ESSAYS IN FINANCIAL ECONOMETRICS

Aaron Mora Melendez

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2024

Co-Supervisor of Dissertation

Francis X. Diebold

Paul F. and Warren S. Miller
Professor of Economics,
Professor of Finance,
Professor of Statistics

Co-Supervisor of Dissertation

Frank Schorfheide

Christopher H. Browne
Distinguished Professor of Economics

Graduate Group Chairperson

David Dillenberger, Professor of Economics

Dissertation Committee

Nikolai Roussanov, Moise Y. Safra Professor Professor of Finance
Juan Camilo Castillo, Assistant Professor of Economics

To my parents, Silvia and Franklin, who taught me that education and hard work can take you very far.

A mis padres, Silvia y Franklin, que me enseñaron que la educación y el trabajo duro te pueden llevar muy lejos.

ACKNOWLEDGEMENT

This dissertation is part of an individual accomplishment, namely my PhD, but it is undoubtedly the result of a collective effort. I want to start by thanking my advisors, Francis X. Diebold and Frank Schorfheide, for being caring mentors throughout my PhD. Without their support and guidance, this dissertation would not have been possible. First as a student in their graduate courses on econometrics and then as an advisee, I learned what top-notch research entails under their wing. Thanks for always being available to offer your valuable advice and for always pushing me to improve my research to the next level. Special thanks to Frank Diebold for having me as his RA for several years. I owe him a huge debt of gratitude for how formative this experience was as a young researcher, for placing his trust in me, for allowing me to learn as his assistant, and for the collaborations that developed from this experience.

I also want to thank the two other members of dissertation committee, Nick Roussanov and Juan Camilo Castillo. Their involvement in my job market paper was key. Thanks to Nick for teaching me a great deal of what I know about asset pricing and for always reminding me not to lose track of the important questions that need answers. Thanks to Juan Camilo for all his advice related to demand estimation, for also having me as an RA, and for being an inspiration as a successful young researcher.

Furthermore, thanks to Karun Adusumilli and Jesús Fernández Villaverde, who were readers of my third-year paper that eventually evolved into the second chapter of this dissertation. Their valuable comments and support for that project, along with that of my advisors, were very important for that project to be honored with the Hiran C. Haney Fellowship prize in economics for best third-year paper at Penn in 2022.

Additional thanks to Xu Cheng and Wayne Gao for helpful discussions about my job market paper and all the good advice on how to deliver an effective presentation. More generally, I am very thankful to the entire Penn econometrics group for creating a supportive research atmosphere and for letting me be part of it. I would like to thank other fellow graduate students that made my

PhD journey a wonderful experience. In no particular order, I thank Priyanka Goonetilleke, Joao Ritto, Leon Huetsch, Kathleen Hui, Marko Mlikota, Yoshiki Ando, Min Kim, Cung Truong Hoang, Artem Kuriksha, and Boyuan Zhang. Thank you so much for enhancing and enriching my six years at Penn.

Special thanks to Rebecca Jorgensen for being my research confidant, for always being willing to discuss the earliest research ideas, and for always being there to help me tremendously in writing my ideas in a concise and effective way. Thanks for always believing in me and offering your support whenever my research hit a wall.

Finally, I want to thank my family, because ever since I emigrated from Costa Rica with the goal of obtaining a PhD in economics, I have never stopped feeling their warm support no matter the distance.

ABSTRACT

ESSAYS IN FINANCIAL ECONOMETRICS

Aaron Mora Melendez

Francis X. Diebold

Frank Schorfheide

This dissertation presents two essays in financial econometrics. These chapters show how to adapt and use econometric techniques to study various questions in financial economics. The first chapter: "*Revealed Preference for Green Stocks: An Asset Demand Approach*" combines a traditional portfolio construction problem with demand estimation techniques to estimate the demand for green stocks of US institutional investors. The methodology presented innovates along two dimensions with respect to recent influential work on asset demand estimation. First, in this framework investors have heterogeneous portfolios not only through differential beliefs about future returns, but also because they place varying importance on the non-financial characteristics of the portfolios they construct. Second, by using a mixed logit demand specification, we can estimate asset demand that delivers more realistic substitution patterns across assets. This chapter uses data on the environmental performance of firms and quarterly stock holdings data from institutional investors to estimate the demand for stocks accounting for environmental scores and return-related stock characteristics. Estimates show that taste for green stocks fluctuates over time and by investors' assets under management. Finally, this chapter presents a counterfactual exercise to study the equity price effects of a ban on green investing for pension funds. Results show that a portfolio with the top brown stocks is estimated to have positive capital gains due to the policy, while a portfolio with the top green stocks is estimated to have capital losses.

The second chapter: "*Exchange Rate Supervised Topic Modelling*," shows how to use a hybrid of supervised and unsupervised learning models to go from text from news articles to an FX news index that can be used to enhance traditional models from the FX literature. To do so we rely on

Supervised Latent Dirichlet Allocation (sLDA) (Blei and Mcauliffe (2008)) which combines information about a supervising variable with topic extraction over a corpus of text in a single-stage estimation. Although this estimation can be done in two stages, this chapter documents with a Monte Carlo simulation that there are efficiency gains from a single-stage approach. The empirical application suggested is centered around the Monex Market, the main Costa Rican platform for FX trade; accordingly news articles are gathered from the main Costa Rican newspapers. The exchange rate of interest is the Costa Rican Colón (CRC), the local currency, and the United States dollar (USD). Using the CRC/USD exchange rate as the supervising variable this chapter relies on sLDA to extract the topics from the news article corpus that are relevant as covariates for the exchange rate over short frequencies. Results provide evidence that these topics have explanatory power in and out of sample, and can be used to provide interpretability to episodes of high volatility of the exchange rate.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

Revealed Preference for Green Stocks: An Asset Demand Approach

1.1. Introduction

The last decade has experienced a steady rise in sustainable investments, with global funds invested in sustainable funds reaching USD 2.7 trillion in the first quarter of 2023 (Morningstar (2023*a*)). This has been accompanied by growing interest in sustainable investment from asset managers, with 85% of them already implementing or planning to implement sustainable investing (Morgan Stanley (2022)), and a growing supply of sustainable funds available to investors (Morningstar (2023*b*)). Equity markets will play a fundamental role in the transition to an environmentally sustainable economy by providing incentives for listed firms to adopt cleaner technologies and practices. Understanding the demand of investors for green stocks and its consequences for equity prices is a key part of understanding the incentives of listed firms to align their business strategies with an environmentally sustainable economy.

This chapter combines a traditional portfolio construction problem with demand estimation techniques to estimate the demand for green stocks of US institutional investors. To do this, it is necessary to have a model capable of dealing with the empirical fact that investors vary greatly in the portfolios they construct. One largely studied reason for portfolio heterogeneity is that investors construct different portfolios because they exhibit belief heterogeneity over the future returns of the assets they include in their portfolios. Another less studied reason for portfolio heterogeneity is that, even if investors have common beliefs, they assign varying importance to the characteristics of the portfolios they construct. That is, they exhibit taste heterogeneity over portfolio characteristics. For example, investors can tilt their portfolios toward environmentally sustainable, or green assets, for motives unrelated to future returns.

This chapter presents a framework to estimate the demand for assets where both belief and taste heterogeneity play a role. Belief heterogeneity over future returns is codified via investor-specific conditional expectations, while taste heterogeneity allows investors to care about portfolio characteristics beyond those directly related to an expected return-versus-risk trade off. We estimate the demand for stocks accounting for environmental scores and return-related stock characteristics. Then, using the estimated demand, we conduct a counterfactual exercise to study the effects of a ban on green investing for pension funds, a policy discussed in the US Senate (Morgan (2023)), on equity prices and aggregate holdings.

This chapter innovates on the recent influential work by Koijen and Yogo (2019) (KY2019) that shows that an asset demand approach combined with a market clearing condition implies a valid asset pricing model. KY2019 rekindled a classic literature on modeling asset demand and advocate for applying industrial organization (IO) tools to asset pricing models. However, the microfoundations for asset demand presented in KY2019 are based exclusively on belief heterogeneity over future returns, not allowing for taste heterogeneity to influence investors' demand, and the substitution patterns between assets are restricted by the demand specification they use.

The methodological contributions of this chapter are twofold. First, the microfoundations of the asset demand framework presented in this chapter also allow for taste heterogeneity in the portfolio construction problem in addition to belief heterogeneity. Allowing for taste heterogeneity in portfolio characteristics opens the door to studying investor behaviors that do not fit within the traditional expected returns-versus-risk paradigm. Examples of such behaviors include: (i) investment strategies based on Environmental, Social, and corporate Governance (ESG) performance metrics of the companies, which take into account stock characteristics unrelated to returns. (ii) "Sin" stock dis-investing, where investors based on ethical considerations alone reduce their investments or completely avoid stocks that belong to so-called *sin* industries, like alcohol, tobacco, gambling, adult entertainment, or weapon manufacturing, based on ethical considerations alone. (iii) The deletion of a stock from a stock index can mechanically change the demand for the stock if, for example, there are hedge funds or mutual funds designed to track the index, even if the fundamentals of the company remain unchanged.

The second contribution in this chapter is to present and estimate a mixed-logit demand specification for the demand for stocks.[1] In this specification, heterogeneity is captured by investor-specific coefficients that are modeled as functions of investor demographics (see Berry et al. (1995) for the seminal application of this framework, Berry and Haile (2021), Gandhi and Nevo (2021), and Conlon and Gortmaker (2020) for modern practices on demand estimation for differentiated products). KY2019 derive a logit demand system for assets where price elasticities are proportional to portfolio shares. To see why this is a restrictive feature, imagine that the stocks for two companies have the same portfolio weights (market shares in this context), but these companies belong to different industries, for example technology and energy. These companies likely have very different fundamentals; however, under a logit demand system, portfolio holdings of another technology stock would react identically to a price change from both stocks. The richer investor-level heterogeneity captured by mixed logit demand delivers flexible substitution patterns between assets, improving on the restrictive price elasticities of logit demand.

This chapter presents an empirical application that studies the demand for stocks of institutional investors. To estimate this demand, we use data on US institutional investors and their holdings of US stocks combined with characteristics of the stocks they hold. To construct stock characteristics, we combine price and accounting data with information on the environmental performance of listed companies in the form of environmental scores (E-scores) from the MSCI rating agency. In this application we quantify the extent to which investors value the green metrics of the stocks they select for their portfolio while also considering returns-related characteristics. While it is not possible to distinguish whether each characteristic is included in the demand due to belief or taste heterogeneity using only holdings data, survey evidence supports the interpretation of environmental aspects as a taste characteristic.[2]

The estimates rely on quarterly data of the stock holdings of US institutional investors and the corresponding stock characteristics from 2001-Q1 to 2019-Q4, and we perform estimation in two-year windows. The first finding is that the revealed taste for green stocks fluctuates over time.

[1] The mixed logit demand is also referred to as random coefficients demand in the IO literature.

[2] Giglio, Maggiori, Stroebel, Tan, Utkus and Xu (2023) examine a survey of retail investors on the motives for Enviromental Social and Governance (ESG) investing and find that generally investors expect ESG investments to underperform the market and that only 7% of investors in ESG assets were motivated by return expectations.

Throughout the estimation sample we find a positive taste for green stocks as measured by a positive semi-elasticity for E-scores that is increasing in the second half of the sample. Moreover, in the period after the Great Recession (2007-Q4 to 2009-Q2) there is an increase in the range of values for the coefficient on E-scores, showing an increase in the heterogeneity in the sensitivity to green characteristics across investors. This suggests that after periods of economic downturn, some investors may care relatively more about the return-related characteristics of the stocks and relatively less about the environmental-friendliness of the companies underlying the stocks.

We also find that the coefficient corresponding to E-scores for a particular investor is a function of the investor's assets under management. We find that in the last ten years of the sample, institutions with more assets under management have on average a higher taste for green stocks. We repeated the estimation exercise under a logit demand specification where all investors share the same sensitivity to the green metrics. Such estimates exhibit much less variation.

In a counterfactual exercise, we use the estimated demand system for stocks to study the effects of a ban on green investing for pension funds on equity prices and aggregate holdings. On March 1st 2023, the US Senate passed a bill to prevent pension fund managers from basing investment decisions on factors like climate change (Morgan (2023)). President Biden vetoed the bill days later (Thomas (2023)), but various US State Legislatures have approved similar initiatives.[3] Inspired by the policy, the counterfactual exercise makes the demand for stocks of pension funds perfectly inelastic to the environmental performance of the stocks, so for these investor only return-related characteristics are taken into account in their demand for stocks.

Using the data and estimates for 2019-Q1, we find that stock with low E-scores, brown stocks, will benefit the most with higher counterfactual prices. A portfolio in the bottom quintile of green stocks is estimated to have an associated average price increase of 1.1% under the counterfactual. In contrast, the top quintile portfolio has an average price decrease of 1.6%. Results for the counterfactual exercise using a logit demand specification exhibit much smaller price changes due to the

[3] The US House of Representatives later tried to override President Biden's veto but failed to secure the necessary votes for that measure (Foran and Wilson (2023)).

restrictive substitution patterns of logit demand.

Related literature. As mentioned above, this chapter is closely related to Koijen and Yogo (2019). In that paper, the authors also propose a demand system approach to asset pricing and estimate a model that jointly explains asset prices and quantities. This chapter extends the framework from KY2019 in three ways. First, our model can accommodate taste heterogeneity allowing investors to consider stock characteristics beyond those related to returns when forming their portfolios. Second, a demand system with a mixed logit demand specification provides flexible substitution patterns between assets, improving on the restrictive elasticities of the logit demand model used in KY2019. Third, KY2019 define the market as pools of investors, while I use a more natural market definition of the US stock market in a quarter. Estimation at the market level facilitates dealing with the endogeneity of prices and allows us to consider instrumental variables inspired by the IO literature.

More broadly, this chapter fits into the literature that models asset demand from investors. Classic works include Brainard and Tobin (1968), Rosen (1974), and Lucas (1978). This literature has recently received renewed attention with the use of new stock holdings data and strategies to tackle endogeneity problems. Recent examples include KY2019 as well as Koijen and Yogo (2020). Koijen and Yogo (2020) builds on the tools presented in KY2019 to study a demand system for financial assets that includes currencies, bonds, and stocks across several countries. Jiang et al. (2020) use a demand approach to portfolio construction to study global imbalances in net foreign assets across countries. Another example of a demand system approach is in Han et al. (2021), where the authors use KY2019's demand approach to quantify the impact of underperforming mutual funds on the overpricing of high-beta stocks. This chapter contributes to these recent papers by studying more than return characteristics as determinants of investors' demand curves, the use of a mixed logit demand specification, and the use of modern instrumental variables suggested by the IO literature.

A very recent but growing literature has studied how the environmental performance of stocks matters for equity holdings and prices. Theoretical approaches include Pástor et al. (2021), where the authors develop a model where there are non-pecuniary benefits from investing in green assets, and such benefits lead to lower expected returns on green assets due to investors pushing up their prices. Empirical approaches include Baker et al. (2022) that, using an asset demand approach, interpret the fees for ESG funds to find that investors are willing, on average, to pay 20 basis points more per annum to invest in a fund with an ESG mandate as compared to an otherwise identical mutual fund without an ESG mandate. Pastor et al. (2023) also employ an asset demand approach to study the degree investors tilt their portfolios between green and brown stocks; they find that on aggregate institutional investors have become increasingly green, exhibiting a positive green tilt, while non-institutional investors have become browner, exhibiting a negative green tilt. Koijen et al. (2023) use a demand system with stock characteristics related to enviromental performance to study the impact of climate-related induced shifts on equity prices. They study a counterfactual exercise where there is an increased sensitivity for green stocks and find this implies capital gains for passive investment institutions and capital losses for active investment institutions. With respect to these recent papers studying the demand of green stocks, this chapter contributes by studying the effects of a ban on green investing on pension funds and employing the methodological contributions mentioned above.

The rest of this chapter is organized as follows: section 1.2 presents the investor portfolio problem and how a demand for assets with a tractable logit functional form can be obtained from its solution. Section 1.3 presents how mixed logit demand can be estimated in the context of demand for stocks. Section 1.4 presents the empirical application that estimates the demand for green stocks from the institutional investors in the US. Section 1.5 shows the counterfactual exercise that studies the effects of a ban on green investing for pension funds. Finally, section 1.6 concludes and discusses future avenues of work.

1.2. Asset Demand with Taste Heterogeneity

In this section we show how a demand for assets with an empirically tractable logit functional form can be obtained from the solution of an traditional portfolio problem. In this set up we allow for investor belief and taste heterogeneity; we discuss how it relates to traditional portfolio problems in the asset pricing literature as well as recent key contributions that model asset demand. The exposition is divided in four subsections. The first one presents the investor portfolio problem and the second one its solution. Importantly, the third subsection presents the assumptions needed to obtain the empirically tractable demand for assets; and the fourth subsection presents the market clearing condition that pin downs equilibrium prices.

1.2.1. Investor Portfolio Problem

In this subsection we present the portfolio construction problem investors solve. The key assumption is that investors differ in their beliefs about future returns and assign varying importance to non-financial characteristics of the portfolios they construct.

To facilitate exposition for the remainder of the chapter we consider the assets available to investors to be stocks, but the framework in this section applies to asset classes other than stocks, and of course to combinations of assets from different classes. Let $t = 1, ..., T$ denote the stock market in given period. In our application the market definition corresponds to the US stock market at a quarter t. In each of these markets there are I_t investors indexed by $i = 1, ..., I_t$, and each investor i has to allocate A_{it} dollars of assets under management (AUM) in market t among J_t available stocks and an outside option. Stocks are treated as differentiated investment products that are demanded by investors. Let $j = 1, ..., J_t$ index one of the J_t available stocks and $j = 0$ denote the outside option.[4] In our context the outside option denotes the possibility that investors allocate a fraction of their AUM into none of the stocks in J_t.

Let R_{t+1} denote a J_t-vector of gross returns between t and $t+1$ for the stocks available in period t; similarly R^0_{t+1} denotes the gross return on the outside asset. For each stock j the gross return between t and $t+1$ is computed as $R_{t+1,j} = \frac{V_{t+1,j}}{P_{t,j}}$, where $V_{t+1,j}$ is the payoff per share of stock j in $t+1$, and $P_{t,j}$ is the price per share of stock j in t.[5]

Each investor solves a two-period problem between the current period (t) and the next period ($t+1$)

[4]When there is no possibility of confusion, we use J_t to denote the set of inside goods and the cardinality of the set itself, that is $J_t = \{1, 2, ..., |J_t|\}$.

[5]In many contexts $V_{t+1,j}$ is divided as sum of the price per stock of stock j in $t+1$, $P_{t+1,j}$ plus the dividends per stock in $t+1$, $D_{t+1,j}$; however, in our setting investors only care about the overall future payoff $V_{t+1,j}$, and not whether is was generated by capital gains or dividends.

where they construct a portfolio by choosing portfolio weights w_{it} (a J_t-dim vector), such that

$$\max_{w_{it}} \quad E_{it}[\log(A_{i,t+1})] + a_i' C_t' w_{it} \tag{1.1}$$

$$\text{s.t.} \quad A_{i,t+1} = A_{it}\left[R^0_{t+1} + w_{it}'[R_{t+1} - R^0_{t+1}1]\right] \tag{1.2}$$

$$w_{it} \geq 0; \quad 1' w_{it} < 1. \tag{1.3}$$

In this problem investors differentiate portfolios according to the characteristics they offer. We separate the characteristics of the portfolio into tomorrow's terminal wealth $A_{i,t+1}$, and additional characteristics C_t of the stocks that composed portfolio w_{it}. The part of utility that comes from tomorrow's dollar value for the portfolio enters through a log utility, while the current value for investor i of other portfolio characteristics enters with linear weights a_i, an investor-specific K_C-vector. The values of a_i capture investor preferences over the characteristics included in C_t. The log specification in the utility for tomorrow's portfolio wealth follows KY2019 and a long tradition that dates back to Samuelson (1969).[6]. If the entries of a_i are set to zero, taste heterogeneity is irrelevant and we are in a context where only pecuniary factors matter for portfolio construction

The matrix C_t denotes a $J_T \times K_C$ matrix where row j contains a K_C-vector c_{jt} of characteristics for stock j that are relevant for the profile of portfolio w_{it}. $E_{it}[\cdot]$ denotes the conditional expectation for investor i at time t, that is $E_{it}[\cdot] \equiv E[\cdot|\mathcal{I}_{it}]$, where $\mathcal{I}_{it}$ denotes the information set of investor i at time t. In this model it is assumed that investors do not learn from the actions of others investors; this assumption is commenly referred in the literature as *investors agree to disagree.* The first constraint in (1.2) denotes the evolution of the portfolio's wealth by choosing portfolio w_{it} and the constraints in (1.3) impose short-sale restrictions and that all wealth is invested in either stocks or the outside option.[7]

Including stock characteristics into the value of selecting a portfolio w_{it} may be relevant to empirically capture investment decisions that do not entirely fit into an expected return-versus-risk

[6]In a multi period setup, assuming Log utility collapses the portfolio problem into a two-period problem, as in our setup.

[7]For a paper that relaxes short sale constraints see Tian (2022).

investment paradigm. Examples of this type of investment behavior include: (i) green stock investing where investor value the enviromental performance of the stocks they include in their portfolios, (ii) more generally in investment strategies based on environmental, social, and corporate governance (ESG) metrics of the companies that not only take into account returns and wealth accumulation when selecting stocks to invest. (iii) "Sin" stock dis-investing, where investor avoid including stocks that belong to a so-called "sin" industry like tobacco, alcohol, gambling, adult entertainment or guns; despite the returns these stocks may offer. (iv) Finally, addition or deletion of a stock into an stock index (e.g. S&P 500, Russell 1000 or The Dow Jones Industrial Average) can mechanically induce demand for the stock, for example from hedgefunds and mutual funds designed to track the index, despite stock fundamentals may remain unchanged. This modeling choice generalizes the setup in KY2019, where next's period portfolio wealth, $A_{i,t+1}$, is the only relevant characteristic to construct a portfolio.

Notice that the problem in (1.1) accommodate two sources of heterogeneity. First, it captures belief heterogeneity over future returns, which is codified via differential information on the expectations operator, $E_{it}[\cdot]$. Under homogenous beliefs it would be the case that $E_{it}[\cdot] = E_t[\cdot]$ for all i. And second, it captures taste heterogeneity over stock characteristics via the weights a_i, if all investors value portfolio characteristics equally then $a_i = a$ for all i. Together, the objective function in (1.1) accommodates these two channels for heterogeneity.

1.2.2. Optimal Portfolio Weights

This subsection presents the solution to the portfolio construction problem. The approximate solution for positive portfolio weights has a traditional form that depends on the first two moments of expected returns but also takes into account investor preferences over non-financial characteristics.

In order to provide the solution to the investor's problem we introduce some notation. Denote by r^x_{t+1} the vector of excess log returns, $r^x_{t+1} = \log(R_{t+1}) - \log(R^0_{t+1})1$. Moreover, let $\tilde{\Sigma}_{it}$, a $J_t \times J_t$ matrix, denote the variance-covariance matrix

$$\tilde{\Sigma}_{it} = E_{it}\left[\left(r^x_{t+1} - E_{it}[r^x_{t+1}]\right)\left(r^x_{t+1} - E_{it}[r^x_{t+1}]\right)'\right],$$

and let $\tilde{\mu}_{it}$, a J_t-vector of conditional expectations adjusted by variance:

$$\tilde{\mu}_{it} = E_{it}[r^x_{t+1}] + \frac{\tilde{\sigma}^2_{it}}{2},$$

where $\tilde{\sigma}^2_{it}$ is a vector of the diagonal elements of $\tilde{\Sigma}_{it}$. Furthermore, without loss of generality, partition the asset space between the J^1_t assets with positive weights on the investor's problem, that is those assets for which short sale constraints are not binding so we can rewrite $\tilde{\Sigma}_{it}$ and $\tilde{\mu}_{it}$ as

$$\tilde{\mu}_{it} = \begin{pmatrix} \mu_{it} \\ \mu^{(2)}_{it} \end{pmatrix}, \qquad \tilde{\Sigma}_{it} = \begin{pmatrix} \Sigma_{it} & \Sigma^{(1,2)}_{it} \\ \Sigma^{(2,1)}_{it} & \Sigma^{(2,2)}_{it} \end{pmatrix},$$

where μ_{it} is a J^1_t-vector and Σ_{it} a $J^1_t \times J^1_t$ matrix, both corresponding to the assets with positive weights. The following proposition parallels Lemma 1 in the KY2019 framework in characterizing the solution for the optimal portfolio but accounts for the extra term that allows for taste heterogeneity.

Proposition 1. *The solution to the investor problem in (1.1)-(1.3), w^*_{it}, is characterized by the Euler equation*

$$E_{it}\left[\left(\frac{A_{it+1}}{A_{it}}\right)^{-1} R_{t+1}\right] = 1 - \left(I - 1w^{*'}_{it}\right)\left(\Lambda_{it} - \lambda_{it}1 + C_t a_i\right), \tag{1.4}$$

where Λ_{it} and λ_{it} are Lagrange multipliers on (1.2) and (1.3) respectively. Moreover, the positive optimal portfolio weights can be approximated (over a short period horizon) as

$$w^*_{it} \approx \Sigma^{-1}_{it}\left(\mu_{it} - \lambda_{it}1 + C_t a_{it}\right). \tag{1.5}$$

The proof for Proposition 1 is shown in Appendix A.1. The Euler equation in (1.4) generalizes the set up in KY2019, since the case with no taste heterogeneity, $a_i = 0$, results in the Euler equation presented in KY2019's model.[8] Furthermore, as in KY2019, if investors do not face short-sale constraints ($\Lambda_{it} = 0$ and $\lambda_{it} = 0$) and have homogeneous beliefs ($E_{it}[\cdot] = E_t[\cdot]$ for all i), then the

[8] Lemma 1 in Koijen and Yogo (2019), page 1481.

Euler equation becomes

$$E_t\left[\left(\frac{A_{it+1}}{A_{it}}\right)^{-1}R_{t+1}\right]=1,$$

which is a moment condition commonly tested in the literature on consumption-based asset pricing. The message in the second part of proposition 1, is that investor i's demand for stocks, given by portfolio weights w_{it}^*, is determined by three components: uncertainty around next period returns (Σ_{it}), expected returns (via μ_{it}) and taste sensitivity ($C_t a_{it}$). All else equal, if investor i has more uncertainty around next period returns for some assets then the portfolio weights on those assets will be relative smaller. Similarly, keeping Σ_{it} fixed, stock holdings are increasing on expected returns. More generally, in our setup stock holdings vary according to the additional value stocks contribute the investors' utility derived from portfolio characteristics. If an stocks j contribute positively to a portfolio characteristics k valued positively by investor i, $(a_{it,k}c_{jt,k}) \geq 0$, then increasing such characteristic for stock j will imply that investor i holds relative more stocks of j.

1.2.3. An empirically tractable demand for stocks

Despite the fact that equation (1.5) provides a clear intuition of the determinants of the optimal positive portfolio weights, it is not very tractable empirically.[9] This subsection presents the set of assumptions that are needed to derive a form of equation (1.5) that uses characteristics of the stocks in a logit form to provide an empirical tractable function for the portfolio weights.

To compute the conditional moments that determine portfolio weights in equation 1.5, namely Σ_{it} and μ_{it}; we need to be explicit about how next period's excess log returns are modeled and how information varies across investors. Recall that by definition $R_{t+1,j} = V_{t+1,j}/P_{t,j}$, where $V_{t+1,j}$ stands for next's period payoff for stocks j and $P_{t,j}$ is the equilibrium price per share. Similarly, $R_{t+1,0} = V_{t+1}^0/P_0$; and without loss of generality we normalize the price of the outside good to one,

[9] It requires obtaining the first two investor-specific conditional moments over excess log returns; which is composed of a large cross section of stock returns.

$P_0 = 1$. If we take logs we obtain that the vector of excess log returns is given by

$$
\begin{aligned}
r^x_{t+1} &= \log(V_{t+1}) - \log(V^0_{t+1})1 - \log(P_t) \\
&= v_{t+1} - v^0_{t+1}1 - p_t \\
&:= v^x_{t+1} - p_t.
\end{aligned}
$$

Given that the excess payoff in $t+1$ is unknown at period t, we assume investors have a prior in period t about its value such that V_{t+1} and V^0_{t+1} follow a lognormal distribution:

Assumption 1. *Distribution of next period's payoff*
The J_t-vector of next period's payoff, V_{t+1}, follows the lognormal distribution

$$V_{t+1} \sim lognormal(\mu_{vt}, \Sigma_{vt}),$$

where μ_{vt} is a J_t-vector and Σ_{vt} a $J_t \times J_t$ matrix. Moreover, the outside option next period's payoff also follows a lognormal distribution

$$V^0_{t+1} \sim lognormal(\mu^0_{vt}, \Sigma^0_{vt}),$$

where μ^0_{vt} and Σ^0_{vt} are scalars. The values of $\mu_{vt}, \mu^0_{vt}, \Sigma_{vt}, \Sigma^0_{vt}$ are common knowledge to investors.

Since next period's payoffs are bounded from below by zero, the log normality assumptions is an appropriate modeling choice and it is a traditional assumption in asset pricing. Assumption 1 implies that the excess log returns can be written as

$$r^x_{t+1} = \mu_{vt} - \mu^0_{vt}1 - p_t + e_v, \tag{1.6}$$

where

$$e_v \sim N(0, \Sigma_{xt})$$
$$\Sigma_{xt} = E_t[(r^x_{t+1} - \mu_{xt})(r^x_{t+1} - \mu_{xt})']$$
$$\mu_{xt} = \mu_{vt} - \mu^0_{vt} 1 - p_t,$$

and the (j,k)-entry of Σ_x is given by:

$$\Sigma_{xt,jk} = \Sigma_{vt,jk} + \Sigma^0_{vt} - cov(V_{t+1,j}, V^0_{t+1}) - cov(V_{t+1,k}, V^0_{t+1}).$$

Notice that the value of excess log returns depends on the vector of prices, p_t, which need to be pinned down in equilibrium. We also assume that equilibrium prices are observed by all investors. Then conditional on public information, excess log returns can be viewed as having a distribution inherited from the distribution assumed for next period excess payoffs.[10] In this case $r^x_{t+1} \sim N(\mu_{xt}, \Sigma_{xt})$, but both μ_{xt} and Σ_{xt} depend on p_t which is endogenous to the model. The next assumption states a factorization for the matrix Σ_{xt}.

Assumption 2. *Factorization for the matrix Σ_{xt}*
The matrix Σ_{xt} admits the representation

$$\Sigma_{xt} = \Gamma_{xt}\Gamma'_{xt} + \sigma^2_e I. \tag{1.7}$$

where Γ_{xt} is a J_t-vector of factor loadings and σ^2_e is common variance across stocks.

This assumption is consistent with a factor structure for the vector of log excess returns where r^x_{t+1} admits the following single factor representation:

$$r^x_{t+1} = \mu_{xt} + \Gamma_{xt} F_{t+1} + e_{t+1},$$

[10]Unfortunately, the setup in KY2019 is not explicit about to what extent returns are endogenous or exogenous to their setup, we believe that being explicit about this and modeling how information is different across investors contribute to the clarity of the microfoundations of asset demand.

the single factor F_{t+1} is distributed as $N(0,1)$ and $e_{t+1} \sim N(0,\sigma_e^2 I)$, with e_{t+1} independent of F_{t+1}.[11]

In order to compute moments that depend on the information set each investor have, we model how information varies across investors. We adapt a traditional setup from the asymmetric information literature in asset pricing (see for example Grossman (1976)).

Assumption 3. *Information Technology*

Each investor i receives a signal about next period's excess log returns, r^x_{t+1}, denote by s_{it}, a J_t-vector and given by

$$s_{it} = \alpha_i r^x_{t+1} + \varepsilon_{it}, \tag{1.8}$$

where $\varepsilon_{it} \sim N(0,\sigma_\varepsilon^2 I)$ and is independent of r^x_{t+1}. The scalar constant α_i is privately known to investor i.

In this set up, each investors knows s_{it} and α_i but does not know the value of ε_{it} that produced s_{it} so they cannot back out immediately the value of r^x_{t+1} and has to update her expectations over r^x_{t+1} by conditioning on the information set $\mathcal{I}_{it} = \{s_{it}, \alpha_i, \mu_{xt}, \Sigma_{xt}, \sigma_\varepsilon^2\}$. Notice that $s_{it}|(r^x_{t+1},\alpha_i) \sim N(\alpha_i r^x_{t+1}, \sigma_\varepsilon^2 I)$ and $r^x_{t+1} \sim N(\mu_{xt},\Sigma_{xt})$ so by Bayes theorem we can compute the distribution of $r^x_{t+1}|(s_{it},\alpha_i)$ which corresponds to the distribution of $r^x_{t+1}|\mathcal{I}_{it}$. The posterior distribution is given by

$$r^x_{t+1}|\mathcal{I}_{it} \sim N(\mu_{r|s_i}, \Sigma_{r|s_i}) \tag{1.9}$$

$$\text{with} \qquad \Sigma_{r|s_i} = \left[\alpha_i^2(\sigma_\varepsilon^2)^{-1} I + \Sigma_{xt}^{-1}\right]^{-1} \tag{1.10}$$

$$\mu_{r|s_i} = \Sigma_{r|s_i}\left[(\sigma_\varepsilon^2)^{-1}\alpha_i s_{it} + \Sigma_{xt}^{-1}\mu_{xt}\right]. \tag{1.11}$$

Details on the derivation of the posterior moments is shown in appendix A.1.

The next proposition shows that under assumptions 1 and 2 we can obtain a convenient decomposition of the investor-specific conditional matrix Σ_{it}.

[11] Alternatively, the factorization in Assumption 2 can be obtained by assuming a factor structure on the variance matrix of future payoffs Σ_{vt}. In this case notice that $\sigma_e^2 = \Sigma^0_{vt}$.

Proposition 2. *Under assumptions 1 to 3, the investor-specific conditional matrix Σ_{it} can be written as $\Sigma_{it} = \Gamma_{it}\Gamma'_{it} + \iota_{it}I$ where Γ_{it} is an J_t-vector and ι_{it} a scalar, both investor-specific.*

Proof of Proposition 2 is shown in appendix A.1. Notably, this result is an assumption in KY2019's model.[12] In our set up we are able to obtain the investor-specific decomposition of Σ_{it} by relaying on weaker assumptions. The next assumption relates the first two moment of the vector of excess log returns with individual stock characteristics.

