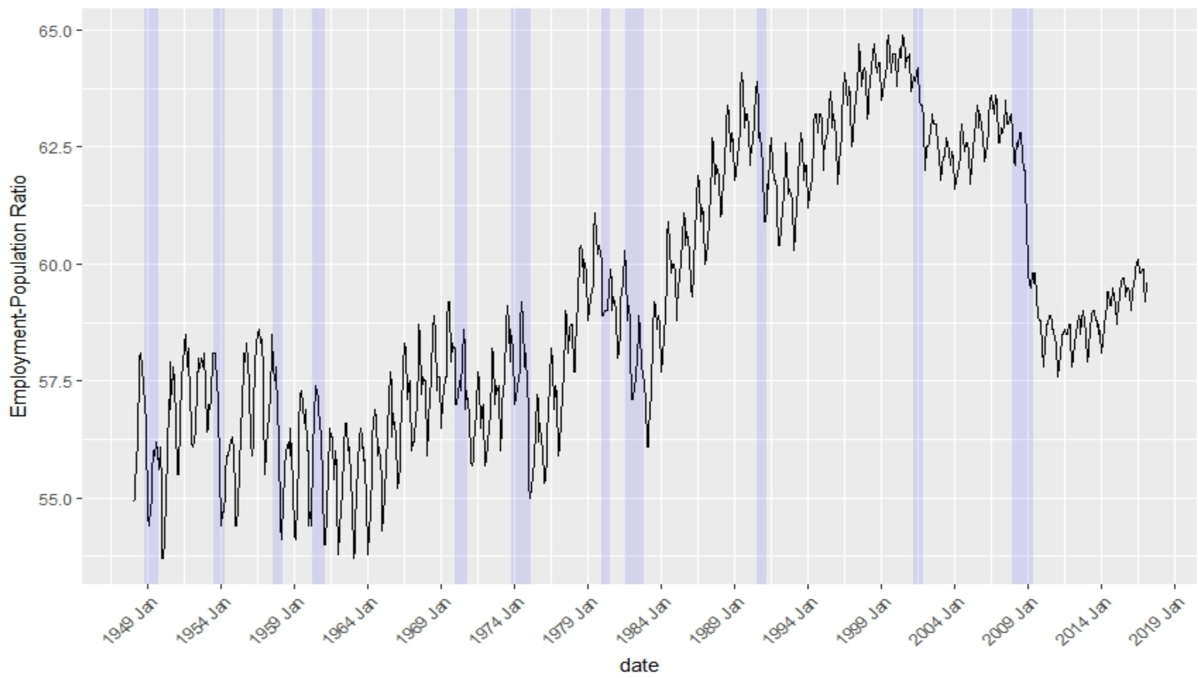


FIGURE 3.3: Total Number of Employees in Manufacturing (in thousands)- Recessions in Blue. Depicted here are the estimated total number of employees in the Manufacturing sector. Manufacturing jobs suffered a massive decline over the last recession, falling by about 30% then.

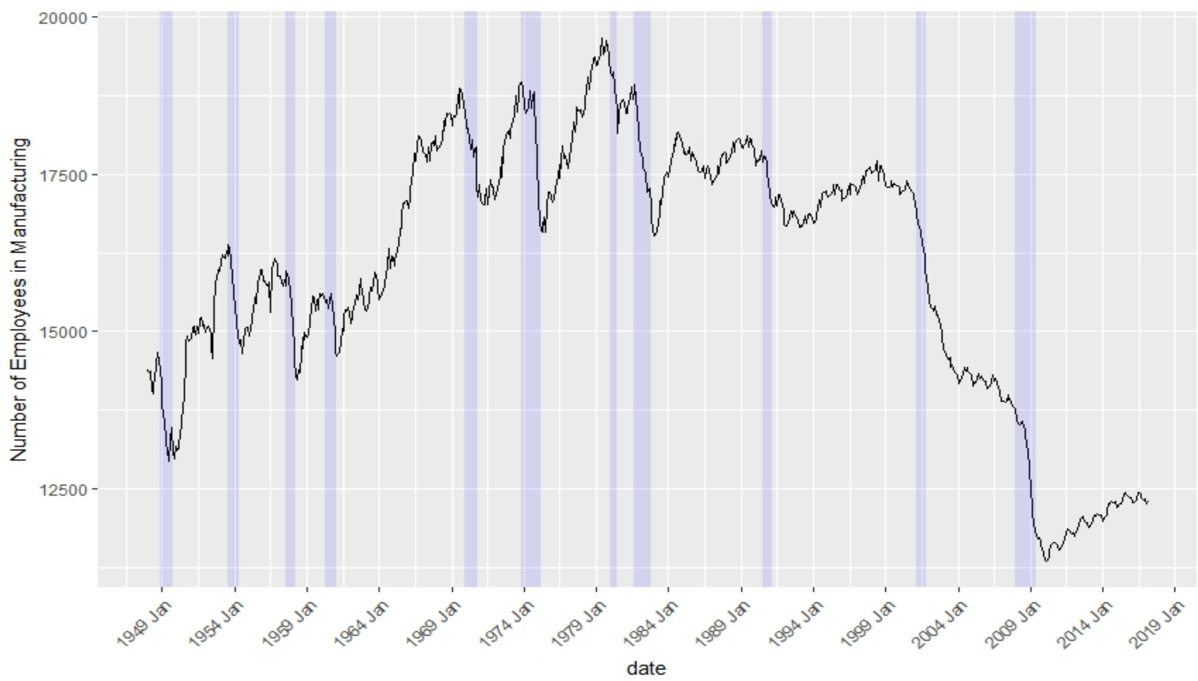


Figure 3.4: ACF/PACF - Original Data. This figure describes the autocorrelation and partial-autocorrelation functions of the original data. At the top, the autocorrelation and partial-autocorrelation functions are presented using all possible lags. The bottom two figures, conversely, display the ACF and PACF including a total of 50 lags.

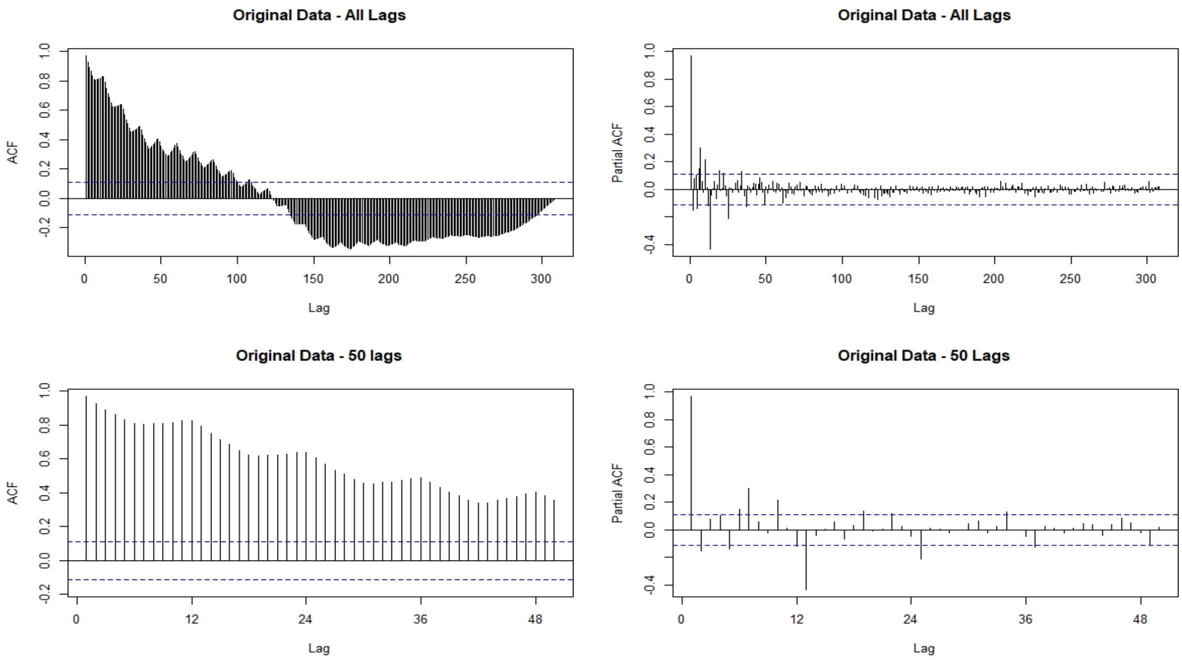


FIGURE 3.5: Detrended Data - Linear Trend. In this figure, the detrended employment-to-population time series is shown, after usage of a linear trend. The pattern suggests that the detrended data (when a linear trend is used) is non-stationary - its variance is not constant. Trend: $Y = -520 + 0.29t$; Augmented Dickey-Fuller Test - p-value = 0.36.

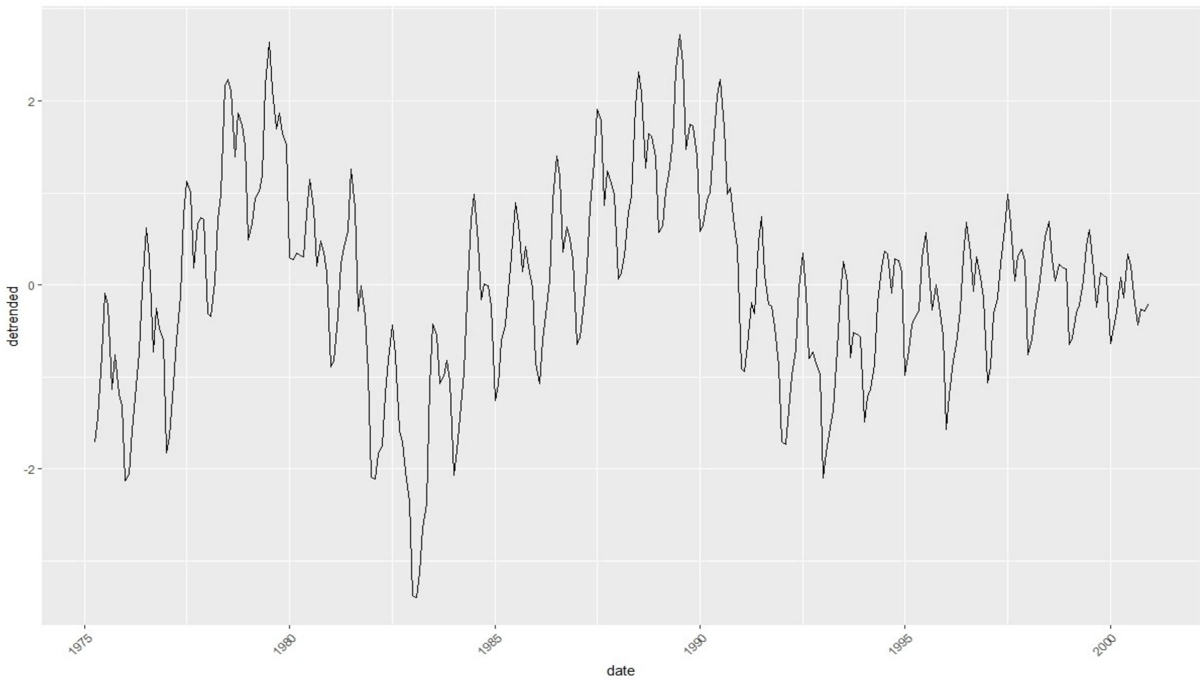


FIGURE 3.6: Smoothing Spline Filter. This figure displays the employment-to-population ratio (red) and a non-linear trend (green), which corresponds to a low frequency component in the series. The low frequency component was estimated through usage of a smoothing spline filter: $f(x) = \alpha_0 + \sum_{j=1}^{p} \alpha_j max(0, x - k_j)$. The solution is found by minimizing: $\sum_{i=1}^{n} ||y_i - \alpha_0 - \sum_{j=1}^{p} \alpha_j max(0, x_i - k_j)||^2 + \lambda \sum_{j=1}^{p} \alpha_j^2$. Observations themselves are the $k_j's$. The low frequency component itself is prone to a cyclical influence/component that displays a long period.

