



FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

**THE IMPACT OF INVESTOR SENTIMENT ON
THE STOCK MARKET RISK PREMIUM:
UNVEILING INFORMATION THROUGH THE
USE OF INFORMATION TECHNOLOGIES**

**Programa de Doctorado en Competitividad Empresarial y
Territorial, Innovación y Sostenibilidad**

Author: Juan José García Petit
Advisor: Doctor D. Antonio Rúa Vieites

July 2019
Madrid

To my parents to whom I owe everything

Abstract

Risk premium analysis becomes a key element in financial literature as it is the driver of investment decisions both in financial and real assets. Most of the literature on risk premium has relied mainly on classical theoretical formulations grounded on the fundamental assumptions of market efficiency and rationality of the agents that altogether would prevent the possibility of abnormal returns. However, the presence of multiple market anomalies have posed the need to open the financial analysis to new perspectives closer to the empirical evidence.

Against this backdrop, behavioral finance poses an alternative or at least a complementary approach to traditional explanations of the risk premium. The presence of agents that do not respond to the assumption of rationality and the incorporation of undisclosed information about agents' preferences, attached to the behavioral theoretical framework, cope with the limitations of traditional finance.

Therefore, the possibility of a symbiosis between both theoretical proposals becomes a milestone. The behavioral advantages of greater proximity to empirical evidence combined with by the advantages of classic solid formulations means a significant improvement to the determination of risk premium and subsequently to asset valuation as a whole.

The research has been conducted as a compendium of three articles expand all the elements that have been previously noted around the risk premium analysis. In fact, the first article, *The journey of behavioral risk premium: a concept map*, outlines a thorough classification of the different research contributions to the definition of risk premium with particular emphasis on the behavioral ones. Additionally, given the role that behavioral elements play

as a source of new information to the market, greater detail will be put on their definition with special focus on the investor sentiment categorization.

The second paper, *How information technologies shape investor sentiment: a web-based investor sentiment index*, focuses on the definition and estimation of investor sentiment. In this regard, we propose a new investor sentiment indicator that combines the use of principal component analysis with information discovery through web searches. This proposal provides economic meaning to the underlying variables, a sound factor structure, and reduces the noise regarding to web searches, when compared to standard search-based sentiment indicators. In fact, our indicator not only confirms the relevance of sentiment for future asset performance and provides greater predictive capacities than standard formulations, but also generates new insights in terms of globalization of investor sentiment and the role that information flows and technology play on that process.

The final article, *Sense and sentiment: a behavioral approach to risk premium modelling*, addresses the need to incorporate behavioral factors into the risk premium estimation process. This paper proposes an alternative methodology to estimate risk premium by including information on agents' intentions into the asset valuation model.

This model will be tested on the American market with the objective of obtaining a more accurate measure of risk premium than the one provided by classical financial approach. Moreover, this methodology makes it possible to offer an alternative explanation to risk-return relationship based on the dynamics of investment sentiment. Finally, the use of behavioral elements into the treatment of the risk premium will also lead to a greater control of market anomalies and, subsequently, to higher efficiency derived from the combination of both classical and behavioral methodologies.

Acknowledgments

I would like to show my sincere appreciation to my advisor Prof. Antonio Rúa Vieites whose expertise, counsel and support have made an invaluable contribution to this doctoral thesis. His guidance and knowledge have gone side by side with the definition and elaboration of the research conducted throughout all these years.

I owe my deepest gratitude to Prof. Esther Vaquero Lafuente for her continuous support and motivation. Her enthusiasm and passion for research have been contagious and encouraging. The success of this research would have not been possible without her outstanding opinions and suggestions.

Overall, I would not be able to think of a better advisory group to guide me in my doctoral studies than the one they have made.

I would also like to thank to my fellow colleagues and professors of the CETIS doctoral program for the comments and suggestions made in the different research seminars in which this research has been presented. This dialogue and exchange of ideas has contributed to improve the quality of the research. In this regard, I would like also to thank to all editors and anonymous referees that have contributed with their comments to improve this work.

Last but not least, my heartfelt gratitude goes to my family. Their patience, understanding, generosity and love have made the greatest contribution to this project. Manane, thanks for being such a star, I couldn't have done it without you.

Contents

| | |
|--|----|
| 1. Introduction | 1 |
| <i>1.1. The risk premium, a driver for the dynamics of financial assets</i> | 3 |
| <i>1.2. Information as a key factor in the determination of the risk premium</i> | 4 |
| <i>1.3. Objective of the thesis: to develop a risk premium valuation model based on investor sentiment</i> | 5 |
| <i>References</i> | 10 |
| 2. The journey of behavioral risk premium: a concept map | 13 |
| <i>2.1. Introduction</i> | 14 |
| <i>2.2. The market risk premium in conventional financial theory</i> | 16 |
| <i>2.2.1. Review of market anomalies and their relationship to the risk premium</i> | 17 |
| <i>2.2.2. Response to market anomalies from conventional financial theory</i> | 20 |
| <i>2.3. Alternative approach to the risk premium: behavioral finance</i> | 21 |
| <i>2.3.1. Theories focused on limitations of conventional financial theories</i> | 24 |
| <i>2.3.2. Theories with psychological fundamentals</i> | 25 |
| <i>2.4. Conclusions</i> | 35 |
| <i>References</i> | 36 |
| 3. How information technologies shape investor sentiment: a web-based investor sentiment index | 43 |
| <i>3.1. Introduction</i> | 44 |
| <i>3.2. Construction of a web-based investor sentiment index</i> | 50 |
| <i>3.2.1. Extracting sentiment from searches</i> | 51 |
| <i>3.2.2. Sentiment index: definition and characteristics</i> | 54 |
| <i>3.3. Empirical analysis</i> | 56 |
| <i>3.3.1. Sentiment and predictability of US stock returns</i> | 56 |
| <i>3.3.1.1. Comparison to other relevant literature</i> | 60 |
| <i>3.3.2. Is globalization affecting sentiment?</i> | 62 |
| <i>3.4. Trading strategy</i> | 68 |
| <i>3.5. Conclusions</i> | 69 |
| <i>Appendix</i> | 71 |
| <i>References</i> | 73 |

| | |
|---|-----|
| 4. Sense and sentiment: a behavioral approach to risk premium modelling | 78 |
| 4.1. <i>Introduction</i> | 79 |
| 4.2. <i>Determination of the risk premium through a valuation model with expanded information</i> | 81 |
| 4.2.1. <i>Implementation of a sentimental asset pricing model</i> | 83 |
| 4.3. <i>Empirical contrast of a conditional sentimental valuation model</i> | 88 |
| 4.3.1. <i>Descriptive analysis of the series</i> | 89 |
| 4.3.2. <i>Modeling of sentimental conditional betas</i> | 92 |
| 4.3.2.1. <i>Empirical evidence from sentimental betas</i> | 97 |
| 4.3.3. <i>Assessing the sentimental risk premium factor</i> | 101 |
| 4.4. <i>Sentiment and market anomalies: the momentum effect</i> | 103 |
| 4.5. <i>Conclusions</i> | 107 |
| <i>Appendix 1</i> | 109 |
| <i>Appendix 2</i> | 112 |
| <i>References</i> | 117 |
| 5. Conclusions, limitations and future research | |
| 5.1. <i>Conclusions</i> | 121 |
| 5.2. <i>Limitations and future research</i> | 124 |
| <i>References</i> | 126 |

Acronym dictionary

ACWI: All Countries World Index

ARIMA: Autoregressive Integrated Moving Average

B/M: Book to Market Ratio

BW: Baker and Wurgler

CAPM: Capital Asset Pricing Model

CCAPM: Consumption-based Capital Asset Pricing Model

DCC: Dynamic Conditional Correlation

EM: Emerging Markets

GARCH: Generalized Autoregressive Conditional Heteroskedasticity

GICS: Global Industry Classification Standard

IT: Information Technology

IR: Impulse-Response

KMO: Kayser-Meyer-Olkin

MSCI: Morgan Stanley Capital International

PC: Principal Component

PCA: Principal Component Analysis

P/E: Price to Earnings Ratio

SARIMA: Seasonal Autoregressive Integrated Moving Average

SD: Standard Deviation

S&P: Standard & Poor's

SVI: Search Volume Index

US: United States

USCONF: United States Consumer Confidence Index

VAR: Vector Autoregression

VIX: Chicago Board Options Exchange Market Volatility Index

WISI: Web-based Investor Sentiment Index

1. Introduction

Over the last two decades, financial theory has gone through such a deep transformation from its classical formulations (prevailing since the 60s) that we can talk about a “behavioral revolution” (Shefrin, 2015, p.95).

Those classical theoretical formulations were based on the fundamental assumptions of market efficiency and rationality of the agents that altogether would prevent the possibility of abnormal returns (Shefrin and Statman, 2003). However, the presence of multiple market anomalies has posed the need to open the financial analysis to new perspectives closer to the empirical evidence.

Therefore, the main objective of this thesis is to address these shortcomings of classical financial theories by opening the study to alternative paradigms that can improve the determination of asset dynamics and valuations.

Precisely, the presence of agents (*noise traders* - De Long et al., 1990) that did not respond to the assumption of rationality (conditioned by biases and heuristics), emphasized that the markets are “substantially driven by psychology” (Shiller, 2014, p.1487) and sociology (Ricciardi and Simon, 2000) that transcended the rational approach underlying existing theories so far.

Against this background, the consideration of the agent’s behavior and preferences arises as a key factor in determining the dynamics of financial assets and explaining the deviations of prices from their fundamental valuations (Kahneman and Tversky, 1979).

Consequently, there are additional informational elements linked to the agent’s individual decision-making processes (particularly, on her future intentions) beyond those already

reflected in asset prices. Then, this unrevealed information is prone to be incorporated into the markets and it will contribute to make them more efficient (Brown et al., 1988).

The incorporation of these behavioral elements into financial theory would allow for a conciliation with classical normative theories (Thaler, 2000), to the extent that they would solve their fundamental weaknesses with a complementary formulation (Singh, 2010). In fact, as pointed out by Malkiel (2003, p.80) "the market is remarkably efficient in its utilization of information" by removing any type of anomaly or pattern as new information comes out.

Accordingly, the key lies in being able to capture agent's intentions properly. In this regard, technological progress, especially in the field of information technologies (IT), contributes significantly to achieving this goal. In fact, there is a growing increase in social connectivity at a global dimension that allows the establishment of a relationship between intentionality of agents and uses of technology.