Assumption 4. *Return-related Stock Characteristics*

Each entry of the J_t-vectors μ_{xt} and Γ_{xt} can be expressed as a polynomial of order M over a K_x-vector of return-related stock characteristics x_{jt}, including price p_{jt}; that is

$$\mu_{xt,j} = y'_{jt}\Phi_\mu + \phi_\mu \tag{1.12}$$

$$\Gamma_{xt,j} = y'_{jt}\Phi_\Gamma + \phi_\Gamma, \tag{1.13}$$

where Φ_μ and Φ_Γ are matrices of coefficients, ϕ_μ and ϕ_μ scalars and y_{jt} is K_y-vector with $K_y = \sum_{m=1}^{M} K_x^m$ and

$$y_{jt} = \begin{pmatrix} x_{jt} \\ x_{jt} \otimes x_{jt} \\ x_{jt} \otimes x_{jt} \otimes x_{jt} \\ \vdots \end{pmatrix},$$

and $\otimes$ stands for the Kronecker product.

The motivation for the previous assumption is twofold. First, modeling the mean and covariance matrices of asset returns as functions of assets characteristics is more empirically tractable than estimating the mean and covariance matrices of asset returns directly from past observations (see for example Brandt et al. (2009)).[13] Second, KY2019 show the empirical the relevance

[12]The first part of Assumption 1 in Koijen and Yogo (2019), p. 1483.

[13]Since the mean and covariance matrices of asset returns are hard to estimate and very likely time-varying, this literature moved from estimation using historical samples of returns to using functions that map assets characteristics to their returns, and hence the first two moments of returns. The rationale is that asset characteristics are more stably related to expected returns than company names in a large cross section.

of characteristics-based demand (their appendix B), by showing that expected returns and factor loadings are well captured by a few asset characteristics, and that this approach better estimates the mean-variance portfolio compared to a benchmark that uses sample estimates of the first two moments of returns.

Notice that Assumption 4 introduces a second set of asset's characteristics. On one hand the vector x_{jt} of stock characteristics, including prices p_{jt}, that are relevant for the conditional expected returns (μ_x) and factor loadings (Γ_x), we called them *return-related characteristics*. On the other hand c_{jt}, that denotes the stock characteristics that are relevant the portfolio characteristics valued by investors, we called them *taste characteristics*. Empirically, this suggest to include observable stock characteristics known to be relevant for the cross section of return into x_{jt} and those relevant for investment decisions but not directly related to stock returns into c_{jt}.

The following proposition uses all previous assumptions and results to characterize optimal portfolio weights as polynomials on asset characteristics with investor specific coefficients, and then obtain the empirically tractable logit form that relates portfolio holdings with stock characteristics.

Proposition 3. *Under Assumptions 1 to 4:*

(i) The j-th entry of the vectors Γ_{it} and μ_{it} can be written as polynomial functions on x_{jt} with investor-specific coefficients.

(ii) The optimal portfolio weights for each asset j with positive weight, $w_{it,j}$ can be written as polynomial function on asset characteristics (x_{jt}, c_{jt}) with investor specific coefficients:

$$w_{it,j} = \tilde{y}_{jt}' \Phi_{w,i} + \phi_{w,i}, \tag{1.14}$$

where $\Phi_{w,i}$ is vector of coefficients, $\phi_{w,i}$ is a scalar, and $\tilde{y}_{jt}$ is a $K_{\tilde{y}}$-vector with $K_{\tilde{y}} =$

$\sum_{m=1}^{2M}(K_X + K_c)^m$, and

$$\tilde{y}_{jt} = \begin{pmatrix} \tilde{x}_{jt} \\ \tilde{x}_{jt} \otimes \tilde{x}_{jt} \\ \tilde{x}_{jt} \otimes \tilde{x}_{jt} \otimes \tilde{x}_{jt} \\ \vdots \end{pmatrix},$$

where $\tilde{x}_{jt} = (x'_{jt} \quad c'_{jt})'$ is a $(K_X + K_C)$-vector.

(iii) Moreover if the polynomial order M goes to infinity, $M \to \infty$, then a restriction of parameters implies that the optimal portfolio weights for investor i can be written as:

$$w_{it,j} = \frac{\exp\left(x'_{jt}b_{it} + c'_{jt}\gamma_{it} + \xi_{jt}\right)}{1 + \sum_{j=1}^{J_t}\exp\left(x'_{jt}b_{it} + c'_{jt}\gamma_{it} + \xi_{jt}\right)}, \tag{1.15}$$

and the portfolio weight for the outside option is given by

$$w_{it,0} = \frac{1}{1 + \sum_{j=1}^{J_t}\exp\left(x'_{jt}b_{it} + c'_{jt}\gamma_{it} + \xi_{jt}\right)}, \tag{1.16}$$

where b_{it} and γ_{it} are investor-specific coefficients on observed returns-related characteristics x_{jt}, that include price p_{jt}, and taste characteristics c_{jt}. The term ξ_{jt} represents an index of unobserved (by the econometrician) return-related characteristics.

It is possible that there are asset characteristics unobserved by the econometrician but relevant for investor's portfolios. Without loss of generality we can assume those unobserved characteristics are summarized in an index ξ_{jt} and that this index is a characteristic included in x_{jt} as in part (iii) of proposition 3. In the previous the coefficient on the unobserved characteristic ξ_{jt} is normalized to 1 and the portfolio value of the outside option is normalized to 1. Part (i) of this proposition is an assumption in KY2019's framework, as in proposition 2, we are able to obtain this result by relaying on weaker assumptions. Proof of proposition 3 is presented in Appendix A.1.[14]

This proposition tell us that the optimal weight on asset j for investor i in market t is directly

[14] Once part (i) is proved, the proof of parts (ii) and (iii) proceed similarly as in proposition 1 of KY2019.

explained by the characteristics of asset j, investor-specific coefficients (b_{it}, γ_{it}), and the value that investor i assigs to j relative to all the other assets available in J_t.

1.2.4. Market Clearing Condition

The portfolio weights in (1.15) represent the asset demand curves for investors taking stock characteristics, including price as given. In this subsection we pair the demand system with a supply side to obtain a market clearing conditions that pin downs equilibrium prices. The key assumption is that the number of shares outstanding is fixed in the short run. In the empirical application of this chapter we work with the stock market at a quarterly frequency so we consider this assumption reasonable.

Let S_{jt} denote the number of shares outstanding for stock j in market t; in the short run this is assumed to be a fixed number which can be interpreted as an inelastic supply of the stock. If we multiply S_{jt} by P_{jt}, the price per share of j in t, we obtain the market equity for stock j, denote by ME_{jt}.

Since each stock of j should be held by an investor in the market, then the following market clearing should hold:

$$ME_{jt} = \sum_{i=1}^{I} A_{it} w_{it,j}. \tag{1.17}$$

This market clearing condition says that all the money invested by investors in stock j, should equal the market equity of the stock. This condition is a re-statement in dollar value, instead of quantities, that prices in equilibrium are such that in the aggregate, supply equals demand:

$$S_{jt} = \frac{\sum_{i=1}^{I} A_{it} w_{it,j}}{P_{jt}}.$$

Notice that the right hand side of (1.17) depends on prices, since p_{jt} enters demand via de the return-related characteristic x_{jt}. As noted in KY2019, the market clearing condition implies a fixed point equation in p_{jt}. Let p_t be the J_t-vector of log prices for the stocks in market t and define

$f : \mathbb{R}^{J_t} \rightarrow \mathbb{R}^{J_t}$ as:

$$f(p_t) = \log\left(\sum_{i=1}^{I} A_{it} w_{it,j}(p_t)\right) - log(S_{jt}), \tag{1.18}$$

so using equation 1.18 we can solve for the equilibrium prices, by looking for p_t^* such that $p_t^* = f(p_t^*)$. KY2019 state the conditions for an unique fixed point to exist and present an algorithm for the determining such fixed point.

1.3. Demand Specification and Estimation

1.3.1. Demand Specification

Using data on portfolio holdings, $\{w_{it,j}\}$ and stock characteristics $\{x_{jt}, c_{jt}\}$ the goal is to estimate the asset demand coefficients defined in (1.15). This chapter uses a mixed logit demand specification (Berry et al. (1995)), also known as a random coefficients demand (RC) specification, where the investor-specific coefficients in b_{it} and γ_{it} follow the structure:

$$b_{it} = b_0 + \Pi_b d_{it} + \Sigma_b^{1/2} v_{b,it} \tag{1.19}$$

$$\gamma_{it} = \gamma_0 + \Pi_\gamma d_{it} + \Sigma_\gamma^{1/2} v_{\gamma,it}, \tag{1.20}$$

where $b_{0,k}$ is a component of $b_{it,k}$, common to all investors; d_{it} denotes an L-vector of observable investor demographics that are relevant to characterize $b_{it,k}$, along with the corresponding coefficients in Π_b, a $K_X \times L$ matrix. The term $v_{b,it}$ is a K_X-vector of investor-specific taste shocks that are scaled by common variance covariance coefficients, Σ_b. The $v_{b,it}$ can also be interpreted as unobservable (by the econometrician) investor demographics relevant for b_{it}. Analogous interpretations apply to γ_0, Π_γ, Σ_γ and $v_{\gamma,it}$.

If we restrict to zero the matrices $\Pi_b, \Pi_\gamma, \Sigma_b$ and Σ_γ, we have that $b_{it} = b_0$ and $\gamma_{it} = \gamma_0$ for all i. This is the highly studied logit demand. In this specification all investors have the same coefficients and is the one used during estimation in KY2019.[15] Logit demand is highly tractable but delivers restrictive substitution patterns. This is because in Logit demand price elasticities are determined

[15] In KY2019, a logit demand is estimated investor by investor, so they obtain a set of estimated coefficients by investor. However, at the investor level the substitution patters between stocks are those of logit demand.

by market shares. Leading authors in the demand estimation literature call this "*a bug not a feature*" (Berry and Haile (2021), pg. 19). The limitations imply that for small portfolio weights, own-price elasticities are approximately proportional to the coefficient corresponding to price. Moreover, two stocks with similar portfolio weights would react identically to the price change of any other stock.[16] Imagine two stocks that have the same portfolio weights, but the companies belong to different sectors, for example technology and energy. Is easy to imagine these stocks would respond to different fundamentals, yet under a logit demand system, portfolio holdings of another technology stock would react identically to a price change from both stocks.

The restrictive substitution patterns of logit demand are a manifestation of the Independence of Irrelevant Alternatives (IIA) assumption (see, e.g. Arrow (1951), Ray (1973), McFadden (1974)), that states that the relative likelihood of choosing between two options will not change on whether a third alternative is present. Because of IIA, logit demand will fail to capture substitution patterns between close substitutes, a feature that has been widely studied in the industrial organization literature. Extensions of logit demand like the nested logit demand (Cardell (1997)) and the mixed logit demand relax the IIA assumption. In nested logit the product choice is sequential, first individuals choose a product nest, and then they choose a product within the nest. Models of this type can be easily accommodated in a mixed logit demand by including the nest-defining characteristics as one of the characteristics in the demand specification.

Price Elasticities. Let $Q_{it,j} = \frac{A_{it} w_{it,j}}{P_{t,j}}$ denote the number of shares of stock j held by investor i in market t. The elasticity of stock j holdings when the price of stock k changes, denoted by $\eta_{it,jk}$ is given by:

$$\begin{aligned}\eta_{it,jk} &= \frac{\partial Q_{it,j}}{\partial P_{t,k}} \frac{P_{t,k}}{Q_{it,j}} = \frac{\partial \log(Q_{it,j})}{\partial \log(P_{t,k})} \\ &= \frac{\partial \log(w_{it,j})}{\partial \log(P_{t,j})} - 1\{j = k\} := e_{it,jk} - 1\{j = k\}.\end{aligned}$$

[16] Berry and Haile (2021) develop their argument further: "*These restrictions do not come from economics but from assumptions chosen for simplicity or analytical convenience. Models must, of course, abstract from reality, and finite samples require appropriate parsimony. But good modeling and approximation methods should aim to avoid strong a priori restrictions on the very quantities of interest unless those restrictions can be defended as natural economic assumptions.*" (pg. 19).

The term $e_{it,jk}$ is also an elasticity but with respect to portfolio weights and hence depends on the demand specification. As mentioned above log prices, p_{jt} is one of the stock characteristics included in x_{jt}. Without loss of generality, let $k = 1$ denote index for the coefficient corresponding to log prices, $b_{it,1}$. Following this notation, the term $e_{it,jk}$ under logit demand ($b_{it,1} = b_{0,1}$ for all i) is given by:

$$e^{Logit}_{it,jk} = \frac{\partial \log(w_{it,j})}{\partial \log(P_{t,k})} = b_{0,1}(1\{j = k\} - w_{it,k}).$$

The corresponding term for RC demand, $e_{it,jk}$, is given by

$$e^{RC}_{it,jk} = \frac{\partial \log(w_{it,j})}{\partial \log(P_{t,k})} = \left(b_{0,1} + \sum_{\ell=1}^{L} \pi_{1,\ell} d_{\ell it} + \sigma_{t,1} v_{i,1}\right)(1\{j = k\} - w_{it,k}).$$

The term $e_{it,jk}$, represents the price elasticity given the idiosyncratic preferences of investors as capture by the price coefficient $b_{it,1}$. However, this expression is not feasible to compute given that the taste shocks, $v_{i,1}$, are not observed. To compute the price elasticity conditional on observable data is necessary to integrate over the distribution of the taste shocks. If F_v denotes the distribution of v_{it} we can numerically integrate out its role on portfolio holdings using F_v and computing $e^{RC}_{it,jk}$ using $E_{F_V}[w_{it,j}]$:

$$\begin{aligned} e^{RC}_{it,jk} &= \frac{\partial \log\left(E_{F_V}[w_{it,j}]\right)}{\partial \log(P_{t,k})} = \frac{\partial E_{F_V}[w_{it,j}]}{\partial P_{t,k}} \frac{P_{t,k}}{E_{F_V}[w_{it,j}]} \\ &= \frac{1}{E_{F_V}[w_{it,j}]} \int \left[\left(b_{0,1} + \sum_{\ell=1}^{L} \pi_{1,\ell} d_{\ell it} + \sigma_{t,1} v_{i,1}\right) \tilde{w}_{it,j}(v_i)(1\{j = k\} - \tilde{w}_{it,k}(v_i))\right] dF(v_i). \end{aligned}$$

Comparing the terms $e^{Logit}_{it,jk}$ and $e^{RC}_{it,jk}$ shows, as mentioned at the beginning of this section, that a logit demand specification is limited in the substitutions patterns it can accommodate when stock prices change. RC demand, on the other hand, by employing a richer structure on the parameters, can deliver more flexible substitution patterns.

1.3.2. Estimation

The relevant equations for estimation are the mapping between asset characteristics and asset holdings, equation (1.15), paired with the random coefficients specifications that relate investor-specific coefficients with their demographics in (1.19) and (1.20). For convenience we present again these equations:

$$\begin{aligned} w_{it,j} &= \frac{\exp\left(x'_{jt}b_{it} + c'_{jt}\gamma_{it} + \xi_{jt}\right)}{1 + \sum_{j=1}^{J_t} \exp\left(x'_{jt}b_{it} + c'_{jt}\gamma_{it} + \xi_{jt}\right)} \\ b_{it} &= b_0 + \Pi_b d_{it} + \Sigma_b^{1/2} v_{b,it} \\ \gamma_{it} &= \gamma_0 + \Pi_\gamma d_{it} + \Sigma_\gamma^{1/2} v_{\gamma,it}. \end{aligned}$$

The data is composed of asset holdings, investor demographics and assets under management $\{w_{it}, d_{it}, A_{it}\}_{i=1,\dots,I_t}$ and asset characteristics $\{x_{jt}, c_{jt}\}_{j=1\dots,J_t}$. The parameters to estimate are the components that form the coefficients corresponding to return and taste characteristics; namely $\{b_0, \Pi_{b,}, \Sigma_b\}$ for the return characteristics x_{jt} and $\{\gamma_0, \Pi_\gamma, \Sigma_\gamma\}$ for the taste characteristics c_{jt}.

For exposition convenience of the estimation steps, let's rewrite observed asset characteristics into the $K = (K_X + K_C)$ vector $X_{jt} = (x'_{jt}, c'_{jt})'$, and accordingly define the coefficients vector $\beta_{it} = (b'_{it}, \gamma'_{it})'$ such that

$$\beta_{it} = \beta_0 + \Pi d_{it} + \Sigma^{1/2} v_{it}, \tag{1.21}$$

where $\beta_0 = (b'_0, \gamma'_0)'$; Π is a $K \times L$ matrix with the demographics coefficients (Π_b, Π_γ); $v_{it} = (v_{b,it}, v_{\gamma,it})'$ a K-vector, and $\Sigma^{1/2}$ a $K \times K$ matrix composed of $(\Sigma_b^{1/2}, \Sigma_\gamma^{1/2})$. With this notation the parameters to estimate can be denoted as $\theta := (\beta_0, \Pi, \Sigma)$. In the IO literature for demand estimation $\theta_1 = \beta_0$ is commonly referred as "linear parameters" and $\theta_2 := (\Pi, \Sigma)$ as "non-linear parameters", due to the way these parameter enter the estimation procedure.

With this notation we can write the exponents in the expression for $w_{it,j}$ as

$$\begin{aligned} x'_{jt}b_{it} + c'_{jt}\gamma_{it} + \xi_{jt} &= X'_{jt}\left[\beta_0 + \Pi d_{it} + \Sigma v_{it}\right] + \xi_{jt} \\ &= \underbrace{X'_{jt}\beta_0 + \xi_{jt}}_{:=\delta_{jt}} + \underbrace{X'_{jt}\left[\Pi d_{it} + \Sigma v_{it}\right]}_{:=h_{ijt}(\theta_2, v_{it})} \\ &= \delta_{jt} + h_{ijt}(\theta_2, v_{it}). \end{aligned}$$

The term δ_{jt} is referred as the "mean utility" for option j in market t, as it is a common component for all investors; while the term $h_{ijt}(\theta_2, v_{it})$ captures investor-specific heterogeneity. Furthermore, we can write $w_{it,j}$ as

$$w_{it,j} = \frac{\exp\left(\delta_{jt} + h_{ijt}(\theta_2, v_{it})\right)}{1 + \sum_{j=1}^{J_t} \exp\left(\delta_{jt} + h_{ijt}(\theta_2, v_{it})\right)}. \tag{1.22}$$

The next step is to obtain aggregate market shares for each stock. In this chapter we perform estimation at the market level. This is motivated because market-level estimation facilitates dealing with the endogeneity of prices, as discussed below, so we can use instrumental variables for prices that have been suggested in the IO literature of demand estimation. This choice is also consistent with the market definition presented in the microfoundations, namely we consider a market to be US stock market in a given a quarter.

We can construct aggregate market shares from the market clearing condition (1.17). Let ME_{0t} denote the aggregate investment in the outside option: $ME_{0t} = \sum_{i=1}^{I_t} A_{it}w_{i0t}$; and denote by ME_t denote the aggregate value of the market in t: $ME_t = \sum_{j=0}^{J_t} ME_{jt}$. By the market clearing condition it has to be the case that the aggregate value of the market is equal to the aggregate assets under management across investors, so we have that $ME_t = \sum_{i=1}^{I_t} A_{it} := A_t$. If we divide the market clearing condition (1.17) by ME_t (or A_t equivalently) we obtain that

$$
\begin{aligned}
ME_{jt} &= \sum_{i=1}^{I_t} A_{it} w_{it,j} \\
\Rightarrow s_{jt} := \left(\frac{ME_{jt}}{ME_t}\right) &= \sum_{i=1}^{I_t} \left(\frac{A_{it}}{At}\right) w_{it,j}.
\end{aligned} \tag{1.23}
$$

This equation tell us that in the aggregate we will compare the observed stock market shares of a "market value portfolio" (s_{jt}) with the model implied shares of a "wealth-adjusted aggregate portfolio" $\left(\sum_{i=1}^{I_t}\left(\frac{A_{it}}{At}\right) w_{it,j}\right)$. To compute the right hand side of (1.23), the model-implied shares, one challenge is that the investor-specific taste shocks v_{it} are not observed by the econometrician. To deal with this problem it is common to assume a prior distribution on these latent variables. If F_v denotes the distribution of v_{it} we can numerically integrate out its role on portfolio holdings using F_v. The model-implied shares, denoted by $\tilde{s}_{jt}$, are

$$
\tilde{s}_{jt} := \sum_{i=1}^{I_t} \left(\frac{A_{it}}{A_t}\right) \int \frac{\exp\left(\delta_{jt} + h_{ijt}(\theta_2, d_{it}, v_{it})\right)}{1 + \sum_{j=1}^{J_t} \exp\left(\delta_{jt} + h_{ijt}(\theta_2, d_{it}, v_{it})\right)} dF_v(v_{it}). \tag{1.24}
$$

It is standard practice in the IO literature to assume F_v to be a multivariate normal with zero mean and variance-covariance matrix S_v, that is $v_{it} \sim N(0, S_v)$; then the integral in (1.24) can be numerically approximated for example by monte carlo simulation or using Gauss-Hermite quadrature procedures. The $\tilde{s}_{jt}(\cdot)$ functions define a demand system

$$
\tilde{s}(\delta_t, \theta_2; d_t, X_t, J_t) = \left(\tilde{s}_1(\delta_t, \theta_2; d_t, X_t, J_t), ..., \tilde{s}_{J_t}(\delta_t, \theta_2; d_t, X_t, J_t)\right). \tag{1.25}
$$

The aggregation across investors in (1.24) with wealth-based weights (A_{it}/A_t) makes this demand system different from the standard RC demand system (for example the canonical BLP demand system of Berry et al. (1995)). However, we can prove this system is *invertible* in the sense that given $(\theta_2; d_t, X_t)$ there is an unique vector δ such that for all j:

$$\tilde{s}_{jt}(\delta, \theta_2; d_t, X_t, J_t) = s_{jt}.$$

Proposition 4. *Demand Inversion*

The demand system in (1.25) is invertible such that given $(\theta_2; d_t, X_t, J_t)$ and nonzero market shares s_{jt} with $\sum_{j=1} s_{jt} < 1$, there exists an unique vector δ such that $\tilde{s}_{jt}(\delta, \theta_2; d_t, X_t, J_t) = s_{jt}$, for all j.

The proof is included in Appendix A.1 and it follows by verifying the conditions of the Berry's inversion theorem (Berry (1994)). Such vector δ can be found as the fixed point of the following contraction mapping. Let $f : \mathbb{R}^J \to \mathbb{R}^J$ that for fixed $(\theta_2; d_t, X_t, J_t)$ is given by

$$f(\delta) = \delta + \log(s_t) - \log(\tilde{s}_t(\delta, \theta_2; d_t, X_t, J_t)), \tag{1.26}$$

notice that if δ^* is such that $f(\delta^*) = \delta^*$ then $\log(s_t) = \log(\tilde{s}_t(\delta^*, \theta_2; X_t, J_t))$. The fact that the demand system is invertible allow us to operationalize an estimation strategy centered around the term ξ_{jt} being interpreted as an structural error term.

Using ξ_{jt} as an structural error term for estimation is predicated on the assumption that unobserved stock characteristics should be conditionally mean zero with respect to a vector z_{jt} of observable stock characteristics,

$$E[\xi_{jt}|z_{jt}] = 0. \tag{1.27}$$

In particular this condition implies that unobservable stock characteristics in z_{jt} should not be correlated with unobserved characteristics ξ_{jt}. In the context of asset demand, as KY2019 mention, this moment condition is motivated by the literature of asset pricing in endowment economics (Lucas (1978)) that assumes that shares outstanding and asset characteristics other than price are exogenous.

The fact that the demand system (1.23) is invertible, allow us to exploit moment (1.27) for estimation. Given θ_2 we can "invert the demand" to find $\hat{\delta}_t(\theta_2)$ such that model-implied shares $\tilde{s}_t(\hat{\delta}_t, \theta_2)$

match the observed shares, s_t. With $\hat{\delta}_t(\theta_2)$ we can construct $\xi_{jt}(\theta)$ given by $\xi_{jt}(\theta) = \hat{\delta}_{jt}(\theta_2) - X'_{jt}\theta_1$ and then we can select $\theta = (\theta_1, \theta_2)$ that minimizes a GMM objective function based on (1.27). Formally, in the GMM strategy for estimation of the demand system in (1.25) we look for $\hat{\theta}_{GMM}$ in market t that solves:

$$\min_{\theta} \quad g\left(\xi_t(\theta)\right)' W_t g\left(\xi_t(\theta)\right) \tag{1.28}$$

$$\text{s.t.} \quad g\left(\xi_t(\theta)\right) = \frac{1}{J_t}\sum_{j=1}^{J_t} z_{jt}\xi(\theta)_{jt} \tag{1.29}$$

$$\xi(\theta)_{jt} = \delta_{jt}(\theta_{2t}) - X'_{jt}\theta_1 \tag{1.30}$$

$$\log(s_{jt}) = \log\left(\tilde{s}_{jt}(\delta_t, \theta_2; d_t, X_t, J_t)\right) \tag{1.31}$$

$$\tilde{s}_{jt}(\delta_t, \theta_2; d_t, X_t, J_t) = \sum_{i=1}^{I_t}\left(\frac{A_{it}}{At}\right)\int \frac{\exp\left(\delta_{jt} + h_{ijt}(\theta_2, d_{it}, v_{it})\right)}{1 + \sum_{j=1}^{J_t}\exp\left(\delta_{jt} + h_{ijt}(\theta_2, d_{it}, v_{it})\right)} dF_v(v_{it}), \tag{1.32}$$

where W_t is a $K_Z \times K_Z$ weight matrix. Using the first order conditions of the GMM problem is possible to show that the parameter search can be simplified to be just over θ_2 (see Berry et al. (1995) and Nevo (2000) for useful derivations). To see why, it suffices to notice that the first order conditions with respect to θ_1 requires

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta_1}\left(g\left(\xi(\theta_t)\right)' W_t g\left(\xi(\theta_t)\right)\right) \\ \Leftrightarrow 0 &= -\frac{2}{J_t^2}[\delta_t(\theta_2) - X_t\beta_{0t}]' Z_t W_t Z'_t X_t \\ \Leftrightarrow \theta_1 &= [X'_t Z_t W_t Z'_t X_t]^{-1}[X_t Z_t W_t Z'_t]\delta_t(\theta_2), \end{aligned} \tag{1.33}$$

so given a value of θ_2, there is a corresponding value for θ_1 according to the GMM objective function. If the estimation is carried out using data from multiple market periods but the parameters in θ are common across such periods, the GMM strategy would compute (1.30) to (1.32) each period t and then stack the moment analogs (1.29) for each market before computing the objective function in (1.28). Algorithm 1 sketches the steps necessary to implement the GMM estimation of the random coefficients demand.

Algorithm 1: Mixed Logit Demand Estimation

Input: Stock characteristics $\{X_{jt} = (x'_{jt}, c'_{jt})'\}$, aggregate stock holdings $\{s_{jt}\}$ and instrumental variables $\{Z_{jt}\}$ for $j = 1, ..., J_t$; assets under managament and demographics $\{A_{it}, d_{it}\}$ for $i = 1, ..., I_t$ for a given market t.

Output: A set of estimated parameters $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ with $\hat{\theta}_1 = \hat{\beta}_0$ and $\hat{\theta}_2 = (\hat{\Pi}, \hat{\Sigma})$.

Initialize: Pick initial values for $\theta_1^{(0)}$ and $\theta_2^{(0)} = (\Pi^{(0)}, \Sigma^{(0)})$.

During step $r \geq 1$ of the optimization routine, do:

i. Compute an initial value for $\delta_t^{(r,0)}$. If $r = 1$ the initial value can be $\delta_t^{(r,0)} = X'_{jt}\theta_1^{(0)}$.

ii. Given the current value for $\theta_2^{(r)}$, *invert* the demand system using the contraction mapping in (1.26).
Iterate until convergence an update for δ_t where the h-th update is given by:

$$\delta_t^{(h)}\left(\theta_2^{(r)}\right) = \delta_t^{(h-1)} + \log(s_t) - \log\left(\tilde{s}_t\left(\delta_t^{(h-1)}, \theta_2^{(r)}; X_t, J_t\right)\right)$$

Use $\delta_t^{(r,0)}$ for the first update. The resulting vector will be a function of $\theta_{22}^{(r)}$ and aggregate stock holdings s_t, denoted by $\delta_t^{(r)}$.

iii. Update the value of the linear parameters using (1.33)

$$\theta_1^{(r)} = [X'_t Z_t W_t Z'_t X_t]^{-1}[X_t Z_t W_t Z'_t]\delta_t^{(r)}.$$

iv. Use $\delta_t^{(r)}$ and $\theta_1^{(r)}$ to compute the GMM error term:

$$\xi_t^{(r)} = \delta_t^{(r)} - X'_t\theta_1^{(r)},$$

and the GMM moment function, notice that g is a function of the parameters in step r:

$$g\left(\theta^{(r)}\right) = \frac{1}{J_t} Z'_t \xi_t^{(r)}.$$

v. Evaluate the GMM objective function at $\theta^{(r)}$. If the objective function has converged report $\hat{\theta}_{GMM} = \theta^{(r)}$. If no convergence has been achieved update $\theta_2^{(r)}$ according to the optimization algorithm used (e.g. a Newton-Rapson update). Label this update as $\theta_2^{(r+1)}$.

vi. Repeat steps i. to v. until convergence of the GMM objective function.

Price Endogeneity. Since the unobserved stock characteristics ξ_{jt} are part of investors demand they will also be a determinant of equilibrium prices. This means that prices and functions of prices will be correlated with ξ_{jt} and cannot be included in z_{jt} for estimation. To solve for the endogeneity of prices with respect to ξ_{jt} we need instrumental variables (IVs) correlated with prices but exogenous with respect to ξ_{jt}.

We consider instrumental variables of the style of Gandhi and Houde (2019). For each stock j denote with $J_t(j)$ the set of stocks that belong to j's industry. Next, for each exogenous dimension k in X_{jt}, we compute a metric of j's isolation with respect to other stocks in $J_t(j)$:

$$X_{jt,k}^{GH} = \sum_{\tilde{j} \in J_t(j)} (X_{jt,k} - X_{\tilde{j}t,k})^2. \tag{1.34}$$

Then the vector of instrumental variables z_{jt} used for estimation will be composed of the exogenous characteristics in X_{jt} plus the Gandhi-Houde IVs (GH-IVs) constructed from such exogenous characteristics. If the dimensions considered to be exogenous with respect ξ_{jt} are in fact exogenous, then the GH-IVs would be uncorrelated with ξ_{jt} by construction since they rely on the values of an exogenous characteristic for j and the corresponding values for those stocks in j's industry.

The case for relevance is more interesting; since stocks are considered as differentiated investment products, they compete on the characteristics they offer to investors. Those stocks with more attractive characteristics to investors will have a relatively higher demand, all else equal. Then, metrics of j's isolation with respect to other stocks in $J_t(j)$ along an exogenous characteristic would capture stock j's ability to compete on such characteristic against alternative stocks in j's industry. If the alternatives of stock j in its industry offer more (less) of a characteristic positively value by investors relative to j, then there will be more (less) demand for alternatives of stock j, less (more) for stock j itself and that will decrease (increase) the price of stock j. Hence metrics of stock j's isolation with respect to other stocks in $J_t(j)$ will be correlated with the price of stock j.

Logit demand estimation. When the non-linear parameters θ_2 are restricted to zero, $\theta_2 = (\Pi, \Sigma) = (0, 0)$, we are in the case of logit demand. In this case the demand system given by $\tilde{s}_{jt}$ in

(1.24) becomes:

$$\begin{aligned}
\tilde{s}_{jt} &= \sum_{i=1}^{I_t} \left(\frac{A_{it}}{A_t}\right) \int \frac{\exp\left(\delta_{jt} + h_{ijt}(0, d_{it}, v_{it})\right)}{1 + \sum_{j=1}^{J_t} \exp\left(\delta_{jt} + h_{ijt}(0, d_{it}, v_{it})\right)} dF_v(v_{it}). \\
&= \frac{\exp\left(\delta_{jt}\right)}{1 + \sum_{j=1}^{J_t} \exp\left(\delta_{jt}\right)} \sum_{i=1}^{I_t} \left(\frac{A_{it}}{A_t}\right) \\
&= \frac{\exp\left(\delta_{jt}\right)}{1 + \sum_{j=1}^{J_t} \exp\left(\delta_{jt}\right)},
\end{aligned}$$

in the second line we used the fact that when $\theta_2 = 0$ then $h_{ijt} \equiv 0$, and in the third line we use the fact that wealth weights should sum up to one. In the case of logit demand we can perform the demand inversion analytically, since $\tilde{s}_{0t} = 1/(1 + \sum_{j=1}^{J_t} \exp\left(\delta_{jt}\right))$ then:

$$\log\left(\tilde{s}_{jt}/\tilde{s}_{0t}\right) = \delta_{jt} = X'_{jt}\beta_0 + \xi_{jt},$$

the first equation tell us that if we set the value of δ_{jt} to $\log\left(s_{jt}/s_{0t}\right)$ then observed shares will match the (logit) model-implied shares. The second equation which is the definition of δ_{jt} tell us how to construct ξ_{jt} to use it in a GMM estimation strategy base on moment (1.27). Specifically, logit estimates will be obtained by linear IV-GMM based on (1.27) and using the Gandhi-Houde IVs described above.

1.4. Demand for Green Stocks

In this section we start by presenting the random coefficients demand specification we'll use for estimation. Then we present the data sources and finally we present the demand estimation results.

Following the notation of equation (1.15), we estimate a demand specification given by

$$w_{it,j} = \frac{\exp\left(x'_{jt}b_0 + c_{jt}\gamma_{it} + \xi_{jt}\right)}{1 + \sum_{j=1}^{J_t} \exp\left(x'_{jt}b_0 + c_{jt}\gamma_{it} + \xi_{jt}\right)}, \tag{1.35}$$

where the vector of return-related characteristics is given by

$$x_{jt} = (1,\ \text{mktBeta}_{jt},\ \text{lat}_{jt}, \text{lbme}_{jt},\ \text{profitability}_{jt},\ \text{investment}_{jt}),$$

that includes an intercept, the stock's market beta (mktBeta_{jt}), log total assets (lat_{jt}), log book-to-market equity (lbme_{jt}), stock's profitability ($\text{profitability}_{jt}$) and stock's investment (investment_{jt}).[17] These return characteristics are motivated by the Fama-French five-factor model (Fama and French (2015)) that offer sensible dimensions to characterize the cross section of returns.[18]

The vector of taste characteristics c_{jt} is composed of the environmental scores of company j in t, $c_{jt} = (\text{Escore}_{jt})$. Moreover, the coefficients corresponding to the return-related characteristics are treated as homogeneous across investors.[19] The coefficient for environmental scores is modeled as heterogeneous across investors following the structure:

$$\gamma_{it} = \gamma_0 + \kappa d_{it} + \sigma v_{it}, \tag{1.36}$$

where the parameters $(\gamma_0, \kappa, \sigma)$ are common across investors but each investor has a different sensitivities to the environmenal scores of the stocks because they differ in their observed demographics d_{it} and unobserved demographics, or taste shocks, v_{it}. In this demand specification we use the investor's assets under management as observed demographics, $d_{it} = \log(\text{AUM})_{it}$; and we assume unobserved demographics follow an standard normal distribution independent and identically distributed across investors. This distributional assumption is common practice in the demand estimation literature and facilitates the numerical approximation of the integral in the definition of model shares, $\tilde{s}_{jt}$ in (1.24), during estimation we approximate such integral using a Gauss-Hermite quadrature approximation of order 20.[20]

[17] Profitability is measure as operating profits to book equity. Investment is measure as annual log growth of total assets.