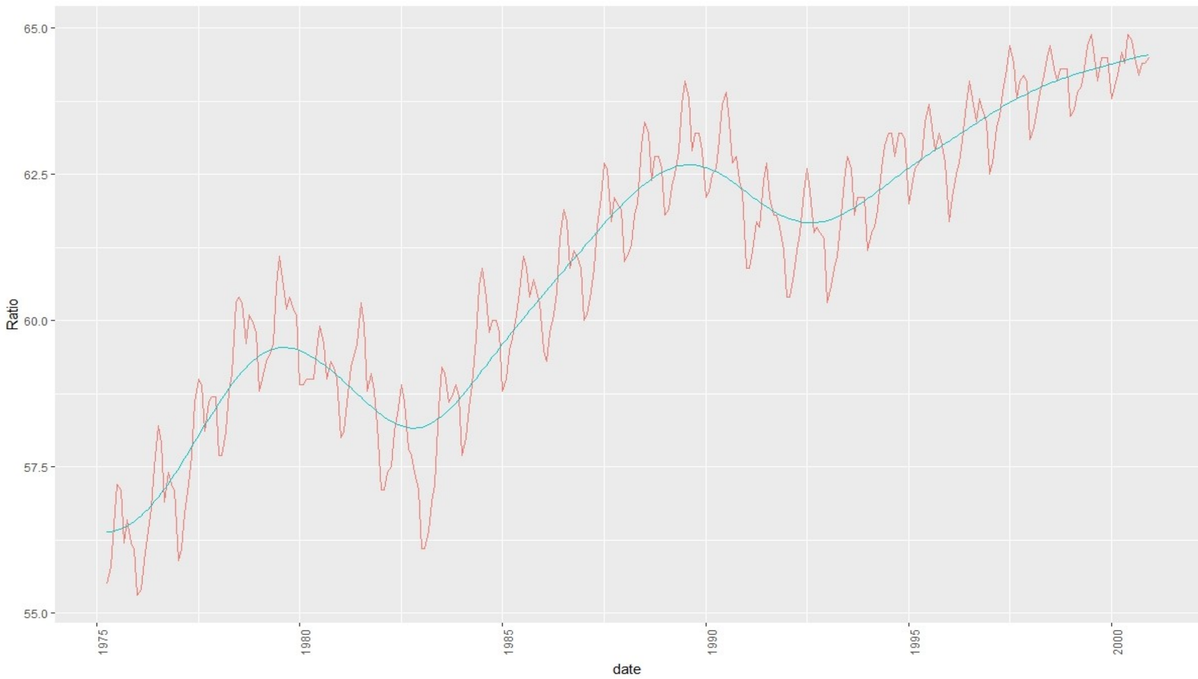


FIGURE 3.7: Filtered Data - Smoothing Splines. This graph depicts the detrended employment-to-population time series, obtained after a filter based on smoothing splines is applied to the original data. Following removal of its low frequency component - non-linear trend - the series displays an apparent stationary behavior. An Augmented Dickey-Fuller test conducted on it allows us to reject the null hypothesis that there exists a unit root in this series, at a significance level of 0.01.

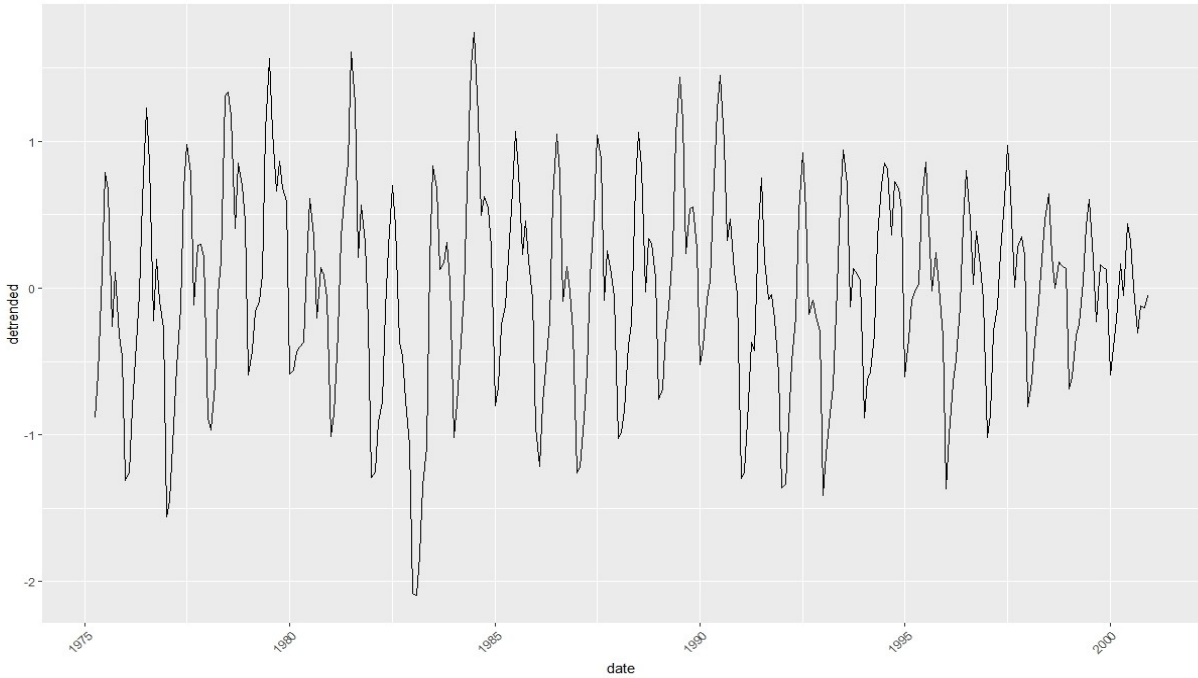


FIGURE 3.8: ACF/PACF - Filtered Data. This figure describes the autocorrelation and partial-autocorrelation functions of the filtered data. At the top, the autocorrelation and partial-autocorrelation functions are presented using all possible lags. The bottom two figures, conversely, display the ACF and PACF including a total of 50 lags. This figure illustrates that filtered/detrended series is influenced by cyclical influences.

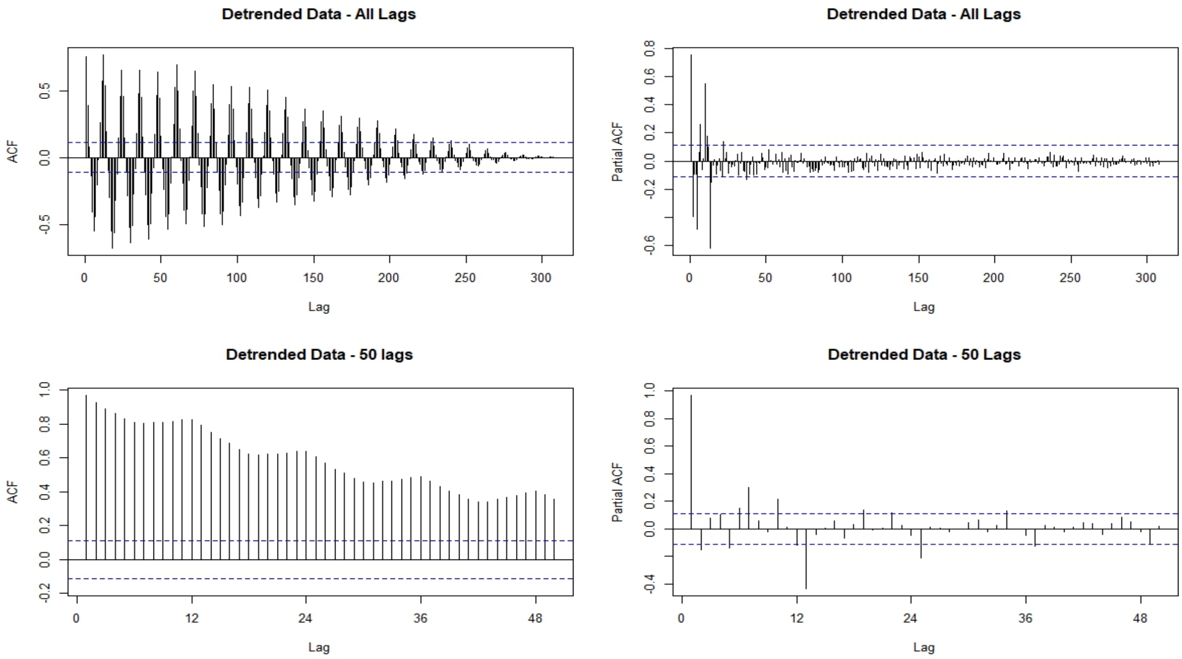


Figure 3.9: Power Spectrum - Raw Periodogram. This figure shows the raw periodogram of the filtered/detrended series. Peaks are located at $\omega = 0.188, 1.013, 1.988, 3$. These correspond to the following periods (in years), respectively: $5.33, 0.99, 0.50, 0.33$. A large proportion of the total variation in the data can be explained by the 1.013 frequency. The red dotted line indicates lower bound of the confidence interval for the spectrum corresponding to $\omega = 1$. All of the abovementioned peaks are statistically significant at a five percent level of confidence.

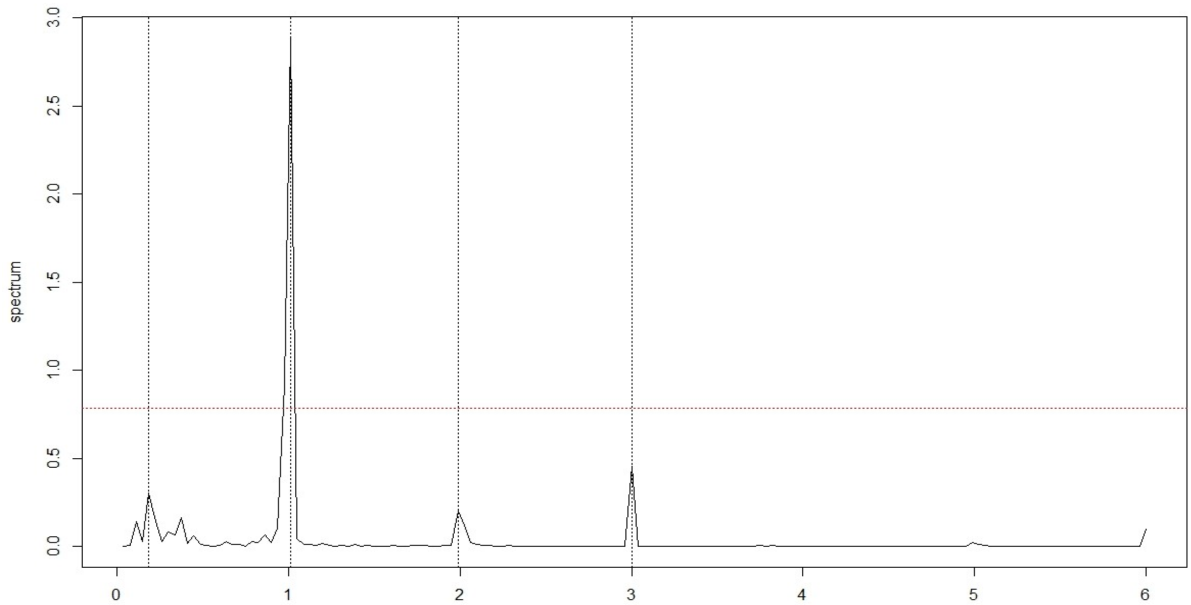


FIGURE 3.10: Yearly Seasonal Component. This figure depicts the impact in employment of its one-year cyclical component, measured after application of our smoothing spline filter. After removal of the non-linear trend through our filtering procedure, we see the prevalence of a strong employment season from March to May, followed by a weak season from October to January. The yearly cyclical component contributes to increasing employment rates by approximately one percent in April, lowering it by roughly the same amount in October. The annual seasonal component is negative from September to February, remaining positive for the remainder of the year.

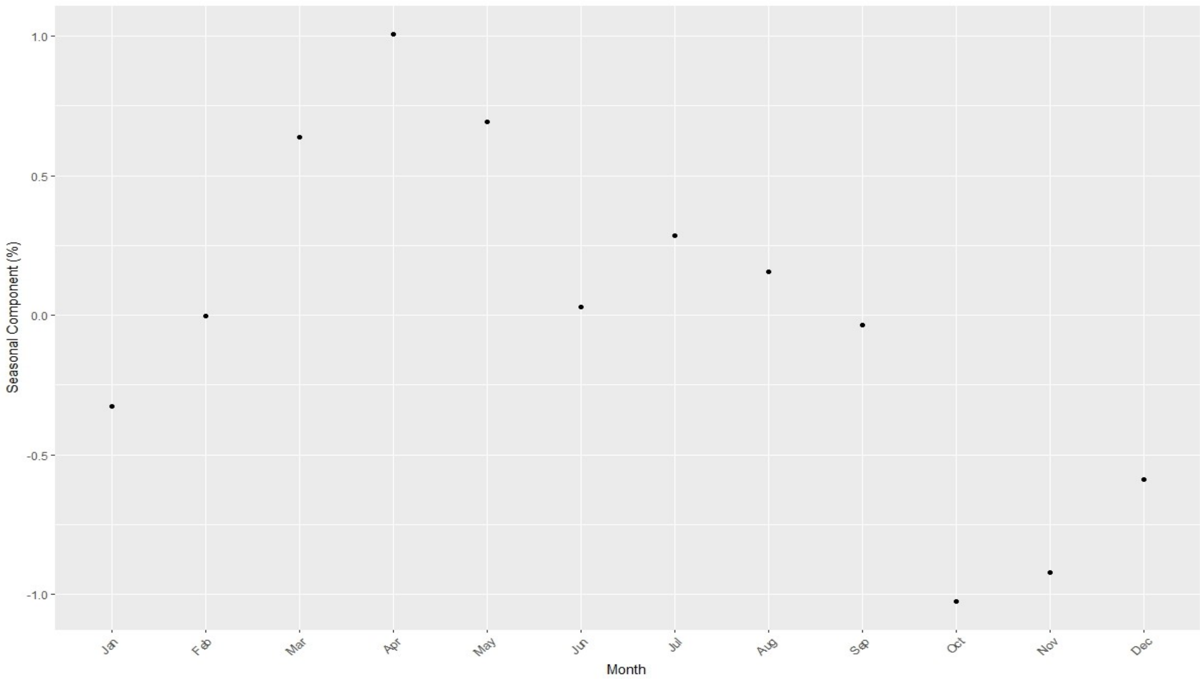


Figure 3.11: Power Spectrum - Smoothed Periodogram - Modified Daniell(2,2). In this graph, a smoothed periodogram, obtained through usage of a modified Daniell (2,2) kernel, of the detrended series is shown. As anticipated, this periodogram is smoother than the raw one. Peaks are located at the same frequencies identified earlier ($\omega = 0.118, 1.013, 1.988$ and 3), confirming the occurrence of cyclical components associated with them.