In particular, this approach considers that the driver of agent's expectations (or sentiment) is the information available at any time (Hoffman and Post, 2015). In this sense, IT show a dual role:

- (i) Firstly, they focus on individual agents as direct providers of information. This approach is superior to any attempt to reflect sentiment through economic proxies, unable to capture undisclosed intentions (Ghysels et al., 2007).
- (ii) Secondly, IT growing penetration, both geographically and at user level, favors the construction of sentiment indicators as a reflection of individual intentions, making them representative of agent's true preferences. That is, collective behavior can be obtained straight from the study of data coming out of the use of these technologies by

individuals (Bentley et al., 2014. Curme et al., 2014). Moreover, this collective behavior can be thought as being global, to the extent that the speed of transmission of information (and, therefore, behavior) is increasingly higher (Hirshleifer, 2015).

In short, it is possible to build a theoretical framework around the concept of investment sentiment based on new information obtained through the intensive use of IT that, simultaneously, would improve the assessment of the dynamics of financial assets.

1.1. The risk premium, a driver for the dynamics of financial assets

The risk premium, understood as the compensation required by investors in exchange for taking systematic risk (Gagliardini et al., 2016), is the key element in setting the valuation of a financial asset.

The existing literature shows a continuous evolution in the quantification of that risk premium but, simultaneously, it also allows for alternative approaches to implement that quantification that, in some cases, might end up being complementary (Section 2 shall extend widely on this subject).

Hence, the classical financial approach has been able to establish a solid quantitative basis that facilitates a first reading of the risk component. However, this measurement is incomplete mainly due to base assumptions that depart from the empirical evidence.

Although different *ad hoc* proposals have been made within this line of research with the aim to further improve the explanatory power of their models, any real improvement in the understanding and quantification of the risk elements demands an opening of the paradigm towards different formulations closer to reality (Cochrane, 2011).

In this context, behavioral finances are positioned, *a priori*, as an alternative proposal to understand how investment processes work from the agent's perspective, bringing the concept of risk premium closer to the one that arises from the decision-making process in the market (Malkiel, 2003; Lo, 2004; Kyriacou et al., 2004).

However, even considering its obvious improvement with respect to traditional formulations, especially regarding the incorporation of assumptions that are closer to what is empirically observed, behavioral finance is not exempt from constraints either. In fact, we are not facing a unified theoretical body but a set of simultaneous explanations to the same problems discussed from the conventional approach (Barberis and Thaler, 2003).

Given this situation, the possibility of a symbiosis between both theoretical proposals becomes relevant. Accordingly, the behavioral advantages of greater proximity to empirical evidence would be complemented by the advantages of having a classic solid formulation that has proved useful when defining asset valuation processes.

These capabilities are clearly complementary with a common objective of extracting as much information as possible (and, certainly, beyond that already present in the market) and improving the understanding of the dynamics of financial assets. As Thaler (2000, p.140) notes, "It seems logical that basing descriptive economic models on more realistic conceptions of economic agents is bound to increase the explanatory power of the models".

1.2. Information as a key factor in the determination of the risk premium

In line with what was previously noted, any improvement in the definition of the risk premium should be aimed at favoring a greater incorporation of information in its estimation processes. For instance, Bertella et al. (2014), find that the risk premium is

continuously underestimated (they link this result to the presence of an overconfidence bias) and that an additional contribution in terms of information should translate into higher risk levels¹.

However, at this point an additional question can be posed: what type of information is relevant to the determination of the risk premium? In this sense, the information must have transcendence, be as broad as possible and reveal investor sentiment².

The use of information technologies can make a significant contribution when it comes to extract new information that is subsequently incorporated into the processes for estimating the risk premium. The evidence of new information obtained with the use of information technologies highlights that market prices are not reflecting all existing information (Da et al., 2011). So, there is a need to develop the extraction and processing mechanisms of such information, which in turn should lead to greater accuracy in assessing the risk premium.

1.3. Objective of the thesis: to develop a risk premium valuation model based on investor sentiment

As mentioned before, it can be suggested that a symbiosis of theories should result in a significant improvement in the determination of the risk premium both from a theoretical point of view and from its practical estimation. Therefore, the main objective of this thesis

¹ And, especially, in negative market movements that is when the set of available information is reduced and that would go in line with the results of Fox et al. (2016) who show less market efficiency as the available information decreases.

² For example, although there is empirical evidence of the effect of the news on asset returns (Amin and Ahmad, 2013), the presence of asymmetry in their impact is also observed (Lee and Mauck, 2014). Therefore, a repurchase of shares may indicate that these shares are undervalued, but if the announcement of repurchases is made on a regular basis, it ceases to have an effect and long-term excess returns are no longer observed (Yook, 2010).

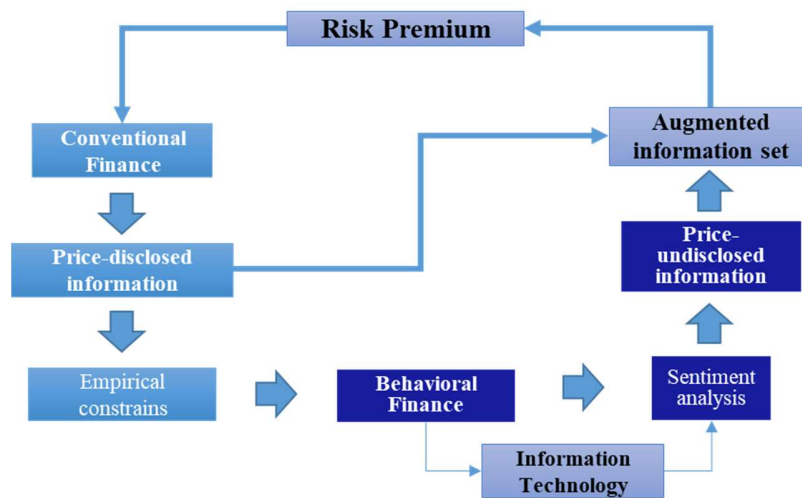
focuses on the incorporation of investor sentiment to the valuation models supported by classic formulations in such a way that a more accurate estimate of the stock market risk premium is obtained. The final achievement of this goal would result in a relevant contribution to the existing literature on the valuation of financial assets and would push for further development of mixed methodologies within this field.

This objective goes through the attainment of two intermediate steps that will make up the core of this research:

- a) The quantification of investor sentiment. This element is the cornerstone for the achievement of the objective of this thesis and requires both the definition of a descriptive theoretical framework establishing a solid relationship between sentiment and information, and a methodology that favors the generation of a broadly representative measure of investor sentiment.
- b) The elaboration of a sentimental risk premium measure. This requires the development of a methodology that allows the incorporation of investor sentiment to a traditional risk premium valuation model.

In order to achieve the aforementioned objective, this work will be structured around a compendium of three research papers that put together make up for an alternative formulation of risk premium (figure 1.1).

Figure 1.1. Thesis Workflow



The first article, *The journey of behavioral risk premium: a concept map*, deals not only with a comprehensive literature review on the risk premium concept but on the behavioral contributions that can build on it. In fact, this review will show how behavioral factors might become key drivers to set a proper risk premium definition.

Accordingly, the paper fulfills a dual objective: firstly, it will thoroughly describe the different research contributions to risk premium from the classical to behavioral theories. Precisely, we create a conceptual map that depicts both the limitations of classical theories (eg market anomalies) and the potential benefits of considering behavioral contributions as an alternative or at least complementary explanation to the risk premium concept.

Secondly, and considering the role of behavioral factors as an alternative to the limitations of classical finance theories, research on investor sentiment will be mapped in greater detail. Accordingly, and building on existent psychological proposals (eg. Igual and Santamaria, 2017), we will set up a new classification that will emphasize the informational content (undisclosed information on investor’s intentions) as the key driver for shaping sentiment. The validation of this sentiment-information relationship will contribute to the

asset pricing literature by overcoming the constraints of prevailing classical financial theories.

The second paper, *How information technologies shape investor sentiment: a web-based investor sentiment index* (*Borsa Istanbul Review*, Forthcoming, 2019), focuses on the definition and estimation of investor sentiment. Given the information-sentiment relationship, selecting the appropriate information set is crucial to capture agents' expectations accurately (Bank and Brustbauer, 2014). Provided that those expectations arise in the information search process itself, the intensive use of information technologies can help to uncover those expectations.

Against this backdrop, this paper proposes a new investor sentiment indicator that combines the use of principal component analysis (widely used in economic-based models of sentiment- e.g. Baker and Wurgler, 2006) with web searches. This proposal provides economic meaning to the underlying variables, a sound factor structure, and reduces the noise regarding to web searches, when compared to standard search-based sentiment indicators (e.g. Da et al., 2015).

In fact, our indicator not only confirms the relevance of sentiment for future asset performance and provides greater predictive capacities than standard formulations (Fang et al., 2014; Hoffman and Post, 2015), but also generates new insights in terms of globalization of investor sentiment and the role that information flows and technology play on that process.

Moreover, sentiment arises as a powerful source of information. It even captures information commonly attached to other market variables. Simultaneously, it challenges some general beliefs present in the literature such as the prevalence of a local bias, the

greater impact of sentiment in developed markets or the fact that institutional investors are fully rational agents.

The third article, *Sense and sentiment: a behavioral approach to risk premium modelling*, addresses the need to incorporate behavioral factors into the risk premium estimation process. In fact, estimates of risk premium as of classical financial theory have consistently shown deviations from observed levels. To a substantial extent, these limitations have been linked to the theoretical rational foundations that rely on asset prices as the main information source.

Therefore, this paper proposes an alternative methodology to estimate risk premium as it includes information on agents' intentions. Accordingly, a new asset valuation model is proposed in such a way that incorporates investor sentiment as a source of additional information (undisclosed) to market prices.

On this basis, the proposal will be tested on the American market with the objective of obtaining a more accurate measure of risk premium than the one provided by classical financial approaches. Moreover, this methodology allows for a dynamic measure of the risk premium that is more realistic and overcomes the limitations of traditional valuation models that measure risk premium for the whole period.

This approach will also offer an alternative explanation to risk-return relationship (positive for the period of study) based on investment sentiment (Bams et al., 2015).

Finally, the use of behavioral elements into the treatment of the risk premium will also lead to a greater control of market anomalies. This outcome shows the greater efficiency achieved by combining classical and behavioral methodologies into the risk premium estimation.