[18] There is a growing literature in the asset pricing questioning whether the Fama-French five-factor characteristics are sufficient to explain the cross section of returns (e.g. Han et al. (2021)). However, considering alternative returns characteristics other than those in the Fama-French five-factor is left for future research.

[19] Allowing for heterogeneity on return-related characteristics is left for future research.

[20] This guarantees the integral is exact for polynomial functions of degree up to 49.

1.4.1. Data

There are two main sources of data: stock characteristics and portfolio holdings. Data on portfolio holdings comes from the Thomson Reuters Institutional Holdings Database that contains data on institutional investors that file the Form 13F from the Securities and Exchange Commission (SEC). Investment institutions that manage more than $100 million are required to disclose stock holdings in the Form 13F. These institutions can be banks, insurance companies, mutual, hedge, and pension funds, as well as other 13F institutions like foundations, nonfinancial corporations, and endowments.

Price and stock characteristics data for this chapter comes from the Compustat and Center for Research in Security Prices (CRSP) datasets which we combine to obtain fundamentals for publicly traded companies in the US stock market. Data for stock prices, dividends, returns and shares outstanding can be obtained from the CRSP Monthly Stock Database. Accounting data from the Compustat North America Fundamentals Annual and Quaterly Databases are combined with CRSP data to construct asset characteristics.

The data uses common stocks (with share codes 10, 11, 12, 18) that trade in the New York Stock Exchange, the American Stock Exchange and Nasdaq (exchange code 1, 2, 3 respectively); those stocks with missing data on returns or prices are filter out. Data on the CRSP database are merge with Compustat database records most recent of at least 6 months, and no more of 18 months prior to trading date. This is to guarantee that accounting data were public on the trading date.

Data on environmental performance of listed companies comes from the MSCI rating agency.[21] MSCI is a pioneer rating agency in the construction of scores that evaluate the Environmental, Social, and Governance (ESG) performance of the firms they rate.[22] We obtained a dataset of firm-level annual ESG scores from 1991 to 2019. One key advantage of the dataset from MSCI is the availability of granular data, for each of the three pillars: E, S and G, the dataset includes a series of

[21] There is a growing number of data providers of ESG scores and a growing literature studying to what extend scores from different vendors are consistent in their evaluations (e.g. Billio et al. (2021)). Using data from other vendors would not alter the methodology here presented, but comparing how results would change when using different ESG score are used is left for future research.

[22] Moreover, as mentioned in Pástor et al. (2022), MSCI has been voted "Best firm for SRI research" in the The Extel & SRI Connect Independent Research in Responsible Investment (IRRI) Survey each year from 2015 to 2019 (see https://www.msci.com/zh/esg-ratings).

performance indicators used to construct the final score. For most vendors of ESG scores, the final score is the result of translating raw data into a numerical score using a proprietary algorithms; with the MSCI granular data we can construct scores directly from the raw data guided by the application at hand.

We focuse on MSCI variables from the "Environmental pillar score", where firms are evaluated in several indicators that capture either positive or negative environmental performance. Positive indicators include appropriate waste management, product carbon footprint, and energy efficiency, while negative indicators include regulatory compliance, toxic emissions and waste, water stress; see Appendix A.2 for a full list of indicators.[23] Using these indicator variables on environmental performance we construct a E-score that compares firm cross-sectionally each year according to their environmental performance.

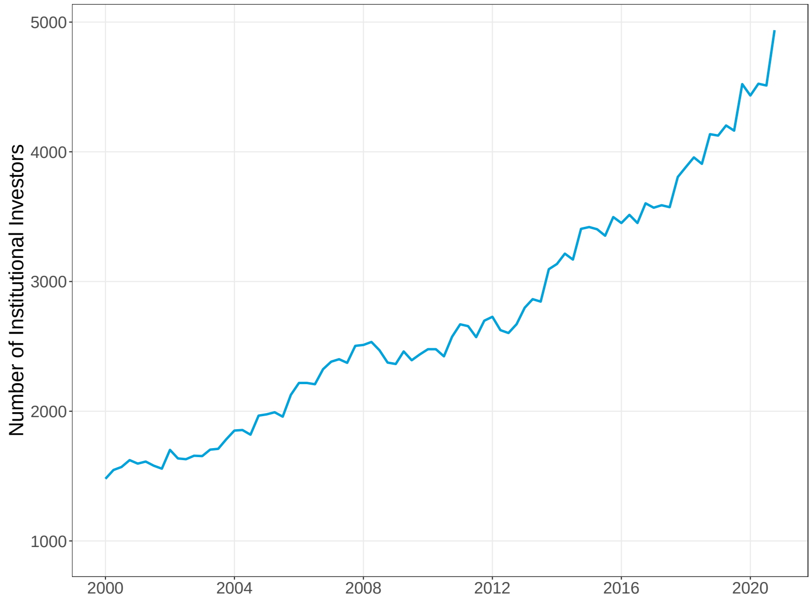
Figure 1.1 presents the evolution of the number of stocks and institutional investors on the sample period. Panel (b) shows that 13F institutions have grown in importance over the sample period. Towards the end of the sample institutional investor collectively managed around 70% of the US stock market from around 53% at the beginning of the sample. See Appendix A.2 for the distribution of institutional investors according to their type over the sample period.

Moreover in each market, we construct a residual investor labeled as the *household sector*. The stock holdings of the household sector are defined as difference between shares outstanding and the sum of shares held by 13F institutions. The introduction of the household sector is necessary for the market clearing to hold. The outside option, $j = 0$, will be set to include be stocks that are foreign (code 12), real estate investment trusts (code 18) or have missing characteristics or returns.

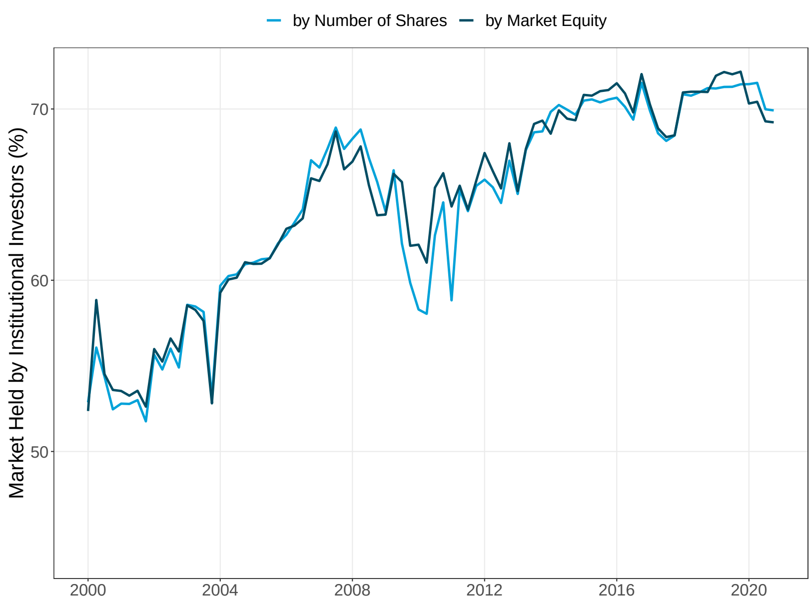
Data Construction. The construction of the return-related characteristics in the vector x_{jt} follows KY2019. It includes five characteristics: market beta, log total assets, log book-to-market equity, a metric of firm's profitability and a metric of firm's investment. Market beta for each stock j is compute on a 60-month rolling window where we the market return comes from Ken-

[23]The threshold for satisfying an indicator are determined by MSCI and are not disclosed with the data.

Figure 1.1: Institutional Stock Holdings



(a) Number of institutional investors



(b) Market held by institutional investors

Notes: Panel (a) shows the evolution on the number of institutional investors in the dataset from 2000-Q1 to 2020-Q1. Panel (b) shows the how much of the U.S. stock market is held by institutional investors from 2000-Q1 to 2020-Q1. First, by number of shares shows the percentage of shares outstanding that is held by institutional investors across all stocks. Second, by market equity shows the dollar value of stocks held by institutional investor as a percentage of the total market equity across all stocks.

neth French's website and the risk-free rate from 3-month Treasury bills.[24] Log market equity is computed summing log price per share at the end of the quarter with the log of shares outstanding expressed in millions. To compute the profitability metric we follow Fama and French (2015) and compute operating profits to book equity. Operating profits in turn are computed as total revenue ($revt$) minus the sum of cost of goods sold ($cogs$), selling, general and admin expenses ($xsga$), and interest and related expense-total ($xint$), or $profit = (revt - cogs - xsga - xint)$. The investment metric is the 1-year log growth of total assets. Table A.1 in Appendix A.2 shows summary statistics for the return characteristics in x_{jt} grouped by sample's year. To reduce the impact of outliers on some variables, profitability, market beta and investment are windsorized at the 2.5th and 97.5th percentile. This windsorizing is done in a quarterly fashion.

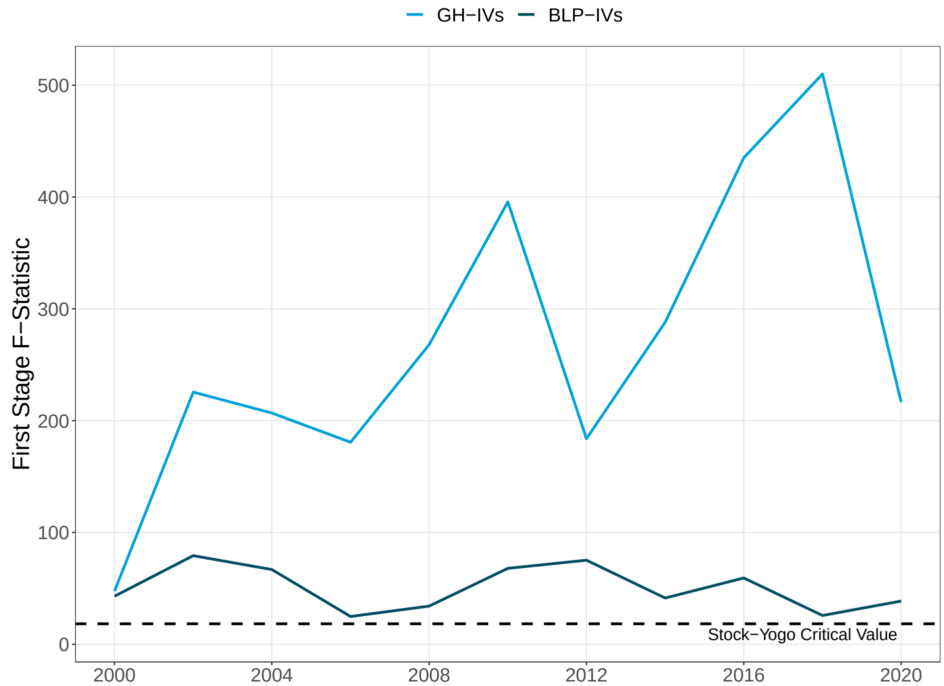
The environmental scores are constructed following Engle et al. (2020) and Hong and Kostovetsky (2012). For each firm in the MSCI dataset we count the number of positive indicators and subtract from it the number of negative indicators, we can this difference raw E-scores. Then, after merging the raw score into the dataset of stock holdings and asset characteristics for each quarter, we rank the raw E-scores cross-sectionally and standardize to range in the interval between -1/2 and 1/2; this are the E-scores used for estimation. In this standardization the median raw score is mapped to zero, 1/2 corresponds to the stock with the highest environmental performance, the *greenest* stock, and -1/2 corresponds with the lowest environmental performance, the *brownest* stock. The data on each firm on the MSCI dataset is updated at least once a year, but not all firm scores get update at the same in a given year. To ensure E-scores are public on the trading date, we merge stock holdings and return characteristics in period t with the E-scores from the calendar year prior to t.[25]

For estimation, the return-related characteristics other that do not depend on price directly are assumed to exogenous (with respect to ξ_{jt}), that is the market beta, log book equity, profitability and investment. The E-score is also considered as exogenous. On these five characteristics we construct the Gandhi-Houde IVs described above, according to (1.34). Figure 1.2 shows the results of a first stage F-test for the null of weak instruments across the sample period. The figure also

[24] The data can be consulted at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[25] For example, in 2018-Q1 and 2018-Q4 we use E-scores from 2017, whereas in 2019-Q1 we use E-scores from 2018.

Figure 1.2: First Stage F-stat of the Instrumental Variables



Notes: First stage F-statistic on the instruments for log Book to Market Equity. We present results for the Gandhi-Houde IVs and the BLP-type IVs. Stock and Yogo (2005) critical value (18.37) for 1 endogenous regressor, 5 instrumental variables and 0.05 bias of two stage least squares relative to OLS. Quarterly sample from 2000-Q1 to 2020-Q4.

shows the F-statistics when using "BLP"-type instruments, another common choice of instrumental variables in the IO literature.[26] In all of the estimation windows the F-statistic of the GH-IVs is above the appropriate critical value to reject the null of weak instruments at 5 percent significance level. Moreover, relatively to BLP-IVs, the null of weak IVs is rejected more easily using GH-IVs.

1.4.2. Estimates

Recall that in estimation the goal is to used data on stock characteristics $\{X_{jt} = (x'_{jt}, c'_{jt})'\}$, aggregate stock holdings $\{s_{jt}\}$, instrumental variables $\{Z_{jt}\}$, and assets under managament and demographics $\{A_{it}, d_{it}\}$ to estimate the parameters in θ, composed of linear parameters $\theta_1 = (b_0, \gamma_0)$ and non-linear parameters $\theta_2 = (\kappa, \sigma)$.

We use twenty years of quarterly data of stocks holdings and characteristics from 2000-Q1 to 2019-

[26]The BLP instruments for the price of stock j are constructed as the sum of the exogenous characteristics of other stocks in j's industry.

Q4. We perform estimation in two-year windows, so for every estimation window we obtain a set of parameter estimates $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ that use data from eight quarters.[27] Moreover during estimation we stardardize the variables log total assets, profitability, investment and E-scores to have mean zero and standard deviation one at each quarterly cross section of stocks. This way the estimated coefficients can be interpreted as the semi-elasticity with respect to the corresponding stock characteristic if multiplied by 100.

Figure 1.3 shows the effective coefficient on E-scores, $\hat{\gamma}_{it} = \hat{\gamma}_0 + \hat{\kappa}\log(\text{AUM})_{it} + \hat{\sigma}v_{it}$, over 2-year estimation windows, and according to estimation based on logit demand or random coefficients demand (RC). The plot uses the mean value, in each window, of log assets under management and shows the 95% confidence interval, of how the coefficient on E-scores varies across investors based on the assumption that unobserved demographics, v_{it}, are normally distributed.

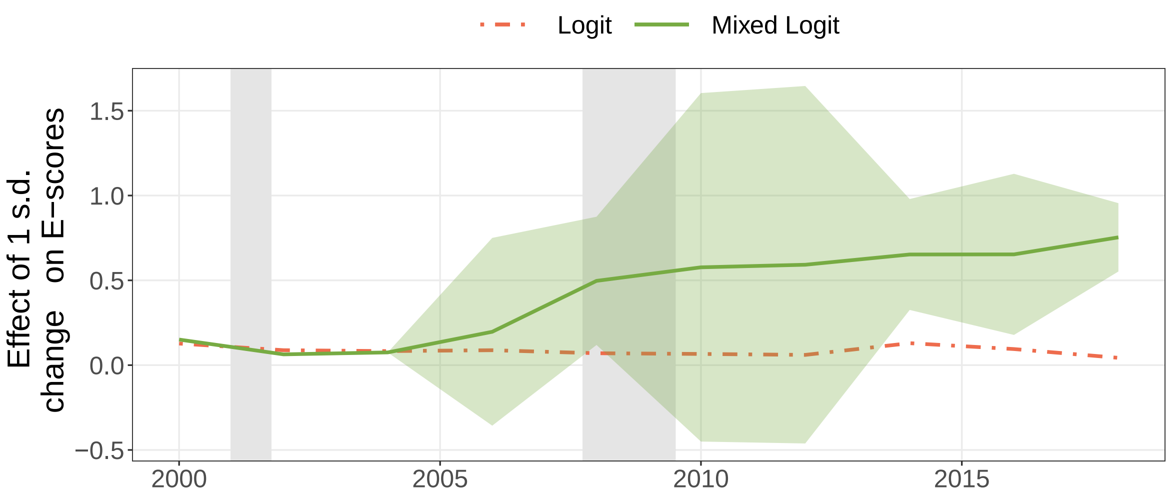
From Figure 1.3 we can see that the sensitivity to E-scores varies over time but it is consistently positive throughout the estimation sample. This is true for both logit and mixed logit estimation. After the Great Recession (2007-Q4 to 2009-Q2) period there is an increase in the range of values for the coefficient on E-scores, due to larger estimated values for $\hat{\sigma}$. This suggests an increase in the heterogeneity in the sensitivity to green characteristics across investors after this period. One possible explanation is that after periods of economic downturn, some investors may be more interested in stocks with higher returns and relatively less interested in the environmental-friendliness of the companies underlying the stocks.

The range of values for the effective coefficient on E-scores after the Great Recession suggests that for some investors the sensitivity is consistent with a preference for brown stocks. For such investors, if there is a determinant of returns not captured by the return-related characteristics of the Fama-French five factor model which is higher for brown stocks, an appetite for returns could explain the preference for brown stocks in this period.[28]

[27]The first estimation window uses data from 2000-Q1 to a 2001-Q4, and the last estimation window uses data from 2018-Q1 to 2019-Q4.

[28]As presented in Pástor et al. (2021), brown stocks can have positive CAPM alphas and higher expected returns than green stocks because they are more exposed to climate risk. Similarly, Bolton and Kacperczyk (2021) find evidence for a carbon premium, in which companies with higher carbon emissions earned higher returns; they also provide evidence that such carbon premium cannot be explained entirely by traditional risk factors. Moreover, In

Figure 1.3: Estimated Coefficients for E-scores



Notes: This plot shows the effect on the demand exponent of one standard deviation change on E-scores based on the associated coefficient for E-scores, $\hat{\gamma}_{it} = \hat{\gamma}_0 + \hat{\kappa}\log(\text{AUM})_{it} + \hat{\sigma}v_{it}$. Estimates are obtained over 2-year estimation windows ranging from 2000-Q1 to 2019-Q4, and according to estimation based on logit demand or random coefficients demand (RC). Multiply this effect by 100 approximates the semi-elasticity of portfolio weights with respect to E-scores. The plot uses the mean value, in each window, of log assets under management and shows the 95% confidence interval of how the coefficient on E-scores varies across investors based on the assumption that unobserved demographics, v_{it}, are normally distributed. Recession periods of the US economy are shown as shaded gray regions. The case of logit demand $\hat{\gamma}_{it} = \hat{\gamma}_{0t}^{logit}$.

From 2005 until the end of the sample, the mixed logit estimates on the coefficient for E-scores is increasing. For example, in the estimation window 2018-2019, the estimated semi-elasticity on the holdings of a stock after a 1 standard deviation increase in its E-score would result in a 73% increase in its holdings, compared with a 48% increase in the 2008-2009 estimation window.

Figure 1.3 also shows the coefficient corresponding to E-scores if we perform estimation under a logit demand specification. In the logit case, there is no heterogeneity in the coefficient and all investors share the same sensitivity to the score, $\hat{\gamma}_{it} \equiv \hat{\gamma}_0^{logit}$ for all i. It is clear from the figure that the logit estimates exhibit much less variation and they don't increase in the second part of the sample.

Another takeaway from estimation is that the sensitivity to E-scores depends on the investor's assets

a related study for mutual funds, Das et al. (2018) found that in the three year period after the Great Recession socially responsible mutual funds exhibited a negative and significant alpha with respect to the Fama-French five factor model. Moreover, they found that funds with lower ESG scores outperformed the fund with high ESG scores during this period.

under management. An alternative version of Figure 1.3 is presented in Appendix A.3. It shows the effective coefficient on E-scores plotted not only at the mean value but also at the 25th and 75th percentile of log assets under management in each estimation window. The main trends and features of the E-scores discussed above do not change by plotting the coefficient under various values of log assets under management; of course in a given estimation window the effective coefficient γ_{it} varies with the log assets under management of investor i according to the coefficient $\hat{\kappa}$. The average across estimation windows in the period 2010-2019 for $\hat{\kappa}$ is positive, so larger investors will are more sensitive to E-scores and will have higher demand for green stocks keeping other coefficients and stock characteristics fixed. This is consistent with Koijen et al. (2023) that find that large investors have a higher demand for stocks with higher environmental scores.[29] This is also consistent with the finding in Pastor et al. (2023) that larger investors tend to tilt their portfolio towards green stocks relative to smaller investors.

In estimation we also consider return-related characteristics in the demand for stocks. Figure 1.4 shows the coefficients corresponding to the return characteristics. For most periods, these characteristics have corresponding coefficients with the same sign. Characteristics like market beta have negative coefficients, which is consistent with the interpretation that market beta captures a basic dimension of risk and that risk is disliked by investors. An estimated negative coefficient for book-to-market equity suggests a preference for growth stocks. In periods of low interest rates, like following the Great Recession, growth stocks may be preferred by investors to value stocks and have larger equity valuations. The sensitivity to profitability and investment peak in periods where we also observe low sensitivity to market beta. This could correspond to a change in investor preferences valuing forward-looking aspects of the firms, such as profitability and investment, relatively more, and valuing backward-looking aspects of the firms, such as market beta, relatively less. Notably the 2008-2009 estimation window that includes the Great Recession period is where we observe the largest confidence intervals around most of the estimates. General uncertainty about stock market performance could be reflected in the relatively large standard errors for that period.

[29] The value of the average across estimation windows is 0.303. The magnitude of this average is not directly comparable with the one reported in Koijen et al. (2023) due to the type of data and demand specification they use.

Table 1.1: Example of Estimated Price Elasticities

		Elasticities (%)	
Price Change	Portfolio Weight (%)	Logit	Mixed Logit
Apple	2.83	-0.5686	-0.5809
Alphabet	1.02	-0.0045	-0.0023
Exxon Mobile	1.02	-0.0045	-0.0013

Notes: Estimated elasticities of the aggregate holdings Apple with respect to the price change of selected stock prices. Data and estimates for 2019Q2 under logit and mixed logit estimation.

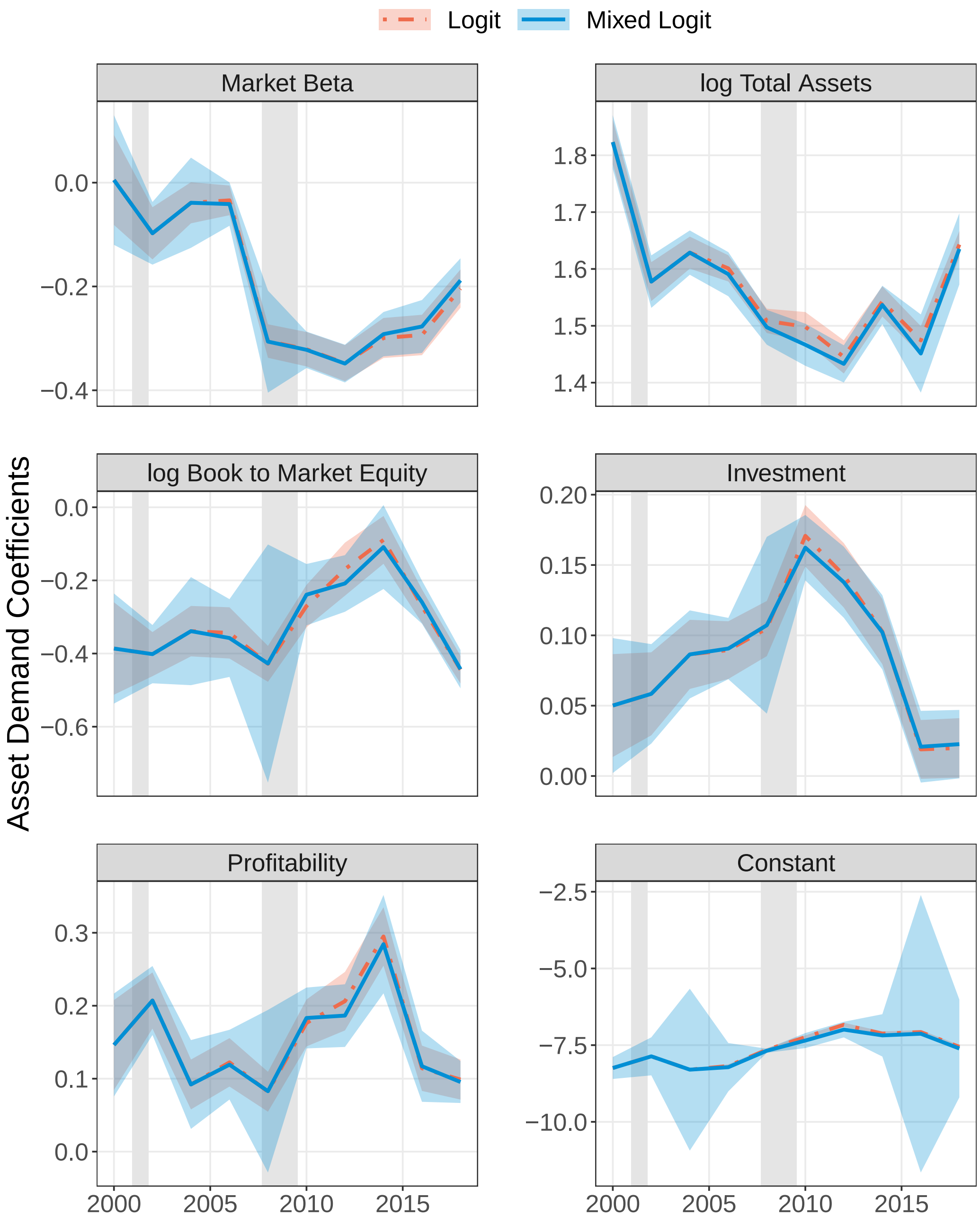
We also report an example of the estimated price elasticities. Table 1.1 shows the estimated price elasticties of the holdings of Apple according to the market value portfolio with respect to the price change of selected stocks. This example was chosen because in 2019Q2, the market value portfolio has similar weights for Alphabet and Exxon Mobile, approximately 1.02%. As discussed in section 1.3, under logit demand, the cross price elasticities are proportional to portfolio weights; hence under a logit demand system, portfolio holdings of another technology stock would react identically to a price change from both stocks. This is precisely what the estimates for logit show; despite Apple and Alphabet belonging to the same industry while Apple and Exxon Mobile belong different industries. On the other hand, the mixed logit estimates are flexible enough to show a larger degree of complementarity between Apple and Alphabet than between Apple and Exxon Mobile.

1.5. Ban of Green Investing for Pension Funds

In this section we use the estimated demand for green stocks to study the effects of a ban of green investing for pension funds on aggregate holdings and equity prices. This counterfactual policy exercise is motivated by policy initiatives discussed in the US Senate at the beginning of 2023. On March 1st 2023, the US Senate passed a bill to prevent pension fund managers from basing investment decisions on factors like climate change (Morgan (2023)). The bill was eventually vetoed by President Biden 19 days later (Thomas (2023)) but many similar initiates have been approved in various US States legislatures.

To implement a ban on green investing for pension funds, we first identify the institutional investors that are pension funds and counterfactually make their demand for stocks perfectly inelastic to

Figure 1.4: Estimated Coefficients for Return-related Characteristics



Notes: This plot shows the estimated coefficients corresponding to the return-related characteristics over 2-year estimation windows ranging from 2000-Q1 to 2019-Q4. Shaded regions represent 95% confidence intervals. Recession periods of the US economy are shown as vertical shaded gray regions.

E-scores. To identify pension funds we use the classification of institutional investors from KY2019. This classification groups institutional investors into 6 categories: banks, insurance companies, mutual funds, pension funds, investment advisors (including hedge funds) and other institutions like foundations, nonfinancial corporations, and endowments. Once the pension funds have been identified, we define the following counterfactual demand curves for a stock j:

$$\tilde{w}_{it,j} = \begin{cases} \frac{\exp\left(\hat{\delta}_{jt} - \hat{\gamma}_0 c_{jt}\right)}{1+\sum_{j=1}^{J_t} \exp\left(\hat{\delta}_{jt} - \hat{\gamma}_0 c_{jt}\right)} & \text{if i is a pension fund} \\ \frac{\exp\left(\hat{\delta}_{jt} + h_{ijt}(\hat{\theta}_2, d_i, v_i)\right)}{1+\sum_{j=1}^{J_t} \exp\left(\hat{\delta}_{jt} + h_{ijt}(\hat{\theta}_2, d_i, v_i)\right)} & \text{otherwise.} \end{cases}$$

In these counterfactual demand curves pension funds are inelastic to E-scores, since in their counterfactual demand the non linear parameters are set to zero and the component corresponding to E-scores in $\hat{\delta}_{jt}$ is offset to zero. In this case pension funds adjust their demand curves as if they no longer derive value from the environmental performance of the stocks. As a consequence of the demand change, prices observed in the data would not clear the market and we need to find counterfactual prices that clear the market. We rely on a market clearing condition (1.17) to find the new counterfactual prices. Recall that the market clearing conditions states that

$$ME_{jt} = \sum_{i=1}^{I} A_{it} w_{it,j},$$

and this condition can be expressed as a fixed point in the log vector of prices. Using the counterfactual demand curves, $\tilde{w}_{it}$, we solve for a vector of log prices p^c such that

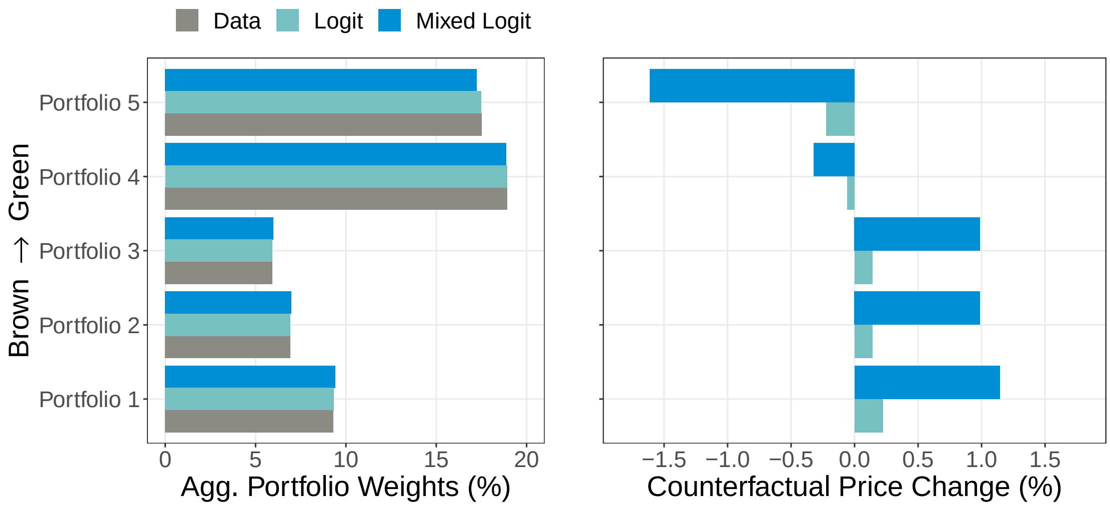
$$p^c = f(p^c) := \log\left(\sum_{i=1}^{I} A_{it} \tilde{w}_{it}(p^c)\right) - log(S_t). \tag{1.37}$$

KY2019 show that a sufficient (but no necessary) condition for this fixed point to exist is that the coefficient accompanying prices has an absolute value less than one. For the period 2019-Q1 where we do the counterfactual exercise the coefficient that corresponding to prices is the coefficient on log book-to-market equity which is -0.443 so it satisfies the sufficient condition for a fixed point to exist.

In computing counterfactual prices that satisfy the market clearing condition the following aspect are assumed to be fixed. First, the number of shares outstanding of each stock is assumed to be fixed, so we have an inelastic supply of the stocks and price changes are determined by demand shifts. Second, the assets under management for each investor A_i is also assumed to remain constant during the counterfactual. That is each investor still decides how to allocate A_{it} dollars into the available stock given that prices change to counterfactual prices and in the case of pension funds they are now inelastic to the E-scores. Third, is it assumed that the coefficients associated to return-related characteristics are fixed as well as the coefficient on E-scores for institutional investors other than pension funds. Fourth, the estimated unobserved stock characteristics, $\hat{\xi}_{jt}$ is also assumed be remain constant during the counterfactual. It can be argued that for the last two assumptions a Lucas (1976) critique applies to the extent that coefficients on return characteristics and unobserved stock characteristics change with the policy. This critique apply to most counterfactual exercises in the asset demand literature and exploring ways to circumvent the critique is left for future research.

Figure (1.5) shows the results of the counterfactual exercise using data and estimates for 2019-Q1. The portfolios in the figure were constructed by sorting stocks by their E-score, and grouping them according to quintiles. Portfolio 1 contains the 20% of stocks with lowest E-scores, the most brown stocks, while Portfolio 5 contains the 20% of stocks with highest E-scores, the most green stocks. The left panel shows aggregate portfolio weights for each portfolio, that is the sum of the market shares for the stocks contained in each portfolio. The aggregate portfolio weights shown are those observed in the data, under the counterfactual policy using estimates from a random coefficients (RC) demand specification and using estimates of a logit demand specification. Results show that the aggregate holdings in the data are very similar to the counterfactual holdings of logit demand, exhibiting little change. Compared to the counterfactual holdings under the random coefficients demand we see larger differences. Portfolio 5, which is composed of the stocks with the highest E-scores, shows a reduction in aggregate holdings under the policy, whereas Portfolios 1, 2, and 3 increase their aggregate holdings. This means that the relative importance of the stocks in Portfolio

Figure 1.5: Counterfactual Holdings and Price Changes of E-score sorted Portfolios



Notes: This figure shows the effect of a ban of green investing for pension funds on aggregate holdings and equity prices in a counterfactual exercise using data and estimates for 2019-Q1. The portfolios in the figure were constructed by sorting stocks by their E-score, and grouping them according to quintiles. Portfolio 1 contains the 20% of stocks with lowest E-scores, while Portfolio 5 contains the 20% of stocks with highest E-scores. The left panel shows the sum of the market shares for the stocks contained in each portfolio. It shows the aggregate portfolio weights observed in the data, under the counterfactual policy using estimates from a random coefficients (RC) demand specification and using estimates of a logit demand specification. The right panel shows the value-weighted average price change for each portfolio comparing the prices observed in the data with the counterfactual prices under RC demand and under logit demand.