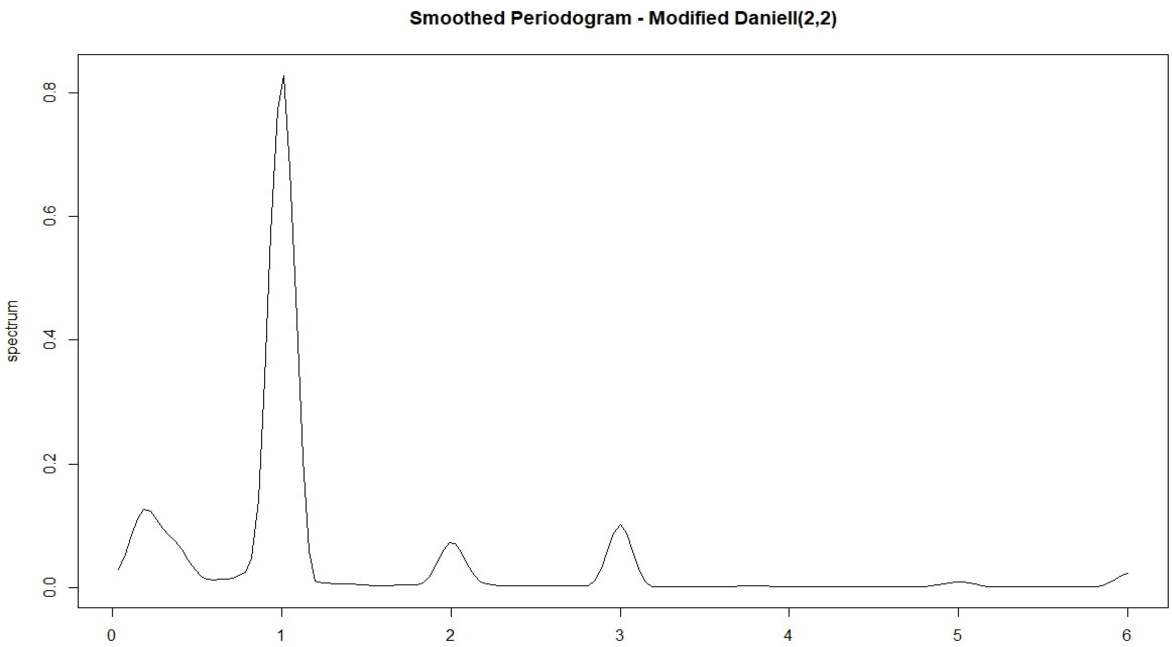


FIGURE 3.12: Autoregressive Spectral Estimator (1). This figure demonstrates what are the AIC and BIC values for autoregressive processes fit to the detrended series. The horizontal axis shows the number of lags - p - of the conjectured autoregressive process. This figure indicates that the optimal number of lags to use with an AR spectral estimator, when AIC is employed as a criteria, is 39.

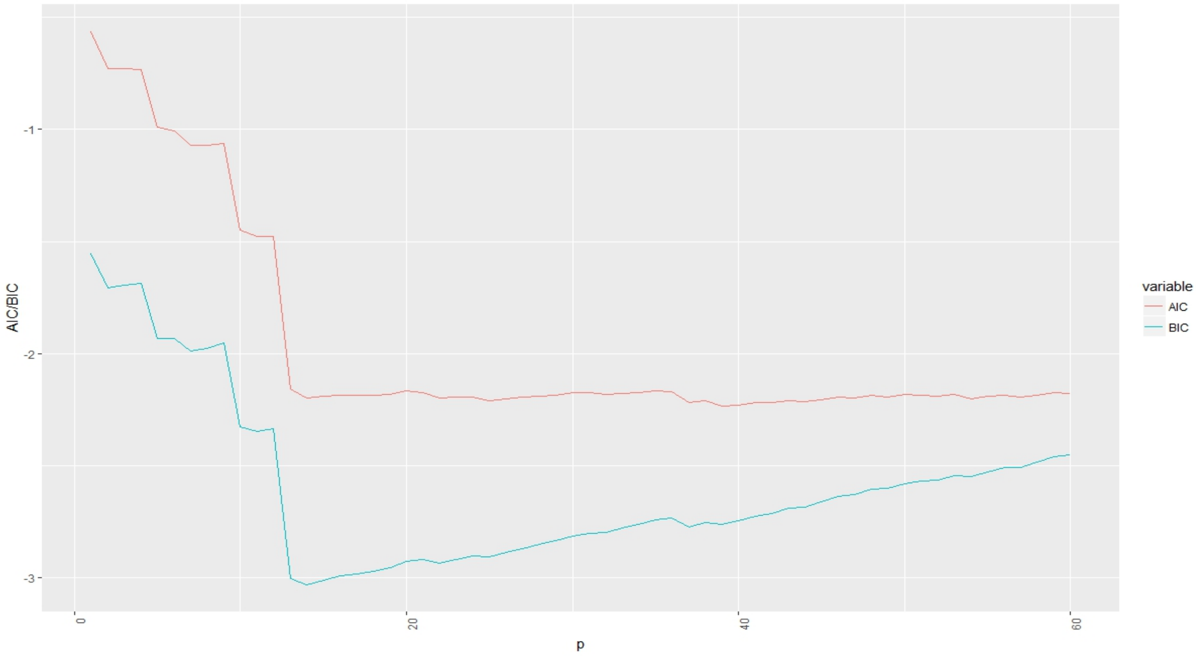


FIGURE 3.13: Autoregressive Spectral Estimator (2). This figure shows the autoregressive spectral estimator for the employment-to-population series using a model selected by AIC ($p = 39$). There are clear peaks at $\omega = 0.192$ and at $\omega \approx 1, 2$ and 3.

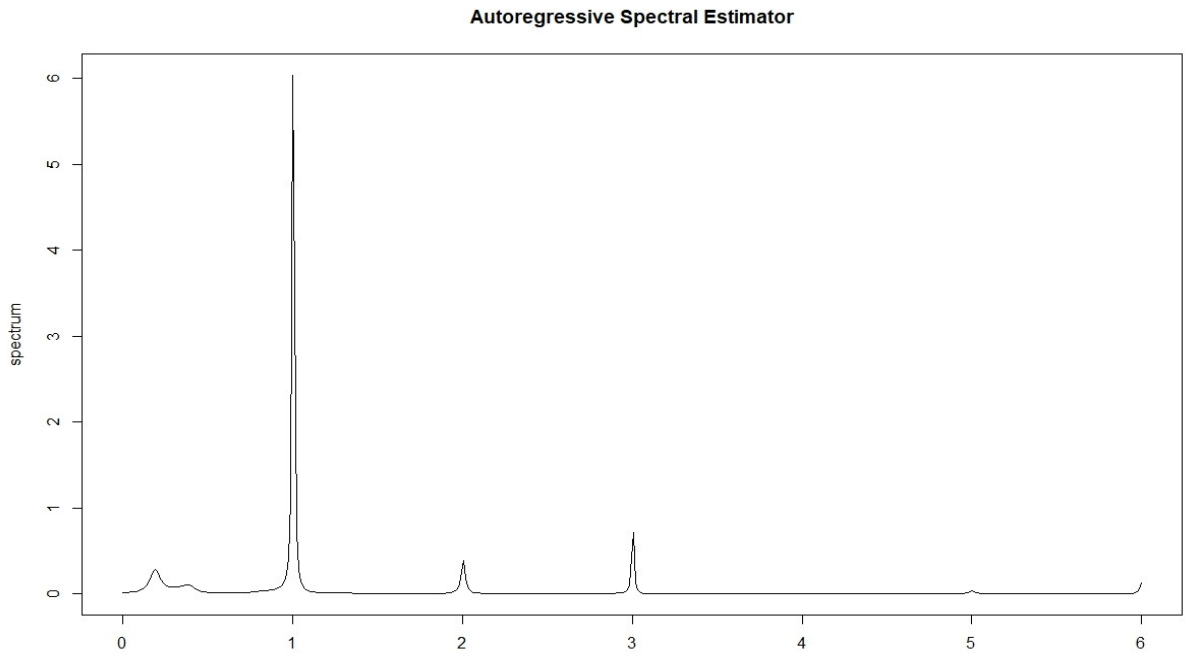


FIGURE 3.14: ACF and PACF of Detrended Data, After Removal of Select Seasonal Components. At the top, the ACF of the detrended series, after removal of the one-year seasonal component is shown. At the bottom, we depict the ACF of the series that results from removing the non-linear trend component, one-year cycle, and 5.33 years cycle. Even when influences from these two cycles are removed, the resulting series exhibits cyclical behavior.

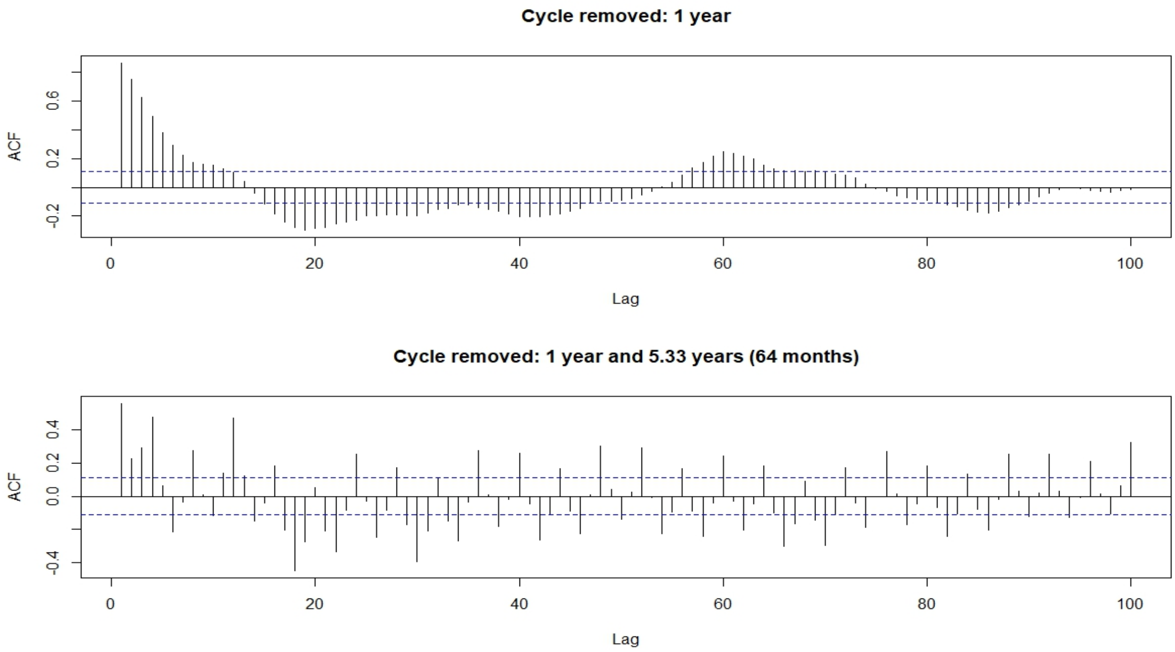


Figure 3.15: ACF and PACF of Residual Data (Trend and All Seasonal Components Have been Removed). This figure describes the autocorrelation and partial-autocorrelation functions of the series that is obtained after a non-linear trend component and seasonal influences identified via spectral analysis have been removed. At the top, the autocorrelation and partial-autocorrelation functions are presented using all possible lags. The bottom two figures, conversely, display the ACF and PACF including a total of 50 lags. While the ACF seems to tail off, the PACF cuts off at lag 14 or 15, suggesting fitting either an AR(14) or AR(15) model to the residual data.