References

- Amin, R., & Ahmad, H. (2013). Does Investor Attention Matter ' S ? *Journal of Public Administration, Finance and Law*, 4, 111–125.
- Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, LXI(4), 1645–1680.
- Bams, D., Honavar, I., & Lehnert, T. (2015). From Time Varying Risk-Aversion to Anomalies in Market Moments' Risk Premia. <http://dx.doi.org/10.2139/ssrn.2692983>.
- Bank, M., & Brustbauer, J. (2014). Investor Sentiment in Financial Markets. <https://www.uibk.ac.at/ibf/mitarbeiter/bank/investor-sentiment-in-financial-markets.pdf>.
- Barberis, N., & Thaler, R.H., (2003). A survey of behavioral finance. In Constantinides, G., Harris, M., & Stulz, R. (Eds.), *Handbook of the Economics of Finance* (pp.1052-1090) Amsterdam: North-Holland.
- Bentley, R. A., O'Brien, M. J., & Brock, W. A. (2014). Mapping collective behavior in the big-data era. *The Behavioral and Brain Sciences*, 37(1), 63–76.
- Bertella, M.A, Pires, F.R., & Feng, L. (2014). Confidence and the stock market: an agent based approach. *PLoSone* 9(1): e83488. <https://doi.org/10.1371/journal.pone.0083488>.
- Brown, K.C., Harlow, W.V., & Tinic, S.M. (1988). Risk aversion, uncertain information, and market efficiency. *Journal of Financial Economics*, 22(2), 355-385.
- Cochrane, J. H. (2011). Presidential Address: Discount Rates. *Journal of Finance*, 66(4), 1047–1108.
- Curme, C., Preis, T., Stanley, H. E., & Moat, H. S. (2014). Quantifying the semantics of search behavior before stock market moves. *Proceedings of the National Academy of Sciences of the United States of America*, 111(32), 11600–11605.
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *Journal of Finance*, 66, 1461–1499.
- Da, Z., Engelberg, J., & Gao, P. (2015). The Sum of All FEARS Investor Sentiment and Asset Prices. *Review of Financial Studies*, 28(1), 1–32.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98, 4, 703-738.
- Hoffmann, A. O. I., & Post, T. (2015). How return and risk experiences shape investor beliefs and preferences. *Accounting and Finance*, 57, 3, 759-788.

Fang, L. H., Peress, J., & Zheng, L. (2014). Does media coverage of stocks affect mutual funds' trading and performance? *Review of Financial Studies*, 27, 12, 3441-3466.

Fox, E. G., Fox, M. B., & Gilson, R. J. (2016). Economic crisis and the integration of law and finance: The impact of volatility spikes. *Columbia Law Review*, 116 (2), 325-408.

Gagliardini, P., Ossola, E., & Scaillet, O. (2016). Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets. *Econometrica*, 84(3), 985–1046.

García Petit, J.J., Vaquero Lafuente, E., & Rúa Vieites, A. (2019). How information technologies shape investor sentiment: a web-based investor sentiment index. *Borsa Istanbul Review*. <https://doi.org/10.1016/j.bir.2019.01.001>.

García Petit, J.J., Vaquero Lafuente, E., & Rúa Vieites, A. (2019). The journey of behavioral risk premium: a concept map. Working paper.

García Petit, J.J., Vaquero Lafuente, E., & Rúa Vieites, A. (2019). Sense and sentiment: a behavioral approach to risk premium modelling. Working paper.

Ghysels, E., Plazzi, A., & Valkanov, R. (2007). Valuation in US commercial real estate. *European Financial Management*, 13, 472–497.

Hirshleifer, D. (2015). Behavioral Finance. *Annual Review of Financial Economics*, 7(1), 133–159.

Igual, M.G., & Santamaría, T.C. (2017). Overconfidence, loss aversion and irrational investor behavior: a conceptual map. *International Journal of Economic Perspectives*, 11(1), 273-290.

Kahneman, D., & Tversky, A. (1979). Prospect Theory - An Analysis Of Decision Under Risk. *Econometrica*, 47, 2, 263-291.

Kyriacou, K., Madsen, J., & Mas, B. (2004). The equity premium. https://www.researchgate.net/publication/4924220_The_equity_premium.

Lee, B. S., & Mauck, N. (2014). Information Asymmetry and the Market Response to Open Market Share Repurchases. https://mpra.ub.uni-muenchen.de/54066/1/MPRA_paper_54066.pdf.

Lo, A. W. (2004). The Adaptive Markets Hypothesis. *The Journal of Portfolio Management*, 30(5), 15–29.

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17, 1, 59-82.

Ricciardi, V., y Simon, H. K. (2000). What is Behavioral Finance ?, *Business, Education & Technology Journal*, 2, 2, 1–9.

Shefrin, H. (2015). The behavioral paradigm shift. *Revista de AdministraÇao de Empresas*, 55(1), 95–98.

Shefrin, H., & Statman, M. (2003). The Contributions of Daniel Kahneman and Amos Tversky. *The Journal of Behavioral Finance*, 4, 2, 54-58.

Shiller, R. J. (2014). Speculative asset prices. *American Economic Review*, 104, 6, 1486-1517.

Singh, R. (2010). Behavioural finance studies: emergence and developments. *Journal of Contemporary Management Research*, 4(2), 1–9.

Thaler, R. H. (2000). From Homo Economicus to Homo Sapiens. *Journal of Economic Perspectives*, 14(1), 133–141.

Yook, K. C. (2010). Long-run stock performance following stock repurchases. *Quarterly Review of Economics and Finance*, 50, 323–331.

2. The journey of behavioral risk premium: a concept map

Juan José García Petit, Esther Vaquero Lafuente, Antonio Rúa Vieites*

Abstract

Risk premium analysis has been traditionally grounded on two key assumptions: first, the presence of efficient markets. Secondly, the existence of rational investors who cannot systematically beat the market. However, in the real world are not many fully rational agents and the information is not perfect.

In this respect, behavioral finance deals not only with market constraints, but also with the presence of irrational investors that respond to changes in sentiment. Therefore, sentiment factors become key drivers to a proper risk premium assessment and deserves larger attention than the one provided by traditional formulations.

Accordingly, this paper presents a twofold objective: first, mapping the different research contributions to risk premium with special focus on behavioral ones.

Secondly, evidencing the role of investor sentiment as the key factor that can cope with the limitations of classical finance theories around risk premium estimation. Sentiment will be considered as a set of undisclosed information about investors' future intentions that by the time it gets revealed (using information technologies) contributes to increase market efficiency and therefore to better risk premium estimates.

Keywords: risk premium, investor sentiment, behavioral finance, information technologies, market efficiency.

JEL: G00, G40.

*As of July 2019 this paper is under review at *Cuadernos de Gestión*.

2.1. Introduction

Risk premium analysis becomes a key element in financial literature as it is the driver of investment decisions both in financial and real assets (Cochrane, 2011; Shefrin, 2015).

But, what is meant by risk premium? Widely, it can be understood as “the incremental return that the shareholders require to hold risky equities rather than risk-free securities” (Dimson et al., 2003, p.28).

However, far from being a universally accepted definition there are difficulties to pin it depending on its size and scope. In this regard, the first approaches to the study of the risk premium focused mostly on empirical observation of existing excess returns in equity markets relative to risk-free assets (primarily in the US market). In other words, focus was placed on identifying factors of risk aversion. Among these first proposals the US Federal Reserve earning yield gap model can be highlighted for its relevance in the financial industry³. This model implicitly assesses the market risk premium with respect to a risk-free asset (generally speaking, government bonds).

But despite its wide penetration in the financial industry, empirical analysis has shown that the level of risk premium observed for the American market was too high to be linked only to risk aversion (Mehra and Prescott, 1985).

These divergences between theoretical results and real observations push research forward and open novel approaches that take asset dynamics⁴ or economic cycle⁵ into account.

³This model is based on the observation of yield differentials between the US stock and 10-year Treasury bond (maturity that shows a similar duration to that of the stock markets). A detailed description can be found in Estrada (2006).

⁴For example, discounted cash flow models (Copeland and Weston; 1992, Brealey and Myers, 2000)

⁵Rietz (1988) suggests that the risk premium incorporates the probability of a crash in the production or consumption cycles. Brown et al. (1995) proposed a model of market survival that would be unique for the American market. This would justify the presence of a higher risk premium.

Nevertheless, despite progress made discrepancies in risk premium estimates persist, even in relation to the estimation period. For instance, Ibbotson and Chen (2003) estimated a premium slightly above 5.9% for the period 1926-2000. But, Fama and French (2002) found levels between 4.17-4.4% for the period 1951-2000.

With these premises, it is confirmed that the approach to set a proper definition of risk premium is highly complex and it is influenced by theoretical positioning. So, it makes sense to conduct a review of existing literature around it in order to define its scope and implications. This review will begin with the contributions made by the classical financial theory and will up continue to consider behavioral finance as an alternative explanation to those divergences present between theoretical estimates and empirical observations.

Against this backdrop, the objective behind this paper is twofold: first, to develop an updated classification of different contributions in the risk premium literature beyond pure classical formulations. As the matter of fact, special emphasis will be placed on behavioral approaches and simultaneously attempting to sort out the heterogeneity existent in this area building on recent proposals of classification. Specifically, we will extend on the proposals made within the psychological based explanations (e.g. Igual and Santamaría, 2017) by paying further attention into agents' preferences and intentions.

Secondly, we will present a theoretical framework for risk premium estimation on the grounds of behavioral contributions and particularly on investor sentiment. Sentiment will rest on information criteria and the fact that it may be considered as a representation of private undisclosed information which, in turn, will interact with the idea of market efficiency finally impacting risk premium estimates as it gets revealed.

2.2. The market risk premium in conventional financial theory

The observed differences between estimates and reality pave the way to greater sophistication in risk premium analysis. This will lead to developing more formal theoretical bodies that will be supported by mathematical models not only focus on asset returns but on their risk component too.

In line with the proposed framework and given the importance of the risk/return ratio for financial theory, much of academic efforts have focused on quantifying this relationship. Given this objective there has been an array of alternative formulations although one stands among all for its wide acceptance and validity: Capital Asset Pricing Model (CAPM). This model was developed almost simultaneously by Sharpe (1964), Lintner (1965), Mossin (1966) and Treynor (1965). This formulation has even generated a fundamental axiom for financial theory: the expected asset returns are determined by their level of risk.

While the CAPM was widely accepted in the academic community, supported by numerous empirical tests (Black, 1972; Black and Scholes, 1972; Fama and MacBeth, 1973), uncertainties about its explanatory power did not take long to rise. Two elements were particularly under scrutiny: first, given the results provided by the model and although there was evidence of a significant relationship between average yields and estimated betas, it was also found that the estimated constant was higher. Besides, the slope was lower than predicted and marginally important to explain the differences between average returns (Fama and McBeth, 1973; Reinganum, 1981).

A second set of questions about the validity of the CAPM as a reference model is linked to the concept of market efficiency (in Fama's terminology⁶) and its interaction with the model itself⁷.

In fact, by the time of testing the model and given that only ex - post data was available, it was assumed that the observed returns reflected all the existent information in market equilibrium. Therefore, any empirical test of market efficiency would require the model to be valid and vice versa.

This circular reasoning has proved to be problematic as many studies have been documenting the presence of irregularities in stock prices (Elton and Gruber, 1995). These anomalies are inconsistent with the notion of efficient capital markets (basic assumption of this model). In efficient markets, any abnormal returns would be arbitrated between competing investors holding the same set of information.