5 diminished while the relative importance of the stocks of portfolios 1, 2 and 3 increased. This suggests that with the policy the aggregate investment share on green stocks was reduced in favor of brown stocks.

The right panel of Figure (1.5) shows the value-weighted average price change for the stocks in each portfolio comparing the prices observed in the data with the counterfactual prices under RC demand and under logit demand. Results show that the changes under a logit demand are much smaller than in the random coefficients demand case, this is due to the restrictive elasticities of logit demand where as mentioned before, own-price elasticities are proportional to market shares. The results for RC demand show that Portolio 5 experienced the most negative change, with an average counterfactual price change of -1.6%, while Portfolio 1 exhibited the biggest positive change with an average counterfactual change of 1.1%.[30] These result have to keep a consistency with the changes

[30]The corresponding changes for logit demand are -0.2% for Portfolio 5 and 0.2% for Portfolio 1.

in aggregate shares, the price of green stocks will decrease under the new counterfactual prices that clear the market, because there is less demand for green stocks, but the decrease will happen up to a point where the reduction in price no longer encourages more demand of the green stocks and the market clears. From the right panel of Figure (1.5), as with aggregate shares, the prices of green stocks decrease the most while the price of brown stocks increased with the policy.

The magnitude of the price changes in Figure (1.5) are commensurate to price changes observed in the data. For example, in the quarter following the data use for the counterfactual, the value-weighted price change of the stocks in Portfolio 1 between 2019-Q1 and 2019-Q2 was 1.1%; a table version of Figure (1.5) showing the price change over this period for all portfolios is included in Appendix A.3. Similarly, as studied in Rudebusch et al. (2023), policy announcements that substantially affect green and brown stocks can trigger price changes on stock indices of brown and green stocks ranging from -2% to 11% in a manner of days, as in the case of the Inflation Reduction Act approved in 2022.

1.6. Conclusions

This chapter combines a traditional portfolio construction problem with demand estimation techniques to estimate the demand for green stocks of US institutional investors. In the framework presented, both belief and taste heterogeneity play a role. In addition to investor heterogeneity through differential beliefs about future returns, our framework allows for investors to care about the characteristics of the portfolio they are forming beyond those characteristics related directly to an expected return-versus-risk trade off. We use this framework to measure the preference for green stocks while considering return-related stock characteristics.

For estimation this chapter uses a mixed logit demand specification in contrast with the logit demand specification more commonly used in the recent asset demand literature. In a logit demand specification, price elasticities are proportional to portfolio shares which restricts the substitution patterns between assets. In a mixed logit demand specification, investor heterogeneity is captured by investor-specific coefficients that are modeled as functions of investor demographics. This richer investor-level heterogeneity delivers more flexible substitution patterns between assets, and it is the modern workhorse model of demand estimation in the IO literature. By doing estimation at the market-level we can not only implement the mixed logit demand specification, but it facilitates dealing with the endogeniety of prices. Specifically, this allows us to consider instrumental variables for prices that have been studied in the IO literature such as BLP-type instruments or Gandhi-Houde price instruments, which have not been used in the asset demand literature.

The empirical exercise uses quarterly data on the stock holdings of institutional investors in the US. We pair this holdings data with return-related characteristics inspired by the Fama-French five factor model and data on the environmental performance of the listed companies in the form of E-scores. We find that the revealed taste for green stocks fluctuates over time. For both logit and mixed logit demand estimation, we find a positive taste for green stocks throughout the estimation sample. However, for mixed logit estimation the semi-elasticity for E-scores increases in the second part of the sample. Moreover, in the period after the Great Recession (2007-Q4 to 2009-Q2) there is an increase in the range of values for the coefficient on E-scores, showing an increase in the heterogeneity in the sensitivity to green characteristics across investors.

In a counterfactual exercise, we use the estimated demand system for stocks to study the effects of a ban on green investing for pension funds on equity prices and aggregate holdings. Using the data and estimates for 2019-Q1, we find that brown stocks will benefit the most in terms of counterfactual pricing. A portfolio with the bottom quintile of green stocks is estimated to have an associated average price change of 1.1% under the counterfactual, while the top quintile portfolio has an average price change of -1.6%.

Future work. Three avenues of work are left for future research. The first one deals with nontraditional stock characteristics. There is a great amount of textual information about listed companies that can be informative to investors. It is possible to extract "topics of risk" by applying a topic model to the text of regulatory risk fillings from listed companies (see, e.g. Lopez-Lira (2023)) and using the corresponding topic loadings as a stock characteristic in the demand curves of investors. This would help enrich the demand model with nontraditional but sensible characteristics related to risk. Related work includes Lopez-Lira and Roussanov (2023) that explores whether traditional common factors are enough to explain the cross section of returns.

A second avenue of future research lies at the intersection of asset pricing and corporate finance. As discussed in Brunnermeier et al. (2021) asset demand curves are flexible enough to include firm characteristics such as leverage, innovation, investment, and payout policies as relevant features signaling growth expectations and risks associated with future cash flows. Adding a model of firm corporate policies would complement asset demand systems with models of corporate decision making.

Third, dynamic considerations are at the very frontier of the asset demand estimation literature. Time-conditional statements in asset pricing are paramount and explicitly modeling the time dimension in the demand for stocks would be an important contribution to the literature. It is easy to argue that in practice portfolio optimization in one period is directly related to the stocks held in the previous period, calling for *inventory* considerations when modeling asset demand curves over time. Modeling asset demand dynamically also requires understanding how asset characteristics evolve over time, how investor funds flow in and out of the stock market, and how investor beliefs update over time. All these issues make dynamic asset demand challenging yet exciting for future research.

CHAPTER 2

Exchange Rate Supervised Topic Modelling

2.1. Introduction

Standard macroeconomic theory supports the view that exchange rates are determined mainly by fundamental macroeconomic variables such as, interest rate differentials, productivity differences, terms of trade, and others. However, empirical evidence has repeatedly shown that such models lack explanatory power when dealing with the short run dynamics of the exchange rate.

Under short periods of intense activity on FX markets, fundamentals can be silent for a practical reason: data on them is published at a much longer frequency, meaning that when confronted with data at short frequencies, fundamentals do not have enough variation to act as covariates in explaining the exchange rate. Despite the lack of new information on fundamentals, it could be that agents are acting on the expectations they have over fundamentals which are essentially latent variables changing at a much faster frequency than published data.

News articles could be an important source of information capable of causing agents to revise their expectations. If the expectation formation process is a signal extraction problem (for example Beaudry and Portier (2014)), some news articles contain new information related to fundamentals, so when agents manage to filter a positive signal correctly and act accordingly, fundamentals should eventually reflect their position. On the other hand, when agents respond positively to a signal that turns out to be noise, fundamentals will prove them wrong, and agents should revise their expectations. True news is therefore information that should have predictive power for future developments in the economy.

This idea suggests *filtering* news articles to extract the signal that is relevant for predicting developments in the economy. This chapter argues that this can be done by extracting the topics a news article writes about and determining which topics correlate with a macroeconomic variable of interest tracking an aspect of the economy. For example, when considering an exchange rate, to the extent that certain topics in news articles provide a relevant description of the economy, the more intensely such topics are represented in the newspaper at a given point in time the more likely it is that such topics represent something important for the economy's development and may cause revised expectations for some agents. This suggests that the information in news articles might be relevant in explaining the short term dynamics in the exchange rate.

This chapter gathers evidence on whether news articles have explanatory power for exchange rates in short term frequencies. The empirical evidence will be centered around the Monex Market, the main Costa Rican platform for FX trade. Accordingly, news articles will be gathered from the main Costa Rican newspapers. The exchange rate of interest is that between the Costa Rican Colón (CRC), the local currency, and the United States dollar (USD). As the United States is Costa Rica's main trade partner, the CRC/USD exchange rate receives almost exclusive attention in FX matters from the general Costa Rican public and the monetary authorities in Costa Rica.

Methodologically, the strategy in this chapter uses a hybrid of unsupervised and supervised learning models to go from text data in the news articles to an FX news index that can be used to enhance traditional models from the FX literature. To do so we rely on Supervised Latent Dirichlet Allocation (sLDA) (Blei and Mcauliffe, 2008) which combines information about a supervising variable with topic extraction over a corpus of text in a single-stage estimation.

From a methodological perspective, it is possible to divide the problem into two stages: a first stage that extracts a low-dimensional feature from the unstructured text data and a second stage that assesses how this low-dimensional feature correlates with the response variable. For example, in the context of exchange rates, we could divide the problem into a first stage that uses standard Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan, 2003) to extract topics from the news articles and a second stage that uses the estimated topic frequencies as regressors on a metric y of the exchange rate (for example the daily log return). However, via a Monte Carlo simulation we provide evidence that opting for a single-stage estimation, like sLDA, has efficiency gains over a two-stage approach.

Another strategy to process text from newspapers into indices is sentiment extraction. Many standard sentiment metrics from news articles are computed based on positive, negative, or particular word counts according to a pre-specified dictionary. Results can be strongly driven by the subjective choice of this dictionary that dictates which words should count. We argue that a topic extraction approach over the news articles, like in a sLDA model, is a less restrictive choice for producing news measures, since topics are extracted in an agnostic fashion. News measures based on positive or negative counts implicitly assume a two topic structure (positive and negative) and are more prone to a *market's animal spirits* interpretation where changes in expectations can be self-fulling and not necessarily rooted in economic fundamentals.

Using the estimates from sLDA it is possible to summarize the information from the supervised extracted topics into an FX news index. This FX news index in turn can be studied as a covariate controlling for public information from news articles in traditional models looking at FX dynamics. Construction of the index relies on the parameter estimates from a single-stage approach, which improves parameter precision. With a similar objective, Mora (2020) uses a two-stage approach combining LDA and a supervised model with dynamic variable relevance to study the role of news articles in the CRC/USD exchange rate from the Monex market.[31] Results show that the inclusion of an FX news index improves a traditional order flow regression. Here, the order flow remains an important driver in short-term dynamics for the exchange rate after. However, by controlling for a proxy of public information, in the form of the FX news index, the results from this exercise provide evidence to interpret the order flow as driven by private information.

Related literature This chapter is related to at least four strands of literature. The first one is about the use of text data in economics and finance. Gentzkow et al. (2019) survey different approaches to incorporate textual analysis into answering economic questions as well present applications of such approaches in recent papers.

One example of a paper which relies on topic models is Larsen and Thorsrud (2019) which pursues a similar, though methodologically different, goal as in this chapter. They use news articles from Norwegian news outlets and assess their information content to predict a set of macroeconomic

[31] This is a previous version of this chapter. It can be consulted at https://www.dropbox.com/s/4s4g7owbwdu266p/Mora_TextIntoFX.pdf?dl=0.

variables in Norway. Thorsrud (2020) builds on this dataset of Norwegian news articles to construct a daily business cycle index combining the text data with quarterly GDP growth. Using news articles from the Wall Street Journal, Bybee et al. (2020) combine a LDA topic model with a vector autoregressive model to track the state of the U.S. economy via text analysis of business news.

Examples of text models in finance include Ke et al. (2019) who use a supervised sentiment extraction model on labeled Dow Jones Newswires to predict asset returns. Lopez-Lira (2020) uses a LDA model on annual reports to elicit the risk factors that firms themselves identify in their annual reports. It is worth remarking that all of the papers mentioned above follow a two-stage approach to estimation where the first stage focuses on extracting numerical features of text data that will be subsequently used in a second stage that relates such features to economics variables of interest.

The second related literature as mentioned in the introduction deals with studying short term dynamics of the exchange rate. This chapter is related to the literature of market microstructure in exchange rates and the role of the order flow, the difference between buyer-initiated and seller-initiated trades, as a high frequency covariate for the exchange rate. Many of the key insights of this literature can be found in Lyons (1995) and Evans and Lyons (2002), surveys of the literature include Vitale (2007) and King et al. (2013). Many papers have suggested explanations for the empirical relevance of the order flow in exchange rate dynamics (e.g. Bacchetta and Van Wincoop (2006), Vitale (2011) Evans and Lyons (2012)). However, despite the fact that these papers propose different hypotheses for why the order flow has explanatory power for short-run exchange rate dynamics, a definite answer backed by empirical evidence remains an open question.

A third related literature has to do with topic modeling. The LDA model from Blei, Ng and Jordan (2003) has become an industry standard in recent years that has sprouted a series of variations extending different aspects of the original model. Blei, Griffiths, Jordan and Tenenbaum (2003) propose hierarchical LDA (hLDA) to include topic hierarchies by using the Chinese restaurant process as a prior for such topic hierarchies.[32] Andrzejewski et al. (2009) also extends LDA at

[32] In this extension, topics follow a L-dimensional tree structure that represent their hierarchy. Then a document is drawn by first choosing an L-level path through the restaurants and then drawing the words from the L topics which are associated with the restaurants along that path.

the level of topic priors to included domain knowledge, that is external information about the composition of words that should have high or low probability in various topics. Jagarlamudi et al. (2012) propose seedeed LDA that also incorporates domain knowledge on the topic priors to guide topic models and learn about topics of specific interest to a user by providing sets of seed words that a user believes are representative of the underlying topics in a corpus. Ramage et al. (2009) also extend LDA to incorporate external information and propose Labelled LDA to study the problem of credit attribution when documents have user tags in addition to a topic structure. A dynamic extension of LDA is presented in Blei and Lafferty (2006), where topic and topic membership for each document evolves over time as a function of previous periods' values. All these papers reflect the convenient modularity of LDA, demonstrating how its basic topic structure can be part of richer models that work with text corpora with different features. In this chapter we introduce slight modifications to sLDA so it fits our goal of introducing text data into a regression of the exchange rate; details are shown in Section 2.2.

Finally a fourth related literature has to do with approximate inference in models with latent variables, in particular with the variational inference methods. Topic models, and in particular sLDA, posit latent features (variables) in the text structure. Given the high-dimensional nature of text data, it is common to resort to approximate inference. sLDA, as presented below, performs estimation using variational inference with a mean-field approximation. A good introduction to the techniques of variational inference can be found in Blei et al. (2017); more through treatments can be found in Bishop (2006) and Wainwright and Jordan (2008).

The rest of this chapter is organized as follows: section 2.2 shows how to extract supervised topics from a database of news articles using a sLDA model. Section 2.3 shows the simulation study that documents the efficiency gains of using sLDA as opposed a two-stage strategy of extracting topics using the traditional LDA and then using its output in a regression. Section 1.4.1 presents the data for our empirical application, the CR News database and the Monex Market, the source of the exchange rate data. Section 2.5 presents the strategy for constructing the FX news index. Finally, Section 1.6 concludes.

2.2. Supervised Topic Extraction

Given a corpus of news articles, our premise is that each word in each document has a latent assignment into one of K possible topics and moreover this latent assignment determines a response variable y. In this context each document is seen as a bag of words, that is a sequence of words $\{w_n\}_{n=1}^{N}$ where each word w_n is indexed in a dictionary of size V and N is the total number of words in a document. Once indexed, a document becomes a sequence of positive integers. To make this sequence human-readable, we would need to replace index $w_n \in \{1, ..., V\}$ with the word corresponding to the w_n-th index in the dictionary.

Topics are considered as clusters of semantic content and modeled as unknown distributions over a dictionary V. To estimate these latent assignments and their relationship with the response variable y, we will use a modified version of Supervised Latent Dirichlet allocation (Blei and Mcauliffe (2008)) (sLDA) that builds on the industry standard methodology of Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan (2003)) to include in a single-stage estimation information of the response variable to improve topic extraction. In a nutshell, sLDA is based on three premises:

i. A topic is a distribution over words.

ii. A document is a mixture of topics. That is, there is mixed membership of documents into topics.

iii. The empirical topic membership of a document along with covariates W_t, determine a response y_t, a label for the document.

In this context data is composed of document-response triplets $\{(w_{t,1:N_t}, y_t, W_t)\}_{t=1}^{T}$. In our exchange rate application, t would index daily dates: y_t will represent the daily log return using the exchange rate between day t and day $t-1$. The variable $w_{t,1:N_t}$ stands for the news articles available at the beginning of day t, hence produced in day $t-1$, pooled into a single document. In this aggregation we are exploiting the mixed membership feature of LDA models.[33] Finally, W_t are

[33]Alternatively we can keep each news article as a document and then pool the empirical topic frequencies of each document with timestamp t into a pooled topic frequency for day t.

other covariates known to be relevant to determine short-term dynamics of the exchange rate, like the order flow.

Two modifications are done with respect to sLDA as presented in Blei and Mcauliffe (2008). The first modification is to include covariates other than text in the supervising equation; which is important for the exchange rate application. The second, is to include a Dirichlet prior for the vector of term frequencies of each topic. This facilitates the estimation process and it is line with the *smooth*-LDA (see Blei, Ng and Jordan (2003) section 5.4) implemented in most programming packages. The next subsection presents the generative for the modified sLDA.

2.2.1. Generative Model

Given a corpus of T documents indexed by $t = 1, ..., T$ compose of words from a dictionary of size V. The supervised model used assumes that there are K underlying topics that generate the text in the following fashion:

i. Draw K vectors $\beta_k \sim Dir(\eta)$ so $\beta_k \in \Delta^{V-1}$.

ii. Draw T vectors $\theta_t \sim Dir(\alpha)$ so $\theta_t \in \Delta^{K-1}$

iii. For each word $w_{t,n}$ with $n = 1, ..., N_t$:

- Draw topic assignment $z_{t,n}|\theta_t \sim Mult(\theta_t)$. So $z_{t,n} \in \{1, ..., K\}$.

- Draw word $w_{t,n}|z_{t,n}, \beta_{z_{t,n}} \sim Mult(\beta_{z_{t,n}})$. So $w_{t,n} \in \{1, ..., V\}$.

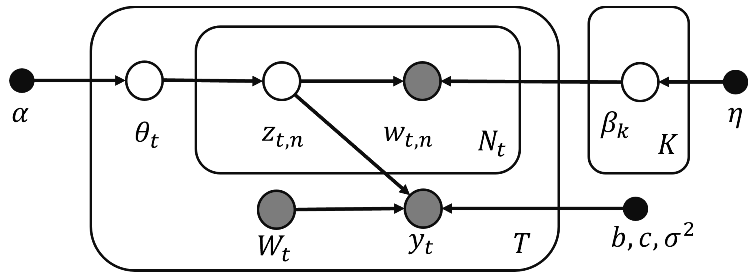
iv. Compute the empirical topic frequency $\bar{Z}_t = \frac{1}{N_d} \sum_{n=1}^{N_t} z_{t,n}$; with $z_{t,n}$ written as an one-hot vector of size K.

v. Given exogenous regressors W_d and parameters generate a response y_d as

$$y_t = b'\bar{Z}_t + c'W_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2). \tag{2.1}$$

The variables dependencies can be summarized in the graphical model depicted in Figure 2.1. In

Figure 2.1: sLDA Graphical Model



Notes: Graphical model for the generative model of Supervised Latent Dirichlet Allocation.

specifying the response equation (2.1) there are several modeling choices. Here we use a basic Gaussian response; however, Blei and Mcauliffe (2008) show how this response equation can be generalized to other members of the exponential family. A response with time varying parameters and with possible stochastic volatility is left for future research.

In the generative model we have assumed that the topic structure determines the response. It is possible to write a version of equation (2.1) that uses θ instead of $\bar{Z}$. This would imply that documents and responses are exchangeable in the sense that once θ is drawn we can first generate the document or the response. It makes more sense in the context of an application that seeks to extract information from news articles capable of changing market behavior and hence market outcomes to assume that documents are determined first followed by responses. Notice that so far (and in the details below) that the strategy would be the same if y is a response containing information for other economic variables, suggesting that we can think of supervised topic extraction with GDP, inflation, trade, unemployment, and other macroeconomic variables.

2.2.2. Variational Inference

The next natural step is to wonder how when given data we can use the previous framework to learn/estimate the parameters involved. The main challenge is the presence of latent variables, namely θ that controls document topic membership, $\{z_n\}$ the words' topic assignments and $\beta_{1:K}$

the term frequencies for each topic. The posterior distribution of the latent variables is given by:

$$p(z_{1:N},\theta,\beta_{1:K}|w_{1:N},y,W;b,c,\sigma^2;\alpha,\eta) = \\ \frac{p(y|z_{1:N},W,b,c,\sigma^2)\left[\prod_{n=1}^{N}p(w_n|z_n,\beta_{1:K})p(z_n|\theta)\right]p(\theta|\alpha)p(\beta_{1:K}|\eta)}{\sum_{z_{1:n}}\int_\theta\int_{\beta_{1:K}}\left\{p(y|z_{1:N},W,b,c,\sigma^2)\left[\prod_{n=1}^{N}p(w_n|z_n,\beta_{1:K})p(z_n|\theta)\right]p(\theta|\alpha)p(\beta_{1:K}|\eta)\right\}d\theta d\beta}.$$

The normalizing term in the posterior distribution of the latent variables is the likelihood of the data $p(w_{1:N},y|\pi,W)$ where $\pi=\{b,c,\sigma^2,\alpha,\eta\}$. Unfortunately this likelihood is not efficiently computable, as documented in Blei, Ng and Jordan (2003), making the posterior of the latent variables also not efficiently computable.[34] If this posterior were computationally tractable we could use a standard EM algortihm (Dempster et al. (1977)) to perform maximum likelihood and estimate the parameters.

Given this limitation, we resort to approximate maximum-likelihood estimation using a variational expectation maximization (VEM) strategy where we treat parameters π as constants to be estimated rather than random variables. The VEM strategy follows the spirit of the traditional EM algorithm but we use a variational approximation to the complete likelihood to find a maximizing set of parameters as our estimates. Notice the hybrid approach to parameters in this strategy: latent variables that can be thought of as observation-specific parameters that have a Bayesian structure with priors on them, whereas common latent variables or global parameters are treated as unknown constants. For a discussion on when to use variational methods versus exact inference strategies like MCMC see Blei et al. (2017).

2.2.3. Evidence Lower Bound and Mean Field Approximation

To obtain an approximation of the complete likelihood in the presence of latent variables, we resort to the Evidence Lower Bound (ELBO), a key object in variational inference. Let $q(\theta,z_{1:N},\beta_{1:K})$ denote the *variational distribution* for our latent variables (topic memberships, word assignments and topic term frequencies). By appropriately selecting the family of the variational distribution,

[34] Notice that in the the normalizing term, just the sum is over K^N terms.

we are able to approximate the likelihood. The log likelihood of the observed data is given by:

$$\begin{aligned}
\log p(w_{1:n}, y|\pi, W) &= \log \left[\sum_{z_{1:n}} \int_\theta \int_{\beta_{1:K}} p(w_{1:n}, y, z_{1:n}, \theta, \beta_{1:K}|\pi, W) d\theta d\beta_{1:K} \right] \\
&= \log \left[\sum_{z_{1:n}} \int_\theta \int_{\beta_{1:K}} p(w_{1:n}, y, z_{1:n}, \theta, \beta_{1:K}|\pi, W) \frac{q(z_{1:n}, \theta, \beta_{1:K})}{q(z_{1:n}, \theta, \beta_{1:K})} d\theta d\beta_{1:K} \right] \\
&\geq E_q \left[\log p(w_{1:n}, y, z_{1:n}, \theta, \beta_{1:K}|\pi, W) \right] - E_q \left[\log q(\theta, z_{1:N}, \beta_{1:K}) \right].
\end{aligned}$$

Expectations in the last line are with respect to $q(\theta, z_{1:N}, \beta_{1:K})$. This last line is the Evidence Lower Bound (ELBO). Denoting the entropy of the variational distribution as $H(q) = -E_q[\log q(\theta, z_{1:N}, \beta_{1:K})]$, the ELBO expands into:

$$\begin{aligned}
\text{ELBO} &= E_q[\log p(w_{1:N}, y, \theta, z_{1:N}, \beta_{1:K}|\pi, W)] + H(q) \\
&= E_q[\log p(\theta, z_{1:N}, \beta_{1:K}|w_{1:N}, y, \pi)] + E_q[\log p(w_{1:N}, y|\pi, W)] + H(q) \\
&= E_q[\log p(\theta|\alpha)] + \sum_{k=1}^{K} E_q[\log p(\beta_k|\eta)] + \sum_{n=1}^{N} E_q[\log p(z_n|\theta)] \\
&+ \sum_{n=1}^{N} E_q[\log p(w_n|z_n, \beta_{1:K})] + E_q[\log p(y|z_{1:N}, W, b, c, \sigma^2)] + H(q) \\
&+ \log p(w_{1:N}, y|\pi).
\end{aligned}$$

Notice that the ELBO is just a function of data $(w_{1:N}, y)$ and the variational distribution q. This bound is tight when q is the posterior of the latent variables, since maximizing the ELBO is equivalent to minimizing the Kullback-Leibler divergence between the posterior and q, (see Blei et al. (2017)). However, this posterior is computationally intractable so we approximate it.

Next, we take a stance on the family of the variational distribution and use mean-field variational inference to specify a fully factorized distribution:

$$q(\theta, z_{1:N}, \beta_{1:K}|\gamma, \phi_{1:N}, \lambda_{1:K}) = q(\theta|\gamma) \prod_{n=1}^{N} q(z_n|\phi_n) \prod_{k=1}^{K} q(\beta_k|\lambda_k). \tag{2.2}$$

Within this family, our goal is to find $(\gamma, \phi_{1:N}, \lambda_{1:K})$ that maximize the ELBO. Since the last term of the ELBO does not depend on the variational distribution our objective function is:

$$\begin{aligned} \mathcal{L}(\gamma, \phi_{1:N}, \lambda_{1:K}) = E_q[\log p(\theta|\alpha)] + \sum_{k=1}^{K} E_q[\log p(\beta_k|\eta)] + \sum_{n=1}^{N} E_q[\log p(z_n|\theta)] \\ + \sum_{n=1}^{N} E_q[\log p(w_n|z_n, \beta_{1:K})] + E_q[\log p(y|z_{1:N}, W, b, c, \sigma^2)] + H(q). \end{aligned} \tag{2.3}$$

Detailed derivations of how to maximize the ELBO can be found in Appendix B.1 The optimization is done by block coordinate-ascent variational inference (CAVI) (see Bishop (2006) for an exposition), which implies iteratively maximizing $\mathcal{L}$ with respect to each variational parameter. In this maximization we have the additional constraint that each multinomial parameter vector ϕ_n should sum up to one. Optimization boils down to iteratively updating $\{\gamma, \phi_{1:N}, \lambda_{1:K}\}$ as:

$$\gamma^{new} \leftarrow \alpha + \sum_{n=1}^{N} \phi_n \tag{2.4}$$

$$\phi_n^{new} \propto \exp\left[E_q[\log\theta] + E_q[\log\beta_{w_n,\cdot}] + \frac{\tilde{y}b}{N\sigma^2} - \frac{1}{2\sigma^2 N^2}[2(b'\phi_{-n})b + b\cdot b]\right], \tag{2.5}$$

with $\tilde{y} := y - c'W$; $\phi_{-n} := \sum_{m\neq n}^{N} \phi_n$ and proportionality means to compute each entry and then normalize to one, and finally $\cdot$ stands for element-wise multiplication; details can be found in Appendix B.1.2. The variational parameters for each topic, λ_k, are updated as

$$\begin{aligned} \lambda_{k,v} &\leftarrow \eta_v + \sum_{t=1}^{T}\sum_{n=1}^{N_t} \mathbb{I}\{w_{t,n} = v\}\phi^*_{t,n,k} \\ \text{or} \quad \lambda_k &\leftarrow \eta + \sum_{t=1}^{T}\sum_{n=1}^{N_t} \phi^*_{t,n,k}\omega_{t,n}, \end{aligned} \tag{2.6}$$

with $\omega_{t,n}$ a one-hot representation of $w_{t,n}$ as a vector of size V. Using the update above for $\{\gamma, \phi_{1:N}, \lambda_{1:K}\}$ we can maximize the ELBO for each document in the corpus to obtain a corpus-level ELBO:

$$
\begin{aligned}
L(b, c, \sigma^2) &= \sum_{t=1}^{T} E_{q_t^*}[\log p(\theta_t, z_{t,1:N_t}, \beta_{1:K}, w_{t,1:N_t}, y_t | W_t, \pi)] + H(q_t) \\
&= \sum_{t=1}^{T} \mathcal{L}(\gamma_t^*, \phi_{t,1:N}^*, \lambda_{1:K}^* | b, c, \sigma^2),
\end{aligned} \tag{2.7}
$$

where t indexes the sample data, that is the document-response pair for day t, and q_t^* stands for the document-specific variational distribution with corresponding maximizing $\{\gamma_t^*, \phi_{t,1:N}^*\}$ and $\{\lambda_{1:K}^*\}$. Notice that the corpus-level ELBO is a just a function of parameters from the supervising equation, this is because (α, η) are treated as hyperparameters selected by the econometrician.

2.2.4. Parameter Estimation

We now look for parameters $\pi = \{b, c, \sigma^2; \alpha, \eta\}$ that maximize the corpus-level ELBO in (2.7); this is done via the VEM algorithm, that iteratively performs the E-Step and M-Step described below.

E-step In the E-step given a set of parameters from the previous iteration, we approximate the likelihood of the data by maximizing each document-specific ELBO, and then summing them into $L(b, c, \sigma^2)$ as described in the previous section. In this step the goal is to find a collection $\{\gamma_t^*, \phi_{t,1:N}^*, \lambda_{1:K}^*\}$ that maximizes each document-specific ELBO using CAVI. This amounts to iteratively perform the updates in (4)-(6).

M-step In the the M-step we maximize the corpus ELBO with respect to the sLDA parameters $\pi = \{c, b, \sigma^2; \alpha, \eta\}$. Detailed derivations for this step are shown in Appendix B.1.3 The updates in this step boil down to:

$$
b^{new} \leftarrow \left(\sum_{t=1}^{T} E_{q^*}[\bar{Z}_t \bar{Z}_t'] \right)^{-1} \left(\sum_{t=1}^{T} E_{q^*}[\bar{Z}_t] \tilde{y}_t \right) \tag{2.8}
$$

$$
c^{new} \leftarrow \left(\sum_{t=1}^{T} W_t W_t' \right)^{-1} \left(\sum_{t=1}^{T} W_t (y_t - E_{q^*}[\bar{Z}_t]) b^{new} \right) \tag{2.9}
$$

$$
\sigma^{2,new} \leftarrow \frac{1}{T} \sum_{t=1}^{T} \left(y_t^2 - y_t E_{q^*}[\bar{Z}_t'] b^{new} - y_t W_t' c^{new} \right), \tag{2.10}
$$

where q^* stands for the variational distribution with the optimal variational parameters from the previous E-step. As mentioned above, the topic membership hyper parameter, α, and the hyperparameter for topic term frequencies, η, are treated as a hyper parameter chosen by the econometrician. For example it is common to set $\alpha = (1/K) \cdot \mathbf{1}_K$ and $\eta = 0.1 \cdot \mathbf{1}_V$ guided by the suggestion in Griffiths and Steyvers (2004).

The VEM algorithm iteratively alternates these steps until convergence in the corpus-level ELBO in (2.7) is achieved. Algorithm 2 sketches the steps for estimating the modified sLDA model and obtaining parameter estimates $\hat{\pi} = \{\hat{b}, \hat{c}, \hat{\sigma}^2; \alpha, \eta\}$.

2.2.5. Prediction

In sLDA given estimates for topics, $\hat{\beta}_{1:K}$, parameters $\hat{\pi}$ and a new document $w_{1:n}$ with covariates W, we can predict the response variable y associated with the new document. Using a quadratic loss function, the prediction simplifies to computing the conditional expected value:

$$
\begin{aligned}
E[y|w_{1:N}, W, \hat{\pi}, \hat{\beta}_{1:K}] &= E[E[y|w_{1:N}, z_{1:n}, \theta, W, \hat{\pi}, \hat{\beta}_{1:K}]] \\
&= \hat{b}' E[\bar{Z}|w_{1:N}, \theta, \hat{\pi}, \hat{\beta}_{1:K}] + \hat{c}' W.
\end{aligned}
$$

To compute the last conditional expectation we resort again to variational inference and compute the posterior mean of $\bar{Z}$. The goal is to find a variational distribution $\tilde{q}(\theta, z_{1:N})$ over the latent variables that minimizes the KL divergence between this distribution and the posterior distribution $p(\theta, z_{1:n}|w_{1:n}, W, \hat{b}, \hat{c}, \hat{\sigma}^2, \hat{\beta}_{1:K})$; this is, again, equivalent to maximizing the corresponding ELBO. This is similar to the ELBO maximization in sLDA but the difference is that terms involving the response variable y will not appear since this variable is averaged out. The variational updates are the same as in an unsupervised LDA model for $(\theta, z_{1:N})$.

Using a fully factorized variational distribution, the updates are given by:

$$\tilde{\gamma}^{new} \leftarrow \alpha + \sum_{n=1}^{N} \tilde{\phi}_n$$
$$\tilde{\phi}_n^{new} \propto \exp\left[E_{\tilde{q}}[\log \theta] + \log \hat{\beta}_{w_n,\cdot}\right].$$

Once the variational parameters have been optimized, then we obtain a predicted response $\hat{y}$ as:

$$\hat{y} = \hat{b}' E[\bar{Z}|w_{1:N}, \hat{\pi}, \hat{\beta}_{1:K}] + \hat{c}'W \approx \hat{b}' E_{\tilde{q}}[\bar{Z}] + \hat{c}'W = \hat{b}' \left(\frac{1}{N} \sum_{n=1}^{N} \tilde{\phi}_n \right) + \hat{c}'W. \tag{2.11}$$

Algorithm 3 summarizes the steps for predicting a response given a document and sLDA estimates.

2.3. Monte Carlo Study

When modeling how text data can be used as a covariate for a supervising variable of interest, the problem can divided into two stages: a first stage that extracts a low-dimensional feature from the unstructured text data and a second stage that assesses how this low-dimensional feature correlates with the response variable. For example, in our context of exchange rates, we could divide the problem into a first stage that uses LDA to extract topics from the news articles and a second stage that uses the estimated topic frequencies as regressors on a metric y of the exchange rate (for example the daily log return). In a more complex two stage example, Larsen and Thorsrud (2019) use LDA in the first stage followed by a Latent Dynamic Threshold model (LTM) (Nakajima and West (2013)) in the second stage to assess how news article text can be filtered into different indices over macroeconomic variables.