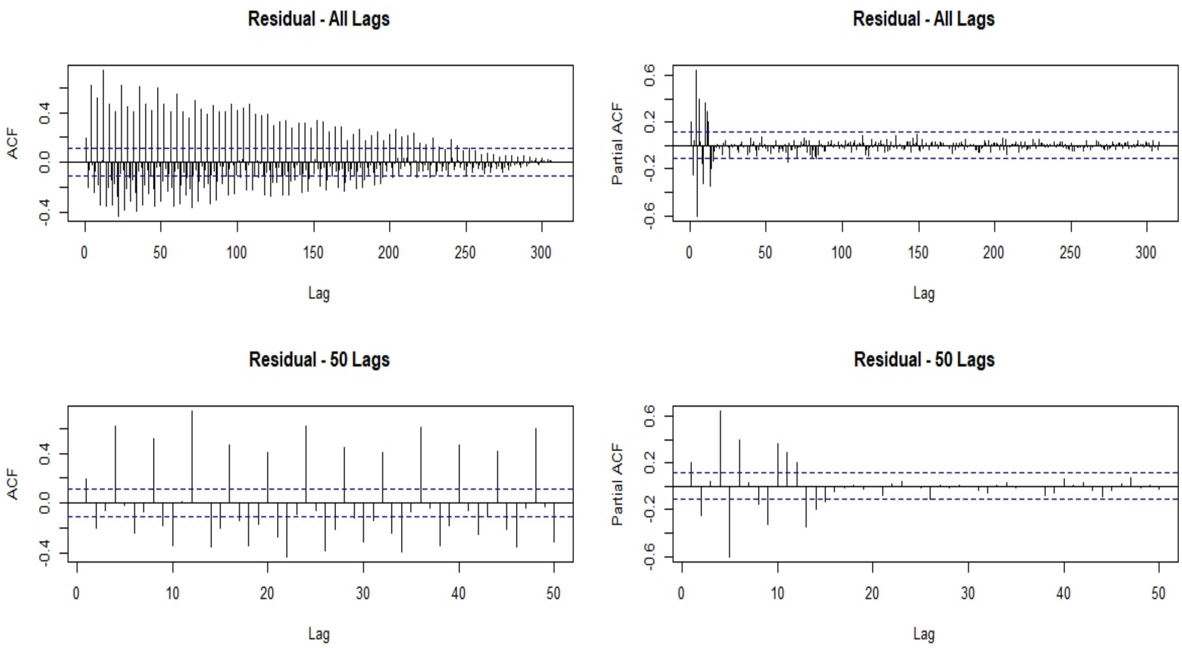


FIGURE 3.16: AIC and BIC of possible AR models. Minimum AIC at p = 14 (AIC = -2.33). This figure demonstrates what are the AIC and BIC values for autoregressive processes fit to the series obtained after a non-linear trend component and seasonal influences identified via spectral analysis have been removed. The AIC criteria suggests fitting an AR(14) model to such series.

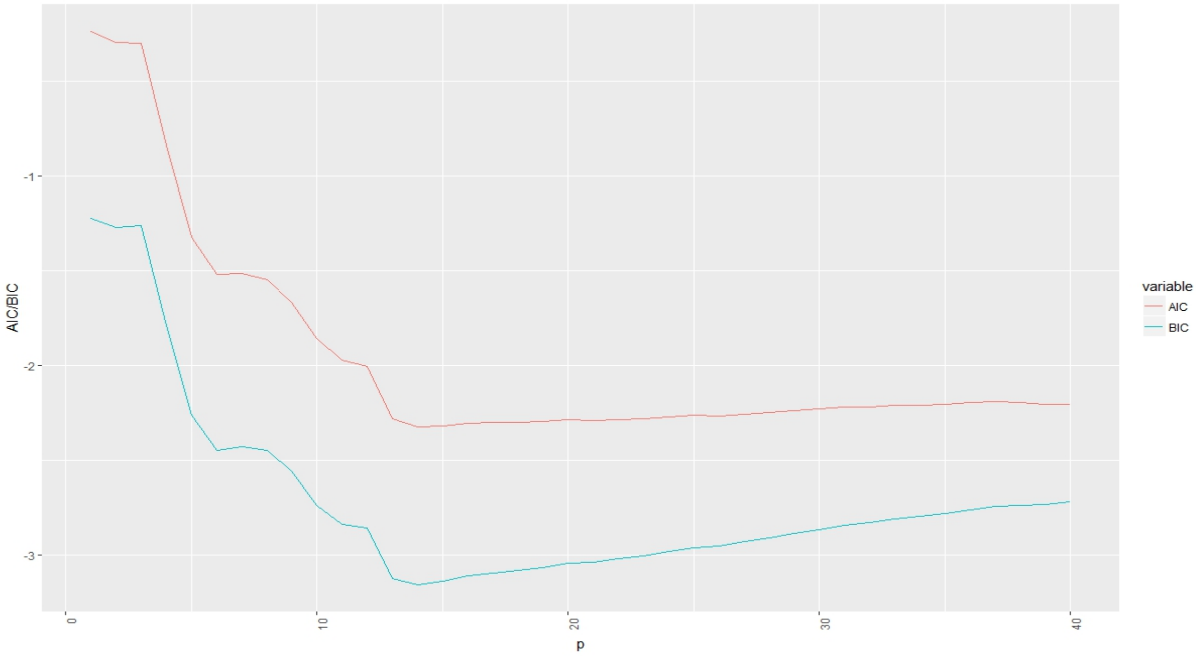


FIGURE 3.17: Diagnostics - AR(14) Model. This figure displays model diagnostics obtained after the original series is decomposed into varied components. An ARMA(14) was used to model the residual series generated following the removal of non-linear trend and cycles. A Gaussian Q-Q plot and a histogram of full model residuals are presented at the top of this figure. The ACF and a graph depicting Ljung-Box test statistics for different lags are showcased at its bottom. Although full model residuals appear to be normally distributed, Ljung-Box test statistics indicate the occurrence of serial correlation in the final residuals, when an AR(14) model is utilized as the last modeling step.

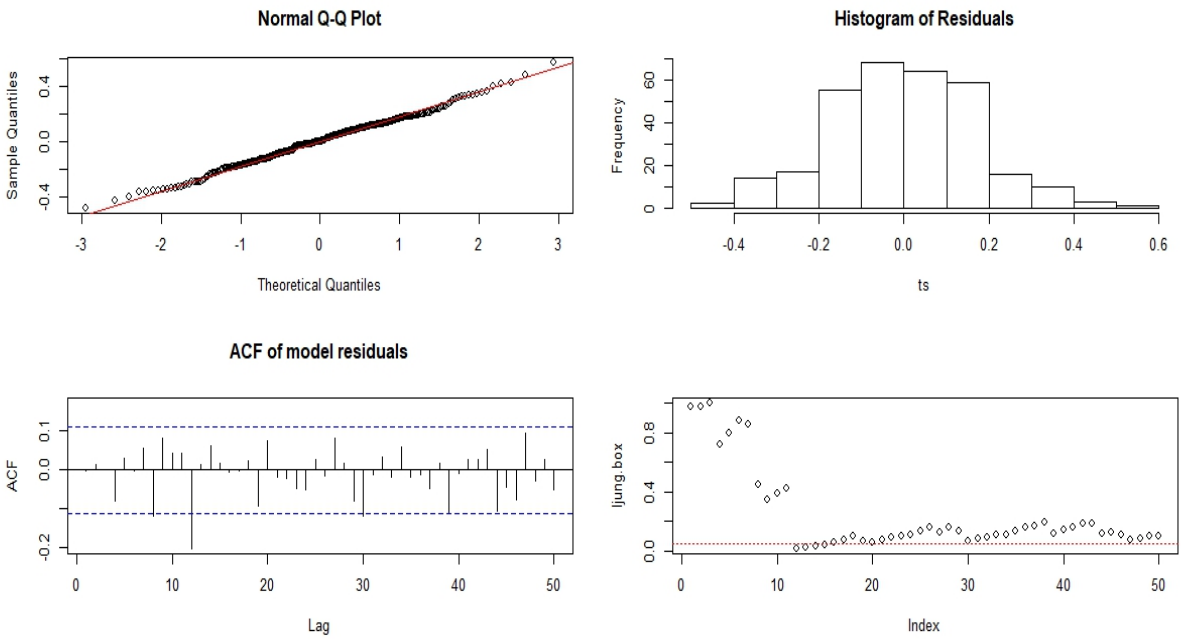
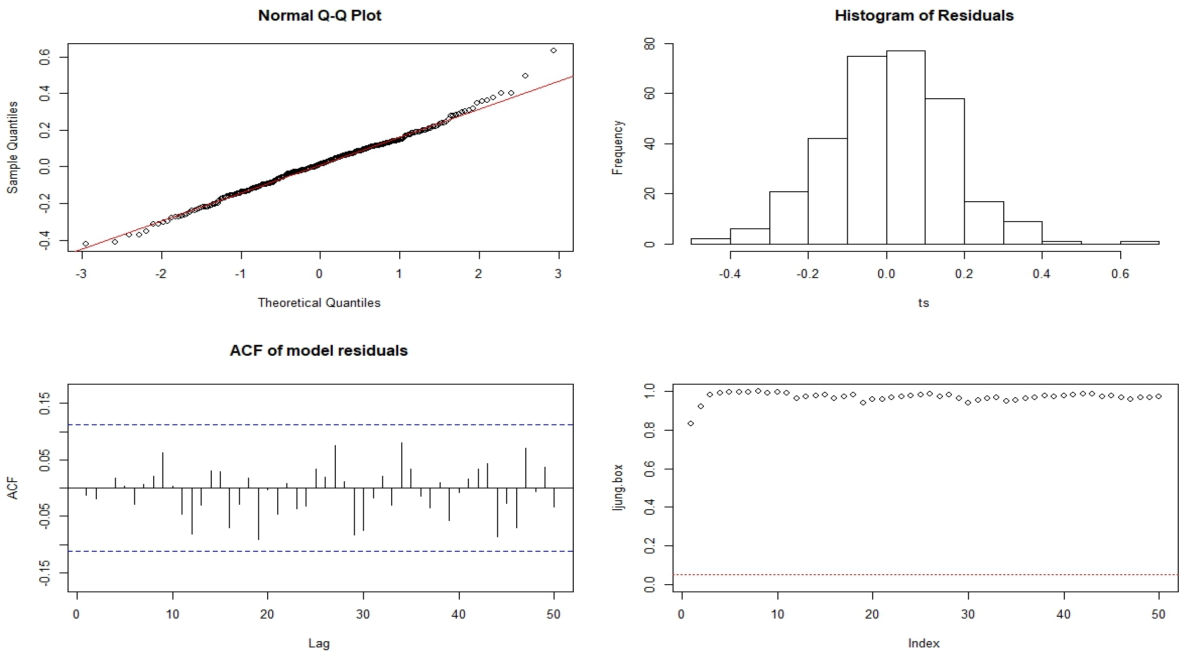


Figure 3.18: Diagnostics - ARMA(14,9) Model. This figure displays model diagnostics obtained after the original series is decomposed into varied components. An ARMA(14,9) was used to model the residual series generated following the removal of non-linear trend and cycles. A Gaussian Q-Q plot and a histogram of full model residuals are presented at the top of this figure. The ACF and a graph depicting Ljung-Box test statistics for different lags are showcased at its bottom. Full model residuals resemble normally-distributed white noise. When conducting a Shapiro-Wilk normality test, we fail to reject the null that the data comes from a normal distribution at a 5% level of significance; the ACF plot suggests that residuals engendered by the final model specification resemble a white noise process; and Ljung-Box tests indicate no presence of serial correlation in them.



CHAPTER 4

The Evolution of Global Financial Integration: A Multivariate Analysis of Currencies and Equities

(with Mahyar Kargar)

4.1 Introduction

The main objective of the study presented in this chapter is to investigate whether financial markets in general have become more integrated in recent times. One would naturally assume that the technological developments seen in the last two decades have led to increased integration between countries. Yet, some studies, such as Bekaert et al. (2009), imply that this may not necessarily have been the case. Pukthuanthong and Roll (2009), hereafter PR, propose an integration measure based on the explanatory power of a multi-factor model, using it empirically to investigate recent trends in integration between global equity markets. Our work is closely related to theirs. In the analysis that follows, we utilize the methodology developed by PR (which we outline below) to assess whether currency markets, as opposed to equity markets, have become more integrated in modern times. We also employ a simple-to-use approach to evaluating integration between different financial markets, which relies on an old and powerful statistical technique: canonical correlation analysis. An implementation of our suggested canonical correlation method is presented.