The study of these anomalies deserves more attention as they pose one of the strongest criticisms of classical models and support the impetus needed for the consolidation of alternative formulations of risk premium.

2.2.1. Review of market anomalies and their relationship to the risk premium

Among the first studies on market anomalies we can mention the one conducted by Basu (1977) who sought to determine whether the stock performance was linked to the price to earnings ratio (P/E). Forming portfolios with similar P/E for the NYSE he found that during

⁶Fama (1970) defines three categories of market efficiency categories according to the type of information available: 1) weak efficiency, it would consist of historical prices. 2) Semi-strong efficiency, including information which is publicly available. 3) Strong efficiency considers information that is not publicly available.

⁷This problem goes back to Ball (1978).

the period April 1957 to March 1971 low P/E portfolios obtained higher yields on average than high P/E portfolios (both in absolute and risk-adjusted terms).

The interpretation of this result relates to the risk premium factor to the extent that these returns would not be properly adjusted for risk in the equilibrium model. The P/E factor would act as an additional risk indicator⁸ and it would be associated with abnormal returns (should a model such as CAPM be used as benchmark). Behind this anomaly there would be failures in the information transmission process to the market (Basu, 1977).

An additional anomaly was evidenced by Banz (1981). He noticed the presence of a size effect in the market. He demonstrated how small companies (classified according to their relative market capitalization) showed a better systematic performance than large firms. The conclusion from this size effect would be that the CAPM is not well specified⁹.

Further analysis (Keim, 1988; Roll, 1983) showed that this size effect was mostly concentrated in January and precisely in the first two weeks of January. Also, that its magnitude was insensitive to stock's beta size (Fama and French, 1992).

In an attempt to explain the origins of this anomaly, different hypotheses have been suggested. First, those related to business characteristics as a proxy for an additional risk variable that when considered would eliminate the relationship between characteristics of those firms and their returns (Banz, 1981).

Other hypotheses suggest that the size effect linked to small firms is related to the fact that they have little coverage by financial analyst (Arbel and Strebel, 1982). Since the amount

⁸ Ball (1978) notes that would be a proxy for unidentified factors.

⁹ Roll (1983), points to the poor beta specification as a proof of the presence of such size effect. Moore and College (1998) found an inverse relationship between size and stock return. Reinganum (1981) found abnormal earnings systematically deviated from those predicted by the CAPM.

of information collected from these companies and the effort that goes into when analyzing their behavior are reduced, opportunities may arise for arbitrage leading to abnormal returns.

Additionally, it has also been found a so-called value effect. According to this anomaly returns would be anticipated by the relative evolution of the market value of the company against its fundamental measures (e.g. book value or book to market - B/M - Basu, 1983). Different studies (Stattman, 1980; De Bondt and Thaler, 1987; Keim, 1988; Fama and French, 1992) have documented a significant negative relation between B/M and equity returns. Fama and French (1992) jointly assess the impact of market beta, leverage, size, P/E and B/M on the NYSE, AMEX and NASDAQ average returns between 1963 and 1990. Their findings suggest that: 1) market beta is not a factor explaining average returns. 2) The combination of size and B/M is greater than the combination of leverage and P/E when explaining average returns.

Moreover, even after bringing former anomalies under control there are still some persistent ones such as the momentum effect. De Bondt and Thaler (1985) studied NYSE stocks and found that those stocks that suffered the greatest losses (gains) for 3-5 years obtained on average higher (lower) returns for a similar length period.

With a short-term perspective, Jegadeesh and Titman (1993) showed that strategies that buy past winner shares and sell past losers generated risk-adjusted positive returns for a 3-12 month tenure period ¹⁰.

¹⁰In other markets besides the US this anomaly is also observed. For example, Muga and Santamaría (2007), using non-parametric methods on the Spanish equity market show that winning portfolios stochastically dominate losing portfolios.

2.2.2. Response to market anomalies from conventional financial theory

In order to respond to this succession of anomalies compared to theoretical results, a new line of research focused on factor modelling. The use of factors helps to capture systematic risk beyond market risk (Koedijk et al., 2016) but has the disadvantage of just being ad hoc responses. They miss the formulation of a full theoretical framework. Among the early contributions Fama and French (1993) stand out with their three-factor model¹¹ that will set the pace for an increasing literature on factors¹².

However, while these formulations achieve noticeable improvements on explaining excess returns, they represent only partial solutions to specific market anomalies.

Simultaneously, it must be considered that these studies are referenced to the same series and similar time periods what might lead to accidental patterns and spurious results (Lo and MacKinlay 1990). Thereof, a clear example is the size effect. During the 15 years after its first publication smaller stocks have shown underperformance (Campbell, 2000).

All this said, factorial research in response to return anomalies has grown exponentially. For instance, Harvey et al. (2016) have recorded more than 300 factors discussed in academic publications.

But it has also been growing criticism around the possibility of empirically testing these models considering that the benchmark has an issue as well: there is no a universal definition of what to be considered as a market¹³.

¹¹These factors are: 1) the performance on a broad equity index; 2) the excess return from a portfolio of small capitalization over a large capitalization portfolio; 3) excess return of a high B/M vs a low B/M portfolio.

¹² For instance, Cahart (1997) building on Fama and French (1993) model adds a portfolio of stocks with high returns over the past 12 months as a proxy for momentum.

¹³ Roll (1977) maintained that only an index that would contain all risky assets (not only financial ones) could be considered as representative of the market.

Therefore, alternative explanations within the framework of the conventional financial theory have come out. They focus on the deviations of models compared to empirical evidence and on the need to achieve a market indicator as complete as possible.

Among these alternatives the one made by Jagannathan and Wang (1993, 1996) can be particularly highlighted. Besides the market factor they consider an additional factor linked to human capital (labor income) which improves traditional valuation models.

This approach has led to incorporate into the models, elements that come closer to agents' behavior with further definition of their preferences, heterogeneity (Campbell and Chocane, 1999; Miller, 2001¹⁴) or rationality (consumption-based capital asset pricing models or CCAPM – Engsted et al., 2000¹⁵).

While these formulations help to improve the explanatory power of the risk-return relationship they are not completely able to cope with some of the most significant anomalies¹⁶ and will point to the need of alternative explanations (De Long et al., 1990).

2.3. Alternative approach to the risk premium: behavioral finance

The consideration of both efficient markets (where prices reflect all available information) and rational investors (with homogeneous expectations) that cannot systematically beat the market is questionable (Fama and French, 2008). In real world, there are not many fully rational agents, and this implies the need to relax those assumptions embedded in

¹⁴Some of the fundamental assumptions relaxed are: 1) investors have the same expectations as to the expected returns. 2) Investors can make unlimited short sales with full use of the proceeds.

¹⁵ These authors, focusing on the Danish market, observe an improvement in explaining risk premium using these models which introduce factors linked to consumer preferences.

¹⁶ This is the case, for example, the momentum effect (Fama and French 1996) considering factors such as data processing or survival bias (Kothari et al., 1995).

conventional financial theory or as Thaler (2000, p.139) says "homo economicus will become more emotional".

Additionally, considering the anomalies previously described, they are prone to be more relevant in those phases of high sentiment when investors overestimate assets (Bams et al, 2015; Stambaugh et al., 2012).

Therefore, there is a clear need to open the financial analysis to new perspectives in which psychological and sociological factors must have a key role in explaining the evolution of financial assets beyond their exposure to systematic factors (Cochrane, 2011).

It is at this point where behavioral approaches emerge as an alternative to conventional theories. Behavioral finance deals not only with market constraints but with the fact that there are irrational investors who respond not only to utilitarian factors (Statman, 1999) but to changes in sentiment (De Long et al., 1990). Given the limits inherent to conventional theories in explaining market anomalies this line of research positions itself as a sound alternative. The interest in these explanations can be seen through the strong growth in the literature produced on behavioral issues in recent years (e.g., the number of citations on some of the most important studies on the subject has boomed- Table 2.1).

Table 2.1. Evolution of quotations from works in behavioral finance

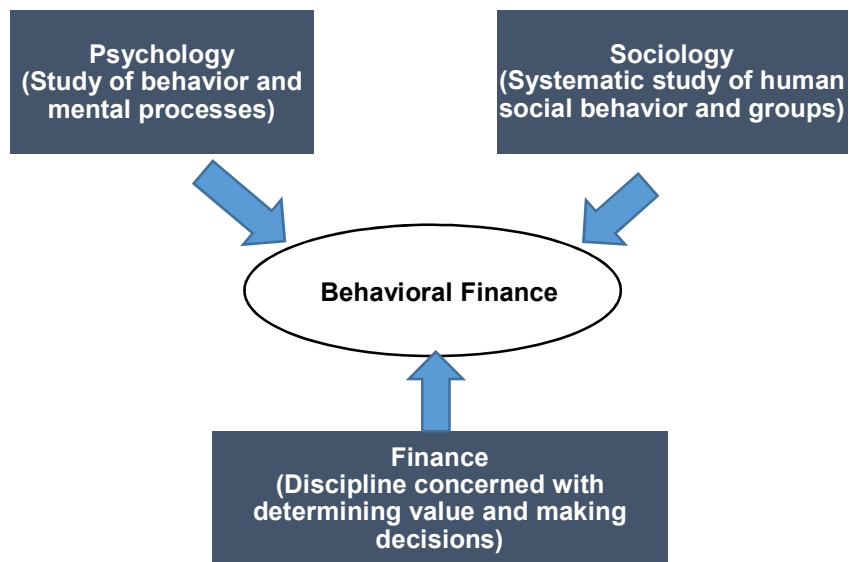
| Year | Author / s | Title | Dating to 2012 | Dating to 2015 |
|------|-------------------------|---|----------------|----------------|
| 1999 | Thaler | The End of Behavioral Finance | 282 | 432 |
| 2000 | Mullainathan and Thaler | Behavioral Economics | 342 | 545 |
| 2000 | Ricciardi and Simon | What is Behavioral Finance | 33 | 94 |
| 2000 | Shefrin | Beyond Greed and Fear: understanding Behavioral Finance and | 798 | 1255 |

| | | | | |
|------|------------------|--|------|------|
| | | the Psychology of Investing | | |
| | | Inefficient Markets: An Introduction to Behavioral Finance | 1796 | 2597 |
| 2000 | Shleifer | | | |
| | Barberis and | A Survey of Behavioral Finance | 1473 | 2373 |
| 2003 | Thaler | | | |
| | | Behavioral Finance and Wealth Management | 24 | 78 |
| 2006 | Pompian | | | |
| 2006 | Zaleskiewicz | Behavioral Finance | 5 | 18 |
| | | Behavioral Finance: Theories and Evidence | 9 | 17 |
| 2008 | Byrne and Brooks | | | |

Source: Park and Sohn (2013, p.32-34) and authors

Unlike conventional theories focused exclusively on asset dynamics, behavioral finance incorporates additional dimensions. As the matter of fact, it considers the presence of individuals in the market and how their decisions and interactions impact such assets (figure 2.1).