In this section we argue, via a Monte Carlo simulation, that opting for single-stage estimation has efficiency gains over a two-stage approach; namely the Mean Square Errors (MSE) on the estimated parameters and the Mean Square Prediction Error (MSPE) over the test sets are smaller when using the sLDA versus the two stage approach of LDA followed by a regression.

Algorithm 2: VEM algorithm for sLDA estimation

Input: Document and Response variables $\{w_{t,1:N_t}, y_t, W_t\}$ for $t = 1, ..., T$; and hyperparameters (α, η).
Output: A set of estimated parameters $\hat{\pi} = \{\hat{b}, \hat{c}, \hat{\sigma}^2\}$.
Initialize: π^0, $\{\gamma_t^0, \phi_{t,1:N_t}^0\}$ for $t = 1, ..., T$ and $\{\lambda_k^0\}$ for $k = 1, ..., K$.

```
while Corpus ELBO has not converged do
    /* E-Step */
    for t = 1 to T do
        while Document-Response ELBO has not converged do
            /* Update Variational Parameters */
            γ^new ← α + Σ_{n=1}^N φ_n ;
            φ^new ∝ exp[ E_q[log θ] + E_q[log β_{w_n,·}] + ỹb/(Nσ²) − 1/(2σ²N²)[2(b'φ_{−n})b + b·b] ];
        end
    end
    for k = 1 to K do
        /* Update Topic Variational Parameters */
        λ_{k,v} = η_v + Σ_{t=1}^T Σ_{n=1}^{N_t} 𝕀{w_{t,n} = v} φ*_{t,n,k}
    end
    /* M-Step: Update Parameters */
    b^new ← (Σ_{t=1}^T E_{q*}[Z̄_t Z̄'_t])^{−1} (Σ_{t=1}^T E_{q*}[Z̄_t] ỹ_t) ;
    c^new ← (Σ_{t=1}^T W_t W'_t)^{−1} (Σ_{t=1}^T W_t (y_t − E_{q*}[Z̄_t]) b^new) ;
    σ^{2,new} ← (1/T) Σ_{t=1}^T (y_t² − y_t E_{q*}[Z̄'_t] b^new − y_t W'_t c^new)
    /* Compute Corpus ELBO */
    for t = 1 to T do
        L(π^new) += 𝓛(γ*_t, φ*_{t,1:N_t}, λ*_{1:K}; w_{t,1:N_t}, y_t, W_t; π^new)
    end
    if Corpus ELBO meets convergence criteria then
        Stop Execution
    end
end
```

Algorithm 3: VEM algorithm for sLDA prediction

Input: Document $w_{1:N}$, covariates W, hyperparameters α, η and sLDA estimates $\{\hat{b}, \hat{c}, \hat{\sigma}^2, \hat{\beta}_{1:K}\}$

Output: A predicted response $\hat{y}$.

Initialize: Variational parameters $(\tilde{\gamma}_0, \tilde{\phi}_{0,1:N})$.

while *Document ELBO has not converged* **do**

$\tilde{\gamma}^{new} \leftarrow \alpha + \sum_{n=1}^{N} \tilde{\phi}_n$;

$\tilde{\phi}_n^{new} \propto \exp\left[E_{\tilde{q}}[\log\theta] + \log\beta_{w_n,\cdot}\right]$;

if *Document ELBO meets convergence criteria* **then**

$\hat{y} = \hat{b}'\left(N^{-1}\sum_{n=1}^{N}\tilde{\phi}_n\right) + \hat{c}'W$;

Return: $\hat{y}$

end

end

2.3.1. Simulation Design

We start by describing how to generate a labeled corpus. We fix the dictionary size a $V = 200$, which means that words are just indices $w \in 1, 2, 3, ..., 200$. We also match these indices with the 200 most common words in English to facilitate comparison between the results. We consider $K = 5$ underlying topics and the topic membership hyperparameter α_0 is set to $\alpha_0 = (1/K) \cdot 1_K$ where 1_K is a vector of ones of size K. Topics are drawn according to $\beta_k^0 \sim Dir(\eta_0)$ with $\eta_0 = 0.1 \cdot 1_V$. Each is a vector of size $V = 200$ with entries summing to one. We put these topics into a (V, K) matrix $\beta_{1:K}^0$. The supervised weights b were set to $b_0 = [2, 1, 0, -1, -2]'$. Next we generate $T = 1000$ document-response pairs following the blueprint described in the generative model in section 2.2:

(i) Draw how many words are in the document: $N_t \sim Unif(100, 300)$.

(ii) Draw topic membership $\theta_t \sim Dir(\alpha_0)$.

(iii) For each word in the sample $n = 1, ..., N_t$:

- Draw topic assignment: $z_n \sim Mult(\theta_t)$. Hence $z_n \in \{1, ..., K\}$.
- Draw a word index $w_n \sim Mult(\beta_{z_n})$. Hence $w_n \in \{1, ..., V\}$.

(iv) Rewrite z_n in one-out-K form and compute the empirical topic frequency $\bar{Z}_t \in \mathbb{R}^K$ for this document:

$$\bar{Z}_t = \frac{1}{N_t} \sum_{n=1}^{N_t} z_n.$$

(v) Generate $\varepsilon_t \sim N(0, \sigma_0^2 = 1)$ and using $b_0 \in \mathbb{R}^K$ generate a response:

$$y_t = \bar{Z}_t' b_0 + \varepsilon_t.$$

The previous steps generate a dataset of document-response pairs $(w_{t,1:N_t}, y_t)$ with $t = 1, ..., 1000$. These documents-response pairs will be split into a training set formed by 900 pairs and a test set composed of the 100 remaining pairs. In this simulation we have assumed no complementary covariates W_t determining y_t along with $w_{t,1:N_t}$ as we are mainly interested in contrasting the estimates related to text in single versus two-stage estimation procedures.

2.3.2. Simulation Results

We generated a hundred datasets, each containing $T = 1000$ document-response pairs. Using each of these datasets, we perform single-stage and two-stage estimation on the training set and prediction over the test set. Single-stage estimation refers to sLDA using the VEM algorithm whereas two-stage estimation refers to running LDA with the VEM algorithm followed by regressing the dependent variable y on the empirical topic frequencies per document extracted from LDA. Details about the execution of the VEM algorithms are shown in Appendix B.2.

Table 2.1 shows the bias, variance and MSE for this repeated exercise for the supervised parameters b and σ^2. These metrics for single-stage estimation present smaller numbers and larger relative efficiency, defined as the MSE from two-stage estimates divided the MSE from single stage estimates, for all the parameters. In particular for σ^2 the MSE from two-stage is 5.8 times larger than in single-stage estimation. These results suggest that even though topics are generated before responses and that both LDA and sLDA model topics the same way, having supervised information improves the topic extraction process.

During prediction we also observe efficiency gains. Figure 2.2 shows distributions of Mean Square

Table 2.1: Bias, MSE and Efficiency for Supervised Parameters

Parameter	Estimation	Bias	Variance	MSE	RE
σ^2	Two-Stage	0.111	0.023	0.035	5.835
	Single-Stage	0.038	0.005	0.006	
b_1	Two-Stage	-0.376	0.074	0.214	3.557
	Single-Stage	-0.185	0.026	0.060	
b_2	Two-Stage	-0.250	0.097	0.158	2.276
	Single-Stage	-0.134	0.052	0.069	
b_3	Two-Stage	0.021	0.058	0.058	1.815
	Single-Stage	0.033	0.031	0.032	
b_4	Two-Stage	0.163	0.041	0.067	1.652
	Single-Stage	0.090	0.033	0.041	
b_5	Two-Stage	0.337	0.066	0.179	2.585
	Single-Stage	0.189	0.034	0.069	

Notes: Results for 100 Monte Carlo replications. Each generated dataset has $K = 5$ underlying topics for a dictionary composed of the top $V = 200$ most common nouns in English. Topics were drawn as $\beta_k \sim Dir(0.1 \cdot \mathbf{1}_V)$ and each simulated training set is composed of $T = 900$ documents, each with a number of words drawn as $N_t \sim Unif(200, 300)$. RE is the efficiency of Single-Stage estimation relative to Two-Stage estimation: RE = MSE(Two-Stage)/MSE(Single-Stage).

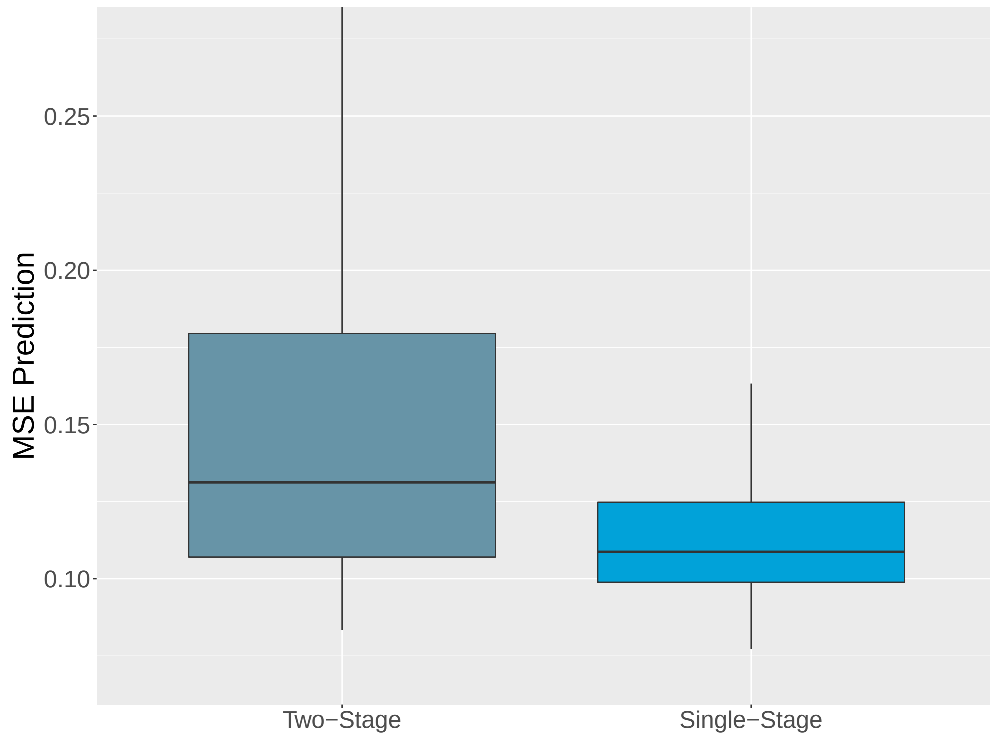
Prediction Errors (MSPE) for both strategies over a test sample. Specifically, each MSPE is computed over a test set of 100 documents, for which the response variable y was predicted using estimates from a training set of 900 document-response pairs. This figure shows that MSPE from single-stage forecasts generally take smaller values than the two-stage predictions; this is consistent with efficiency gains in favor of the single-stage strategy during estimation.

2.4. Data Sources

2.4.1. News Articles

For this chapter more than 56,000 news articles were collected from La Nación (LN) (Grupo-Nación-LN (2021)), the main Costa Rican newspaper, and from El Financiero (EF) (Grupo-Nación-EF (2021)), the main Costa Rican business newspaper. Article dates range from January 2015 to March 2021, with almost all articles written in Spanish. In this period of 2,271 days there is a median of 21

Figure 2.2: MSE Prediction



Notes: Boxplots for the Mean Square Prediction Error (MSPE) over 100 Monte Carlo replications. Each MSPE is computed over a test set of 100 documents, for which the response variable y was predicted using estimates from a training set of 900 document-response pairs.

news articles, 15,193 words and 13 newspaper sections per day. For every article, the section, the author (if available), publishing date up to milliseconds, subject tags, header, subheader, and the main text of the article were recorded.

It is important to mention that despite the fact that this database contains only news articles from a Costa Rican newspaper, news about the United States economy is closely followed. As Costa Rica is a small open economy, the US is its main trade partner, and the USD is the vehicle currency for international trade, the CRC/USD exchange rate is a macroeconomic variable closely followed by the media. Hence news about developments in the United States will typically be covered in the newspapers.

2.4.2. Text Preprocessing

Before the text in each news article can be used as an input to estimate a topic model it needs to be preprocessed. This is done to clean the text and reduce its dimension to facilitate the extraction of semantic content. This is done in several steps: tokenization, filtering out special terms, stemming, removing stop words, and a final frequency filter.

Tokenization requires transforming a long string of text into a vector with just one word in each entry; for example a phrase comprised of 10 words would be tokenized into a vector of size 10 with each word of the phrase occupying one entry. In a second step we filter out special terms; this removes numbers, punctuation, words not in Spanish, and special characters. The next step is to stem each word, this is done with Porter (2001)'s `snowball` stemmer. This step reduces the dimension of the dictionary by *collapsing* different words with the same underlying concept into a single stem, for example the words: *bank, banks, banking* will be stemmed into *bank.*[35]

Once words have been stemmed, we further reduce the dimension of the dictionary by removing stop words, which are common words such as connectors, that are so generic that removing them does not change the semantic content of a phrase but reduces its dimension. This step is done by filtering out words from a list of terms considered to be stop words. Spanish stop words come from `Python`'s package `stopwords-ISO` (Diaz (2020)). A final frequency filter is applied based on the *tf-idf* measure. This metric ranks each term in the dictionary according to the multiplication of the term frequency (tf) and the inverse document frequency (idf), and will assign a low ranking to terms that are *too common* or *too rare* in the corpus, and hence they do not provide representative semantic content to distinguish topics.[36] Using the *tf-idf* metric we only keep the top $50,000$ stems in the dictionary.

Figure 2.3a shows a wordcloud with the size of the word representing how frequently that word appears in the CR News database; note that many economic terms can be identified. Figure 2.3b

[35]For instance, in a language like Spanish where verbs are conjugated differently according to each grammatical tense and subject, stemming simplifies the dictionary of the corpus while preserving the main concepts.

[36]For each term v, *tf* is the number of times term v appears in document d divided by the number of terms in document d. *idf* is defined as the logarithm of the number of documents in the corpus minus the logarithm of the number of documents in the corpus which contain term v. Finally *tf-idf* is given by $tf \times idf$.

shows an example of a news paragraph before and after preprocessing it using the steps described above. After preprocessing the CR News database of articles, we have over 56,000 news articles comprising more than 15 million stems and $50,000$ unique stems.

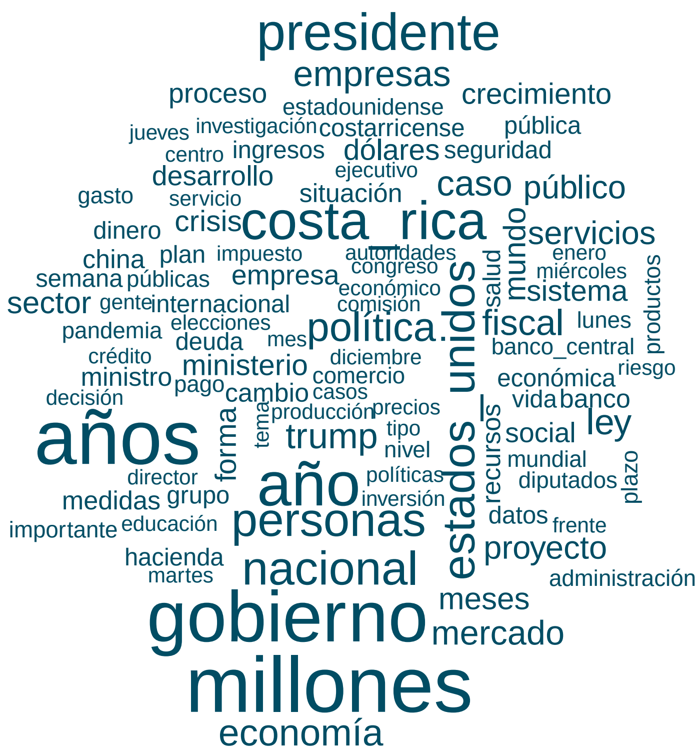
2.4.3. Costa Rica's Monex Market

The Monex market is a digital platform created and operated by the Central Bank of Costa Rica (BCCR) and it constitutes the official wholesale market where FX dealers exchange Costa Rican colónes (CRC), the local currency, against United States dollars (USD). As the United States is Costa Rica's main trade partner, the CRC/USD exchange rate receives almost exclusive attention in FX matters from the general public and the monetary authorities in Costa Rica. Moreover, the role of the USD as a vehicle currency for international trade is exacerbated in Costa Rica: almost all exports and imports in Costa Rica are invoiced in USD dollars despite that approximate 40% of the exports are destined to the United States and around 42% of the imports come from the United States (Boz et al., 2020).

The nature and design of the Monex market in Costa Rica is inherently related to the FX regime operating in the country. In 2006 the Monex market was created to play a fundamental role in the crawling band regime which ran 2006 to 2015. This chapter focuses on the managed floated regime that began in February 2015, where in principle dealers' incentives are not explicitly tied to the level of the exchange rate, as may happen in a crawling band regime when the exchange rate hits the bands.

The Monex market can be considered a second tier of the Costa Rican FX market; the first tier is constituted by over-the-counter windows (OTC Windows) between the general public and financial institutions. The second tier, reserved for FX interdealer trades, accounts for around 10% of all USD trades registered in the Costa Rican economy while the rest is dispersed through the OTC windows of financial institutions. As Cubero et al. (2019) remarks, approximately 90% of the transactions occurring through the OTC windows with the general public are for amounts less than 1000 USD, which is the minimum trade amount of the Monex market. This suggests that the Monex market is reserved for relatively specialized trade and therefore its turnover share.

Figure 2.3: CR News Database



(a) Word cloud for most frequent words in the news articles

Original string

```
"De los nuevos recursos que prestó el sector financiero costarricense en el p
rimer trimestre de este año, el Gobierno se llevó el 60%, 14 puntos porcentua
les más que en el primer trimestre del 2015. Lo que se llevó de más el Gobier
no se había prestado el año pasado a personas o empresas privadas. La partici
pación privada pasó de 52,6% a 35,3%, según los datos que publica el Banco Ce
ntral en su página web. La mayor cantidad de recursos que obtuvo el Gobierno
son para financiar el gasto que no cubre con los ingresos de los impuestos. P
or otro lado, el crédito al sector privado se dirige a consumo y producción."
```

After Preprocessing

```
["de", "los", "nuevos", "recursos", "que", "presto", "el", "sector", "financi
ero", "costarricense", "en", "el", "primer", "trimestre", "de", "este", "año"
, "el", "gobierno", "se", "llevó", "el", "60%", "14", "puntos", "porcentuales
", "más", "que", "en", "el", "primer", "trimestre", "del", "2015", "lo", "que
", "se", "llevó", "de", "más", "el", "gobierno", "se", "había", "prestado", "
el", "año", "pasado", "a", "personas", "o", "empresas", "privadas", "la", "pa
rticipación", "privada", "pasó", "de", "52,6%", "a", "35,3%", "según", "los",
"datos", "que", "publica", "el", "banco", "central", "en", "su", "pagina", "w
eb", "la", "mayor", "cantidad", "de", "recursos", "que", "obtuvo", "el", "gob
ierno", "son", "para", "financiar", "el", "gasto", "que", "no", "cubre", "con
", "los", "ingresos", "de", "los", "impuestos", "por", "otro", "lado", "el",
"credito", "al", "sector", "privado", "se", "dirige", "a", "consumo", "y", "p
roduccion"]
```

(b) Example of preprocessing outcome

Notes: Panel (a) a word cloud with the most frequent terms in the CR News Database; the size of the word represents its relative frequency in the corpus. Panel (b) shows the before and after text preprocessing for a news article extract.

Despite the fact that the Monex market is reserved for specialized trade, it has a crucial role in the price discovery process of the CRC/USD exchange rate. As Osler et al. (2011) hypothesizes, this could be though a 3-stage process: in the first stage the FX dealers extract information from their OTC transactions with informed agents, in the second stage the FX dealers will trade on this information in the interdealer market, in our case the Monex market, where this new information would be incorporated into the bid/ask exchange rates. Finally, after such private information is progressively shared and compounded into the equilibrium exchange rate of the interdealer market, in the third stage all FX dealers would react to this new information by updating their quotes in OTC windows. This suggests that the Monex market is a critical place where information is incorporated into the equilibrium exchange rate. Furthermore the Monex does not only signal private information coming to a specific customer segment served by participating FX dealers, but as it is also the market where BCCR's interventions take place, it can potentially signal current or future information on fundamental variables.

Participants in the Monex market are mainly financial institutions authorized by the BCCR to act as FX dealers. In order to be authorized to participate in the Monex market, dealers should comply with the ROOC (*Reglamento para las Operaciones Cambiarias de Contado*) regulation promulgated by the BCCR, that specifies requirements on behalf of FX dealers such as appropriate accounting systems, reports on FX activity to the BCCR and compliance with risk measures related to FX dealership (BCCR (2017a)).

Furthermore, due to regulation on FX dealers that puts prudential limits on their open net FX positions, both at absolute levels and in daily variation rates, any excess demand or supply of USD that FX dealers receive through their OTC windows outside these limits is transferred to the Monex market. Thus, in addition to being a place for price discovery, there is an incentive to participate in the Monex market, since it also represents a valuable tool for financial institutions to comply with FX regulation.

The number of participants in the Monex market varies from year to year depending on how FX dealers weight the costs and benefits of keeping their authorization from the BCCR. In the managed float period, 38 FX dealers can be identified as regular participants in the Monex market: the BCCR, the CD platform, 4 state-own banks, 12 private banks, 2 credit unions, 5 cooperatives, 2 currency exchange offices, 8 brokerage firms, and 3 financial firms.[37] If the general public that participates through the CD platform is disaggregated, then number of participants during the managed float period, is on average 77, with a minimum of 47 and a maximum of 137 participants.

The Monex market trades an average of 15.3 million USD daily with a median of 161 transactions.[38] This chapter makes use of access to proprietary data from the Monex Market, including every transaction made in the Monex market from February 2nd 2015 until February 26th 2021, with more than 255,000 observations on 1520 trading days. Information about every transaction records: name of both parties involved, the amount of USD traded, the corresponding exchange rate, whether the trade is buyer or seller initiated, and a timestamp up to milliseconds.

2.5. Exchange Rate Supervised Topics and FX News index

In this section we use text data from news articles to augment a standard model from the FX literature. We work with an specification that relates the daily exchange rate return and the order flow from currency transactions. This market-microstructure approach has the advantage that it can be estimated at relative higher frequencies than traditional fundamentals-based models, and hence shed more light about short term dynamics of the exchange rate.

Viewing FX activity using market microstructure tools, in particular the order flow, allows us to articulate a story based on the informational content of exchange rates and market frictions in order to map changes in fundamentals into new equilibria. Evidence in favor of an order flow with explanatory power even in short frequencies in the Monex market suggests an informational imbalance on the part of the FX dealers. In qualifying this informational imbalance, a news index with explanatory power is relevant.

We consider a classic order flow regression inspired by Evans and Lyons (2002), but augmented with

[37] A "public" dealer for the general public introduced in 2009 where any individual or firm with registered accounts on the CD platform can put their offers in the Monex Market as long as the currencies obtained are not used for FX dealership (BCCR (2010)).

[38] These statistics are computed using transactions data from the start of the managed float period, February 2nd 2015, until February 26th 2021, for a total of 1520 trading days.

the text data, given by the following supervising equation:

$$\begin{aligned} y_t &= b'\bar{Z}_t + c'W_t + \varepsilon_t \\ &= b'\bar{Z}_t + c' \underbrace{[\mathcal{O}_t \quad \Delta rd_t \quad N_t \quad m_t]'}_{W_t} + \varepsilon_t, \end{aligned} \tag{2.12}$$

where y_t is the daily log-return $y_t := \log(ER_t) - \log(ER_{t-1})$; and ER_t is the amount-weighted average exchange rate (expressed as CRC per USD) among all the trades in the Monex market for day t. $\bar{Z}_t$ are the empirical topic frequencies that follow the sLDA generative model in section 2.2.1 using news articles available at the beginning of day t; that is, news articles produced in $t-1$ and up to day t (excluding day-t news).[39] The covariates in W_t include: $\mathcal{O}_t$ as the order flow, the difference between the number of buyer-initiated and seller-initiated trades. $\Delta rd_t = \Delta(i_t^* - i_t)$ is the first difference in the interest rate differential using benchmark interest rates for Costa Rica and the United States. In particular i_t stands for the Costa Rican TBP (*Tasa básica pasiva*, an index of passive short-term interest rates paid in Costa Rica) and i_t^* for the U.S. prime rate, both at daily level. N_t number of daily transactions. Finally, m_t is net amount traded (total purchases minus total sales in USD).

In our application, the estimation of sLDA will yield an output composed of (i) exchange rate supervised topics, (ii) topic memberships that can be interpreted as time series on the daily attention each topic receives in the newspapers and (iii) parameters estimates from the supervising equation. Looking at these three outputs together will be key to interpret the supervised extracted topics.

2.5.1. Base Results

The base results are based on 31,000 news articles from the CR News database ranging from January 2018 to December 2020. The daily number of terms ranges from $3,000$ to $30,000$ with an average $12,000$ terms and the corpus is composed of 8.6 million terms after preprocessing. Pairing these

[39] Since the Monex trades on weekdays, $\{w_{t,1:N_t}\}$ will be one or three days worth of news articles. For example, if date t happens to be a Monday, w_t will include news articles from the previous Friday, Saturday and Sunday, that is news available at the beginning of Monday. Notice that in concatenating articles into a single article for day t we are exploiting the mixed membership feature in the LDA part of sLDA.

text data with FX data produces $T = 750$ document-response triplets that will be used in this base estimation. The algorithm is run with $K = 25$ topics, a dictionary size of $V = 10,000$ and hyperparameters set to $\alpha = (1/K) \cdot \mathbf{1}_K$ and $\eta = 0.1 \cdot \mathbf{1}_V$.[40]

Figure 2.4 shows word clouds, attention series and words associate with the most frequent terms for two of the extracted supervised topics. In the sLDA model each topic is represented as a vector of dimension $V = 10,000$, where each entry represents the probability of sampling the term in that entry according to the topic. A common way to represent topics is to use word clouds with the most probable terms in each topic; the size of the word represents its relative term frequency according to the topic.

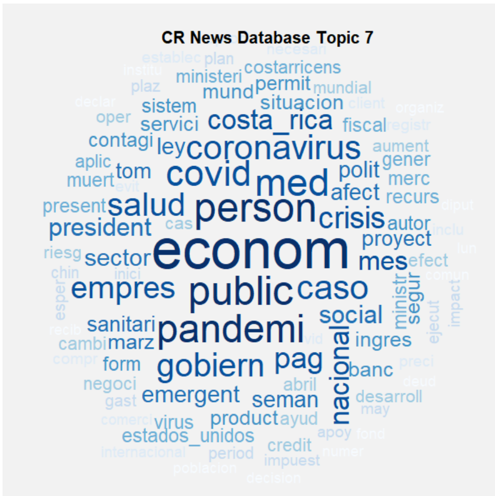
Panels (a) and (c) from Figure 2.4 show word clouds with the most frequent terms for the Exchange Rate Supervised Topics 7 and 20. Similarly, Panels (b) and (c) show daily series of topic memberships for these topics; such time series can be interpreted as the attention level for these topics in the newspapers. Additionally a set of the most probable words corresponding to the most probable terms (word stems) in each topic are listed. In conjunction, the word cloud of most probable terms, the attention series and the list of most probable words suggest that topic 20 refers to the 2018 Tax Reform Discussion in Costa Rica, whereas Topic 7 refers to the Covid-19 Pandemic. Looking at the attention series validates this interpretation: we see that in the case of Topic 20, its attention peaks around the start of strikes in Costa Rica triggered by a tax reform debated and approved in 2018. For topic 7, the most probable words and the time when its attention peaks supports labeling it as the Covid-19 Pandemic topic.

Notice that the labeling is innocuous for estimation; what identifies topics are the corresponding $(\beta; \theta)$ mixtures; labels are just a convenient way to refer to topics when presenting results. A complete list of labeled topics extracted from the sLDA estimation with their top most probable stems in Spanish and their English translations can be found in Appendix B.2.1.

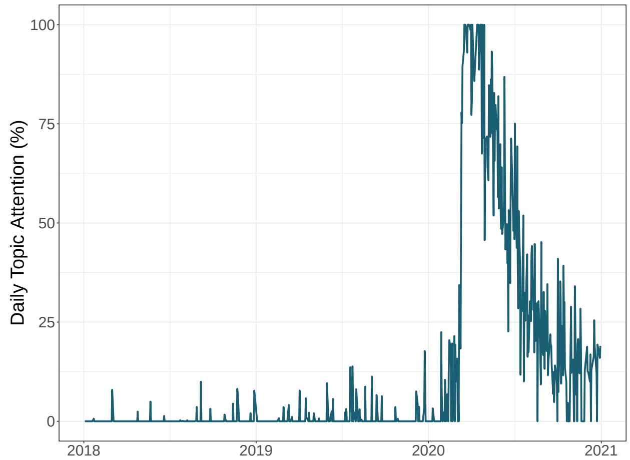
[40] Other model specifications with 30, 35 and 40 topics were estimated but $K = 25$ yielded the best balance regression fit and interpretable topics.

Figure 2.4: Selected Exchange Rate Supervised Topics

Topic 7-Covid-19 Pandemic



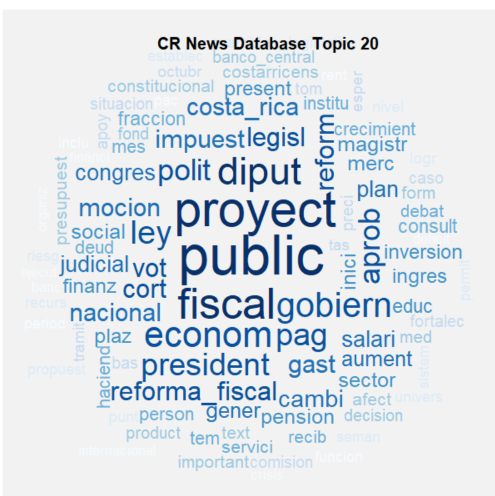
(a) Word cloud Topic 7



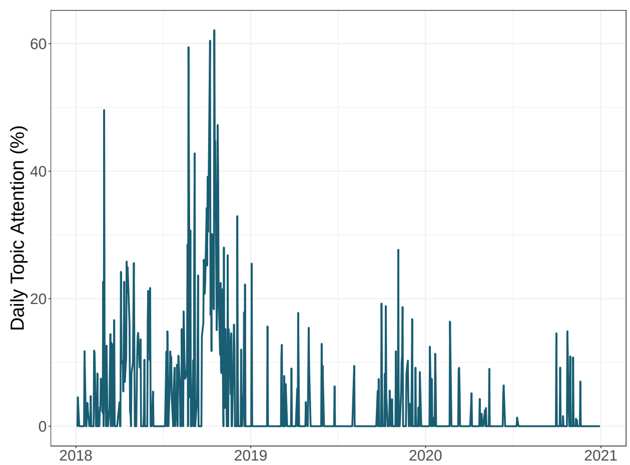
(b) Attention Series Topic 7

Topic 7 Top words: economy, medicine, pandemic, person, covid, coronavirus, public, firm, government, health, crisis, national, emergency, month.

Topic 20 - 2018 Tax Reform



(c) Word cloud Topic 20



(d) Attention Series Topic 20

Topic 20 Top words: public, project, fiscal_reform, government, president, national, law, services, social, judicial, case, legislature.

Notes: Panels (a) and (c) show word clouds with the most frequent terms for the Exchange Rate Supervised Topics 2 and 20; the size of the term represents its relative frequency according to each topic. Panels (b) and (c) show daily topic memberships for Topic 2 and 20; these time series can be interpreted as the attention level for these topics in the newspapers. Additionally a set of the most probable words corresponding to the most probable terms (word stems) in each topic is listed. In conjuction, the word cloud, the attention series and the list of most probable words suggest how to interpret each topic. See text for details.

Table 2.2 shows sLDA estimates for the supervising equation in (2.12). Results for a traditional order flow regression without text data are shown in column (1). Results for the sLDA model with the order flow regression supervising equation are shown in column (2). Supervising parameters for selected extracted topics in column (2) show the average effects in basis points for the daily log return in the exchange rate if the topics receives a 100% of the attention in the newspaper. Positive (negative) values of these coefficients represent a partial effect consistent with short-term depreciations (appreciation) of the CRC/USD exchange rate.

In the case of the selected topics we observe the expected coefficient sign. For example, topic 20 and 25 refer to tax reform that was debated and approved in Costa Rica in 2018, respectively. The discussion of tax reform was put forward to correct the rapid growth of the government debt and restore investors' confidence that the Costa Rican government would be able to meet its debt payments. Topic 25 refers to the general discussion and strikes that were triggered during the debate of the tax reform. This topic has a positive coefficient which supports the interpretation that the strike jeopardized the likelihood of approval of the tax reform and clouded investors' trust on the long term sustainability of the CRC currency. Investors in face of such uncertainty will have a relative greater taste for holding USD and the exchange rate will depreciate. Topic 25 refers to the coverage in the newspaper of the approval of tax reform towards the end of 2018. This topic has a negative coefficient which is consistent with the interpretation that the approval of the tax reform eased concerns about the fiscal sustainability of the Costa Rican Government which in turn will return confidence in the CRC currency and an appreciation followed.

Table 2.2 also shows that the inclusion of text data improves the fit of the order flow regression. R^2 almost double compared with an specification that does not include text data from news articles. This pattern is also present in the adjusted-R^2 of the model which improves the fit in more than 10% with respect to an order flow that does not controls for the role of public information in determining the short-term dynamics of the exchange rate.

Once the sLDA have been estimated it is possible to summarize the supervised informational signal from news articles into an FX news index that can be used in further anaylis and even as a covariate in other model specifications of the FX literature. With sLDA estimates $\hat{\pi} = \{\hat{b}, \hat{c}, \hat{\sigma}^2; \alpha, \eta\}$ and $\{\hat{\gamma}_t, \hat{\phi}_{t,1:N_t}, \hat{\beta}_{1:K}\}$; we can construct a FX news index as:

$$NI_t := \hat{b}'\hat{Z}_t = \hat{b}'\left(\frac{1}{N_t}\sum_{i=1}^{N_t}\hat{\phi}_{t,i}\right). \tag{2.13}$$

Notice that we estimate $\bar{Z}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} z_{t,i}$ with the expected value under the estimated variational distribution $E_{\hat{q}_t}[\bar{Z}] = \frac{1}{N_t}\sum_{i=1}^{N_t}\hat{\phi}_{t,i}$. Positive (negative) values of the index can be interpreted as a positive correlation for a depreciation (appreciation) for the CRC/USD exchange rate (larger daily return for holding USD) for period t. The larger the magnitude of the FX news index, the larger the correlation for a depreciation or an appreciation on the CRC/USD exchange rate.