Overall, we find evidence suggesting that countries became more integrated from the mid-nineties through the early years of the twenty-first century – although the positive trend in integration verified during such period is no longer present in recent times. As PR note, "capital mobility and free trade are hallmarks of cross-country market integration". In that regard, the results shown in this study corroborate the notion that even though we have transitioned to a setting characterized by lower worldwide protectionism levels and by increased mobility of capital and goods, concrete advances on this front have been surprisingly absent lately. With the usage of multivariate statistical techniques, we quantify the magnitude of these phenomena.

This article proceeds as follows: first, we briefly outline the methods we have used to evaluate integration between countries and financial markets; next, we provide a description of the data used in the analysis, and then present and analyze our results; subsequently, we discuss possible limitations associated with this study and recommend avenues of future work related to the theme at hand; remarks containing a short summary of our conclusions are shown at the end.

4.2 Overview of the Empirical Study

Though a myriad of techniques directed towards quantifying economic/financial integration have been suggested, most studies on this subject focus on evaluating integration among stock markets (equities) - see, e.g., PR. This paper is different from these in that: (i) we evaluate integration between currencies (exchange-rates); (ii) we also assess how the relationship between these and equities has evolved through time (i.e., we analyze equity-currency integration, or "cross-market integration"). These warrant clarification.

4.2.1 Integration Among Currencies

To evaluate integration between currencies, we employ an adaptation of the algorithm suggested by PR. Put simply, PR claim that two markets are completely integrated if their returns are entirely driven by the same factors. As they explain, "A sensible intuitive quantitative measure of financial market integration is the proportion of a country's returns that can be explained by global factors. If that proportion is small, the country is dominated by local or regional influences. But if a group of countries are highly susceptible to the same global influences, there is a high degree of integration". The core idea here – from our perspective – is that, to identify integration between currencies, one should look at the factor structure of currency returns. This is precisely what we will do. In essence, for every year in the period spanning 1994 to 2016, we will extract latent factors using principal component analyses from covariance matrices of currency returns; we will then regress, on immediately subsequent years, each individual time series of currency returns on *out-of-sample* constructed latent factors. The average adjusted $R-$squared obtained from all regressions in a given year will be our measure of currency integration for that year. Note that we could have opted to use a standard factor analysis approach when extracting global latent factors. However, also notice that, as mentioned by Campbell et al. (1997), "Factor analysis represents only one statistical method of forming factor portfolios. An alternative approach is principal component analysis". To better conform with the extant literature on financial integration (and in particular with the algorithm suggested by PR), we have opted to use the latter method. A thorough description of our methodology is presented further below in Section 4.4.

4.2.2 Integration Between Equities and Currencies

We have utilized canonical correlation analyses to investigate integration between global equity markets (stock market indices) and currencies[1]. Our interest lies in evaluating the evolution of the "first" canonical correlation extracted from two sets of variables: (i) latent factors driving global equity returns; and (ii) latent factors underlying currency (exchange rate) returns. Such correlation, as we see it, can be interpreted as a measure for how much "equity factors" and "currency factors" are intrinsically related to each other.

[1]A comprehensive description of this approach is shown in Section 4.4.

High values for it should indicate that equity factors and currency factors are potentially closely associated - this, in turn, would suggest a large level of integration between these two asset classes[2]. Though this measure of integration is far from perfect, we believe its simplicity and the possible insights that can be drawn from it justify it being used in studies of this type.

4.3 Description of Data

4.3.1 Currencies

The exchange rates we have used were obtained from DataStream, a division of Thomson Reuters (please note that, henceforth, we will refer to the terms “currency” and “exchange rate” interchangeably). DataStream is a news-data system widely used in financial studies – it provides its users with historical global coverage of stock markets, derivatives, currencies, bond markets and economic data. Such exchange rates, we should note, were all in units of foreign currency per one U.S. dollar. Thirteen currencies were included in the study (outlined below). In order to evaluate whether currencies have become more integrated with each other in recent times, daily prices from 12/31/1993 to 12/29/2017 were used[3].

Time series of dollar-denominated daily currency (exchange rate) prices for the following thirteen countries/regions were all included in our study: Australia, Canada, the Eurozone, India, Japan, New Zealand, Norway, Singapore, South Africa, South Korea, Sweden, Taiwan and the United Kingdom. Collectively, these currency pairs account for nearly seventy percent of the global foreign exchange market (BIS (2016)). Over the

[2]The central idea behind canonical correlation analysis is related to explaining the relationship between two sets of variables. As such, we believe this is a relevant tool for dealing with market integration issues such as the one investigated here.

[3]WM/Reuters rates were utilized in our study. With the exception of spot rates pertaining to the Euro or British pound, WM/Reuters exchange rates are available as of 12/31/1993 for the currencies used in our main data set.

sample period, from 12/31/1993 to 12/29/2017, there are approximately 81,000 daily observations of exchange rate data for the countries/regions in our currency data set.

An important aspect of our sample is that none of the currencies included in it were subject to an official strict peg relative to the U.S. Dollar during the sample period. Another aspect of it, we note, is that it is composed of currencies from countries/regions that are geographically dispersed.

4.3.2 Equities

Our equity-currency integration study involves one currency and one stock market from each of the aforementioned thirteen countries. The equity data we have used thus consists of thirteen time series of prices (levels) of stock market indices from around the globe. As pointed out by PR, "DataStream, a division of Thomson Financial, provides stock market indices for the most countries and longest time periods". Thus, we have also opted to use DataStream as our source for stock market prices. The equity indices and countries associated with them are shown in Table 4.1. All equity prices are daily and correspond to the period that goes from 12/31/1993 to 12/29/2017. Notice that some countries have numerous equity indices associated with them in DataStream. Again, following PR, to determine which index to choose for each country, we relied on a simple data availability criteria: the stock index with the longest period of data availability in DataStream was chosen to represent each country.

4.4 Analysis and Results

4.4.1 Integration Among Currencies

4.4.1.1 Methodology

Initial Comments

In this study, we use principal component analysis as a statistical method for identifying orthogonal latent factors driving currency returns. As we have alluded to earlier, we do so to better conform with the literature on financial market integration (standard factor analysis could also have been used towards that same purpose). The exact approach we use here (described in the next subsection) was suggested by PR. Note that the usage of PCA as a factor extraction technique is commented in numerous books, including Campbell et al. (1997) and Tsay (2010). According to Campbell et al. (1997), "Principal components is a technique to reduce the number of variables being studied without losing too much information in the covariance matrix. In the present application (i.e., using it as a latent factor extraction tool), the objective is to reduce the dimension from N asset returns to K factors. The principal components serve as the factors. The first principal component is the (...) linear combination of asset returns with maximum variance. The second principal component is the (...) linear combination of asset returns with maximum variance of all combinations orthogonal to the first principal component, and so on".

Our approach to analyzing global currency integration can be divided into two stages. Each of these is explained in detail below.

Stage 1: Extraction of latent factors from currency returns

We start by computing daily returns for all the thirteen exchange rates in our sample. These were calculated as log differences in prices: i.e., the daily return for exchange rate

i on day t was computed as:

$$\ln(S_{j,t}) - \ln(S_{j,t-1})$$

where $S_{j,t}$ denotes the price (or level) of the exchange rate for country j on day t in units of country j's currency per one U.S. dollar.

Covariance matrices of daily currency returns were used as inputs to our principal component analyses (PCA). One covariance matrix was computed each year; naturally, PCA's were conducted using each of these "yearly" matrices. Figure 4.1 plots the average (across all years) cumulative percentage of variance explained by principal components. As shown, the first principal component explains about 51% of the variation in the data, and three principal components are required to explain just over 75%. This result is analogous to the one verified by PR in their study of stock market integration. It corroborates the notion that exchange rate returns are driven by several latent global factors.

Figure 4.2 plots the cumulative variance explained in each estimation year by each of the first eight principal components. There is some time-series variation, as expected. The total cumulative variance explained by the first six principal components, nevertheless, lies above 90% throughout most of the sample years. In light of this result, we decided to keep the first six principal components as proxies for latent global factors in currency markets – please refer to Section 4.5 for further comments on this choice.

Subsequently, the eigenvectors computed in year $y-1$ were applied to currency returns in year y, thus generating *out-of-sample* principal component estimates (for $y = 1995$ to $y = 2017$). The first six estimated out-of-sample principle components were then taken as latent factors affecting currency markets.

Stage 2: Principal component regressions

The estimated latent global factors (out-of-sample principal components) served as independent variables in our regression analyses. Each available time series of currency returns was regressed on the six out-of-sample principal components for each available year from 1995 to 2017.

Let $r_{j,t}^{FX}$ denote the continuously compounded return of exchange rate j on day t. Let $f_{i,t}$ for $i = 1, 2, \ldots, 6$ be the estimated return for latent factor i on day t.

For each currency in the sample, and for every year, we run the following regression:

$$r_{j,t}^{FX} = \alpha_{j,t} + \beta_{1,j} f_{1,t} + \beta_{2,j} f_{2,t} + \beta_{3,j} f_{3,t} + \beta_{4,j} f_{4,t} + \beta_{5,j} f_{5,t} + \beta_{6,j} f_{6,t} + \epsilon_{j,t} \tag{4.1}$$

where, $r_{j,t}^{FX} = \ln S_{j,t} - \ln S_{j,t-1}$ is the continuously compounded return of exchange rate j from day $t-1$ to day t. Note that $S_{j,t}$ is country j's exchange rate on day t per one U.S. dollar, as defined above.

The average (among all currencies) adjusted $R-$squared from all regressions conducted in a given year is the measure of aggregate global financial integration suggested by PR for that year. We are mainly interested in assessing how our currency-based measure has evolved through time.

4.4.1.2 Results

Table 4.2 shows the time series of average adjusted $R-$squared values from 1995 to 2017. Figure 4.3 depicts these results. There is a general upward trend in average adjusted $R-$squared values in the early portion of our sample period, suggesting economies have become more integrated then. Interestingly, however, we see that such trend appears to have waned and is no longer evident in recent times. We will revisit this issue in Section 4.5, and provide an analysis of these results in conjunction with those from our equity-currency study further below.

4.4.2 Integration Between Equities and Currencies

4.4.2.1 Methodology

A two-stage approach was also used to investigate global equity-currency integration. Our methodology is explicated below.

Stage 1: Extraction of latent factors from currency returns and from equity returns

The method we have used in this "stage" shares many similarities with the one we employed when investigating integration among currencies. Stock market indices from thirteen countries (shown in Table 4.1) and spot exchange-rates from these same countries were included in the analyses. All levels (i.e. prices) associated with these stock indices were translated into a common currency – the U.S. dollar. The thirteen spot exchange rates used in the study were also converted to a "U.S. dollar basis". Daily prices from 12/31/1993 to 12/29/2017 were used[4].

After converting all prices in our data set to U.S. dollar terms, we then computed daily returns for each of the 26 time series[5] in our sample. Similarly as done with currencies (see Section 4.4.1), daily returns were computed as log differences in prices. This procedure resulted in two sets of time series of dollar-denominated daily returns, spanning over two decades – one containing returns from 13 stock market indices and one with returns from 13 corresponding currencies. Hereafter we'll refer to the former as our set of X variables, and to the latter as our set of Y variables.

Next, for every calendar year, a covariance matrix for the X variables was computed.

[4]Our time series of prices for the Japanese currency, for example, was expressed in terms of Yens per unit of U.S. dollar. Similarly, Japan's stock market index (the TOPIX), was also expressed in terms of Yens per unit of U.S. dollar. The same applies to other exchange rates and equity indices.

[5]13 time series of currency prices and 13 time series of corresponding stock (index) prices.

Principal components were then extracted from each of these covariance matrices. Table 4.3 shows the cumulative percentage of variance explained by each of the first ten principal components obtained from equity returns. On average, six principal components are responsible for explaining a large amount, 85.6%, of the total variation found in the data. In our canonical correlation analyses, we have opted to use the first six principal components (computed every year, and on an out-of-sample basis) as proxies for global latent factors affecting equity returns[6].

A similar analysis was implemented using our set of Y variables. As previously mentioned in Section 4.4.1, six principal components are also capable of largely explaining the variation contained in exchange rates. Our proxies for latent factors affecting currency markets will also be given by six principal components (i.e., eigenvectors associated with the first six PCs obtained from currency returns every year, applied to returns evidenced on immediately subsequent years).

Stage 2: Canonical Correlations

For every year from 1995 to 2017, we conduct a canonical correlation analysis between the six latent factors associated with our X (equity) variables, and the six factors associated with our Y (currency) variables. The correlation associated with the first pair of canonical variates obtained in each year is our measure of global equity-currency market integration for that year.