Figure 2.1. Disciplines that integrate behavioral finance



Source: Authors based on Ricciardi and Simon (2000, p.2)

However, when counter posing this approach to the dynamics of financial markets there is not a unified behavioral theory but several simultaneous explanations to the same problems seen from the perspective of classical financial theory (Barberis and Thaler, 2003).

Despite this initial heterogeneity, two major classifications can be set among these behavioral perspectives: 1) theories that focus on the limitations of classical financial theory; and, 2) theories that are based on psychology and highlight the key role of the agent idiosyncratic factors.

2.3.1. Theories focused on limitations of conventional financial theories

Within this first set of explanations, it is possible to make a more precise categorization:

1) Limits of arbitrage. This approach deals with the operational limitations of models and the relaxation of basic theoretical assumptions. Under this stand, the failure of traditional asset pricing models and the subsequent risk premium estimation are hand by hand with the deviations between empirical observations and the theoretical assumptions about how markets work.

The presence of transaction costs can be noticed among the most relevant explanations to this explanation. The presence of these costs introduces friction in some situations such: first, the possibility of risk-free arbitrage to terminate misalignments in asset prices (Singh, 2010). Secondly, the possibility to take unlimited short positions or even should these be feasible that they would be a perfect market risk hedge (Merton, 1987).

It also reveals the existence of both heterogeneity among market players and the presence of behaviors that are not consistent with expectations built on available information or equivalently the presence of segmented markets¹⁷.

2) Criticism of the fundamental assumptions of the conventional theory. This approach is based on the critique to the efficient markets' assumption. Accordingly, it cannot be considered that the market always reflects all available information as this is incompatible with the presence of patterns that recur over time and are not corrected (as the efficiency hypothesis states). However, that does not mean that there is no efficiency at all, but that it is limited in scope and linked to the presence of mechanisms that help to make the information available (Malkiel, 2003).

It might even be thought that the traditional model would not be invalidated but would only have to relax the assumption and consider market efficiency to be adaptative (Lo, 2004). This is, it evolves according to conditions and market participants. This interpretation would explain some distortions in models such as the instability of the risk-return relationship over long time horizons (Antoniou et al., 2016).

2.3.2. Theories with psychological fundamentals

Following these theories, the criticism regarding conventional financial theory focuses on the assumption of rationality¹⁸. In this new framework the drivers of investment decisions would be factors such as life experience (Hoffmann and Post, 2015), habit formation

¹⁷ Cochrane (2011) finds that average returns depend largely on idiosyncratic factors against the systematic component part of the classic explanation of the estimation of the risk premium.

¹⁸ Fama (2008) accepts certain element of irrationality although no to be extrapolated to the whole market.

(Hirshleifer, 2001) or the level of confidence (Kumaran, 2013). Then, further classification can be made in this current subset:

1) Heuristics and biases. Although broad theories¹⁹ start to build up, the fact is that most contributions are limited to: a) the study of different behavioral biases that can be observed in a group and then extrapolate the results to the entire population. b) The analysis of the effect of those biases on asset performance (table 2.2).

Anyway, although a theoretical framework for those biases is still missing, they are relevant when facing the limits of classical theories on explaining risk premiums.

Table 2.2. Most relevant behavioral biases

| Bias | Explanation |
|--------------------------|---|
| Confirmation | Search of selective information to support opinions or interpret the facts in such a way that fit a preconceived idea. |
| Availability / attention | Things that occur most frequently in the media will be remembered more quickly by investors when looking for a suitable investment instrument |
| Home bias | Most investors tend to buy stocks of their country of origin |
| Favorite long-short | People take the riskier strategy because it promises very high profits |
| Anchoring | Investors tend to base their decisions on the price at which the original position was taken |
| Myopic loss aversion | Most investors fear more losses than enjoy profits |
| Mental accounting | Investors make distinctions in their head that do not exist financially |
| Disposition effect | Profits are realized too early and losses too late |
| Overconfidence | Investors overestimate their own abilities |
| Hindsight bias | Give an explanation for everything that happened. Keep us from learning from our mistakes |
| Get-even-itis | After a loss, investors take greater risks to make up for it |
| Representativeness | After a brief period of positive returns, it is thought that the world has changed for better |
| Gambler's fallacy | Effective odds are widely over and underestimated |
| Framing | Decisions are based on how the facts are represented in statistical terms |

Source: Authors based on Hens and Meier (2014, p 16-17)

¹⁹ For example, the theory of planned behavior that demonstrates the importance of attitudes and biases in explaining and predicting investment behaviors (Pascual-Ezama et al., 2014).

For example, Mitroi and Oproiu (2014) analyzed the presence of biases in a group of individuals and how investment decisions were constrained by those biases. Results show that asset returns come their historical long-term average. This would be enough to justify the absence of random walk.

Additionally, Antoniou et al. (2016) found that in those situations where the market dynamics are strongly positive (high sentiment), overconfident investors increase trading underestimating risk. They conclude that in bear markets (low sentiment) with no overconfidence present, the weight of systematic factors increases and, therefore, the explanatory power of traditional valuation models improves.

2) Investor preferences. This is the predominant line within the psychological based methodologies banking on the work of Kahneman and Tversky (1979). These authors developed a new theoretical framework to explain asset price dynamics based on investors' expectations and concluded that investors are mostly risk averse. This risk aversion determines their investment behavior that could be estimated through a loss function. This feature would underscore how investors should be overcompensated to make decisions that would imply a potential loss.

Following this factor-modelling linked to investor preferences²⁰ a completely new research trend emerges focusing on investor idiosyncratic characteristics.

It is at this point that progress is made in setting a new classification of different research areas within the psychological theories. Thereof, this paper is builds on the work of Igual and Santamaría (2017) who, while performing an exhaustive classification of beliefs and heuristics point to a third alternative line of research focused on herd behavior. But, herd

²⁰ That means to dismiss the homogeneous expectations assumption embedded in conventional financial theory.

behavior at the end is no more than a single case of a larger caseload that can be grouped under the heading of investor sentiment (Moskowitz and Grinblatt, 1999).

Investor sentiment is not an easy concept to define especially considering that: a) sentiment is not an objective variable that can be measured directly (Gao, Ren and Zhang, 2016). b) Reflects heterogeneous expectations (Duch and Kellstedt, 2011).

There has been a surge in the literature focused on the estimation of investor sentiment. Within the research on sentiment analysis and following Heston and Sinha (2016) two major groups can be set: 1) direct measures. They try to quantify investor sentiment by studying investor characteristics (usually measured with surveys). However, in spite of the investor being the direct provider of information this method has characteristics that limit presents relevant constrains by the time of setting a broad sentiment measure. Among those potential failures for the survey-based methods we can highlight: a) timing (Da et al., 2015). b) Limited confidence at showing investor real intentions (Stivers, 2015). c) Most work conducted under this specification has shown poor explanatory power²¹.

2) Indirect measures. This line of research focuses on proxies that should reflect sentiment in aggregate (with estimators supported by market data). Although this is the way that shows greater growth in the literature also has limitations. Precisely, the fact that the sentiment is not observable directly (Gao et al., 2016) which makes it difficult to estimate.

Following this measurement problem two major subgroups can be identified: 1) models based on economic data. 2) Models based on search for information.

²¹ Heston and Sinha (2016) find that the University of Michigan consumer sentiment index can only explain a contemporaneous relationship with US financial assets and is not able to make good predictions.

These models have two common features: first, they are top-down models. That is, a measure of sentiment is constructed with the assumption that it is representative for the entire population (Baker and Wurgler, 2007).

Second, no single measure is considered as the true representation of sentiment among all alternatives coming out from any of the two subgroups (Corredor et al., 2013).

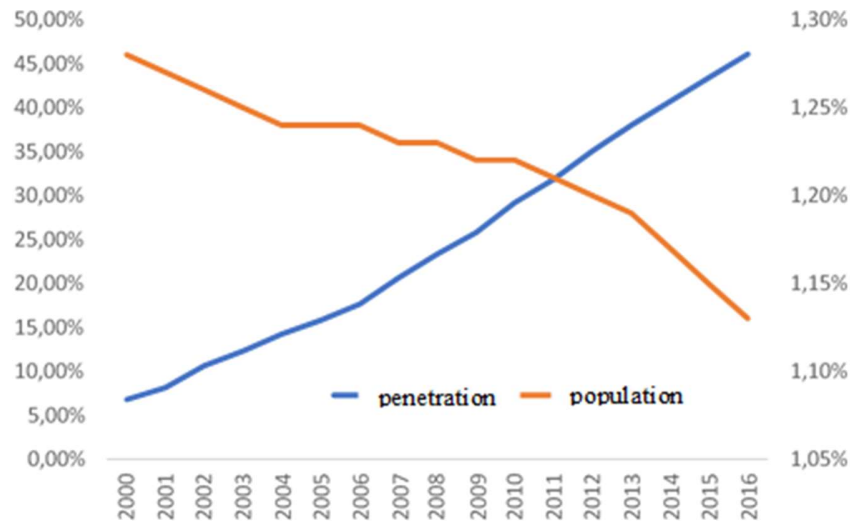
As for the differences, these are mainly methodological. Economic based models assume that sentiment reflects a set of expectations about the future and precisely about economic future. Therefore, sentiment measures are the result of a combination of variables linked to the economy with the assumption that these accurately reflect expectations (Baker and Wurgler, 2006).

By contrast, models based on the search for information take a different perspective. They consider that both the available information how this information affects the investor mindset are the key elements shaping investor sentiment (Hoffman and Post, 2015).

Accordingly, the key issue is capturing investor intentions properly. In this sense, technological advances especially in the field of information technologies (IT) contribute to achieving this goal. In fact, in recent decades, there has been a sharp increase in global social connectivity that allows the establishment of links between agents' intentions and their use of technology.

In this regard, Figure 2.2 shows how the internet penetration rates in the global population are showing an exponential growth in contrast to the evolution of population growth itself, reinforcing the idea of booming global networking.

Figure 2.2. Internet penetration rate (% global population) vs global population growth rate



Source: Authors based on Internet Live Stats

In this context, information technology fulfills a dual role:

- (i) First, it focuses on individuals as direct providers of information. This individually generated information exceeds any attempt to approach sentiment through aggregate economic factors. These are poor when trying to capture undisclosed intentions (Ghysels et al., 2007).
- (ii) Second, the widespread use of IT, both geographically and at user level allows the generation of aggregated sentiment indicators as a direct result of individual interactions. Those measures can be considered as representative of investors' real intentions or preferences. That is, collective behavior can be obtained from the study of individual data generated using IT technologies (Dietzel et al, 2014; Bentley et al, 2014; Curme et al, 2014.). Also, this can be done at a global dimension to the extent that the speed of transmission of information is increasingly higher as time goes (Hirshleifer, 2015).