Figure 2.5 shows the FX news index constructed from the sLDA estimation. Two dynamics are worth remarking from the evolution of the index. The first one is the appreciation of the CRC towards the end of 2018, as mentioned above it is likely related to the approval of the tax reform and a restore confidence in the currency. The second is the sustained depreciation trend for most of 2020 and that started around March 2020. This depreciation trend is consistent with the interpretation that investors, in times of generalized global economic uncertainty, substitute away currencies from emerging market economies, in this case the CRC to hold hard currencies like the USD. In this case such generalized global economic uncertainty was triggered by the onset of the Covid-19 pandemic.

2.5.2. Prediction

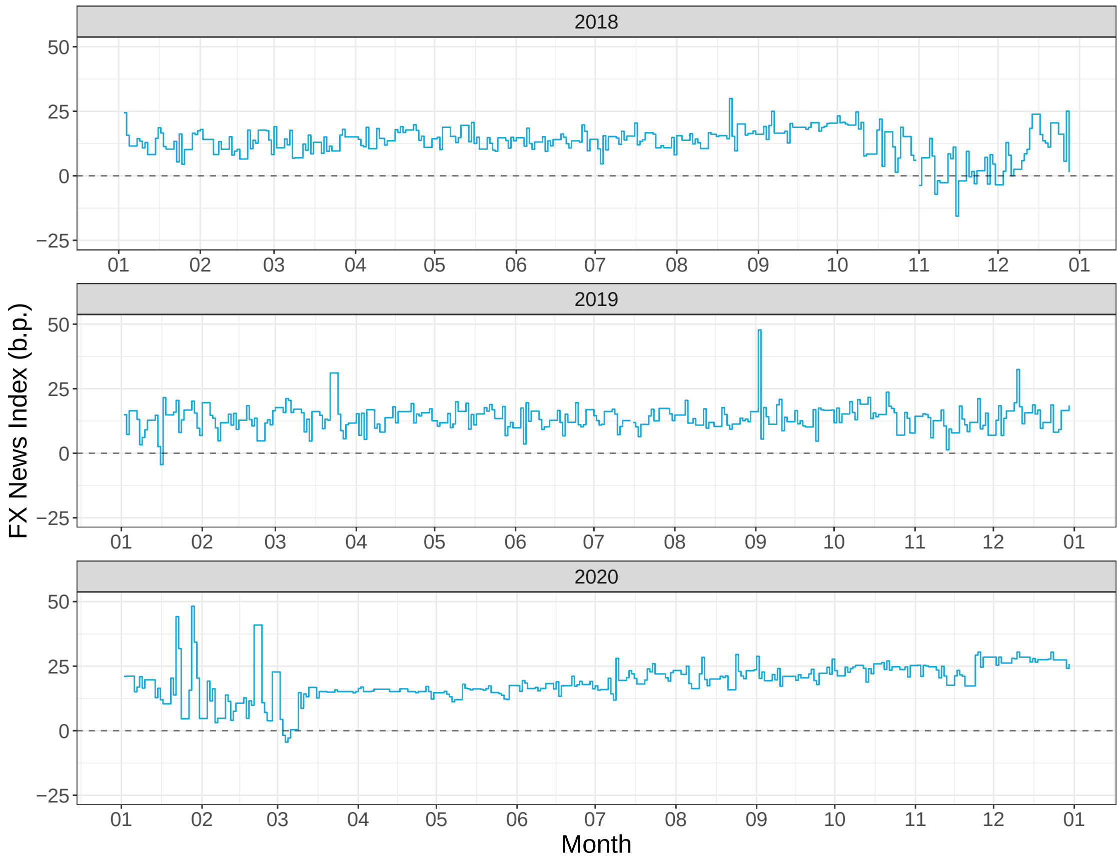
We can also assess the predictive power that adding text data from news articles has on an order flow regression. In this out-of-sample prediction exercise we use base the estimates from models presented in Table 2.2 to forecast the daily log return of the CRC/USD exchange rate over period of two months. Specifically we used observed data from January and February 2021, in this period we compare the out-of-sample prediction and order flow regression will make with the forecast that the augmented model with text data from news articles will make. Given that we observe the realized daily return for the exchange rate we can compare prediction errors under both specifications.

Table 2.2: Supervising Equation Parameter Estimates

	Dependent Variable: Daily Log Return ER_t (b.p.)	
	Model without Text	Model with Text
Δrd_t (b.p)	0.133	0.140
	(0.091)	(0.094)
Order Flow	0.277	0.248
	(0.039)	(0.028)
N. Transactions	0.019	-0.070
	(0.007)	(0.021)
Selected Topics (b.p. for 100% attention)		
Topic 7: Covid 19 Pandemic		15.165
		(5.077)
Topic 20: 2018 Tax Reform Discussion		25.904
		(10.325)
Topic 25: 2018 Tax Reform Approval		-15.558
		(12.739)
Mean Daily Log Return	1.056	1.056
σ	23.051	21.213
Number of Topics	-	25
Observations	750	750
R^2	0.159	0.288
R^2-Adjusted	0.156	0.258
Aikaike Information Criterion (x10^6)	291.989	247.020
Bayesian Information Criterion (x10^6)	291.989	247.020

Notes: Results for a traditional order flow regression are shown in the column labelled as "Model without text". Results for an order flow augmented with text data from newspapers and estimated with sLDA are shown in the column "Model with Text". The sLDA model is estimated with $K = 25$ underlying topics and a dictionary size of $V = 10,000$. Supervising parameters for selected extracted topics are shown in column (2); these coefficients show the average effects in basis point for the daily log return in the exchange rate if the topics receives a 100% of the attention in the newspaper. Standard errors are compute as $\widehat{var}(\hat{\beta}) = T^{-1}(T^{-1}\sum_{t=1}^{T} x_t x_t')^{-1}(T^{-1}\sum_{t=1}^{T} \hat{u}_t^2 x_t x_t')(T^{-1}\sum_{t=1}^{T} x_t x_t')^{-1}$ where x_t stands for the vector of regressors included in the regression model and $\hat{u}_t$ the corresponding residuals.

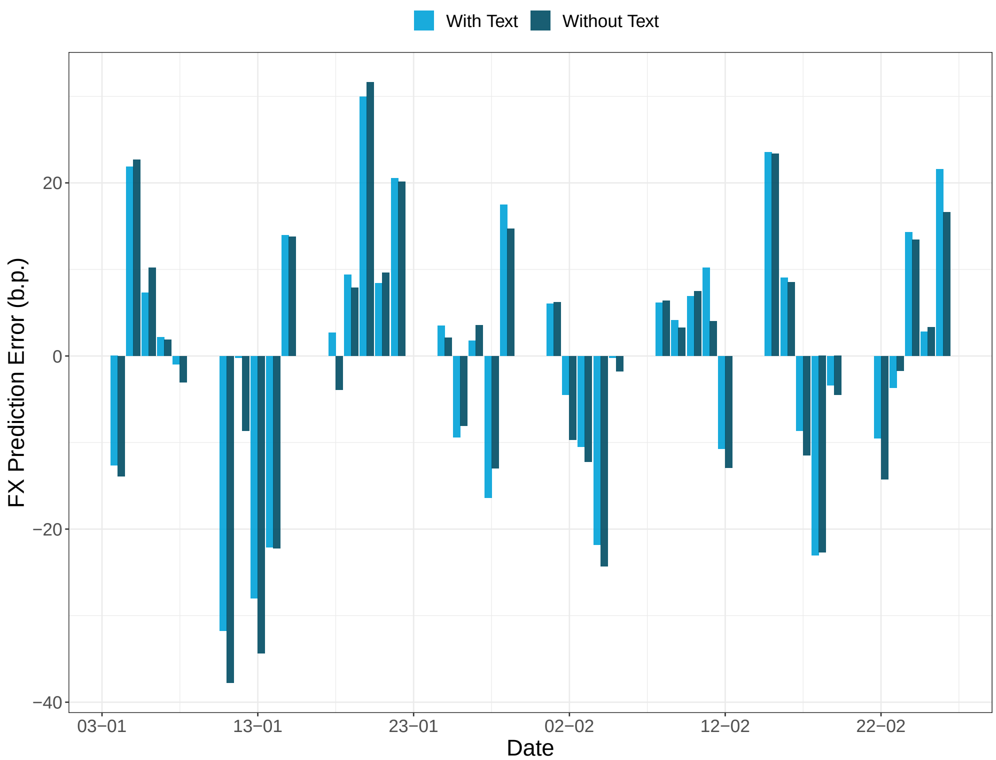
Figure 2.5: FX News Index for the CRC/USD exchange rate



Notes: FX News Index for the CRC/USD exchange rate constructed from the sLDA estimation. It corresponds to the partial effect of news data in the supervising equation of sLDA: $NI_t := \hat{b}'\hat{Z}_t = \hat{b}'\left(\frac{1}{N_t}\sum_{i=1}^{N_t}\hat{\phi}_{t,i}\right)$. Positive values of the index can be interpreted as a positive correlation for a depreciation on the CRC/USD exchange rate (larger daily return for holding USD) for period t. Analogously negative values can be interpreted as a correlation for an appreciation of the CRC/USD exchange rate.

Results for the 40 trading sessions contained in this two-month period are shown in Figure 2.6. The MSPE for the model with no text is 233.99 whereas MSPE the sLDA model is 208.47. These numbers yield a relative efficiency of prediction of 1.12, that is prediction errors of a model without text are 12% higher than the sLDA model. This suggest that incorporating a control for public information in the form of supervised extracted topics from news articles improves the prediction power of an order flow regression in short-term frequencies.

Figure 2.6: Prediction Errors for the Exchange Rate



Notes: Out-of-sample prediction errors from the models presented in Table 2.2. The label 'with text' refers for the sLDA model that uses text data from news articles and the label 'without text' refers to a standard order flow regression that does not uses text data. The out-of-sample period comprises 40 trading sessions in the months of January and February 2022. The MSPE for the model with no text is 233.99 whereas MSPE the sLDA model is 208.47. These numbers yield a relative efficiency of prediction of 1.12, that is prediction errors of a model without text are 12% higher than the sLDA model.

2.6. Conclusions

This chapter shows how to use a hybrid of unsupervised and supervised learning models to go from text from news articles to an FX news index that can be used to enhance traditional models from the FX literature. We relied on Supervised Latent Dirichlet Allocation (sLDA) which combines information about a supervising variable with topic extraction over a corpus of text in a single-stage estimation.

Via a Monte Carlo simulation we document the efficiency gains of single-stage strategies like sLDA over two-stage estimation approaches that perform LDA follow by a regression. The message from this simulation study is threefold: (i) efficiency measured by the ratio of MSE from single-stage to MSE two-stage estimates of supervised parameters are the largest; this is consistent with an explicit acknowledge in sLDA of the connection between text and response. (ii) Even though topics are generated before responses and both LDA and sLDA model topics the same way, having supervised information improves the topic extraction process. This is reflected in the estimation of word probabilities. (iii) MSPE are also generally smaller when using a single-stage approach; the relative more efficient estimation of sLDA also improves prediction.

The empirical application is centered around the Monex Market, the main Costa Rican platform for FX trade and accordingly news articles are gathered from the main Costa Rican newspapers. Using the CRC/USD exchange rate we estimated an sLDA model with a supervising equation form the market-microstructure approach to exchange rates. This equation relates the the daily exchange rate return, the order flow from currency transactions, other control covariates and it is augmented to include text data in the form of topic membership series; all estimated in the single-stage procedure of the modified sLDA used in this chapter. Estimates from sLDA show that the inclusion of text data improves the fit of the order flow regression. R^2 almost double compared with an specification that does not include text data from news articles. Adjusted-R^2 of the sLDA model also shows an improved the fit, with a more than 10% increase with respect to an order flow that does not controls for the role of public information in determining the short-term dynamics of the exchange rate.

The extracted supervised topics show high degree of interpretability. Most of them can be related to events that can be interpreted in connection with exchange rate dynamics. The inclusion of external information during the topic extraction process facilitates a parsimonious text structure. This is reflected in a relative lower number of topics, compared with a unsupervised topic extraction, as our focus is to extract topics that matter for the supervising variable, in this case the exchange rate. The supervised informational signal from news articles can be summarized into an FX news index. We observed that this index correlates with depreciation and appreciation dynamics that can be interpreted in the light of macroeconomic events.

When performing analysis centered around prediction we presented evidence that in an out-of-sample prediction exercise, the forecast errors of a traditional order flow regression without text are 12% higher than the sLDA model. This suggest that incorporating a control for public information in the form of supervised extracted topics from news articles improves the prediction power of an order flow regression in short-term frequencies.

Future research. There are extensions to both data and methodology which could benefit from future research. In terms of expanding the data, next step is to include news articles from US newspapers; they would contain information about the US economy that could influence the demand/supply for USD in the Monex Market. The main challenge with including US newspapers is to ensure topic cohesiveness when the corpus is written in more than one language.

Another data expansion is to consider a panel of countries/currencies, for example when considering the USD vs the pound sterling (GBP), the USD vs the Mexican peso (MXN) and the USD vs the Canadian dollar (CAD), all fit the methodology presented in this chapter: exchange rates in flotation for which news articles can be found in English or Spanish. This collection of exchange rates also receive more attention given that these currencies represent a larger share of international trade.

In term of methodology, there a research opportunity is to consider time-varying parameters in sLDA, specifically the supervised parameters (a time varying version of LDA already exists in Blei and Lafferty (2006)). A first step would be to gather evidence that time-varying sLDA improves empirical applications, for doing so can perform static sLDA with estimation over a rolling window. If results provide evidence of dynamic relevance of topics to explain the exchange rate at a daily frequency, it would be interesting to pursue a time-varying sLDA.

Another methodology aspect that requires further research is to compare the estimation of sLDA using variational inference that approximates the likelihood versus using exact inference in the form of a MCMC strategy. This can be done by assuming priors for the supervise parameters and writing out a corresponding Gibbs sampling strategy; then a Monte Carlo simulation contrasting both approaches can be designed.

APPENDIX A

Chapter 1 Appendices

A.1. Proofs and Mathematical Derivations

A.1.1. Proof of Proposition 1

Proof. The following proof follows the main steps of Lemma 1 proof's in KY2019 with the adaptation for the more general objective function. The function inside the conditional expectation in (1.1) takes the form

$$F_i(A_{i,t+1}, C_{it}, w_{it}) = \log(A_{i,t+1}) + a_i' C_t' w_{it},$$

we can replace the first term with

$$\log(A_{i,t+1}) = \log(A_{it}) + \log\left(\frac{A_{i,t+1}}{A_{it}}\right) = \log(A_{it}) + \log\left(R^0_{t+1} + w_{it}'(R_{t+1} - R^0_{t+1}1)\right)$$

by using the budget constraint (1.2). Then the Lagrangian for the problem is given by

$$\begin{aligned} L(w_{it}, \Lambda_{it}, \lambda_{it}) = \log(A_{it}) + E_{it}\left[\log\left(R^0_{t+1} + w_{it}'(R_{t+1} - R^0_{t+1}1)\right)\right] \\ + a_i' C_t' w_{it} + \Lambda_{it}' w_{it} + \lambda_{it}(1 - 1' w_{it}), \end{aligned}$$

the first order condition with respect to w_{it} given by

$$\begin{aligned} E_{it}\left[\left(R^0_{t+1} + w_{it}(R_{t+1} - R^0_{t+1}1)\right)^{-1}\left(R_{t+1} - R^0_{t+1}1\right)'\right] + \Lambda_{it}' - \lambda_{it}1' + a_i' C_t' = 0 \\ \Rightarrow E_{it}\left[\left(\frac{A_{i,t+1}}{A_{it}}\right)^{-1}\left(R_{t+1} - R^0_{t+1}1\right)\right] = -\left(\Lambda_{it} - \lambda_{it}1 + C_t a_i\right). \qquad \text{(A1)} \end{aligned}$$

Multiplying this equation by $-1w_{it}'$ yields:

$$-E_{it}\left[\left(\frac{A_{i,t+1}}{A_{it}}\right)^{-1} 1w_{it}'\left(R_{t+1}-R_{t+1}^{0}1\right)\right] = 1w_{it}'\left(\Lambda_{it}-\lambda_{it}1+C_t a_i\right)$$
$$\Rightarrow -E_{it}\left[\left(\frac{A_{i,t+1}}{A_{it}}\right)^{-1}\left(\frac{A_{i,t+1}}{A_{it}}-R_{t+1}^{0}\right)1\right] = 1w_{it}'\left(\Lambda_{it}-\lambda_{it}1+C_t a_i\right),$$

using the intertemporal budget constraint. Summing the last expression with (A1) results in Euler equation in Proposition 1:

$$E_{it}\left[\left(\frac{A_{it+1}}{A_{it}}\right)^{-1} R_{t+1}\right] = 1-\left(I-1w_{it}'\right)\left(\Lambda_{it}-\lambda_{it}1+C_t a_i\right).$$

Next, using the intertemporal budget constraint we can write the objective function of the investor problem as:

$$\log(A_{it}) + E_{it}\left[\log\left(R_{t+1}^{0}+w_{it}'(R_{t+1}-R_{t+1}^{0}1)\right)\right] + a_i'C_t'w_{it}.$$

Let R_{t+1}^{p} denote the gross return of the portfolio with weights w_{it}, then

$$R_{t+1}^{p} = R_{t+1}^{0} + w_{it}'(R_{t+1}-R_{t+1}^{0}1),$$

and the log excess return of the portfolio weith respect to the outside option gross return is given by

$$r_{t+1}^{p}-r_{t+1}^{0} = \log\left(\frac{R_{t+1}^{p}}{R_{t+1}^{0}}\right) = \log\left(1+w_{it}'\left(\exp(r_{t+1}-r_{t+1}^{0}1)-1\right)\right).$$

Now consider the function $g(x):\mathbb{R}^J\to\mathbb{R}$ given by $f(x)=\log(1+w'(\exp(x)-1))$, where the $\exp(\cdot)$ applies entry-by-entry, and w is a J-vector constant, then a second order Taylor approximation of g around $x_0=0\in\mathbb{R}^J$ is given by

$$g(x)\approx w'\left(x+\frac{1}{2}x\odot x\right)-\frac{1}{2}w'\left(xx'\right)w,$$

where $\odot$ stands for entry-by-entry multiplication. Applying this approximation to the expression for $r^p_{t+1} - r^0_{t+1}$ yields

$$r^p_{t+1} - r^0_{t+1} \approx w'_{it} \left((r_{t+1} - r^0_{t+1}) + \frac{1}{2}(r_{t+1} - r^0_{t+1}) \odot (r_{t+1} - r^0_{t+1}) \right) - \frac{1}{2} w'_{it} (r_{t+1} - r^0_{t+1})(r_{t+1} - r^0_{t+1})' w_{it}.$$

Next we apply the expectations operator $E_{it}[\cdot]$ and the second term in the objective function can be approximated by

$$\begin{aligned} E_{it}\left[\log\left(R^0_{t+1} - w'_{it}(R_{t+1} - R^0_{t+1}1)\right)\right] &\approx r^0_{t+1} + w'_{it}\left(E_{it}[r_{t+1} - r^0_{t+1}] + \frac{\tilde{\sigma}^2_{it}}{2}\right) - \frac{w'_{it}\tilde{\Sigma}_{it}w_{it}}{2} \\ &= r^0_{t+1} + w'_{it}\tilde{\mu}_{it} - \frac{w'_{it}\tilde{\Sigma}_{it}w_{it}}{2}, \end{aligned}$$

this approximation replaces $E_{it}[(r_{t+1} - r^0_{t+1})(r_{t+1} - r^0_{t+1})']$ with $\tilde{\Sigma}_{it}$ and $E_{it}[(r_{t+1} - r^0_{t+1}) \odot (r_{t+1} - r^0_{t+1})]$ with $\tilde{\sigma}^2_{it}$, and follows from the one presented in Campbell and Viceira (2002) (Eq. 2.23). With this approximation the first order condition for becomes

$$\begin{aligned} &\tilde{\mu}_{it} - \tilde{\Sigma}_{it}w_{it} + \Lambda_{it} - \lambda_{it}1 + C_t a_i = 0 \\ &\Rightarrow w_{it} = \tilde{\Sigma}^{-1}_{it}\left(\tilde{\mu}_{it} + \Lambda_{it} - \lambda_{it}1 + C_t a_i\right), \end{aligned}$$

Partition the asset space between those with non-binding short sale constraint and those binding we write $\Lambda'_{it} = [0' \quad \Lambda^{(2)'}_{it}]$ and $(C_t a_{it}) = [(C_t a_i)'_1 \quad (C_t a_i)'_2]'$, then using the partitions for $\tilde{\Sigma}_{it}$ and $\tilde{\mu}_{it}$ we have that

$$w_{it} = \begin{pmatrix} w^{(1)}_{it} \\ 0 \end{pmatrix} = \begin{pmatrix} \Sigma_{it} & \Sigma^{(1,2)}_{it} \\ \Sigma^{(2,1)}_{it} & \Sigma^{(2,2)}_{it} \end{pmatrix}^{-1} \left(\begin{pmatrix} \mu_{it} \\ \mu^{(2)}_{it} \end{pmatrix} + \begin{pmatrix} 0 \\ \Lambda^{(2)}_{it} \end{pmatrix} - \lambda_{it}1 + \begin{pmatrix} (C_t a_i)_1 \\ (C_t a_i)_2 \end{pmatrix} \right).$$

The inverse of $\tilde{\Sigma}_{it}$ is given by

$$\tilde{\Sigma}_{it}^{-1} = \begin{pmatrix} \left(\Sigma_{it} - \Sigma_{it}^{(1,2)}\Sigma_{it}^{(2,2)-1}\Sigma_{it}^{(2,1)}\right)^{-1} & -\Sigma_{it}^{-1}\Sigma_{it}^{(1,2)}\left(\Sigma_{it}^{(2,2)} - \Sigma_{it}^{(1,2)}\Sigma_{it}^{-1}\Sigma_{it}^{(2,1)}\right)^{-1} \\ -\Sigma_{it}^{(2,2)-1}\Sigma_{it}^{(2,1)}\left(\Sigma_{it} - \Sigma_{it}^{(1,2)}\Sigma_{it}^{(2,2)-1}\Sigma_{it}^{(2,1)}\right)^{-1} & \left(\Sigma_{it}^{(2,2)} - \Sigma_{it}^{(1,2)}\Sigma_{it}^{-1}\Sigma_{it}^{(2,1)}\right)^{-1} \end{pmatrix},$$

then w_{it} becomes

$$\begin{pmatrix} w_{it}^{(1)} \\ 0 \end{pmatrix} = \begin{pmatrix} W_a \\ W_b \end{pmatrix},$$

with

$$\begin{aligned} W_a = &\left(\Sigma_{it} - \Sigma_{it}^{(1,2)}\Sigma_{it}^{(2,2)-1}\Sigma_{it}^{(2,1)}\right)^{-1}\left(\mu_{it} - \lambda_{it}1 + (C_t a_i)_1\right) \\ &- \Sigma_{it}^{-1}\Sigma_{it}^{(1,2)}\left(\Sigma_{it}^{(2,2)} - \Sigma_{it}^{(1,2)}\Sigma_{it}^{-1}\Sigma_{it}^{(2,1)}\right)^{-1}\left(\mu_{it}^{(2)} + \Lambda_{it}^{(2)} - \lambda_{it}1 + (C_t a_i)_2\right), \end{aligned}$$

$$\begin{aligned} W_b = &-\Sigma_{it}^{(2,2)-1}\Sigma_{it}^{(2,1)}\left(\Sigma_{it} - \Sigma_{it}^{(1,2)}\Sigma_{it}^{(2,2)-1}\Sigma_{it}^{(2,1)}\right)^{-1}\left(\mu_{it} - \lambda_{it}1 + (C_t a_i)_1\right) \\ &+ \left(\Sigma_{it}^{(2,2)} - \Sigma_{it}^{(1,2)}\Sigma_{it}^{-1}\Sigma_{it}^{(2,1)}\right)^{-1}\left(\mu_{it}^{(2)} + \Lambda_{it}^{(2)} - \lambda_{it}1 + (C_t a_i)_2\right). \end{aligned}$$

We can multiply the second block by $\Sigma_{it}^{-1}\Sigma_{it}^{(1,2)}$ and sum both blocks to obtain

$$\begin{aligned} w_{it}^{(1)} &= \left(I - \Sigma_{it}^{-1}\Sigma_{it}^{(1,2)}\Sigma_{it}^{(2,2)-1}\Sigma_{it}^{(2,1)}\right)\left(\Sigma_{it} - \Sigma_{it}^{(1,2)}\Sigma_{it}^{(2,2)-1}\Sigma_{it}^{(2,1)}\right)^{-1}\left(\mu_{it} - \lambda_{it}1 + (C_t a_i)_1\right) \\ &= \Sigma_{it}^{-1}(\mu_{it} - \lambda_{it}1 + (C_t a_i)_1). \end{aligned}$$

So following the notation, for the optimal positive weights on the investor's problem can be approximated by

$$w_{it} \approx \Sigma_{it}^{-1}\left(\mu_{it} - \lambda_{it}1 + (C_t a_i)_1\right).$$

To pin down the value of λ_{it}, notice that when constraint (1.3) is binding then

$$1'w_{it} = 1'\Sigma_{it}^{-1}\left(\mu_{it} - \lambda_{it}1 + C_t a_i\right) = 1,$$

then

$$\lambda_{it} = \frac{\max\{1'\Sigma_{it}^{-1}\left(\mu_{it} + C_t a_i\right)1, 0\}}{1'\Sigma_{it}^{-1}1}.$$

□

A.1.2. Derivation of investor-specific posterior moments

Proof. We have that

$$s_{it}|(r_{t+1}^x, \alpha_i) \sim N(\alpha_i r_{t+1}^x, \Sigma_\varepsilon)$$
$$r_{t+1}^x \sim N(\mu_{xt}, \Sigma_{xt}),$$

where in our case $\Sigma_\varepsilon = \sigma_\varepsilon^2 I$. The pdfs of these distributions are given by:

$$p(s_{it}|r_{t+1}^x, \alpha_i) = (2\pi)^{-J_t/2} det(\Sigma_\varepsilon)^{-1/2} \exp\left[-\frac{1}{2}(s_{it} - \alpha_i r_{t+1}^x)'\Sigma_\varepsilon^{-1}(s_{it} - \alpha_i r_{t+1}^x)\right]$$
$$p(r_{t+1}^x) = (2\pi)^{-J_t/2} det(\Sigma_{xt})^{-1/2} \exp\left[-\frac{1}{2}(r_{t+1}^x - \mu_{xt})'\Sigma_{xt}^{-1}(r_{t+1}^x - \mu_{xt})\right].$$

By Bayes theorem $p(r^x_{t+1}|s_{it},\alpha_i) \propto p(s_{it}|r^x_{t+1},\alpha_i)p(r^x_{t+1})$ and

$$\begin{aligned}
\log(p(r^x_{t+1}|s_{it},\alpha_i)) =& -\frac{1}{2}(s_{it}-\alpha_i r^x_{t+1})'\Sigma_\varepsilon^{-1}(s_{it}-\alpha_i r^x_{t+1}) \\
& -\frac{1}{2}(r^x_{t+1}-\mu_{xt})'\Sigma_{xt}^{-1}(r^x_{t+1}-\mu_{xt}) + cons \\
=& -\frac{1}{2}(\alpha_i r^x_{t+1})'\Sigma_\varepsilon^{-1}(\alpha_i r^x_{t+1}) + (\alpha_i r^x_{t+1})'\Sigma_\varepsilon^{-1}(s_{it}) \\
& -\frac{1}{2}(r^x_{t+1})'\Sigma_{xt}^{-1}(r^x_{t+1}) - \frac{1}{2}(r^x_{t+1})'\Sigma_{xt}^{-1}(\mu_{xt}) + cons \\
=& -\frac{1}{2}(r^x_{t+1})'[\alpha_i^2\Sigma_\varepsilon^{-1}+\Sigma_{xt}^{-1}](r^x_{t+1}) + (r^x_{t+1})'[\Sigma_\varepsilon^{-1}\alpha_i s_{it}+\Sigma_{xt}^{-1}\mu_{xt}] + cons \\
=& -\frac{1}{2}\left(r^x_{t+1} - [\alpha_i^2\Sigma_\varepsilon^{-1}+\Sigma_{xt}^{-1}]^{-1}[\Sigma_\varepsilon^{-1}\alpha_i s_{it}+\Sigma_{xt}^{-1}\mu_{xt}]\right)' \\
& \cdot [\alpha_i^2\Sigma_\varepsilon^{-1}+\Sigma_{xt}^{-1}]^{-1}\left(r^x_{t+1} - [\alpha_i^2\Sigma_\varepsilon^{-1}+\Sigma_{xt}^{-1}]^{-1}[\Sigma_\varepsilon^{-1}\alpha_i s_{it}+\Sigma_{xt}^{-1}\mu_{xt}]\right) + cons.
\end{aligned}$$

This is the pdf of a multivariate normal distribution with variance $\Sigma_{r|s_i}$ and mean $\mu_{r|s_i}$ given by

$$\begin{aligned}
\Sigma_{r|s_i} &= [\alpha_i^2(\sigma_\varepsilon^2)^{-1}I + \Sigma_{xt}^{-1}]^{-1} \\
\mu_{r|s_i} &= \Sigma_{r|s_i}[(\sigma_\varepsilon^2)^{-1}\alpha_i s_{it} + \Sigma_{xt}^{-1}\mu_{xt}],
\end{aligned}$$

where we used the fact that $\Sigma_\varepsilon = \sigma_\varepsilon^2 I$. □

A.1.3. Proof of Proposition 2

Proof. Recall that $\Sigma_{it} = \Sigma_{r|s_i} = (\alpha_i^2(\sigma_\varepsilon^2)^{-1}I + \Sigma_x^{-1})^{-1}$. The term Σ_{xt} is given by

$$\Sigma_x = (\Gamma_{xt}\Gamma_{xt}' + \sigma_e^2 I),$$

then by the Woodbury matrix identity

$$\begin{aligned}
\Sigma_{xt}^{-1} &= [\Gamma_{xt}\Gamma_{xt}' + \sigma_e^2 I]^{-1} \\
&= \frac{1}{\sigma_e^2}\left(I - \frac{\Gamma_{xt}\Gamma_{xt}'}{\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt}}\right),
\end{aligned}$$

and by substituting into the expression for Σ_{it} yields

$$\begin{aligned}
\Sigma_{it}^{-1} &= \left(\frac{\alpha_i}{\sigma_\varepsilon^2} + \frac{1}{\sigma_e^2}\right) I - \frac{1}{\sigma_e^2(\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt})}\Gamma_{xt}\Gamma_{xt}' \\
&= \left(\frac{\alpha_i\sigma_e^2 + \sigma_\varepsilon^2}{\sigma_\varepsilon^2\sigma_e^2}\right) I - \frac{1}{\sigma_e^2(\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt})}\Gamma_{xt}\Gamma_{xt}' \\
&= \frac{1}{\sigma_e^2(\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt})}\left[\frac{(\alpha_i\sigma_e^2 + \sigma_\varepsilon^2)(\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt})}{\sigma_\varepsilon^2} I - \Gamma_{xt}\Gamma_{xt}'\right].
\end{aligned}$$

Now let's define $\delta_i = \frac{(\alpha_i\sigma_e^2+\sigma_\varepsilon^2)(\sigma_e^2+\Gamma_{xt}'\Gamma_{xt})}{\sigma_\varepsilon^2}$ then

$$\Sigma_{it}^{-1} = \frac{1}{\sigma_e^2(\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt})}[\delta_i I - \Gamma_{xt}\Gamma_{xt}'],$$

and

$$\begin{aligned}
\Sigma_{it} &= \sigma_e^2(\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt})[\delta_i I - \Gamma_{xt}\Gamma_{xt}']^{-1} \\
&= \sigma_e^2(\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt})\frac{1}{\delta_i}\left[I + \frac{\Gamma_{xt}\Gamma_{xt}'}{\delta_i - \Gamma_{xt}'\Gamma_{xt}}\right] \\
&= \frac{\sigma_\varepsilon^2\sigma_e^2}{\alpha_i\sigma_e^2 + \sigma_\varepsilon^2} I + \frac{\sigma_e^2(\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt})}{\delta_i - \Gamma_{xt}'\Gamma_{xt}}\Gamma_{xt}\Gamma_{xt}' \\
&:= \iota_{it} I + \Gamma_{it}\Gamma_{it}',
\end{aligned}$$

where

$$\begin{aligned}
\iota_{it} &:= \frac{\sigma_\varepsilon^2\sigma_e^2}{\alpha_i\sigma_e^2 + \sigma_\varepsilon^2} \\
\Gamma_{it} &:= \left[\frac{\sigma_e^2(\sigma_e^2 + \Gamma_{xt}'\Gamma_{xt})}{\delta_i - \Gamma_{xt}'\Gamma_{xt}}\right]^{1/2}\Gamma_{xt} = \left[\frac{\sigma_\varepsilon^2\sigma_e^2}{\alpha_i\sigma_e^2 + \sigma_\varepsilon^2}\right]^{1/2}\Gamma_{xt} = \iota_{it}^{1/2}\Gamma_{xt}.
\end{aligned}$$

$\square$

A.1.4. Proof of Proposition 3

Proof. **Part (i)**

First lets work with Γ_{it}, we have that for its j-th entry

$$\Gamma_{it,j} = \iota_{it}^{1/2}\Gamma_{xt,j} = y'_{jt}\Phi_{\Gamma,i} + \phi_{\Gamma,i},$$

with

$$\Phi_{\Gamma,i} = \iota_{it}^{1/2}\Phi_{\Gamma}$$
$$\phi_{\Gamma,i} = \iota_{it}^{1/2}\phi_{\Gamma}.$$

Then we work with $\mu_{it,j}$. Recall that $\mu_{it} = \mu_{r|s_i} + \frac{1}{2}diag(\Sigma_{it})$ and that the term $\mu_{r|s_i}$ is given by

$$\mu_{r|s_i} = \Sigma_{r|s_i}\left[(\sigma_{\varepsilon}^2)^{-1}\alpha_i s_{it} + \Sigma_{xt}^{-1}\mu_{xt}\right].$$