4.4.2.2 Results and Analysis

Figure 4.4 depicts our results. The "first" canonical correlation between equity factors and currency factors has increased from nearly 0.62 in 1995 to about 0.87 in 2017. This pattern indicates that global equity markets and currency markets have become increas-

[6]Further comments on this selection method are presented in the discussion section.

ingly related to each other. Our results show that the association between global equity returns and currency returns has become overall more pronounced over the last twenty three years. In Section 4.4.1, we showed that six latent factors impacting currencies have, generally speaking, more and more been able to explain currency returns around the globe. Here, we see that these same factors have apparently become more intrinsically close to those driving equity returns. Taken together, these results buttress the notion that economies and financial markets have, in general, become more integrated since the mid-nineties.

Furthermore, note that Figure 4.3 and Figure 4.4 are strikingly similar. These two figures suggest that the simple canonical correlation measure utilized in this study may be "as good as" PR's measure, in terms of evaluating patterns in financial market integration. Finally, a somewhat surprising result is that here too we see that the upward trend in global integration has faced headwinds recently (a similar result was obtained in our analysis of integration among currencies). Currencies and global equities shared a strong positive association throughout the initial years present in our sample period. However, the strength of this relationship has quickly fallen over the last years. We discuss this phenomenon further below.

4.5 Discussion

4.5.1 Limitation and Suggestions for Future Work

4.5.1.1 Trend in Integration faces Headwinds

There is evidence of a rather strong recent slowdown in the pace of increase in financial integration, and the positive trend that predominated in the early portion of our sample is no longer evidenced in recent times. Given this is a somewhat surprising result, we have opted to further investigate it by computing the first-order canonical correlation between equity and exchange rate factors from 2007 to 2017 while using an expanded data set. Our expanded data set is composed of the same thirteen countries/regions used previously in the canonical correlation study shown in Section 4.4.2 plus seven additional countries[7]. These, and their respective stock market indices, are shown in Table 4.4.

Table 4.5 shows the cumulative percentage of variance explained by each of the first nine principal components extracted from covariance matrices of daily returns constructed using all 20 equity indices. On average, the first six principal components can explain more than 81% of the total variation in the data.

Figure 4.5 depicts the results of our expanded canonical correlation study. Again, we see no evidence of an upward trend in integration during the later portion of our sample period. Fitting a linear model to our first-order canonical estimates, while using as an explanatory variable simply time - i.e., the year when the estimate was computed - produces a non-significant slope coefficient of -0.005 (p-value $= 0.21$).

To more thoroughly investigate this finding, we also re-conducted the analysis done in Section 4.4.1 using all twenty currencies from countries/regions in the aforementioned expanded data set, utilizing data from 2007 to 2017. Table 4.6 and Figure 4.6 display the outcomes of this approach. Once more, there is evidence of a decrease in financial integration over the more recent portion of our sample period. Understanding the reasons behind such drop should be the subject of future research, as it may potentially help us gain further knowledge into the forces ultimately driving global financial integration.

4.5.1.2 Factor Extraction

Admittedly, one could argue that the number of factors we have used in our analyses was to some extent subjective. Why not keep 7 or 8 factors, for example? First, we note that we have conducted all analyses using 7 and 8 factors, and similar patterns in integration

[7]Similarly as with our main sample, none of the currencies included in our expanded sample were subject to an official strict peg relative to the U.S. Dollar during the period in which such sample is used.

were verified. More importantly, perhaps the best way to address this dimensionality issue would be to use the Bayesian principal components method developed by Bishop (1998). By imposing some structure to our currencies and equities data (e.g., assuming returns are normally distributed), one could tackle such issue in a more structured fashion. This, we believe, constitutes a potentially fruitful area that may be subsequently explored. We are not aware of any study of financial integration that uses Bayesian principal component analysis.

4.5.1.3 Similarities in Results

Again, we note that Figure 4.3 and Figure 4.4 apparently depict similar patters in financial integration. These figures share a striking resemblance, and support the proposition that our approach, assessing the first-order canonical correlation between latent factors extracted from different asset classes, may indeed represent a viable alternative way to effectively measuring global financial integration. This basic algorithm should be subject to further scrutiny in future studies.

4.5.1.4 Country-specific integration

Table 4.7 shows mean adjusted $R-$squared values associated with the "expanded" regressions discussed in Section 4.4.1, above. We see that country-specific integration measures are highest for South Africa, New Zealand and the Eurozone. Not surprisingly, this measure is lowest for China. Indeed, China is well-known for its long-lasting barriers on both investment and trade. Investigating potential causes for these cross-sectional differences in financial integration may represent another way in which our work may be expanded over the coming years.

4.5.2 Final Remarks

Numerous approaches to measuring financial integration have been proposed in the extant literature. These, in turn, have at times led researchers to draw dissonant conclusions about the "size" of the so-called "globalization" phenomenon. The results shown in this chapter, in particular, help substantiate the notion that increases in financial integration have been ubiquitous from the mid-nineties to the early years of the twenty-first century. Our analysis suggests that a small number of global latent factors was then capable of explaining an increasing amount of the variation found in currency returns. The very same latent factors, furthermore, became more and more intrinsically related to the factors driving equity returns across the globe during those years. When combined, these results buttress the notion that economies and financial markets became progressively integrated with each other then. It is important to note, however, that the positive trend in integration that predominated during such period gradually waned and is no longer visible in recent times. This latter result, in our view, should be the focus of future research, as it may provide insights into the driving forces underlying global financial integration.

TABLE 4.1: Countries and DataStream mnemonics associated with the 13 major equity indices used in our analysis. All index values were converted to U.S. dollar values. An index with the designation "RI" is a total return index (with reinvested dividends). The designation "PI" denotes a pure price index.

Country	Index Identification	DataStream mnemonic
Australia	Australia–DS Market $ TOT Return IND	TOTMAU$(RI)
Canada	S&P/TSX Composite Index TOT Return IND (~U$)	TTOCOMP(RI)~U$
Germany	DAX 30 Performance TOT Return IND (~U$)	DAXINDX(RI)~U$
India	S&P BSE (100) National Price Index (~U$)	IBOMBSE(PI)~U$
Japan	Topix TOT Return IND (~U$)	TOKYOSE(RI)~U$
New Zealand	New Zealand–DS Market $ TOT Return IND	TOTMNZ$(RI)
Norway	Norway–DS Market $ TOT Return IND	TOTMNW$(RI)
Singapore	Singapore–DS Market EX TMT-Return IND (~U$)	TOTXTSG(RI)~U$
South Africa	South Africa–DS Market $ TOT Return IND	TOTMSA$(RI)
South Korea	Korea SE Composite (KOSPI) Price Index (~U$)	KORCOMP(PI)~U$
Sweden	OMX Stockholm (OMXS) Price Index (~U$)	SWSEALI(PI)~U$
Taiwan	Taiwan SE Weighed TAIEX Price Index (~U$)	TAIWGHT(PI)~U$
United Kingdom	UK–DS Market $ TOT Return IND	TOTMUK$(RI)

FIGURE 4.1: Average cumulative percentage of variance explained by principal components extracted from daily currency returns. Eigenvalues associated with principal components were sorted from largest to smallest. Average percentages were taken over all years in the period that goes from 1994 to 2016. The first six principal components, on average, can explain approximately 90.3% of the cumulative variance of currency returns. The sample consists of daily currency returns from 13 countries/regions: Australia, Canada, Eurozone, India, Japan, New Zealand, Norway, Singapore, South Africa, South Korea, Sweden, Taiwan and United Kingdom.

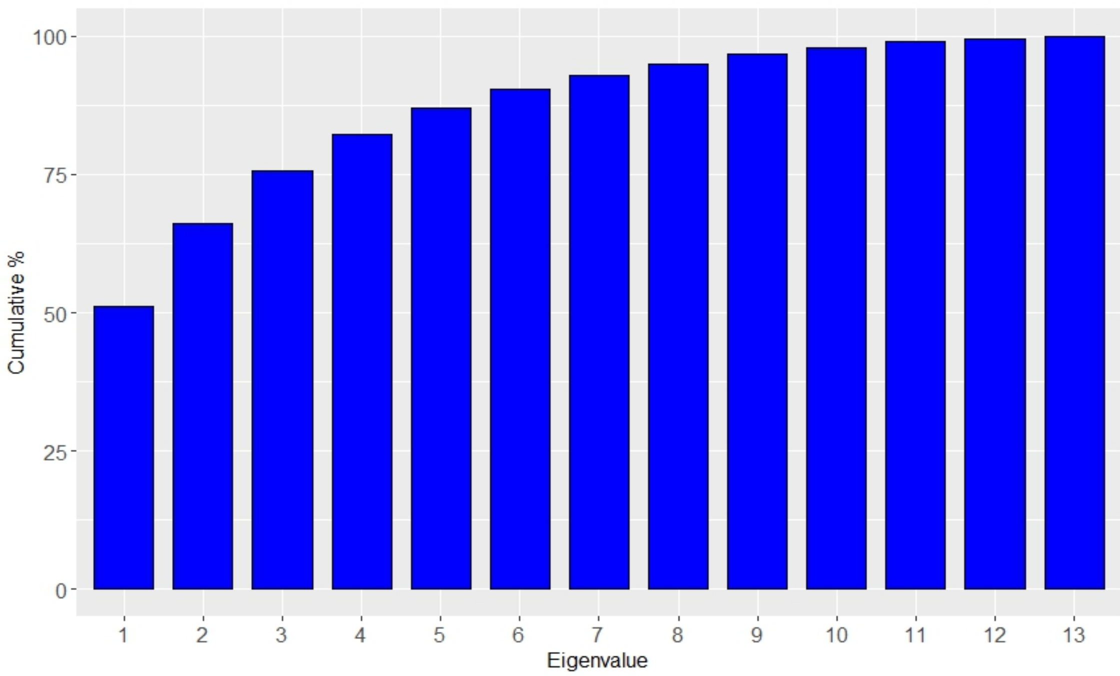


Figure 4.2: Cumulative percentage of variance explained by principal components extracted from covariance matrices of daily currency returns. In most years from 1994 to 2016, the first six principal components can explain at least 90% of the cumulative variation of daily currency returns. The sample consists of daily currency returns from 13 countries/regions: Australia, Canada, Eurozone, India, Japan, New Zealand, Norway, Singapore, South Africa, South Korea, Sweden, Taiwan and United Kingdom.

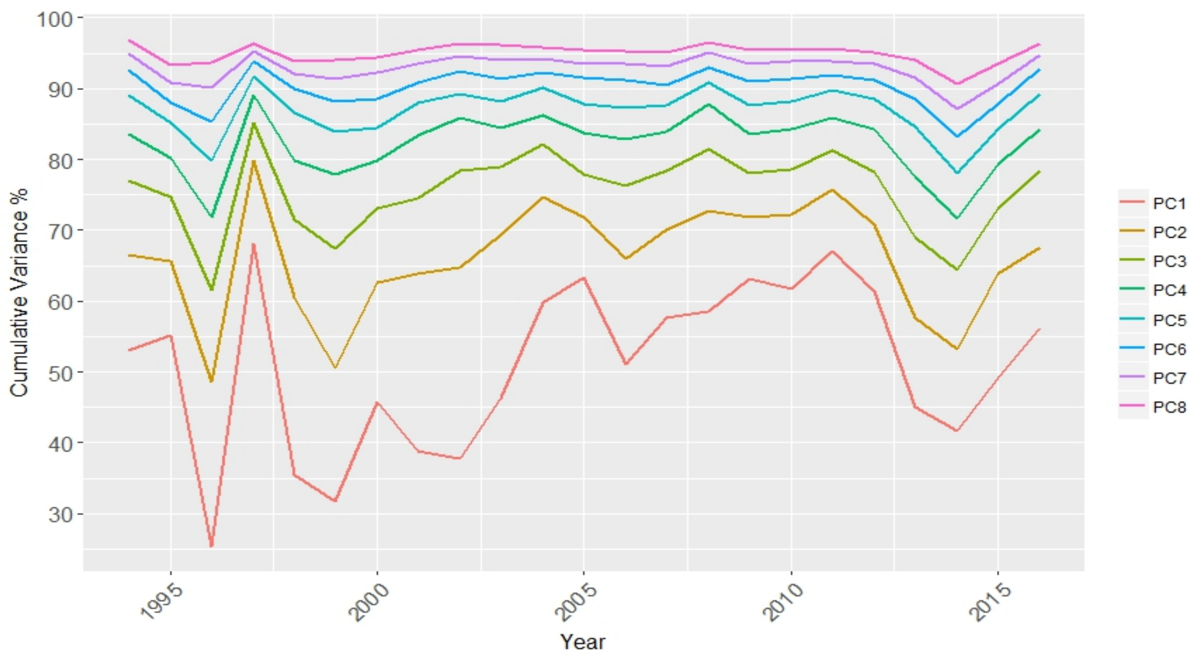


TABLE 4.2: Mean adjusted $R-$squared values from regressions of daily currency returns on 6 latent global factors (see Equation 4.1). Each time series of currency returns was regressed on 6 global factors every year. Adjusted $R-$squared values, computed every year from 1995 to 2017, were then averaged across all 13 countries/regions in our sample. We use this as a measure of integration among foreign exchange markets. The sample consists of daily currency returns from 13 countries/regions: Australia, Canada, Eurozone, India, Japan, New Zealand, Norway, Singapore, South Africa, South Korea, Sweden, Taiwan and United Kingdom.