In turn, three main classifications can be singled out within the search for information-based models depending on the source of information (Ranco et al, 2015.):

a) News and media analysis. These basically look for evidence of investor sentiment through positive or negative biases in the news or in different media, according to a predefined lexicon. Most studies in this field try to prove the presence of a relationship between that lexicon in the news and the returns on financial assets (Tetlock, 2007; Amin and Ahmad, 2013; Lee and Mauck, 2014; Yook, 2010). Precisely, some studies find a positive relationship between those stocks that have media coverage and their returns (Fang et al., 2014; Barber and Odean, 2008; Hirshleifer, et al., 2011).

b) Social networks. This category focuses on the analysis of interactions in social networks as a proxy for investor sentiment. Overall, the studies based on the use of social networks search a relationship between the number or content of the messages and their impact on asset returns. Most of the results on this area of study show a negative relationship between asset returns and peak volume of messages (Dickinson and Hu, 2015) or between the social network activity and trading volumes (Antweiler and Franck, 2004).

However, social networks have two important drawbacks as well: 1) consistency. Users tend to switch networks quite frequently and there is also a significant number of inactive or false profiles that can distort results. 2) Text analysis. Creating a good lexicon is crucial to make correct interpretations of sentiment in textual analysis (Loughran and McDonald, 2011)²².

c) Information search engines. This approach focuses on the relationship between the number of searches for specific terms and the subsequent reaction of economic and

²² According to this work, three-quarters of the words of annual reports of US companies, compared to a negative interpretation according Harvard dictionary, tend to have a positive meaning in a financial context.

financial variables. Google has become the dominant tool in this search for information²³ (Artola and Galán, 2012; Choi and Varian, 2009; Dimpfl and Jank, 2016; Da et al., 2011, 2015, Preis et al, 2013.).

Methodologies grounded on web searches do not face text analysis problems and show larger population coverage²⁴. This should be more efficient in revealing preferences or information about future intentions. In addition, the tools these models use can detail the scope of searches in an area of specific knowledge favoring further precision at estimating investor sentiment (e.g. Finance).

The literature shows mixed results when using web searches to set a link between sentiment and future asset performance. For instance, Preis et al (2013) found no significant relationship between stock returns and searches. On the other hand, they succeeded at showing predictive capabilities regarding trading volumes and by the time of setting correlation with periods of uncertainty, when information becomes scarce (Kristoufek, 2013). But the period of study might a crucial factor to be considered²⁵ and those results should not be taken for granted.

Moreover, Da et al. (2011, 2015) found a significant relationship between Google searches and stock returns. They create a sentiment index based on search volumes and find a negative relationship between sentiment and future performance. The most recent studies on the subject reaffirm this negative relationship between information searches and future asset performance (Gao et al, 2016. Dietzel et al, 2014. Bijl et al, 2016. Fricke et al, 2014.).

²³ The relevance of Google is evident if one considers that, according to StatCounter Global Statistics (www.gsstatcounter.com) it represents 91.5% of the search engine global market share by October 2017.

²⁴ While a social network like Twitter reports 320 million users, 3570 million users of internet services are accounted for (according to figures from Livestats by the end 2016).

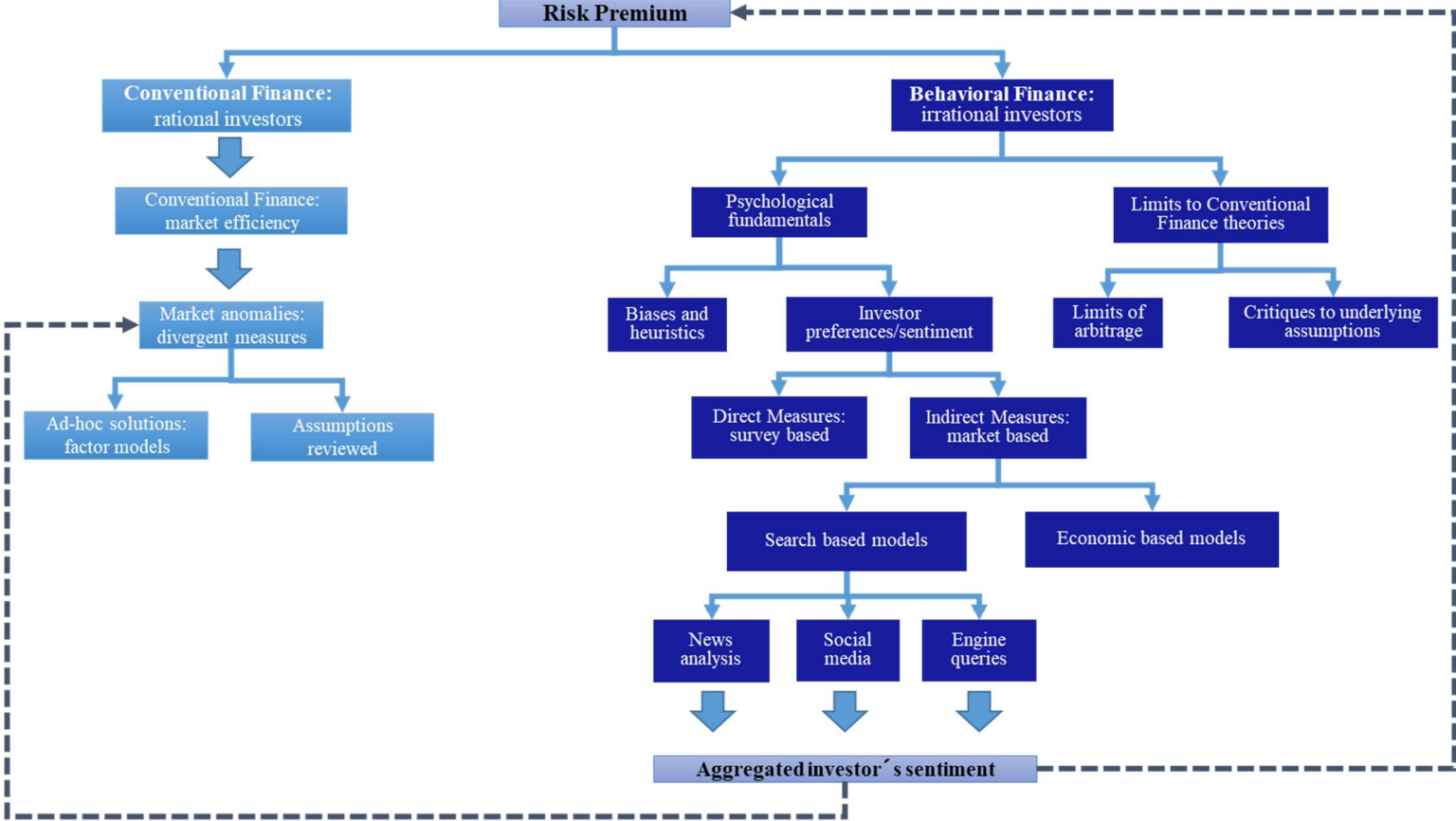
²⁵ Much of the literature on the subject uses data up to 2008 and has been a sharp rise in internet penetration thereafter (Livestats estimated penetration of 23% of the total population in 2008 vs. 46% in 2016).

According to previous evidence, a definition of sentiment is set on the grounds of the search of information: set of expectations about future asset returns that differ from its fundamental value estimated with public available information. Therefore, these expectations incorporate an undisclosed information content that might be captured using information technologies.

All in all, and taking risk premium as the compensation required by investors for taking systematic risk (Gagliardini et al, 2016), there are two consequences for the risk premium when considering the use of informational/sentimental factors on its estimation (figure 2.3):

- 1) It would contribute to increase market efficiency as previously undisclosed information is brought in, granting better identification of the systematic risk component which in turn should improve the risk premium estimates.
- 2) It favors larger control of the market anomalies that, ultimately reflect the lack of sufficient information attached to market prices. Thus, incorporating behavioral factors into the risk premium analysis contributes to reconcile the relationship between market and efficiency as "market is remarkably efficient in its utilization of information" Malkiel (2003, p. 80) and ends up with any anomaly or pattern when additional information is incorporated.

Figure 2.3. Risk premium concept map



2.4. Conclusions

This paper contributes to the literature on risk premium in two ways: first, a concept map is made, showing the limitations of classical theories and the potential benefits of the behavioral contribution as an alternative or at least complementary explanation.

As the matter of fact, the application of behavioral finance to risk premium analysis ends up with a greater control of market anomalies mostly due to the constraints of classical theories, precisely, the assumptions of efficient markets and rational investors. Thus, the incremental information provided by behavioral explanations contributes to increase market efficiency and favors better risk premium estimates.

Secondly, focused on behavioral finance this paper digs further on the heterogeneity of its psychological alternatives. Therefore, building on previous classifications (e.g. Igual and Santamaria, 2017), research on investor sentiment will be mapped in greater detail.

This new classification will emphasize on the informational content (undisclosed information on investor's intentions) as the key driver for shaping investor sentiment.

This will be even lead to an alternative definition of investor sentiment based on information processing as the key element (Bank and Brustbauer, 2014).

In order to make that information available we will focus on the use of information technologies and investor interactions. Increasing global networking will lead to the identification of a proper global collective behavior or global sentiment that will be based on individual interactions using the same technology around the world. Then, as undisclosed information on aggregate investor sentiment gets into the markets (Da et al., 2011) it will contribute to a greater market efficiency as previously noted and finally to get better risk premium estimates than those coming out of information provided only by asset prices.