First we need to write $\mu_{r|s_i}$ as a function of μ_{xt}, Γ_{xt} and Γ_{it}. Notice that the term $\Sigma_{xt}^{-1}\mu_{xt}$ can be written as

$$\begin{aligned}
\Sigma_{xt}^{-1}\mu_{xt} &= [\sigma_e^2 I + \Gamma_{xt}\Gamma'_{xt}]^{-1}\mu_{xt} \\
&= \frac{1}{\sigma_e^2}\left[I - \frac{\Gamma_{xt}\Gamma'_{xt}}{\sigma_e^2 + \Gamma'_{xt}\Gamma_{xt}}\right]\mu_{xt}.
\end{aligned}$$

Next we can approximate s_{it} with $\alpha_i\mu_{xt}$ in order to write it as a component that depends on i and μ_{xt}. Here we are using that $E[s_{it}] = \alpha_i\mu_{xt}$. Then we can write

$$\begin{aligned}
\mu_{r|s_i} &= \Sigma_{r|s_i}\left[(\sigma_{\varepsilon}^2)^{-1}\alpha_i s_{it} + \Sigma_x^{-1}\mu_x\right] \\
&= [\iota_{it} I + \Gamma_{it}\Gamma'_{it}]\left[\left(\frac{\alpha_i^2}{\sigma_e^2} + \frac{1}{\sigma_e^2}\right)\mu_{xt} - \left(\frac{\Gamma'_{xt}\mu_{xt}}{\sigma_e^2 + \Gamma'_{xt}\Gamma_{xt}}\right)\Gamma_{xt}\right] \\
&= \kappa_{1i}\mu_{xt} + \kappa_{2i}\Gamma_{xt} + \kappa_{3i}\Gamma_{it},
\end{aligned}$$

with

$$\kappa_{1i} = \iota_{it}\left(\frac{\alpha_i^2}{\sigma_e^2} + \frac{1}{\sigma_e^2}\right)$$
$$\kappa_{2i} = -\iota_{it}\left(\frac{\Gamma'_{xt}\mu_{xt}}{\sigma_e^2 + \Gamma'_{xt}\Gamma_{xt}}\right)$$
$$\kappa_{3i} = \Gamma'_{it}\left[\left(\frac{\alpha_i^2}{\sigma_e^2} + \frac{1}{\sigma_e^2}\right)\mu_{xt} - \left(\frac{\Gamma'_{xt}\mu_{xt}}{\sigma_e^2 + \Gamma'_{xt}\Gamma_{xt}}\right)\Gamma_{xt}\right];$$

then we have that for asset j

$$\begin{aligned}\mu_{r|s_i,j} &= \kappa_{1i}\mu_{xt,j} + \kappa_{2i}\Gamma_{xt,j} + \kappa_{3i}\Gamma_{it,j} \\ &= \kappa_{1i}(y'_{jt}\Phi_\mu + \phi_\mu) + \kappa_{2i}(y'_{jt}\Phi_\Gamma + \phi_\Gamma) + \kappa_{3i}(y'_{jt}\Phi_{\Gamma,i} + \phi_{\Gamma,i}) \\ &= y'_{jt}\left[\kappa_{1i}\Phi_\mu + \kappa_{2i}\Phi_\Gamma + \kappa_{3i}\Phi_{\Gamma,i}\right] + \left[\kappa_{1i}\phi_\mu + \kappa_{2i}\phi_\Gamma + \kappa_{3i}\phi_{\Gamma,i}\right].\end{aligned}$$

Recall that $\sigma_{it}^2 = diag(\Sigma_{it})$, so $\sigma_{it,j}^2 = \Gamma_{it,j}^2 + \iota_{it}$. Since $\Gamma_{it,j}$ is a polynomial of degree M is clear that $\sigma_{it,j}^2$ is a polynomial of degree $2M$ on x_{jt}. To accommodate this we can define $\bar{y}_{jt}$ so it includes the degree combinations of $y_{jt} \otimes y_{jt}$, that is define

$$\bar{y}_{jt} = \begin{pmatrix} x_{jt} \\ x_{jt} \otimes x_{jt} \\ x_{jt} \otimes x_{jt} \otimes x_{jt} \\ \vdots \end{pmatrix},$$

with $\bar{y}_{jt}$ having dimension $K_{\bar{y}} = \sum_{m=1}^{2M} K_x^m$. Then we define $\Phi_{\bar{y}}$ so $(y'_{jt}\Phi_{\Gamma,i})^2 = \bar{y}'_{jt}\Phi_{\bar{y}}$. With this notation we can write

$$\begin{aligned}\sigma_{it,j}^2 &= \Gamma_{it,j}^2 + \iota_{it} = (y'_{jt}\Phi_{\Gamma,i} + \phi_{\Gamma,i})^2 + \iota_{it} \\ &= ((y'_{jt}\Phi_{\Gamma,i})^2) + y'_{jt}[2\phi_{\Gamma,i}\Phi_{\Gamma,i}] + [\phi_{\Gamma,i}^2 + \iota_{it}] \\ &= \bar{y}'_{jt}\Phi_{\bar{y}} + y'_{jt}[2\phi_{\Gamma,i}\Phi_{\Gamma,i}] + [\phi_{\Gamma,i}^2 + \iota_{it}].\end{aligned}$$

Notice that the term $\mu_{r|s_i,j}$ can be written as a polynomial on $\bar{y}_{jt}$; just pad with zeros the coefficients corresponding to powers of the elements in x_{jt} present $\bar{y}_{jt}$ but not in y_{jt}. Then it is possible to write

$$\begin{aligned}\mu_{it,j} &= \mu_{r|s_i,j} + \frac{1}{2}\sigma^2_{it,j}\\ &:= \bar{y}_{jt}\Phi_{\mu,i} + \phi_{\mu,i},\end{aligned}$$

with $\Phi_{\mu,i}$ exact configuration of padded zeros depends on the dimensions of x_{jt} and the degree M and $\phi_{\mu,i}$ is given by

$$\phi_{\mu,i} = \kappa_{1i}\phi_\mu + \kappa_{2i}\phi_\Gamma + \left(\kappa_{3i} + \frac{1}{2}\right)\phi_{\Gamma,i} + \frac{1}{2}\iota_{it}.$$

Part (ii)

Recall that the positive optimal portfolio weights are given by

$$w_{it} \approx \Sigma_{it}^{-1}\left(\mu_{it} - \lambda_{it}1 + C_t a_{it}\right),$$

using the expression for Σ_{it} we have

$$\begin{aligned}w_{it} &\approx [\iota_{it} + \Gamma_{it}\Gamma'_{it}]^{-1}(\mu_{it} - \lambda_{it}1 + C_t a_{it})\\ &= \frac{1}{\iota_{it}}\left[I - \frac{\Gamma_{it}\Gamma'_{it}}{\iota_{it} + \Gamma'_{it}\Gamma_{it}}\right](\mu_{it} - \lambda_{it}1 + C_t a_{it})\\ &= \left(\frac{1}{\iota_{it}}\right)\mu_{it} - \left(\frac{\lambda_{it}}{\gamma_{it}}\right)1 + C_t\left(\frac{a_{it}}{\iota_{it}}\right) + \kappa_{it}\Gamma_{it},\end{aligned}$$

with

$$\kappa_{it} = -\frac{\Gamma'_{it}(\mu_{it} - \lambda_{it}1 + C_t a_{it})}{\iota_{it} + \Gamma'_{it}\Gamma_{it}}.$$

Now define $\tilde{x}_{jt} = [x'_{jt} \quad c'_{jt}]'$ a $K_{\tilde{x}} = (K_x + K_c)$ vector and define the $K_{\tilde{y}}$ vector $\tilde{y}_{jt}$ with $K_{\tilde{y}} =$

$\sum_{m=1}^{2M}(K_x+K_c)^m$ and

$$\tilde{y}_{jt} = \begin{pmatrix} \tilde{x}_{jt} \\ \tilde{x}_{jt}\otimes\tilde{x}_{jt} \\ \tilde{x}_{jt}\otimes\tilde{x}_{jt}\otimes\tilde{x}_{jt} \\ \vdots \end{pmatrix},$$

next we show that each term in $w_{it,j}$ can be written as polynomial in $\tilde{y}_{jt}$. First

$$c'_{jt}\left(\frac{a_{it}}{\iota_{it}}\right) = \tilde{y}'_{jt}\begin{pmatrix} 0 \\ a_{it}/\iota_{it} \\ 0 \\ \vdots \end{pmatrix} = \tilde{y}'_{jt}\Phi_{C,i},$$

next

$$\begin{aligned}\left(\frac{1}{\iota_{it}}\right)\mu_{it,j} &= \left(\frac{1}{\iota_{it}}\right)\left(\bar{y}'_{jt}\Phi_{\mu,i}+\phi_{\mu,i}\right) \\ &= \bar{y}'_{jt}\left(\frac{1}{\iota_{it}}\Phi_{\mu,i}\right)+\frac{\phi_{\mu,i}}{\iota_{it}} \\ &= \tilde{y}_{jt}\tilde{\Phi}_{\mu,i}+\frac{\phi_{\mu,i}}{\iota_{it}},\end{aligned}$$

where $\tilde{\Phi}_{\mu,i}$ has zeros whenever a term with c_{jt} appears in $\tilde{y}_{jt}$. Finally,

$$\begin{aligned}\kappa_{it}\Gamma_{it,j} &= \kappa_{it}(y'_{jt}\Phi_{\Gamma,i}+\phi_{\Gamma,i}) \\ &= y'_{jt}(\kappa_{it}\Phi_{\Gamma,i})+\kappa_{it}\phi_{\Gamma,i} \\ &= \tilde{y}'_{jt}\tilde{\Phi}_{\Gamma,i}+\kappa_{it}\phi_{\Gamma,i},\end{aligned}$$

and once again $\tilde{\Phi}_{\Gamma,i}$ has zeros whenever there is term in $\tilde{y}_{jt}$ with c_{jt} or has a power of x_{jt} not present in y_{jt}. Collecting terms we have that

$$w_{it,j} \approx \tilde{y}'_{jt}\Phi_{w,it}+\phi_{w,it}$$

with

$$\Phi_{w,it} = \tilde{\Phi}_{\mu,i} + \Phi_{C,i} + \tilde{\Phi}_{\Gamma,i}$$
$$\phi_{w,it} = \frac{\phi_{\mu,i}}{\iota_{it}} + \kappa_{it}\phi_{\Gamma,i} - \lambda_{it}.$$

Part (iii)

Restricting the parameters so $\phi_{w,it} = w_{it0}$ and $\Phi_{w,it}/w_{it0} = [\beta_{it} \quad 1/2vec(\beta_{it}\beta_{it}') \quad \cdots]'$ then

$$\begin{aligned}\frac{w_{ij,t}}{w_{it,0}} &\approx 1 + \tilde{y}_{jt}'\frac{\Phi_{w,it}}{w_{it0}} = 1 + \tilde{x}_{jt}'\beta_{it} + \frac{1}{2}vect(\tilde{x}_{jt}\tilde{x}_{jt}')'vec(\beta_{jt}\beta_{jt}') + \cdots \\ &= \sum_{m=1}^{M}\frac{(\tilde{x}_{jt}'\beta_{it})^m}{m!} \longrightarrow \exp\left[\tilde{x}_{jt}'\beta_{it}\right], \quad \text{as } m \to \infty.\end{aligned}$$

Writing $\beta_{it}' = [1 \quad b_{it}' \quad \gamma_{it}']$ and assuming the first characteristics in x_{jt} is unobserved and denote by ξ_{jt} then we have that

$$\frac{w_{it,j}}{w_{it,0}} \approx \exp\left(\xi_{jt} + x_{jt}'b_{it} + c_{jt}'\gamma_{it}\right).$$

Finally because $w_{it,0} + \sum_{j=1}^{J_t} w_{it,j} = 1$, then $1 + \sum_{j=1}^{J_t} w_{it,j}/w_{it,0} = 1/w_{it,0}$ and

$$w_{it,j} \approx \frac{\exp\left(x_{jt}'b_{it} + c_{jt}'\gamma_{it} + \xi_{jt}\right)}{1 + \sum_{k=1}^{J_t}\exp\left(x_{kt}'b_{it} + c_{kt}'\gamma_{it} + \xi_{kt}\right)},$$

and the weight for the outside option is given by

$$w_{it,0} \approx \frac{1}{1 + \sum_{k=1}^{J_t}\exp\left(x_{kt}'b_{it} + c_{kt}'\gamma_{it} + \xi_{kt}\right)}.$$

□

A.1.5. Proof of Proposition 4

We start by stating Berry's Inversion theorem for demand systems. See Berry (1994) for a full proof. Then the proof consists in verying the conditions of the theorem for the demand system in (1.25).

Berry's Inversion Theorem Consider the metric space $(\mathbb{R}^K, d)$ with $d(x,y) = ||x-y||$ and $||\cdot||$ denoting the sup-norm. Let $f : \mathbb{R}^K \to \mathbb{R}^K$ satisfy:

i. $\forall x \in \mathbb{R}^K$, $f(x)$ is continuously differentiable such that for any j and k:

$$\frac{\partial f_j(x)}{\partial x_k} \geq 0$$
$$\sum_{k=1}^{K} \frac{\partial f_j(x)}{\partial x_k} < 1.$$

ii. $\min_j \inf_x f(x) := \underline{x} > -\infty$

iii. There is a value $\bar{x}$ with the property that if for any j $x_j \geq \bar{x}$ then for some k (not necessarily equal j) $f_k(x) < x_k$.

Then there is a unique fixed point $x_0 \in \mathbb{R}^K$ to f. Moreover, let $\mathcal{X} := [\underline{x}, \bar{x}]^K$ and define the truncated function $\hat{f}_j(x) = \min\{f_j(x), \bar{x}\}$. Then $\hat{f}(x)$ is a contraction of modulus less than one on $\mathcal{X}$.

Proof. Verying the conditions of Berry's Inversion Theorem
Denote by θ the parameters to estimate in market t; s_t the vector of observed aggregate share, and $\tilde{s}_t(\delta_t, \theta_2; d_t, X_t, J_t)$ the vector of model-implied shares. Here the operator $f : \mathbb{R}^{J_t} \to \mathbb{R}^{J_t}$ for which we look a fixed point is given by:

$$f(\delta) = \delta + \log s_t - \log \tilde{s}_t(\delta, \theta_2; d_t, X_t, J_t).$$

On this operator we check the conditions for Berry's inversion. On the following we drop the t index and denote the model implied market share for asset j as $\tilde{s}_j := \tilde{s}_j(\delta, \theta_2; d_t, X_t, J_t)$.

- **Checking i.**

We start by verifying that the first derivatives are non-negative, we have that:

$$\frac{\partial f_j(\delta)}{\partial \delta_k} = \begin{cases} 1 - \frac{1}{\tilde{s}_j}\frac{\partial \tilde{s}_j}{\partial \delta_j} & \text{if } k = j \\ \\ -\frac{1}{\tilde{s}_j}\frac{\partial \tilde{s}_j}{\partial \delta_k} & \text{if } k \neq j. \end{cases}$$

We know that $\tilde{s}_j \geq 0$ and using the definition for the model-implied shares we have that

$$\frac{\partial \tilde{s}_j}{\partial \delta_k} = \frac{\partial}{\partial \delta_k}\left[\sum_{i \in I}\left(\frac{A_i}{A}\right)\int w_{ij}(v_i)dF_v(v_i)\right] = \sum_{i \in I}\left(\frac{A_i}{A}\right)\int \frac{\partial w_{ij}(v_i)}{\partial \delta_k}dF_v(v_i),$$

where

$$\frac{\partial w_{ij}(v_i)}{\partial \delta_k} = \begin{cases} w_{ij}(v_i)(1 - w_{ij}(v_i)) & \text{if } k = j \\ \\ -w_{ik}(v_i)w_{ij}(v_i) & \text{if } k \neq j. \end{cases}$$

From the previous we see that if $k \neq j$ then $\frac{\partial w_{ij}}{\partial \delta_k} \leq 0$ and since $(A_i/A) > 0$ then $\frac{\partial \tilde{s}_j}{\partial \delta_k} \leq 0$. For $k = j$ notice that $\frac{\partial w_{ij}}{\partial \delta_k} \leq w_{ij}$ and then $\frac{\partial \tilde{s}_j}{\partial \delta_k} \leq \tilde{s}_j$ so $\frac{\partial \tilde{s}_j}{\partial \delta_k}$. The next step is to verify that the sum of partial derivatives is less than 1. Notice that:

$$\begin{aligned} \sum_{k=1}^{J_t}\frac{\partial f_j(\delta)}{\partial \delta_k} &= \frac{\partial f_j(\delta)}{\partial \delta_j} + \sum_{k \neq j}\frac{\partial f_j(\delta)}{\partial \delta_k} = 1 - \frac{1}{\tilde{s}_j}\frac{\partial \tilde{s}_j}{\partial \delta_j} + \sum_{k \neq j}\left(-\frac{1}{\tilde{s}_j}\frac{\partial \tilde{s}_j}{\partial \delta_k}\right) \\ &= 1 - \frac{1}{\tilde{s}_j}\left[\sum_{k=1}^{J_t}\frac{\partial \tilde{s}_j}{\partial \delta_k}\right] = 1 - \frac{1}{\tilde{s}_j}\left[\sum_{k=1}^{J_t}\sum_{i \in I}\left(\frac{A_i}{A}\right)\int \frac{\partial w_{ij}(v_i)}{\partial \delta_k}dF_v(v_i)\right] \\ &= 1 - \frac{1}{\tilde{s}_j}\left[\sum_{i \in I}\left(\frac{A_i}{A}\right)\sum_{k=1}^{J_t}\int \frac{\partial w_{ij}(v_i)}{\partial \delta_k}dF_v(v_i)\right]. \end{aligned}$$

Given the definition of $\tilde{s}_j$ it is sufficient to show that $\sum_{k=1}^{J_t}\int \frac{\partial w_{ij}(v_i)}{\partial \delta_k}dF_v(v_i) < \int w_{ij}(v_i)dF_v(v_i)$.

For this we notice that

$$\begin{aligned}\sum_{k=1}^{J_t} \frac{\partial w_{ij}(v_i)}{\partial \delta_k} &= \frac{\partial w_{ij}(v_i)}{\partial \delta_j} + \sum_{k \neq j} \frac{\partial w_{ij}(v_i)}{\partial \delta_k} \\ &= w_{ij}(v_i)(1 - w_{ij}(v_i)) - \sum_{k \neq j} w_{ij}(v_i) w_{ik}(v_i) \\ &= w_{ij}(v_i) \left[1 - \sum_{k \neq j} w_{ik}(v_i) \right] \leq w_{ij}(v_i).\end{aligned}$$

- **Checking ii.**

  There first step is to rewrite the model-implied shares as

$$\tilde{s}_j = \sum_{i \in I} \left(\frac{A_i}{A} \right) \int w_{ij}(v_i) dF_v(v_i) = \exp(\delta_j) \sum_{i \in I} \left(\frac{A_i}{A} \right) D_{ij}(\delta)$$
$$\text{with } D_{ij}(\delta) = \int \frac{\exp(h_{ij}(v_i))}{1 + \sum_{j=1}^{J} \exp(\delta_j + h_{ij}(v_i))} dF_v(v_i)$$

  This implies that $\ln(\tilde{s}_j) = \delta_j + \ln \left(\sum_{i \in I} (A_i/A) D_{ij}(\delta) \right)$ and that

$$f(\delta)_j = \ln(s_j) - \ln \left(\sum_{i \in I} (A_i/A) D_{ij}(\delta) \right).$$

  Now notice that when $\delta_m \to -\infty$ for $m \neq j$ then the term $D_{ij}(\delta)$ tends to $\int \exp(h_{ij}(v_i)) dF_v(v_i)$ so a lower bound for $f(\delta)_j$ is

$$\underline{\delta}_j > \ln(s_j) - \ln \left[\sum_{i \in I} (A_i/A) \int \exp(h_{ij}(v_i)) dF_v(v_i) \right],$$

  so condition ii. is satisfied with $\underline{\delta} := \min_j \underline{\delta}_j$.

- **Checking iii.**

  For this part set $\delta_k = -\infty$ for $k \neq j$ and defined $\bar{\delta}_j$ as the value of δ_j such that $\tilde{s}_0(\delta, \theta_2) = s_0$, that is the value of δ_j that along with $\delta_k = -\infty$ would match the observed shares for the outside good. Moreover let $\bar{\delta} > max_j \bar{\delta}_j$.

If δ is such that $\exists j$ with $\delta_j > \bar{\delta}$ then $\tilde{s}_0(\delta) < s_0$ and hence $\sum_{k=1}^{I_t} \tilde{s}(\delta)_k > \sum_{k=1}^{J_t} s_k$ which means that there is a least one element k such that $\tilde{s}(\delta)_k > s_k$. For such k we have that $f(\delta)_k < \delta_k$ as required in part iii.

□

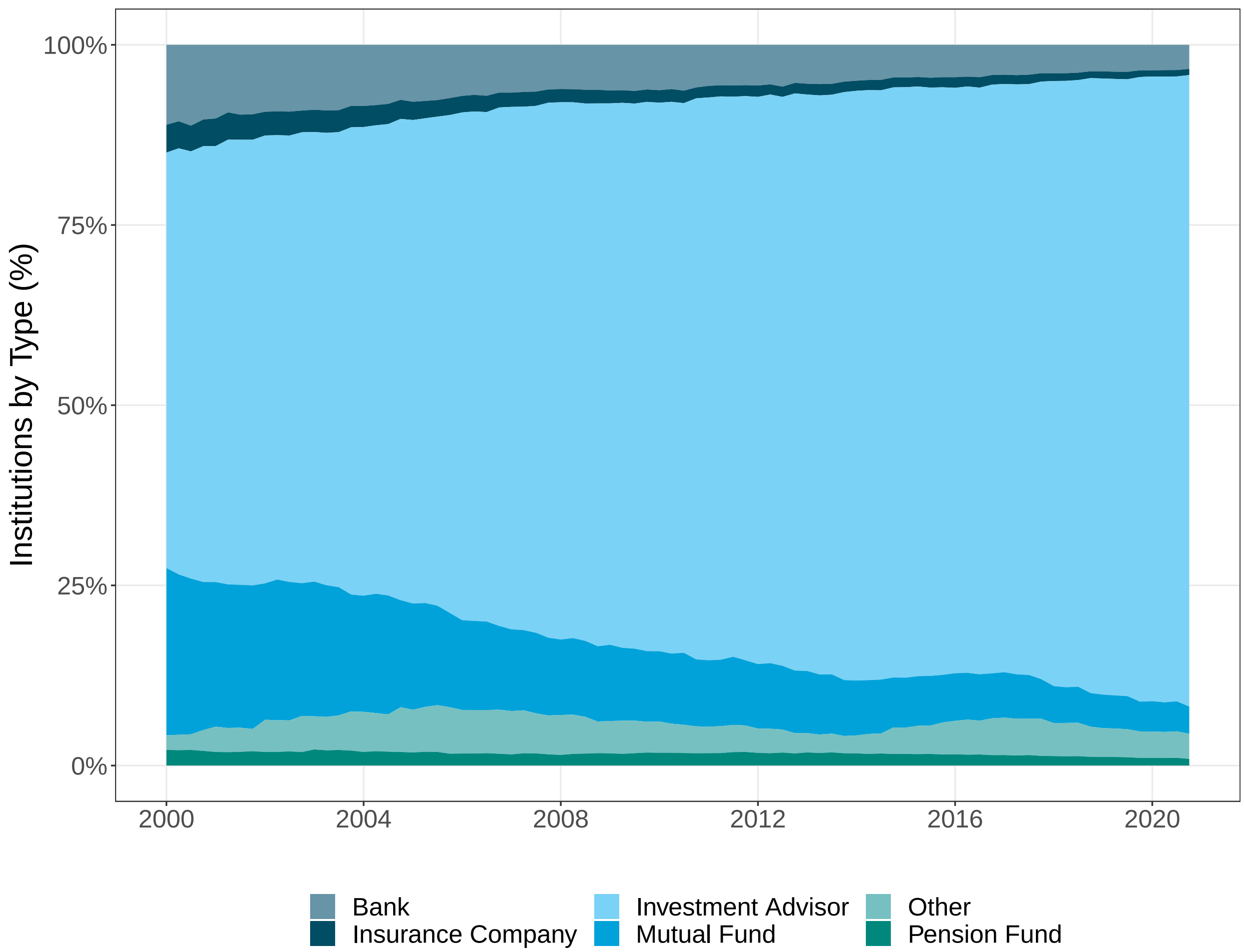
A.2. Data Appendix

Table A.1: Summary Statistics for Stock Characteristics

Variable	N. Stocks	Mean	Median	Std. Dev.	Min	Pc. 25th	Pc. 75th	Max
2000-2004								
Market Beta	499	0.684	0.639	0.554	-0.238	0.302	0.960	3.974
log Market Equity	499	8.009	7.985	1.900	0.598	6.897	9.175	13.256
log Total Assets	499	8.404	8.572	1.858	1.516	7.348	9.751	13.455
log B-M Equity	499	-0.408	-0.360	0.923	-5.521	-0.968	0.206	3.653
Profitability	499	0.211	0.192	0.227	-1.752	0.104	0.319	0.785
Investment	499	0.073	0.051	0.193	-0.811	-0.016	0.132	1.305
E-score	499	0.059	0.000	0.238	-0.500	-0.125	0.333	0.500
2005-2009								
Market Beta	696	1.208	1.099	0.713	-0.163	0.683	1.587	3.974
log Market Equity	696	8.038	7.981	1.804	0.923	6.884	9.299	13.149
log Total Assets	696	8.298	8.261	1.736	2.513	7.159	9.537	14.673
log B-M Equity	696	-0.482	-0.482	0.916	-5.659	-1.073	0.086	4.716
Profitability	696	0.223	0.211	0.251	-2.381	0.121	0.333	0.941
Investment	696	0.072	0.055	0.200	-0.741	-0.015	0.139	0.917
E-score	696	0.048	0.000	0.173	-0.500	0.000	0.200	0.500
2010-2014								
Market Beta	1122	1.294	1.223	0.646	0.009	0.812	1.678	3.182
log Market Equity	1122	8.460	8.558	1.801	1.169	7.345	9.657	13.374
log Total Assets	1122	8.665	8.650	1.869	1.534	7.539	9.871	14.697
log B-M Equity	1122	-0.599	-0.593	0.936	-7.699	-1.150	-0.002	4.218
Profitability	1122	0.227	0.209	0.241	-2.098	0.120	0.319	0.878
Investment	1122	0.062	0.046	0.151	-0.691	-0.007	0.114	0.816
E-score	1122	0.020	0.000	0.168	-0.500	0.000	0.125	0.500
2015-2019								
Market Beta	955	1.150	1.139	0.569	-0.396	0.779	1.487	3.184
log Market Equity	955	8.910	9.064	1.850	1.054	7.632	10.203	14.068
log Total Assets	955	9.003	8.949	1.903	1.534	7.803	10.263	14.780
log B-M Equity	955	-0.842	-0.794	1.061	-9.991	-1.441	-0.160	4.943
Profitability	955	0.261	0.223	0.292	-2.880	0.127	0.359	1.034
Investment	955	0.053	0.033	0.173	-0.633	-0.021	0.095	0.911
E-score	955	0.047	0.000	0.141	-0.500	0.000	0.125	0.500

Notes: Summary statistics for the stock's characteristics used during estimation. Statistics are compute over pooled quarterly observations of the variables every five years. Summary Statistics present mean, median, standard deviation, minimum, 25th percentile, 75th percentile and maximum.

Figure A.1: Types of Institutional Investor



Notes: Evolution of institutional investors by type from 2000q1 to 2020q4. The classification of institutional investors follows the six categories as in Koijen and Yogo (2019).

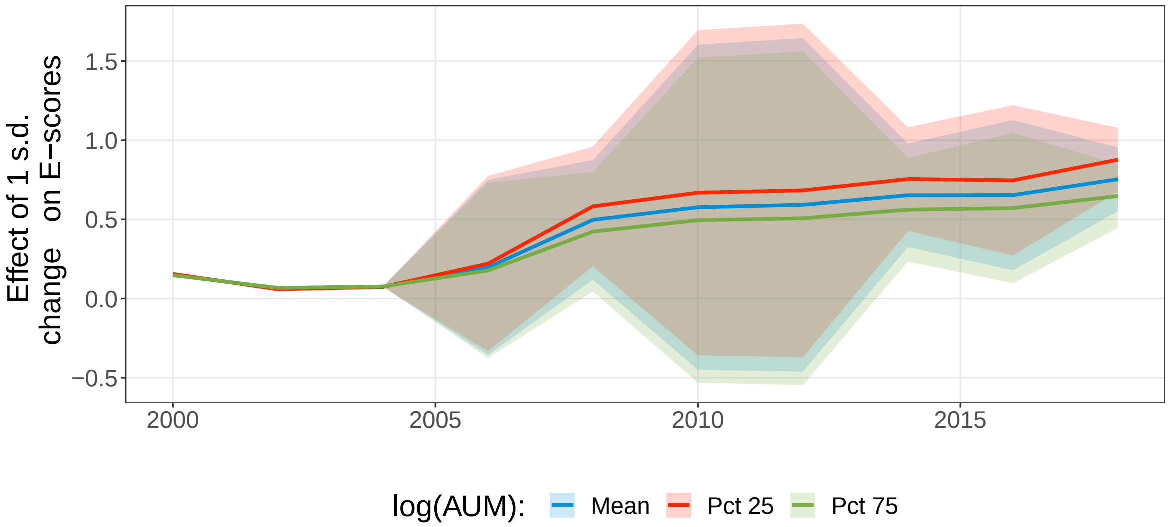
Table A.2: Environmental Indicators from MSCI

Positive Indicators	Negative Indicators
Environmental Opportunities	Hazardous Waste
Waste Management	Regulatory Compliance
Packaging Materials & Waste	Ozone Depleting Chemicals
Climate Change	Toxic Spills & Releases
Environmental Management Systems	Agriculture Chemicals
Water Stress	Climate Change
Biodiversity & Land Use	Impact of Products & Services
Raw Material Sourcing	Biodiversity & Land Use
Natural Resource Use	Operational Waste
Environmental Opportunities - Green Buildings	Supply Chain Management
Environmental Opportunities in Renewable Energy	Water Management
Waste Management - Electronic Waste	Other Concerns
Climate Change - Product Carbon Footprint	
Climate Change - Insuring Climage Change Risk	
Other Strengths	

Notes: List of environmental performance indicators in the MSCI dataset. Each indicator is a dummy variable. The threshold for satisfying an indicator are determined by MSCI and are not disclosed with the data.

A.3. Complementary Results

Figure A.2: Estimated coefficients for E-scores



Notes: This plot shows the effective coefficient on E-scores, $\gamma_{it} = \gamma_0 + \kappa \log(\text{AUM})_{it} + \sigma v_{it}$, over 2-year estimation windows ranging from 2000-Q1 to 2019-Q4. The plot uses varios values, in each window, of log assets under management and shows the 95% confidence interval of how the coefficient on E-scores varies across investors based on the assumption that unobserved demographics, v_{it}, are normally distributed.

Table A.3: Counterfactual holdings and price changes of E-score sorted portfolios

	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4	Portfolio 5
N. Stocks	104	104	104	103	103
ME (USD Bill.)	2847	2115	1814	5789	5356
Agg. Port. Share (%)	9.310	6.917	5.932	18.930	17.515
	Counterfactual Agg. Port Shares (%)				
Logit	9.331	6.926	5.940	18.919	17.476
Mixed Logit	9.417	6.985	5.991	18.869	17.233
	Counterfactual Price change (%)				
Logit	0.222	0.138	0.138	-0.059	-0.220
Mixed Logit	1.142	0.988	0.988	-0.319	-1.611
	Price Change (%)				
2019 Q1 -2019-Q2	-1.046	3.496	2.606	2.950	4.734

Notes: This table shows the effect of a ban of green investing for pension funds on aggregate holdings and equity prices in a counterfactual exercise using data and estimates for 2019-Q1. The portfolios were constructed by sorting stocks by their E-score, and grouping them according to quintiles. Portfolio 1 contains the 20% of stocks with lowest E-scores, while Portfolio 5 contains the 20% of stocks with highest E-scores. The table shows the counterfactual changes in aggregate portfolio holdings and value-weighted prices changes according to a logit demand specification and to a random coefficients (RC) demand specification. The observed value-weighted average price change between 2019-Q1 and 2019-Q2 for each portfolio is also shown.

APPENDIX B

Chapter 2 Appendices

B.1. Computation details for the ELBO

The objective in the E-step of the VEM algorithm is to find the values $\{\gamma_t^*, \phi_{t,1:N_t}^*\}$ for each document $t = 1, ..., T$ and $\{\lambda_k^*\}$ for each topic $k = 1, ..., K$ that maximize the corpus-level ELBO. To do such maximization we start by providing detailed expressions of the ELBO at each document:

$$\mathcal{L}(\gamma, \phi_{1:N}, \lambda_{1:K}) = E_q[\log p(\theta|\alpha)] + \sum_{k=1}^{K} E_q[\log p(\beta_k|\eta)] + \sum_{n=1}^{N} E_q[\log p(z_n|\theta)] \\ + \sum_{n=1}^{N} E_q[\log p(w_n|z_n, \beta_{1:K})] + E_q[\log p(y|z_{1:N}, W, b, c, \sigma^2)] + H(q).$$

B.1.1. Components of $\mathcal{L}(\gamma, \phi_{1:N}, \lambda_{1:K})$

1. $E_q[\log p(\theta|\alpha)]$: Since $\theta|\alpha \sim Dir(\alpha)$ we have that $p(\theta|\alpha) = \left[\frac{\prod_{k=1}^{K}\Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^{K}\alpha_k)}\right]\prod_{k=1}^{K}\theta_k^{\alpha_k - 1}$ then:

$$\log p(\theta|\alpha) = \log\Gamma\left(\sum_{k=1}^{K}\alpha_k\right) - \sum_{k=1}^{K}\log\Gamma(\alpha_k) + \sum_{k=1}^{K}(\alpha_k - 1)\log\theta_k$$
$$\text{and } E_q[\log p(\theta|\alpha)] = \log\Gamma\left(\sum_{k=1}^{K}\alpha_k\right) - \sum_{k=1}^{K}\log\Gamma(\alpha_k) + \sum_{k=1}^{K}(\alpha_k - 1)E_q[\log\theta_k],$$

   where $E_q[\log\theta_k] = \psi(\gamma_k) - \psi\left(\sum_{k=1}^{K}\gamma_k\right)$ and $\psi(\cdot)$ stands for the digamma function.