Year	Mean adjusted $R-$squared
1995	0.673
1996	0.618
1997	0.643
1998	0.700
1999	0.699
2000	0.677
2001	0.686
2002	0.746
2003	0.743
2004	0.800
2005	0.828
2006	0.825
2007	0.758
2008	0.755
2009	0.811
2010	0.846
2011	0.852
2012	0.826
2013	0.794
2014	0.714
2015	0.785
2016	0.823
2017	0.785

Figure 4.3: Evolution of mean adjusted $R-$squared values. This figure plots mean adjusted $R-$squared values from regressions of daily currency returns on 6 latent global factors extracted via PCA (see Equation 4.1). The average adjusted $R-$squared is used as a measure of integration among foreign exchange markets. There is an upward trend in average $R-$squared values from the mid-nineties to the early years of the current century. Such trend is subsequently no longer visible. The sample consists of daily currency returns from 13 countries/regions: Australia, Canada, Eurozone, India, Japan, New Zealand, Norway, Singapore, South Africa, South Korea, Sweden, Taiwan and United Kingdom.

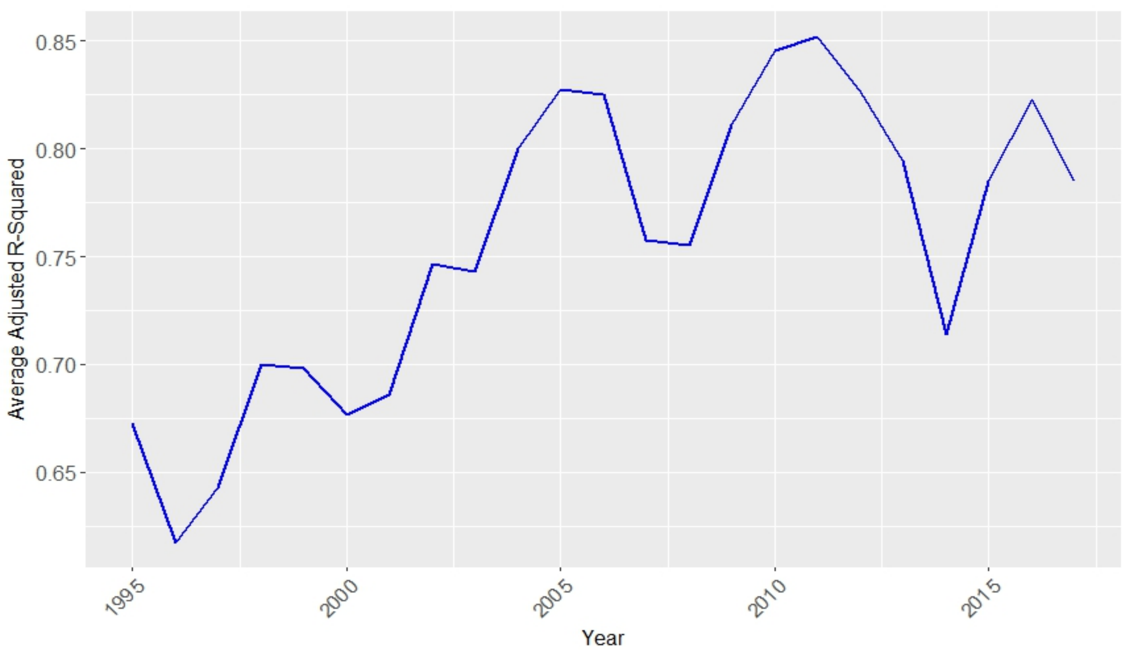


Table 4.3: Cumulative percentage of variance explained by principal components extracted from covariance matrices of daily returns from 13 major equity indices. In every year, principal components were extracted from our equity data set. The sample being used is composed of 23 years of equity returns (1994-2016). Six principal components, for example, explain on average 85.6 percent of the total variation found in the data.

Statistic	Mean	St. Dev.	Min	Max
PC.1	45.578	14.341	21.039	68.121
PC.2	61.117	12.840	37.113	81.538
PC.3	69.622	10.601	50.373	85.692
PC.4	76.192	8.139	59.114	88.822
PC.5	81.361	6.339	67.238	90.980
PC.6	85.591	4.889	74.342	92.924
PC.7	88.960	3.717	80.452	94.320
PC.8	91.781	2.737	85.627	95.569
PC.9	94.217	2.008	89.794	96.779

FIGURE 4.4: Evolution of Canonical Correlation. This figure depicts the evolution of the "first-order" canonical correlation between global equity factors and currency factors through time. Principal components were extracted every year from thirteen time series of daily country equity returns. Eigenvectors corresponding to the "first" six principal components were retained each year from 1994 to 2016, and used to compute out-of-sample principal components on immediately subsequent years. These were used as proxies for latent equity factors. Currency factors were similarly extracted from thirteen time series of corresponding daily exchange rate returns. The first canonical correlation between latent factors affecting country equity returns and factors influencing currency returns was then calculated each year. We use this as a measure of integration between equity and foreign exchange markets. There is a positive linear trend in the earlier portion of our sample, which tapers off and is no longer present in recent times.

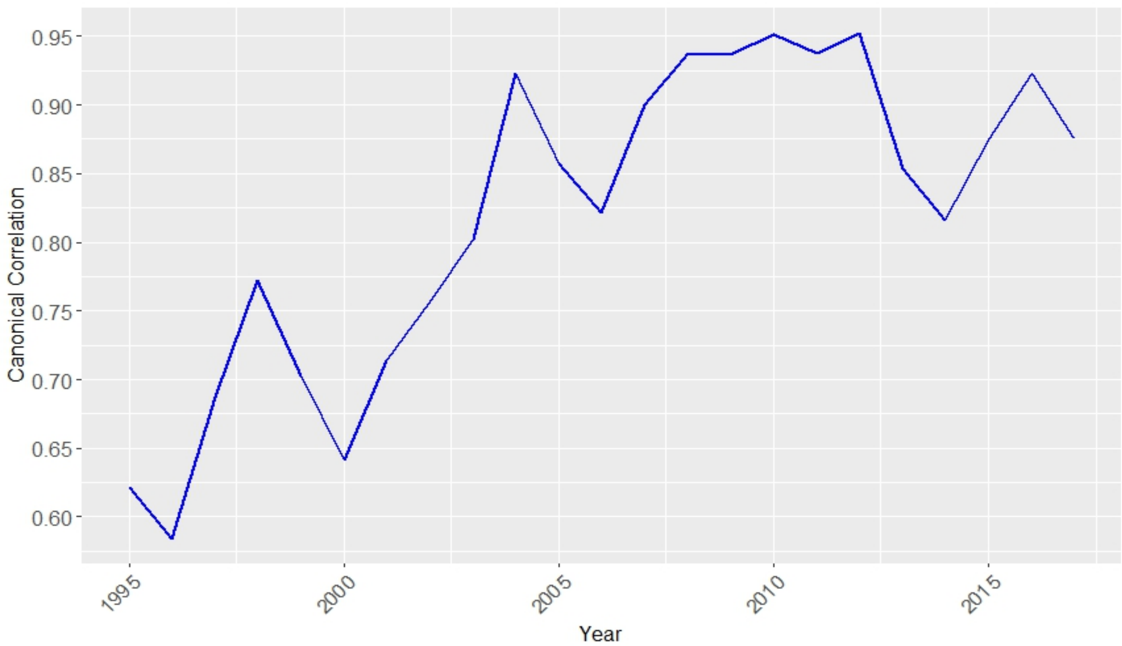


Table 4.4: Countries and DataStream mnemonics corresponding to the 7 equity indices we used in our expanded equity sample. All index values were converted to U.S. dollars. An index with the designation "RI" is a total return index (with reinvested dividends). The designation "PI" denotes a pure price index.

Country	Index Identification	DataStream mnemonic
Argentina	Argentina Merval Price Index (~U$)	ARGMERV(PI)~U$
Chile	Chile Santiago SE General (IGPA) Price Index (~U$)	IGPAGEN(PI)~U$
China	Shenzen SE Composite Price Index (~U$)	CHZCOMP(PI)~U$
Iceland	OMX Iceland All Share Price Index (~U$)	ICEXALL(PI)~U$
Mexico	Mexico IPC (Bolsa) Price Index (~U$)	MXIPC35(PI)~U$
Russia	Russia RTS Index Price Index (~U$)	RSRTSIN(PI)~U$
Thailand	Thailand-DS Market $ TOT Return IND	TOTMTH$(RI)

Table 4.5: This table is analogous to Table 4.3, except that 20 equity indices were used in the analysis. Every year, from 2006 to 2016, principal components were extracted from our expanded daily equity data set. The data set is composed of 11 years of daily equity returns (2006 to 2016) from all 20 indices.

Statistic	Mean	St. Dev.	Min	Max
PC.1	47.150	8.769	34.757	60.476
PC.2	60.845	7.399	48.995	72.718
PC.3	68.574	5.942	57.921	76.606
PC.4	73.879	4.797	65.410	80.408
PC.5	78.042	4.086	70.448	83.438
PC.6	81.490	3.487	74.946	86.168
PC.7	84.477	2.897	79.125	88.786
PC.8	86.923	2.360	82.736	90.836
PC.9	88.975	1.932	85.908	92.427

Figure 4.5: Digging Deeper: Recent Canonical Correlation. This figure shows the first-order canonical correlation between equity and exchange rate latent factors computed with an expanded data set. The latter consists of twenty time series of daily country equity returns and twenty time series of corresponding daily exchange rate returns. Fitting a linear model to the data yields a non-significant time slope coefficient of -0.005 (p-value = 0.21).

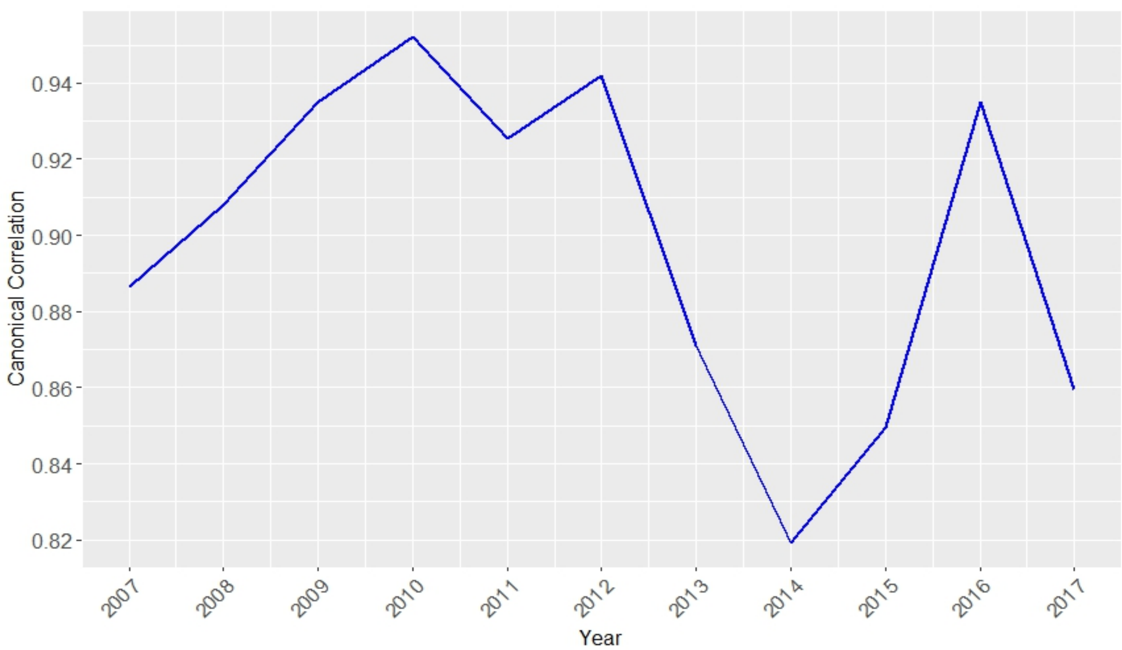


Table 4.6: Digging Deeper: Mean adjusted $R-$squared values from regressions of daily currency returns on 6 latent global factors (see Equation 4.1). Each time-series of currency returns was regressed on all 6 global factors every year. Yearly Adjusted $R-$squared values from 2007 to 2017 were then averaged across countries/regions. We use these as measures of integration among foreign exchange markets. The sample consists of daily returns from currencies of all countries/regions mentioned in Table 4.2, plus returns from the currencies of the following countries: Argentina, Chile, China, Iceland, Mexico, Russia, and Thailand.