References

- Amin, R., & Ahmad, H. (2013). Does Investor Attention Matter ' S ? *Journal of Public Administration, Finance and Law*, (4), 111–125.
- Antoniou, C., Doukas, J. A., & Subrahmanyam, A. (2016). Investor Sentiment, Beta, and the Cost of Equity Capital. *Management Science*, 62(2), 347-367.
- Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of Internet stock message boards. *Journal of Finance*, 59, 1259–1294.
- Arbel, A. & Strebel, P. (1982). The neglected and small firm effects, *Financial Review*, 17, 201-218.
- Artola, C., & Galán, E. (2012). Tracking the Future on the Web: Construction of Leading Indicators Using Internet Searches. <https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosOcasionales/12/Fich/do1203e.pdf>
- Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, LXI(4), 1645–1680.
- Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives, American Economic Association*, 21, 129–152.
- Ball, R. (1978). Anomalies in relationships between securities' yield and yield-surrogates. *Journal of Financial Economics*, 6, 103-126.
- Bams, D., Honarvar, I., & Lehnert, T. (2015). Towards Disentangling Sentiment from Risk Premia. http://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2015-Amsterdam/papers/EFMA2015_0128_fullpaper.pdf.
- Bank, M., & Brustbauer, J. (2014). Investor Sentiment in Financial Markets. <https://www.uibk.ac.at/ibf/mitarbeiter/bank/investor-sentiment-in-financial-markets.pdf>.
- Banz, R.W. (1981). The relationship between return and market value of common stock. *Journal of Financial Economics*, 9, 3-18.
- Barber, B. M., & Odean, T. (2008). All that Glitters: The effect of Attention and news on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, 21(2), 785-818.
- Barberis, N., & Thaler, R.H., (2003). A survey of behavioral finance. In Constantinides, G., Harris, M., & Stulz, R. (Eds.), *Handbook of the Economics of Finance* (pp.1052-1090), Amsterdam: North-Holland.
- Basu, S. (1977). Investment performance of commons stocks in relation to their price/earnings ratios: a test of the efficient market hypothesis. *Journal of Finance*, 32, 663-682.
- Bentley, R. A., O'Brien, M. J., & Brock, W. A. (2014). Mapping collective behavior in the big-data era. *The Behavioral and Brain Sciences*, 37(1), 63–76.

- Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, 45, 150–156.
- Black, F. (1972). Capital market equilibrium with restricting borrowing. *Journal of Business*, 45, 444-455.
- Black, F. & Scholes, M. (1972). The Capital Asset Pricing Model: some empirical tests. In M. Jensen (ed.), *Studies in the Theory of Capital Markets*, Praeger, 451- 471.
- Brealey, R., & Myers, S. (2000), *Principles of Corporate Finance*, 6th. Ed., Irwin.
- Brown, S. Goetzmann, W.N., & Ross, S.A. (1995). Survival. *Journal of Finance*, 50, 853-874.
- Cahart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52, 57-82.
- Campbell, J., & Cochrane, J. (1999). By force of habit: a consumption-based explanation of aggregate stock market behaviour. *Journal of Political Economy*, 107, 205-251.
- Campbell, J. Y. (2000). Asset Pricing at the Millenium. *Journal of Finance*, 55(4), 1515-1567.
- Choi, H., & Varian, H. (2009). Predicting the Present with Google Trends. https://static.googleusercontent.com/media/www.google.com/es//googleblogs/pdfs/google_predicting_the_present.pdf.
- Cochrane, J. H. (2011). Presidential Address: Discount Rates. *Journal of Finance*, 66(4), 1047–1108.
- Copeland, T., & Weston, J. (1992). *Financial Theory and Corporate Policy*. Addison-Wesley Publishing Co.
- Corredor, P., Ferrer, E., & Santamaria, R. (2013). Investor sentiment effect in stock markets: Stock characteristics or country-specific factors? *International Review of Economics and Finance*, 27, 572–591.
- Curme, C., Preis, T., Stanley, H. E., & Moat, H. S. (2014). Quantifying the semantics of search behavior before stock market moves. *Proceedings of the National Academy of Sciences of the United States of America*, 111(32), 11600–11605.
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *Journal of Finance*, 66, 1461–1499.
- Da, Z., Engelberg, J., & Gao, P. (2015). The Sum of All FEARS Investor Sentiment and Asset Prices. *Review of Financial Studies*, 28(1), 1–32.
- De Bondt, W., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, 40, 793-805.
- De Bondt, W., & Thaler, R. (1987). Further evidence on investor overreactions and stock market seasonality. *Journal of Finance*, 42, 557-581.

- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4), 703-738.
- Dickinson, B., & Hu, W. (2015). Sentiment Analysis of Investor Opinions on Twitter. *Social Networking*, 04(03), 62–71.
- Dietzel, M. A., Braun, N., & Schäfers, W. (2014). Sentiment-Based Commercial Real Estate Forecasting with Google Search Volume Data. *Journal of Property Investment & Finance*, 32(6), 540-569.
- Dimpfl, T., & Jank, S. (2016). Can internet search queries help to predict stock market volatility? *European Financial Management*, 22(2), 171–192.
- Dimson, E., Marsh, P., & Staunton, M. (2003). Global Evidence on the Equity Risk Premium. *Journal of Applied Corporate Finance*, 15(4), 27-38.
- Duch, R. M., & Kellstedt, P. M. (2011). The heterogeneity of consumer sentiment in an increasingly homogenous global economy. *Electoral Studies*, 30(3), 399-405.
- Elton, E. J., & Gruber, M.J. (1995). *Modern Portfolio Theory and Investment Analysis*. 5th Ed., John Wiley & Sons Inc.
- Engsted, T., Mammen, E. & Tanggaard, C., (2000). Evaluating the C-CAPM and the equity premium puzzle at short and long horizons: a markovian bootstrap approach. https://econpapers.repec.org/paper/hhbaarfin/2000_5f010.htm.
- Estrada, J. (2006). The Fed model: a note. *Finance Research Letters*, 3, 14-22.
- Fama, E. (1970). Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, 25, 383-417.
- Fama, E., & McBeth, J. (1973). Risk, return and equilibrium: empirical tests. *Journal of Political Economy*, 71, 607-636.
- Fama, E., & French, K. (1992). The cross section of expected stock returns. *Journal of Finance*, 47, 427-465.
- Fama, E., & French, K. (1993). Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics*, 33, 3-56.
- Fama, E., & French, K. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51, 55-84.
- Fama, E., & French, K. R. (2002). The Equity Premium. *The Journal of Finance*, 57(2), 637-659.
- Fama, E., & French, K. R. (2008). Dissecting Anomalies. *Journal of Finance*, 63(4), 1653-1678.
- Fang, L. H., Peress, J., & Zheng, L. (2014). Does media coverage of stocks affect mutual funds' trading and performance? *Review of Financial Studies*, 27(12), 3441-3466.
- Fricke, E., Fung, S., & Goktan, M. (2014). Google search, information uncertainty, and post-earnings announcement drift. *Journal of Accounting and Finance*, 14(2), 11–27.

- Gagliardini, P., Ossola, E., & Scaillet, O. (2016). Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets. *Econometrica*, 84(3), 985–1046.
- Gao, Z., Ren, H., & Zhang, B. (2016). Googling Investor Sentiment around the World. http://www.cuhk.edu.hk/fin/event/symposium/SEFM_2016_paper_58.pdf.
- Ghysels, E., Plazzi, A., & Valkanov, R. (2007). Valuation in US commercial real estate. *European Financial Management*, 13, 472–497.
- Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the Cross-Section of Expected Returns. *Review of Financial Studies*, 29(1), 5-68.
- Heston, S., & Sinha, N. (2016). News versus sentiment: predicting stock returns from news stories. <http://dx.doi.org/10.17016/FEDS.2016.048>.
- Hens, T., & Meier, A. (2014). Behavioral Finance: The Psychology of Investing. https://www.researchgate.net/publication/283488173_Behavioral_Finance_The_Psychology_of_Investing.
- Hirshleifer, D. (2001). Investor psychology and asset pricing. *Journal of Finance*, 56(4), 1533-1597.
- Hirshleifer, D. (2015). Behavioral Finance. *Annual Review of Financial Economics*, 7(1), 133–159.
- Hirshleifer, D. a., Lim, S. S., & Teoh, S. H. (2011). Limited Investor Attention and Stock Market Misreactions to Accounting Information. *Review of Asset Pricing Studies*, 1, 35-73.
- Hoffmann, A. O. I., & Post, T. (2015). How return and risk experiences shape investor beliefs and preferences. *Accounting and Finance*, 57(3), 759-788.
- Ibbotson, R., & Cheng, P. (2003). Long-Run stock returns: participating in the real economy. *Financial Analysts Journal*, 59(1), 88-98.
- Igual, M.G., & Santamaría, T.C. (2017). Overconfidence, loss aversion and irrational investor behavior: a conceptual map. *International Journal of Economic Perspectives*, 11(1), 273-290.
- Jagannathan, R., & Wang, Z. (1993). The CAPM is alive and well. <https://www.minneapolisfed.org/research/staff-reports/the-capm-is-alive-and-well>.
- Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross section of expected returns. *Journal of Finance*, 51, 3-53.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: implication for stock market efficiency. *Journal of Finance*, 48, 65-92.
- Kahneman, D., & Tversky, A. (1979). Kahneman & Tversky (1979). Prospect Theory - An Analysis Of Decision Under Risk. *Econometrica*, 47(2), 263-291.
- Keim, D.B. (1988). Stock market regularities: a synthesis of the evidence and explanation, in E. Dimson (ed.), *Stock Market Anomalies*, Cambridge University Press, 16-39.

- Koedijk, K., Slager, A., & Stork, P. (2016). A Trustee Guide to Factor Investing. *The Journal of Portfolio Management*, 42(5), 28-38.
- Kothari, S., Shanken, J., & Sloan, R. (1995). Another look at the cross section of expected stock returns. *Journal of Finance*, 50, 185-224.
- Kristoufek, L. (2013). Fractal markets hypothesis and the global financial crisis: wavelet power evidence. *Scientific Reports*, 3 (2857), 1-7.
- Kumaran, S. (2013). Impact of Personal Epistemology, Heuristics and Personal Attributes on Investment Decisions. *International Journal of Financial Research*, 4(3), 62-69.
- Lee, B. S., & Mauck, N. (2014). Information Asymmetry and the Market Response to Open Market Share Repurchases. https://mpra.ub.uni-muenchen.de/54066/1/MPRA_paper_54066.pdf.
- Lintner, J. (1965). The valuation of risk assets and the selection of risky investment in stock portfolios and capital budgets. *Review of Economic Statistics*, 47, 13-37.
- Lo, A., & MacKinley, C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies*, 3, 175-205.
- Lo, A. W. (2004). The Adaptive Markets Hypothesis. *The Journal of Portfolio Management*, 30(5), 15–29.
- Loughran, T., & McDonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35–65.
- Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17(1), 59-82.
- Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2), 145-161.
- Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *Journal of Finance*, 42(3), 483-510.
- Mitroi, A., & Oproiu, A. (2014). Behavioral finance : new research trends , sociomics and investor emotions. *Theoretical and Applied Economics*, 21(4), 153–166.
- Moore, C., & College, E. (2003). Unraveling the size effect. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.603.7906&rep=rep1&type=pdf>.
- Moskowitz, T. J., & Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance*, 54(4), 1249-1290.
- Mossin, J. (1966). Equilibrium in capital asset market. *Econometrica*, 34, 768-783.
- Muga, L., & Santamaría, R. (2007). Riesgo asimétrico y estrategias de momentum en el mercado de valores español. *Investigaciones económicas*, 31(2), 323-340.
- Park, H., & Sohn, W. (2013). Behavioral Finance : A Survey of the Literature and Recent Development. *Seoul Journal of Business*, 19(1), 3-42.