2. $\sum_{k=1}^{K} E_q[\log p(\beta_k|\eta)]$: Since $\beta_k|\eta \sim Dir(\eta)$ we have that $p(\beta_k|\eta) = \left[\frac{\prod_{v=1}^{V}\Gamma(\eta_v)}{\Gamma(\sum_{v=1}^{V}\eta_v)}\right]\prod_{v=1}^{V}\beta_{k,v}^{\eta-1}$ then (similar to computation for $\theta|\alpha$):

$$E_q[\log p(\beta_k|\eta)] = \log\Gamma\left(\sum_{v=1}^{V}\eta_v\right) - \sum_{v=1}^{V}\log\Gamma(\eta_v) + \sum_{v=1}^{V}(\eta_v - 1)E_q[\log\beta_{k,v}]$$
$$\text{and } \sum_{k=1}^{K} E_q[\log p(\beta_k|\eta)] = K\log\Gamma\left(\sum_{v=1}^{V}\eta_v\right) - K\sum_{v=1}^{V}\log\Gamma(\eta_v) + \sum_{k=1}^{K}\sum_{v=1}^{V}(\eta_v - 1)E_q[\log\beta_{k,v}],$$

with $E_q[\log \beta_{k,v}] = \psi(\lambda_{k,v}) - \psi\left(\sum_{v=1}^{V} \lambda_{k,v}\right)$.

3. $\sum_{n=1}^{N} E_q[\log p(z_n|\theta)]$: We assumed that $z_n|\theta \sim Mult(\theta)$. Coding z_n in one-in-K form we can write $p(z_n|\theta) = \prod_{k=1}^{K} \theta_k^{z_{n,k}}$, then $\log p(z_n|\theta) = \sum_{k=1}^{K} z_{n,k} \log \theta_k$ and

$$E_q[\log p(z_n|\theta)] = \sum_{k=1}^{K} E_q[z_{n,k}] E_q[\log \theta_k] = \sum_{k=1}^{K} \phi_{n,k} E_q[\log \theta_k]$$
$$\text{and hence} \sum_{n=1}^{N} E_q[\log p(z_n|\theta)] = \sum_{n=1}^{N} \sum_{k=1}^{K} \phi_{n,k} E_q[\log \theta_k].$$

4. $\sum_{n=1}^{N} E_q[\log p(w_n|z_n, \beta_{1:K})]$: Recall that $w_n \in \{1, ..., V\}$ and that $\beta_{1:K}$ is $V \times K$ matrix so we can write $p(w_n|z_n, \beta_{1:K}) = \prod_{k=1}^{K} \beta_{w_n,k}^{z_{n,k}}$ so we have:

$$E_q[\log p(w_n|z_n, \beta_{1:K})] = \sum_{k=1}^{K} \phi_{n,k} E_q[\log \beta_{w_n,k}]$$
$$\text{and} \sum_{n=1}^{N} E_q[\log p(w_n|z_n, \beta_{1:K})] = \sum_{n=1}^{N} \sum_{k=1}^{K} \phi_{n,k} E_q[\log \beta_{k,w_n}].$$

5. $E_q[\log p(y|z_{1:N}, W, b, c, \sigma^2)]$: Given the gaussian response for $y = b'\bar{Z} + c'W + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2)$ we have:

$$p(y|z_{1:N}, W, b, c, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \bar{Z}'b - c'W)^2}{2\sigma^2}\right)$$
$$\Rightarrow \log p(y|z_{1:N}, W, b, c, \sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - c'W - \bar{Z}'b)^2$$

Denoting $\tilde{y} = y - c'W$, we have:

$$E_q[\log p(y|z_{1:N}, W, b, c, \sigma^2)] = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[\tilde{y}^2 - 2\tilde{y}E_q[\bar{Z}]'b + b'E_q[\bar{Z}\bar{Z}']b\right]$$

with

$$
\begin{aligned}
E_q[\bar{Z}] &= \frac{1}{N}\sum_{n=1}^{N}\phi_n \\
E_q[\bar{Z}\bar{Z}'] &= \frac{1}{N^2}\left(\sum_{n=1}^{N}\phi_n\phi'_{-n} + \sum_{n=1}^{N}\mathrm{diag}(\phi_n)\right) \\
\phi_{-n} &= \sum_{m\neq n}\phi_m.
\end{aligned}
$$

6. $H(q)$: the entropy of the variational distribution is given by

$$
\begin{aligned}
H(q) = &-\log\Gamma\left(\sum_{k=1}^{K}\gamma_k\right) + \sum_{k=1}^{K}\log\Gamma(\gamma_k) - \sum_{k=1}^{K}(\gamma_k - 1)E_q[\log\theta_k] \\
&-\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{n,k}\log(\phi_{n,k}) \\
&-\sum_{k=1}^{K}\log\Gamma\left(\sum_{v=1}^{V}\lambda_{k,v}\right) + \sum_{k=1}^{K}\sum_{v=1}^{V}\log\Gamma(\lambda_{k,v}) - \sum_{k=1}^{K}\sum_{v=1}^{V}(\lambda_{k,v}-1)E_q[\log\beta_{k,v}].
\end{aligned}
$$

B.1.2. Maximization of $\mathcal{L}(\gamma, \phi_{1:N}, \lambda_{1:K})$

The optimization is done by block coordinate-ascent variational inference, that means iteratively maximizing $\mathcal{L}$ with respect to each variational parameter. In this maximization we need to include the constraint that each of the multinomial parameters ϕ_n should sum up to one.

First we obtain the partial derivative of $\mathcal{L}$ with respect to γ_k, for $k = 1, ..., K$ we have:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \gamma_k} &= \frac{\partial}{\partial \gamma_i} E_q[\log p(\theta|\alpha)] + \frac{\partial}{\partial \gamma_k} \sum_{n=1}^{N} E_q[\log p(z_n|\theta)] + \frac{\partial}{\partial \gamma_k} H(q) \\
&= (\alpha_k - 1) \frac{\partial E[\log \theta_k]}{\partial \gamma_k} + \sum_{n=1}^{N} \phi_{n,k} \frac{\partial E[\log \theta_k]}{\partial \gamma_k} - \psi \left(\sum_{k=1}^{K} \gamma_k \right) \\
&+ \psi(\gamma_k) - E_q[\log \theta_k] - (\gamma_k - 1) \frac{\partial E[\log \theta_k]}{\partial \gamma_k} \\
&= \left[\alpha_k - \gamma_k + \sum_{n=1}^{N} \phi_{n,k} \right] \frac{\partial E[\log \theta_k]}{\partial \gamma_k} \\
\Rightarrow \frac{\partial \mathcal{L}}{\partial \gamma_k} &= 0 \Leftrightarrow \gamma_k = \alpha_k + \sum_{n=1}^{N} \phi_{n,k}.
\end{aligned}$$

Then we can update the whole vector:

$$\gamma^{new} \leftarrow \alpha + \sum_{n=1}^{N} \phi_n, \tag{A1}$$

notice that this step is exactly the same as in the LDA VEM update. Next we obtain the partial derivative with respect to $[\phi_{n,k}]$ for $k = 1, ..., K$ and $n = 1, ..., N$:

$$\frac{\partial \mathcal{L}}{\partial \phi_{n,k}} = E_q[\log \theta_k] + E_q[\log \beta_{w_n,k}] - \log \phi_{n,k} - 1 + \lambda + \frac{\partial}{\partial \phi_{n,k}} E_q[\log p(y|z_{1:N}, b, c, \sigma^2)],$$

where λ is the Lagrange multiplier for the sum up to one constraint. The last partial derivative requires more work:

$$\frac{\partial}{\partial \phi_{n,k}} E_q[\log p(y|z_{1:N}, b, c, \sigma^2)] = \frac{(y - c'W) b_k}{N\sigma^2} - \frac{1}{2\sigma^2} \frac{\partial}{\partial \phi_{n,k}} \left(b' E_q[\bar{Z}\bar{Z}'] b \right).$$

Let $\phi_{-n} := \sum_{m \neq n} \phi_m$; the trick here is to express $b' E_q[\bar{Z}\bar{Z}']b$ as a function of just ϕ_n:

$$\begin{aligned} f(\phi_n) &= \frac{1}{N^2} b'[\phi_n \phi'_{-n} + \phi_{-n}\phi'_n + diag(\phi_n)]b + \text{const.} \\ &= \frac{1}{N^2}[2(b'\phi_{-n})b'\phi_n + (b \cdot b)'\phi_n] + \text{const.} \\ \Rightarrow \frac{\partial}{\partial \phi_{n,k}} \left(b' E_q[\bar{Z}\bar{Z}']b \right) &= \frac{1}{N^2}[2(b'\phi_{-n})b + b \cdot b]. \end{aligned}$$

Substituting this and noticing that λ acts as a normalization constant then:

$$\frac{\partial \mathcal{L}}{\partial \phi_{n,k}} = 0 \Leftrightarrow \phi_{n,k} \propto \exp\left[E_q[\log \theta_k] + E_q[\log \beta_{w_n,k}] + \frac{(y - c'W)b_k}{N\sigma^2} - \frac{1}{2\sigma^2 N^2}[2(b'_k \phi_{-n})b_k + b_k \cdot b_k] \right].$$

and updating the whole vector is done by:

$$\phi_n^{new} \propto \exp\left[E_q[\log \theta] + E_q[\log \beta_{w_n,\cdot}] + \frac{(y - c'W)b}{N\sigma^2} - \frac{1}{2\sigma^2 N^2}[2(b'\phi_{-n})b + b \cdot b] \right], \qquad \text{(A2)}$$

here the proportional stands for compute each entry and then normalize to one. Finally the derivative with respect to $[\lambda_{k,v}]$ is given by:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_{k,v}} &= \frac{\partial}{\partial \lambda_{k,v}} \sum_{k=1}^{K} E_q[\log(\beta_k | \eta)] + \frac{\partial}{\partial \lambda_{k,v}} \sum_{n=1}^{N} E_q[\log p(w_n | z_n, \beta_{1:K})] + \frac{\partial H(q)}{\partial \lambda_{k,v}} \\ &= \left[\eta_v - \lambda_{k,v} + \sum_{n=1}^{N} \mathbb{I}\{w_n = v\}\phi_{n,k} \right] \frac{\partial}{\partial \lambda_{k,v}} E_q[\log \beta_{k,v}], \end{aligned}$$

so accounting for commonality of $\lambda_{k,v}$ among all documents and equating the derivative to zero yields:

$$\lambda_{k,v} = \eta_v + \sum_{t=1}^{T} \sum_{n=1}^{N_t} \mathbb{I}\{w_{t,n} = v\}\phi_{t,n,k},$$

and updating the vector is done by:

$$\lambda_k = \eta + \sum_{t=1}^{T} \sum_{n=1}^{N_t} \phi_{t,n,k} \omega_{t,n} \tag{A3}$$

with $\omega_{t,n}$ a one-hot representation of $w_{t,n}$ as a vector of size V.

B.1.3. Maximization of $L(b, c, \sigma^2)$

Once we have maximizers $\{\gamma_t^*, \phi_{t,1:N_t}^*\}$ of the ELBO for each given document-response and $\{\lambda_{1:K}^*\}$ for the topics, we are ready to update the global parameters $\pi = \{b, c, \sigma^2\}$ in the M-step by maximizing $L(b, c, \sigma^2)$, the corpus-level ELBO. In this step we want to maximize over π the objective function:

$$L(b, c, \sigma^2) = \sum_{t=1}^{T} \mathcal{L}(\gamma_t^*, \phi_{t,1:N}^*, \lambda_{1:K}^* | b, c, \sigma^2).$$

Here each term in the sum is the $(w_{t,1:N_t}, y_t, W_t)$ specific ELBO evaluated at the corresponding maximizing variational parameters from the E-step. Optimizing with respect to b we have:

$$\begin{aligned}
\frac{\partial L}{\partial b} &= \sum_{t=1}^{T} \frac{\partial}{\partial b} E_{q^*} \left[\log p(y_t | z_{t,1:N_t}, b, c, \sigma^2)\right] \\
&= \sum_{t=1}^{T} \frac{1}{\sigma^2} \left[(y_t - c'W_t) E_{q^*}[\bar{Z}_t]' - b' E_{q^*} \left[\bar{Z}_t \bar{Z}_t'\right]\right] = 0 \\
\Rightarrow b^{new} &= \left(\sum_{t=1}^{T} E_{q^*} \left[\bar{Z}_t \bar{Z}_t'\right]\right)^{-1} \left(\sum_{t=1}^{T} E_{q^*} \left[\bar{Z}_t\right] (y_t - c'W_t)\right).
\end{aligned}$$

Next, optimizing with respect to c:

$$\begin{aligned}
\frac{\partial L}{\partial c} &= \sum_{t=1}^{T} \frac{\partial}{\partial c} E_{q^*} \left[\log p(y_t | z_{t,1:N_t}, b, c, \sigma^2)\right] \\
&= \sum_{t=1}^{T} \frac{1}{\sigma^2} \left[(y_t - E_{q^*}[\bar{Z}_d]' b) W_t' - c' W_t W_t'\right] = 0 \\
\Rightarrow c^{new} &= \left(\sum_{t=1}^{T} W_t W_t'\right)^{-1} \left(\sum_{t=1}^{T} W_t (y_t - E_{q^*}[\bar{Z}_d]' b^{new})\right).
\end{aligned}$$

Finally the response dispersion coefficient σ^2 is optimized by:

$$\frac{\partial L}{\partial \sigma^2} = \sum_{t=1}^{T} \frac{\partial}{\partial \sigma^2} E_{q^*}[\log p(y_t | z_{t,1:N_t}, b, \sigma^2)] \tag{A4}$$

$$= \sum_{t=1}^{T} -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} E_{q^*}[(y_t - \bar{Z}_t' b^{new} - W_t' c^{new})^2] = 0$$

$$\Rightarrow \sigma^{2,new} = \frac{1}{T} \sum_{t=1}^{T} \left(y_t^2 - y_t E_{q^*}[\bar{Z}_t'] b^{new} - y_t W_t' c^{new} \right). \tag{A5}$$

Table B.1: VEM Algorithm Execution Parameters

	Single-Stage	Two-Stage
Max VEM iterations	100	100
Max ELBO iterations by document	25	25
VEM-Corpus Tolerance	10^{-5}	10^{-5}
ELBO-Document Tolerance	10^{-5}	10^{-5}
Documents in the training sample	900	900
Documents in the test sample for prediction	100	100

Notes: Execution parameters used for both versions of the VEM Algorithm on the simulation study. See text for details.

B.2. Additional Details of the Monte Carlo Simulation

In the VEM algorithm for supervised topic extraction there are four key parameters during implementation: (i) the max number of *VEM iterations*, that is the maximum number of alternations between the E-step and the M-step, (ii) the max number of *ELBO iterations*, which is the maximum number of times we use coordinate ascent to maximize the ELBO document by document, (iii) the *VEM-Corpus tolerance*, base on which we would stop the VEM algorithm when the change in the corpus ELBO between subsequent iterations is less than such tolerance and finally in a similar fashion (iv) the *ELBO-Document tolerance*, that controls the maximization of the ELBO document by document. In the single-stage estimation the value for these execution parameters are shown in Table B.1.

The two-stage estimation is done by first using LDA on $\{w_{t,1:N_t}\}$ to estimate $\{z_{t,1:N_T}\}$, $\beta_{1:K}$ and $\{\theta_t\}$. For this step we use the VEM algorithm simplified to handle just the unsupervised LDA, but we need still to set during execution the max VEM iterations, the max ELBO iterations, the VEM-Corpus tolerance and the ELBO-Document tolerance; these are shown in Table B.1. The second stage uses the estimates $\{\hat{z}_{t,1:N_T}\}$ to construct estimates of the empirical topic frequency for each document $\{\hat{Z}_t = N_t^{-1} \sum_{i=1}^{N_t} \hat{z}_{t,i}\}$ that would be used as regressors for $\{y_t\}$. Based on this second-stage regression we would obtain estimates for b and σ^2.

B.2.1. Extracted Topics Details

Table B.2: CR Exchange Rate Supervised Extracted Topics

Topic	Label	Spanish Top Stems	English Top Words
1	Health System Strike	public; polit; ley; huelg; president; ccss; econom; chin; servici; declar;	public; politics; law; strike; president; CCSS; economy; China; service; declare
2	Public Institutions	rabin; suges; lineamient; naranj; marcham; wong; mideplan; soa; mattis; garr;	Rabin; SUGESE; guideline; Naranjo; Auto Insurance; Wong; MIDEPLAN; SOA; Mattis; Garro
3	International News	econom; public; person; nacional; cambi; polit; banc; iran; pag; costa_rica;	Economy; public; person; national; change; policy; bank; Iran; payment; Costa Rica
4	Europe Economic Activity	public; gobiern; econom; costa_rica; gast; polit; nacional; proyect; pag; president;	public; government; economy; Costa Rica; expense; policy; national; project; payment; president
5	IMF Deal Strike	gobiern; public; president; nacional; huelg; servici; fiscal; person; proyect; pag;	government; public; president; national; strike; service; fiscal; person; project; payment
6	State-owned Companies	public; costa_rica; banc; person; empres; nacional; pag; econom; credit; gobiern;	public; Costa Rica; bank; person; firm; national; payment; economy; credit; government

Table B.2: CR Exchange Rate Supervised Extracted Topics (continued)

Topic	Label	Spanish Top Stems	English Top Words
7	Covid-19 Pandemic	econom; public; person; pandemi; med; covid; caso; coronavirus; gobiern; salud;	economy; public; person; pandemic; physician; covid; case; coronavirus; government; health
8	US Presidential Election	public; trump; biden; president; nacional; dialog; econom; sector; octubr; fiscal;	public; Trump; Biden; president; national; dialog; economy; sector; october; fiscal
9	Consumer Finance Policy	credit; tas; banc; person; entidad; comision; financ; tarjet; sugef; merc;	credit; rate; bank; person; entity; comission; finance; card; SUGEF; market
10	Covid-19 Second Wave	president; public; polit; gobiern; moral; econom; nacional; octubr; social; proyect;	president; public; policy; government; Morales; economy; national; october; social; project
11	Tax Discussion	polit; econom; public; person; gobiern; president; social; nacional; estados_unidos; costa_rica;	policy; economy; public; person; government; president; social; national; United States; Costa Rica
12	US Economy News	gobiern; polit; president; trump; econom; estados_unidos; public; empres; chin; nacional;	government; policy; president; Trump; economy; United States; public; firm; China; national
13	Fiscal Policy	subast; contribuyent; lun; compr; deud; abort; ajust; chin; kong; hong;	auction; taxpayer; Monday; buy; debt; abort; adjust; China; Hong Kong

Table B.2: CR Exchange Rate Supervised Extracted Topics (continued)

Topic	**Label**	**Spanish Top Stems**	**English Top Words**
14	Covid-19 Vaccines	president; vacun; gobiern; econom; dolar; estados_unidos; public; pandemi; polit; person;	president; vaccine; government; economy; dollar; United States; public; pandemic; policy; person
15	Middle East News	econom; israel; netanyahu; polit; ley; gobiern; president; estados_unidos; cambi; proyect;	economy; Israel; Netanyahu; politics; law; government; president; United States; change; project
16	Congress Debates	polit; president; gobiern; public; econom; costa_rica; nacional; fiscal; diput; person;	Policy; president; government; public; economic; Costa Rica; fiscal; lawmaker; person
17	Government Economic Policy	public; pag; empres; impuest; econom; servici; fiscal; caso; person; polit;	public; payment; firm; tax; economy; service; fiscal; case; person; policy
18	Brexit	presupuest; gast; ger; corrient; nicol; brexit; vuel; interes; johnson; europe;	budget; expenditure; current; Brexit; flight; interest; Johnson; Europe
19	Industrial Policy	proyect; empres; ley; electr; pension; energ; gobiern; cambi; person; pag;	project; firm; law; electric; pension; energy; government; change; person; payment
20	2018 Tax Reform Discussion	public; proyect; fiscal; diput; econom; gobiern; ley; pag; president; aprob;	public; project; fiscal; lawmaker; economy; government; law; payment; president; approval

Table B.2: CR Exchange Rate Supervised Extracted Topics (continued)

Topic	Label	Spanish Top Stems	English Top Words
21	Introduction of VAT	iva; impact; micas; servici; dueñ; public; empres; pag; caso; impuest;	VAT; impact; service; owner; public; firm; payment; case; tax
22	UPAD Goverment Scandal	dat; polit; president; public; caso; person; democrat; investig; febrer; ministr;	data; politics; president; public; case; person; democrat; investigation; february; minister
23	Exports	econom; gobiern; costa_rica; president; estados_unidos; trump; public; polit; chin; merc;	economy; government; Costa Rica; president; United States; Trump; public; policy; China; market
24	Rent Prices	candidat; polit; public; empres; gobiern; econom; gast; banc; tem; fiscal;	candidate; politics; public; firm; government; economy; economy; expense; bank; topic; fiscal
25	2018 Tax Reform Approval	public; fiscal; polit; econom; person; president; gobiern; merc; caso; pag;	public; fiscal; politics; economy; person; president; government; market; case; payment

BIBLIOGRAPHY

Andrzejewski, D., Zhu, X. and Craven, M. (2009), Incorporating domain knowledge into topic modeling via dirichlet forest priors, *in* 'Proceedings of the 26th annual international conference on machine learning', pp. 25–32.

Arrow, K. J. (1951), *Social choice and individual values*, New Haven: Yale University Press.

Bacchetta, P. and Van Wincoop, E. (2006), 'Can information heterogeneity explain the exchange rate determination puzzle?', *American Economic Review* **96**, 552–576.

Baker, M., Egan, M. L. and Sarkar, S. K. (2022), 'How do investors value ESG?'. National Bureau of Economic Research Working paper. Available at https://www.nber.org/papers/w30708.

BCCR (2010), 'Mercado cambiario en costa rica y pasivos con costo en colones del banco central de costa rica, octubre 2006-marzo 2010', *San José: Banco Central de Costa Rica* . https://bit.ly/2Dx6jbl, Consulted on 9-25-2018.

BCCR (2017a), 'Reglamento de operaciones cambiarias de contado', *San José: Banco Central de Costa Rica* .

Beaudry, P. and Portier, F. (2014), 'News-driven business cycles: Insights and challenges', *Journal of Economic Literature* **52**, 993–1074.

Berry, S. (1994), 'Estimating discrete-choice models of product differentiation', *The RAND Journal of Economics* **23**, 242–262.

Berry, S., Levinsohn, J. and Pakes, A. (1995), 'Automobile prices in market equilibrium', *Econometrica* **63**, 841–890.

Berry, S. T. and Haile, P. A. (2021), Foundations of demand estimation, *in* K. Ho, A. Hortaçsu and A. Lizzeri, eds, 'Handbook of Industrial Organization', Vol. 4, Elsevier, pp. 1–62.

Bezanson, J., Edelman, A., Karpinski, S. and Shah, V. B. (2017), 'Julia: A fresh approach to numerical computing', *SIAM review* **59**, 65–98.

Billio, M., Costola, M., Hristova, I., Latino, C. and Pelizzon, L. (2021), 'Inside the ESG ratings: (dis) agreement and performance', *Corporate Social Responsibility and Environmental Management* **28**, 1426–1445.

Bishop, C. (2006), *Pattern Recognition and Machine Learning*, Springer.

Blei, D. M., Griffiths, T., Jordan, M. I. and Tenenbaum, J. (2003), 'Hierarchical topic models and the nested chinese restaurant process', *Advances in neural information processing systems* **16**.

Blei, D. M., Kucukelbir, A. and McAuliffe, J. D. (2017), 'Variational inference: A review for statisticians', *Journal of the American statistical Association* **112**, 859–877.

Blei, D. M. and Lafferty, J. D. (2006), Dynamic topic models, *in* 'Proceedings of the 23rd international conference on Machine learning', pp. 113–120.

Blei, D. M. and Mcauliffe, J. D. (2008), Supervised topic models, *in* 'Advances in neural information processing systems', pp. 121–128.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003), 'Latent dirichlet allocation', *Journal of machine Learning research* **3**, 993–1022.

Bolton, P. and Kacperczyk, M. (2021), 'Do investors care about carbon risk?', *Journal of financial economics* **142**, 517–549.

Boz, E., Casas, C., Georgiadis, G., Gopinath, G., Le Mezo, H., Mehl, A. and Nguyen, T. (2020), 'Patterns in invoicing currency in global trade', *IMF Working Papers* .

Brainard, W. C. and Tobin, J. (1968), 'Pitfalls in financial model building', *The American Economic Review* **58**, 99–122.

Brandt, M. W., Santa-Clara, P. and Valkanov, R. (2009), 'Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns', *The Review of Financial Studies* **22**, 3411–3447.

Brunnermeier, M., Farhi, E., Koijen, R. S. J., Krishnamurthy, A., Ludvigson, S. C., Lustig, H., Nagel, S. and Piazzesi, M. (2021), 'Review Article: Perspectives on the Future of Asset Pricing', *The Review of Financial Studies* **34**, 2126–2160.

Bybee, L., Kelly, B. T., Manela, A. and Xiu, D. (2020), The structure of economic news, Technical report, National Bureau of Economic Research.

Campbell, J. Y. and Viceira, L. M. (2002), *Strategic asset allocation: portfolio choice for long-term investors*, Clarendon Lectures in Economic. New York: Oxford Univ. Press.

Cardell, N. S. (1997), 'Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity', *Econometric Theory* **13**, 185–213.

Conlon, C. and Gortmaker, J. (2020), 'Best practices for differentiated products demand estimation with pyblp', *The RAND Journal of Economics* **51**, 1108–1161.

Cubero, R., Lankester, V. and Munoz, E. (2019), 'Costa rica: Learning to float', *Foreign exchange interventions in inflation targeters in Latin America* pp. 155–194.

Das, N., Chatterjee, S., Sunder, A. and Ruf, B. (2018), 'ESG ratings and the performance of socially responsible mutual funds: A panel study', *Journal of Finance Issues* **17**, 49–57.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the em algorithm', *Journal of the Royal Statistical Society: Series B (Methodological)* **39**, 1–22.

Diaz, G. (2020), 'Stopwords iso: The most comprehensive collection of stopwords for multiple languages.', https://github.com/stopwords-iso/stopwords-iso.

Engle, R. F., Giglio, S., Kelly, B., Lee, H. and Stroebel, J. (2020), 'Hedging climate change news', *The Review of Financial Studies* **33**, 1184–1216.

Evans, M. D. and Lyons, R. K. (2002), 'Order flow and exchange rate dynamics', *Journal of Political Economy* **110**, 170–180.

Evans, M. D. and Lyons, R. K. (2012), 'Exchange rate fundamentals and order flow', *The Quarterly Journal of Finance* **2**.

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

Foran, C. and Wilson, K. (2023), 'House fails to override Biden veto of resolution to overturn investment rule'. *CNN*, https://www.cnn.com/2023/03/23/politics/veto-override-vote-house/index.html (accessed 10/06/2023), published March 23, 6:19pm EDT.

Gandhi, A. and Houde, J.-F. (2019), 'Measuring substitution patterns in differentiated-products industries'. National Bureau of Economic Research Working Paper, available at https://www.nber.org/papers/w26375.

Gandhi, A. and Nevo, A. (2021), Empirical models of demand and supply in differentiated products industries, *in* K. Ho, A. Hortaçsu and A. Lizzeri, eds, 'Handbook of Industrial Organization', Vol. 4, Elsevier, pp. 63–139.

Gentzkow, M., Kelly, B. and Taddy, M. (2019), 'Text as data', *Journal of Economic Literature* **57**, 535–74.

Giglio, S., Maggiori, M., Stroebel, J., Tan, Z., Utkus, S. and Xu, X. (2023), 'Four facts about ESG beliefs and investor portfolios'. National Bureau of Economic Research Working paper. Available at https://www.nber.org/papers/w31114.

Griffiths, T. L. and Steyvers, M. (2004), 'Finding scientific topics', *Proceedings of the National academy of Sciences* **101**, 5228–5235.

Grossman, S. (1976), 'On the efficiency of competitive stock markets where trades have diverse information', *The Journal of finance* **31**, 573–585.

Grupo-Nación-EF (2021), 'Periódico el financiero'. https://www.elfinancierocr.com/.

Grupo-Nación-LN (2021), 'Periódico la nación'. https://www.nacion.com/.

Han, X., Roussanov, N. L. and Ruan, H. (2021), 'Mutual fund risk shifting and risk anomalies'. Working Paper available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3931449.

Hong, H. and Kostovetsky, L. (2012), 'Red and blue investing: Values and finance', *Journal of Financial Economics* **103**, 1–19.

Jagarlamudi, J., Daumé III, H. and Udupa, R. (2012), Incorporating lexical priors into topic models, *in* 'Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics', pp. 204–213.

Jiang, Z., Richmond, R. and Zhang, T. (2020), 'A portfolio approach to global imbalances and low interest rates preliminary'. Working Paper available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3744340.

Ke, Z. T., Kelly, B. T. and Xiu, D. (2019), Predicting returns with text data, Technical report, National Bureau of Economic Research.

King, M. R., Osler, C. L. and Rime, D. (2013), 'The market microstructure approach to foreign exchange: Looking back and looking forward', *Journal of International Money and Finance* **38**, 95–119.

Koijen, R. S., Richmond, R. J. and Yogo, M. (2023), 'Which investors matter for equity valuations and expected returns?'. *The Review of Economic Studies, forthcoming.*

Koijen, R. S. and Yogo, M. (2019), 'A demand system approach to asset pricing', *Journal of Political Economy* **127**, 1475–1515.

Koijen, R. S. and Yogo, M. (2020), 'Exchange rates and asset prices in a global demand system'. Working paper available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3383677.

Larsen, V. H. and Thorsrud, L. A. (2019), 'The value of news for economic developments', *Journal of econometrics* **210**, 203–218.

Lopez-Lira, A. (2020), 'Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns', *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper* .

Lopez-Lira, A. (2023), 'Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns'. Working paper available at: https://ssrn.com/abstract=3313663.

Lopez-Lira, A. and Roussanov, N. L. (2023), 'Do common factors really explain the cross-section of stock returns?'. Working Paper available at: https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3628120.

Lucas, R. (1978), 'Asset prices in an exchange economy', *Econometrica* **46**, 1429–1445.

Lucas, R. E. J. (1976), Econometric policy evaluation: A critique, *in* 'Carnegie-Rochester conference series on public policy', Vol. 1, North-Holland, pp. 19–46.

Lyons, R. K. (1995), 'Tests of microstructural hypotheses in the foreign exchange market', *Journal of Financial Economics* **39**, 321–351.

McFadden, D. (1974), 'Condition logit analysis of qualitative choice behavior'. In: Zarembka, P. (Ed.) Frontiers in econometrics. Academic Press New York.

Mora, A. (2020), Text into fx: Exchange rate supervised topic extraction., Working paper, University of Pennsylvania, Philadelphia.

Morgan, D. (2023), 'U.S. Congress votes to block ESG investing, Biden veto expected'. *Reuters*, https://www.reuters.com/business/us-senate-poised-consider-blocking-biden-esg-investment-rule-2023-03-01/ (accessed 10/06/2023), published March 1, 7:44pm EDT.

Morgan Stanley (2022), 'Sustainable signals: Opportunities for asset managers to meet asset owner demands'. Morgan Stanley Institute for Sustainable Investing Report. Available at https://www.morningstar.com/lp/global-esg-flows.

Morningstar (2023*a*), 'Global sustainable fund flows: Q1 2023 in review'. Morningstar Manager Research Report. Available at https://www.morningstar.com/lp/global-esg-flows.

Morningstar (2023*b*), 'Sustainable funds U.S. landscape report'. Morningstar Manager Research Report. Available at https://www.morningstar.com/lp/sustainable-funds-landscape-report.

Nakajima, J. and West, M. (2013), 'Bayesian analysis of latent threshold dynamic models', *Journal of Business & Economic Statistics* **31**, 151–164.

Nevo, A. (2000), 'A practitioner's guide to estimation of random-coefficients logit models of demand', *Journal of Economics & Management Strategy* **9**, 513–548.

Osler, C. L., Mende, A. and Menkhoff, L. (2011), 'Price discovery in currency markets', *Journal of International Money and Finance* **30**, 1696–1718.

Pástor, L., Stambaugh, R. F. and Taylor, L. A. (2021), 'Sustainable investing in equilibrium', *Journal of Financial Economics* **142**, 550–571.

Pástor, L., Stambaugh, R. F. and Taylor, L. A. (2022), 'Dissecting green returns', *Journal of*

Financial Economics **146**, 403–424.

Pastor, L., Stambaugh, R. F. and Taylor, L. A. (2023), 'Green tilts'. National Bureau of Economic Research Working paper. Available at https://www.nber.org/papers/w31320.

Porter, M. F. (2001), 'Snowball: A language for stemming algorithms'.

Ramage, D., Hall, D., Nallapati, R. and Manning, C. D. (2009), Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora, *in* 'Proceedings of the 2009 conference on empirical methods in natural language processing', pp. 248–256.

Ray, P. (1973), 'Independence of irrelevant alternatives', *Econometrica* **41**, 987–991.

Rosen, S. (1974), 'Hedonic prices and implicit markets: product differentiation in pure competition', *Journal of Political Economy* **82**, 34–55.

Rudebusch, G. D., Offner, E. A. and Bauer, M. D. (2023), 'The effect of US climate policy on financial markets: An event study of the Inflation Reduction Act'. Hutchins Center on Fiscal and Monetary Policy at Brookings Working Paper 89, available at https://www.brookings.edu/articles/the-effect-of-u-s-climate-policy-on-financial-markets-an-event-study-of-the-inflation-reduction-act/.

Samuelson, P. A. (1969), 'Lifetime portfolio selection by dynamic stochastic programming', *The Review of Economics and Statistics* **51**, 239–246.

Thomas, K. (2023), 'Joe Biden issues first veto, rejecting attempt to block ESG effort'. *The Wall Street Journal*, https://www.wsj.com/articles/biden-issues-first-veto-rejecting-attempt-to-block-esg-effort-63c2f969 (accessed 10/06/2023), published March 20, 3:24pm EDT.

Thorsrud, L. A. (2020), 'Words are the new numbers: A newsy coincident index of the business cycle', *Journal of Business & Economic Statistics* **38**, 393–409.

Tian, D. (2022), 'Do intermediaries matter for stock returns? a dynamic demand system approach'. Manuscript University of Pennsylvania.

Vitale, P. (2007), 'A guided tour of the market microstructure approach to exchange rate determination', *Journal of Economic Surveys* **21**, 903–934.

Vitale, P. (2011), 'The impact of fx intervention on fx markets: a market microstructure analysis', *International Journal of Finance & Economics* **16**, 41–62.

Wainwright, M. J. and Jordan, M. I. (2008), *Graphical models, exponential families, and variational inference*, Now Publishers Inc.