Year	Mean adjusted $R-$squared
2007	0.603
2008	0.642
2009	0.657
2010	0.683
2011	0.730
2012	0.699
2013	0.606
2014	0.561
2015	0.634
2016	0.714
2017	0.569

FIGURE 4.6: Digging Deeper: Recent Adjusted $R-$squared Values. Plotted here are the mean adjusted $R-$squared values obtained from regressions of daily currency returns on 6 latent global factors (see Equation 4.1). Average $R-$squared values were estimated every year from 2007 to 2017 for each individual currency and then averaged across currencies in the sample. The sample consists of daily returns from currencies of all countries/regions mentioned in Table 4.2, plus returns from the currencies of the following countries: Argentina, Chile, China, Iceland, Mexico, Russia, and Thailand. Fitting a linear model to the data yields a non-significant time slope coefficient of -0.003 (p-value = 0.62).

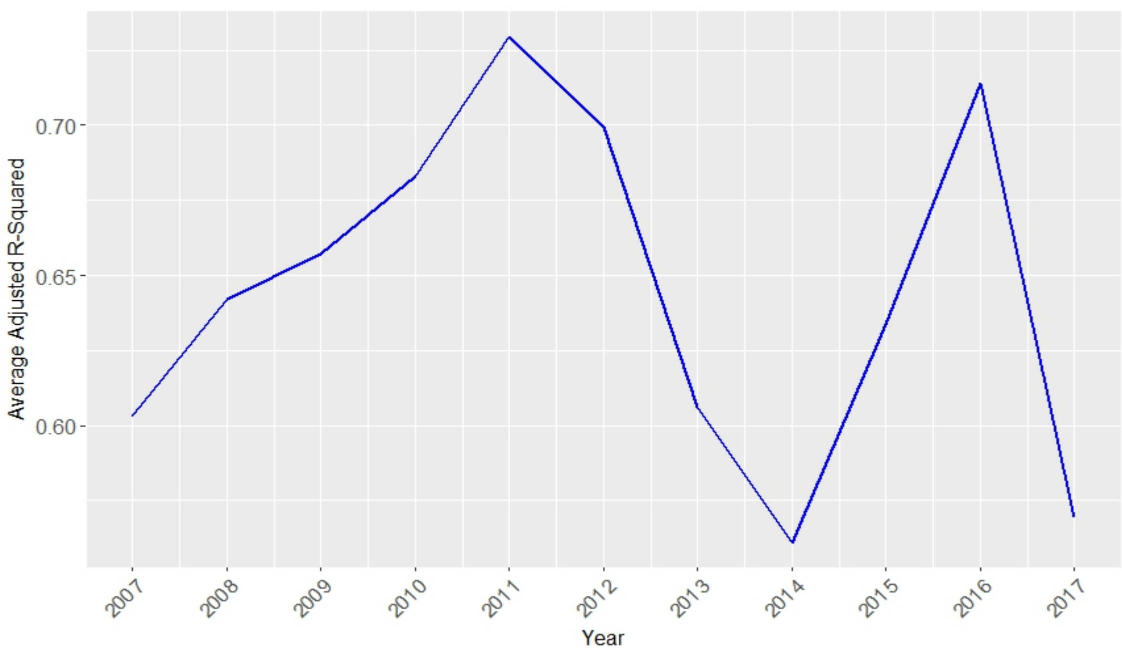


TABLE 4.7: Mean adjusted $R-$squared values from regressions of daily currency returns on 6 latent global factors. The sample consists of daily returns from currencies of the 13 countries/regions in Table 4.2, plus returns from the currencies of the following countries: Argentina, Chile, China, Iceland, Mexico, Russia, and Thailand. Adjusted $R-$squared values were estimated every year from 2007 to 2017 for each individual currency and then averaged across all years.

Country	Mean Adjusted $R-$squared
Argentina	0.300
Australia	0.838
Canada	0.640
Chile	0.672
China	0.096
Eurozone	0.855
Iceland	0.810
India	0.436
Japan	0.725
Mexico	0.741
New Zealand	0.871
Norway	0.831
Russia	0.758
Singapore	0.676
South Africa	0.958
South Korea	0.574
Sweden	0.835
Taiwan	0.347
Thailand	0.388
United Kingdom	0.555

Bibliography

Abraham, Katharine G., and Melissa S. Kearney, 2018, Explaining the decline in the employment-to-population ratio: A review of the evidence, *Working Paper*.

Avramov, Doron, and Tarun Chordia, 2006, Asset pricing models and financial market anomalies, *Review of Financial Studies* 1002–1040.

Bekaert, Geert, Robert J Hodrick, and Xiaoyan Zhang, 2009, International stock return comovements, *The Journal of Finance* 64, 2591–2626.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, 2014, Inference on treatment effects after selection among high-dimensional controls, *Review of Economic Studies* 81, 608–650.

BIS, 2016, Triennial central bank survey, Bank of International Settlements. Available at: url = https://www.bis.org/publ/rpfx16fx.pdf.

Bishop, Christopher M., 1998, Bayesian pca, in *Advances in Neural Information Processing Systems*, volume 11, 382–388 (MIT Press).

BLS, 2018a, Civilian unemployment rate [unratensa], Retrieved from FRED, Federal Reserve Bank of St. Louis; url = https://fred.stlouisfed.org/series/UNRATENSA.

BLS, 2018b, Civilian employment-population ratio [emratio], Retrieved from FRED, Federal Reserve Bank of St. Louis; url = https://fred.stlouisfed.org/series/EMRATIO.

BLS, 2018c, Civilian noninstitutional population [cnp16ov], Retrieved from FRED, Federal Reserve Bank of St. Louis; url = https://fred.stlouisfed.org/series/CNP16OV.

Brennan, M. J., T. Chordia, and A. Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics* 1–33.

Bryzgalova, Svetlana, 2015, Spurious factors in linear asset pricing models, *Doctoral Dissertation.*

Campbell, John Y., Andrew W. Lo, A. Craig MacKinlay, Petr Adamek, and Luis M. Viceira, 1997, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).

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

Daniel, K., and S. Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 1–33.

de Boor, C., 1978, *A Practical Guide to Splines* (New York: Springer).

Fama, Eugene, and Kenneth French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Fama, Eugene, and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *The Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2017, Taming the factor zoo, *Working Paper*.

Ferson, Wayne E., and Campbell R. Harvey, 1999, Conditioning variables and the cross-section of stock returns, *Journal of Finance* 54, 1325–1360.

Fleckenstein, Matthias, Francis Longstaff, and Hanno Lustig, 2014, The tips-treasury bond puzzle, *The Journal of Finance* 69, 2151–2197.

Flinn, Christopher, 2004, Discussion of the gauss-markov theorem, Available at: url = http://www.nyu.edu/econ/user/flinnc/courses/0266/gauss-markov%20theorem.pdf.

Fox, John, and Sanford Weisberg, 2013, Robust regression in r - an appendix to an r companion to applied regression, second edition, Available at: url = http://users.stat.umn.edu/ sandy/courses/8053/handouts/robust.pdf.

French, Kenneth R., n.d., Online data library, Available at: url = http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Frey, Carl Benedikt, and Michael A. Osborne, 2013, The future of employment: How susceptible are jobs to computerization?, *Working Paper*.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, 2010, Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software* 33, Issue 1, Available at: url = http://www.jstatsoft.org/v33/i01/.

Gomes, J., L. Kogan, and L. Zhang, 2003, Equilibrium cross-section of returns, *Journal of Political Economy* 111, 693–732.

Hamilton, James D., 2017, Why you should never use the hodrick-prescott filter, *Working Paper*.

Harvey, C., Y. Liu, and H. Zhu, 2015, ...and the cross-section of expected returns, *Working Paper*.

Hawking, Stephen, 2016, This is the most dangerous time for our planet, Available at: url = https://www.theguardian.com/commentisfree/2016/dec/01/stephen-hawking-dangerous-time-planet-inequality.

Hazlett, C., 2016, Statistical modeling and learning, Los Angeles, California, University of California, Los Angeles, [handout], STAT201B.

Hodrick, Robert, and Edward C. Prescott, 1997, Postwar u.s. business cycles: An empirical investigation, *Journal of Money, Credit, and Banking* 29 (1), 1–16.

Huber, Peter J., 1964, Robust estimation of a location parameter, *The Annals of Mathematical Statistics* 73–101.

Inc., MSCI, 2009, Backtesting gem vs. gem2. msci barra research paper no. 2009-28., Available at SSRN: url = https://ssrn.com/abstract=1452863 or url = http://dx.doi.org/10.2139/ssrn.1452863.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of Finance* 48, 65–91.

Koenker, Roger, 2017, *quantreg: Quantile Regression*, R package version 5.33. Available at: url = https://CRAN.R-project.org/package=quantreg.

Koenker, Roger, and Vasco d'Orey, 1987, Algorithm as 229: Computing regression quantiles, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 36.

Kothari, S. P., J. Shanken, and R. G. Sloan, 1995, Another look at the cross-section of expected stock returns, *Journal of Finance* 185–224.

Leamer, Edward, 2012, Who can save us? wall street, k street, or main street?, *UCLA Anderson Forecast*.

Lo, A. W., and A. C. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–467.

McLean, R. David, and Jeffrey Pontiff, 2015, Does academic research destroy stock return predictability?, *Journal of Finance*.

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

Murphy, Kevin P., 2012, *Machine Learning: A Probabilistic Perspective* (Massachusetts Institute of Technology).

OECD, 2018, Main economic indicators - complete database, Main Economic Indicators (database), url = http://dx.doi.org/10.1787/data-00052-en (Accessed on July 25, 2018).

Pukthuanthong, Kuntara, and Richard Roll, 2009, Global market integration: An alternative measure and its application, *Journal of Financial Economics* 94, 214 – 232.

Pukthuanthong, Kuntara, Richard Roll, and A. Subrahmanyam, 2017, A protocol for factor identification, *Working Paper*.

Roll, R., 1977, A critique of the asset pricing theory's tests: on past and potential testability of theory, *Journal of Financial Economics* 4, 129–176.

Ross, Stephen, 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 341–360.

Ross, Stephen A., 1973, Return, risk and arbitrage, *Wharton Discussion Paper, published in I. Friend and J. Bicksler, eds., Risk and Return in Finance (Cambridge:Ballinger).*

Ryan, Jeffrey A., 2016, quantmod: Quantitative financial modelling framework R package version 0.4.7, url = https://CRAN.R-project.org/package=quantmod.

Seo, Minseok, Kwondo Kim, Joon Yoon, Jin Young Jeong, Hyun-Jeong Lee, Seoae Cho, and Heebal Kim, 2016, Rna-seq analysis for detecting quantitative trait-associated genes, *Scientific Reports* 6, Available at: url = http://dx.doi.org/10.1038/srep24375.

Steven J. Foresti, Chris Tessma, and Andre Minassian, 2015, 2014 active management review, Available at: url =http://https://www.wilshire.com/media/41722/Wilshire_-2014_Active_Management.pdf.

Swedroe, Larry, 2017, Short term momentum and long term reversals can coexist, Available at: url = https://alphaarchitect.com/2017/08/30/short-term-momentum-and-long-term-reversals-can-coexist/.

Tibshirani, Robert, 1996, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)* 267–288.

Trapletti, Adrian, and Kurt Hornik, 2017, tseries: Time series analysis and computational finance, R package version 0.10-38., url = https://CRAN.R-project.org/package=tseries.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman, 2017, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer).

Tsay, Ruey S., 2010, *Analysis of Financial Time Series, 3rd Edition* (Wiley, Hoboken, New Jersey).

Venables, W. N., and B. D. Ripley, 2002, *Modern Applied Statistics with S*, fourth edition (Springer, New York), ISBN 0-387-95457-0.

Wahba, G., 1990, *Spline Models for Observational Data* (Society for Industrial and Applied Mathematics).

Wu, Ying Nian, 2017, Nonparametric function estimation and modeling, Los Angeles, California, University of California, Los Angeles, [handout], STAT204.

Yi, Congrui, 2017, *hqreg: Regularization Paths for Lasso or Elastic-Net Penalized Huber Loss Regression and Quantile Regression*, R package version 1.4. Available at: url = https://CRAN.R-project.org/package=hqreg.

Yi, Congrui, and Jian Huang, 2015, Semismooth newton coordinate descent algorithm for elastic-net penalized huber loss regression and quantile regression, *Working Paper*.

Zou, Hui, and Trevor Hastie, 2005, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 301–320.