- Pascual-Ezama, D., Scandroglio, B., & Gil-Gomez de Liaño, B. (2014). Can we predict individual investors' behavior in stock markets? A psychological approach. *Universitas Psychologica*, 13(1), 25-35.
- Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, 3(1684), 1-6.
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *PloS One*, 10(9), 1-21.
- Reinganum, M. (1981). A misspecification of capital asset pricing: empirical anomalies based on earning yields and market values. *Journal of Financial Economics*, 9, 19-45.
- Ricciardi, V., & Simon, H. K. (2000). What is Behavioral Finance ?, *Business, Education & Technology Journal*, 2(2), 1-9.
- Rietz, T. (1988). The equity risk premium: a solution. *Journal of Monetary Economics*, 22(1), 117-131.
- Roll, R. (1977). A critique of the asset pricing, theory's test, part1: on past and potential testability of the theory. *Journal of Financial Economics*, 4, 129-176.
- Roll, R. (1983). Was it Das? The turn-of-the-year effect and the return premia of the small firms. *Journal of Portfolio Management*, 9, 18-28.
- Sharpe, W. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, 19, 425 – 442.
- Shefrin, H. (2015). Investors' judgments, asset pricing factors and sentiment. *European Financial Management*, 21, 205-227.
- Singh, R. (2010). Behavioural finance studies: emergence and developments. *Journal of Contemporary Management Research*, 4(2), 1-9.
- Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2), 288-302.
- Stattman, D. (1980). Book values and expected stock returns. *The Chicago MBA: A Journal of Selected Papers*, 4, 25-45.
- Statman, M. (1999). Behavioral Finance: Past Battles and Future Engagements. *Financial Analysts Journal*, 55(6), 18-27.
- Stivers, A. (2015). Forecasting Returns with Fundamentals-Removed Investor Sentiment. *International Journal of Financial Studies*, 3(3), 319-341.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62, 1139-1168.
- Thaler, R. H. (2000). From Homo Economicus to Homo Sapiens. *Journal of Economic Perspectives*, 14(1), 133-141.
- Treynor, J. L. (1965). How to Rate Magement of Investment Funds. *Harvard Business Review*, 43(1), 63-75.

Yook, K. C. (2010). Long-run stock performance following stock repurchases. *Quarterly Review of Economics and Finance*, 50, 323–331.

3. How information technologies shape investor sentiment: a web-based investor sentiment index

Juan José García Petit, Esther Vaquero Lafuente, Antonio Rúa Vieites*

Abstract

This paper proposes a new investor sentiment indicator that combines the use of principal component analysis with web searches. This proposal provides economic meaning to the underlying variables, a sound factor structure, and reduces the noise regarding to web searches, when compared to standard search-based sentiment indicators. In fact, our indicator not only confirms the relevance of sentiment for future assets performance and provides greater predictive capacities than standard formulations, but also generates new insights in terms of globalization of investor sentiment and the role that information flows and technology play on that process. Moreover, it challenges some general beliefs present in the literature of sentiment such as the prevalence of a local bias, the greater impact of sentiment in developed markets or the fact that institutional investors are not sensitive to sentiment. Finally, an investment strategy is implemented showing how a sentiment-based investment rule generates above-market returns.

Keywords: investor sentiment, information technology, globalization, search engines, principal components.

JEL: G41, G12, O33

* *Borsa Istanbul Review* (2019), 19 (2), p. 95-105. <https://doi.org/10.1016/j.bir.2019.01>.

3.1. Introduction

Sentiment has become a prominent feature of financial markets literature (Park and Sohn, 2013), explaining from investor positioning (Frazzini and Lamont, 2008) to market anomalies (Stambaugh et al. 2012).

However, setting a definition of investor sentiment is not an easy task, due especially to two elements: 1) sentiment is not an objective and directly measurable variable (Gao, Ren, and Zhang, 2016), and 2) sentiment comprises a heterogeneity of expectations (Duch and Kellstedt, 2011).

In this paper, investor sentiment will be considered as the set of non-revealed information that explains the difference between expected future assets returns and their fundamental values built on publicly available information.

In accordance with this definition, revealing investor's undisclosed information becomes the key element to set any viable measure of investor sentiment (Bank and Brustbauer, 2014). Therefore, the main objective of this paper will be to create a broadly representative measure of investor sentiment founded on non-revealed information about agents' future intentions.

Given that the aim of the sentiment indicator should be to reveal as much information as possible about investor's future intentions, this will be made by making an active use of information technologies to extract new data from agents. The internet has become the greatest source of information, and it has the advantage of being easily and globally accessible. These characteristics facilitate the generation of massive data sets from the web that might be considered representative of collective behavior (Dietzel et al., 2014). Precisely, this paper will rely on the literature on search-based models of sentiment (Heston and Sinha, 2016). These models focus on the agents as direct providers of information and assume that only data coming from information technologies (and no

other economic proxies) can fully reflect all the information needed to shape future intentions (Ghysels et al., 2007).

Following this line of thought, at least three large classifications of search-based models can be set based upon their source of information (Ranco et al., 2015):

1) News and media analysis. This large category basically searches for evidence of investor sentiment through the signal on the news or presence in the media of a pre-defined lexicon. Most of the research conducted on this topic tries to demonstrate the existence of a relationship between the presence of that lexicon in the news and subsequent financial asset returns (Tetlock, 2007; Amin and Ahmad, 2013; Lee and Mauck, 2014; Yook, 2010). In fact, several studies find a positive relationship between those stocks that show media coverage and their future returns (Fang, et al., 2014; Barber and Odean, 2008; Hirshleifer et al., 2011).

2) Social media. This category analyzes the interactions in social networks as a proxy for investor sentiment. Most of the studies based on the use of social networks search a relationship between the number or content of the messages and their impact on future assets returns. Accordingly, most of the results achieved on this area show a negative relationship between asset returns and peak volume of messages (Dickinson and Hu, 2015) or between the social network activity and trading volumes (Antweiler and Frank, 2004). Twitter outstands as the main tool for social media analysis.

These two categories present one element in common: they draw on text analysis to analyze sentiment. Therefore, the creation of a good lexicon is crucial to make correct interpretations of that sentiment. In fact, most of the studies in these areas tend to consider a set of items coming from popular economic dictionaries, precisely the Harvard IV and Laswell dictionaries, mostly following Tetlock (2007). However, as Loughran and McDonald (2011) show, popular dictionaries of terms are not the best instruments.

According to their study, three-quarters of word counts in 10-k filings based on the Harvard dictionary are typically not negative in a financial context, showing a wrong bias. Therefore, term selection and the need to pre-define a sentiment signal to each of them can be consider one flaw to use these methods.

Additionally, and related to the social media models, there is another drawback that might be noted: consistency. Users tend to switch networks quite frequently and there is also a significant number of inactive or false profiles that can distort results.

Finally, there is the question of how these categories can stand for a proper representation of the intentions of the whole population. Precisely, Tetlock (2007) focuses his study of sentiment on the analysis of news in the Wall Street Journal and Ranco et al. (2015) analyzing the effect of Twitter comments on financial markets consider the content of 1.5 million tweets. These magnitudes and scope of sources of information seem limited when trying to achieve a global dimension of sentiment²⁶.

Therefore, and given the limitations of the previous search-based methodologies, we opt for the third option in this classification: search-engine queries. This approach looks for the existence of a relationship between the number of queries of specific terms and the reaction (present or future) of economic or financial variables. Google has become the most prominent tool used here²⁷ (Artola and Galán, 2012; Choi and Varian, 2009; Dimpfl and Jank, 2016; Da et al., 2011, 2015; Preis et al., 2013).

However, most of the literature on queries has shown inconclusive results so far when trying to set a sound relationship between sentiment and future asset performance. In fact, Preis et al. (2013) use Google searches to gather information about specific company

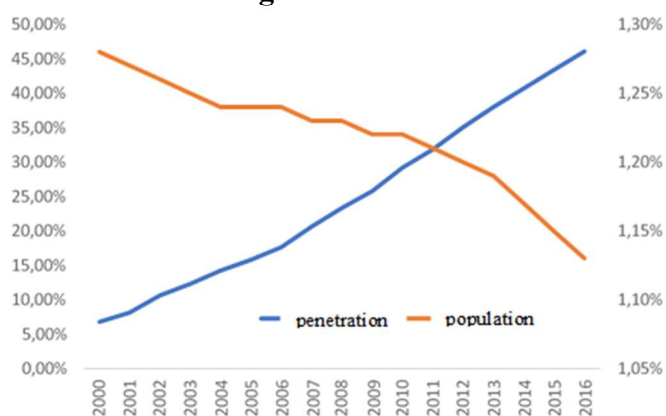
²⁶ While a social network like Twitter reports 320 million users, 3570 million users of internet services are accounted for (according to figures from Livestats by the end 2016).

²⁷ The popularity of this tool is evident considering that Google holds 91.5% of the search engine market share worldwide as of October 2017 (according to Statcounter Global Statistics, www.gsstatcounter.com).

names, and although they succeed at showing their predictive capabilities on trading volume and at setting correlations with periods of uncertainty (Kristoufek, 2013), they do not find a significant relationship between stock returns and those searches.

However, the period of study might be an issue when explaining this, *a priori*, weak evidence. As the matter of fact, most original studies in this approach run data up to 2008, and there has been a surge in the penetration of the internet in more recent times. In this regard, internet penetration rates in the global population are showing an exponential growth in contrast to the evolution of population growth itself, reinforcing the idea of booming global networking, particularly in the last decade (see Figure 3.1). This evidence reinforces the establishment of links between agents' intentions and their use of technology.

Figure 3.1. Internet penetration rate (% global population) vs global population growth rate



Source: Authors based on Internet Live Stats

Following this idea, Curme et al. (2014) find a link between changes in the number of views of financial topics and subsequent large stock market moves. Similarly, Da et al. (2011, 2015) find a relation between Google searches and stock returns. They go even further and build a sentiment index based solely on search volumes, finding a negative relationship between sentiment and subsequent returns. Most recent studies conducted on

this subject reaffirm the negative relationship between internet searches and future market returns (Gao et al., 2016; Dietzel et al., 2014; Bijl et al., 2016; Fricke et al., 2014).

Given this evidence, this paper makes a methodological contribution to the literature on search-based models of sentiment by combining factor analysis with web searches. The use of factor analysis has been widely used in the financial literature to obtain different indicators. For instance, Asness et al. (2013), use the first principal component of a set of variables linked to market and funding liquidity to generate a liquidity risk indicator. Similarly, focusing on the estimation of sentiment, the use of this methodology has been widely used by economic-based measures of sentiment. In fact, Baker and Wurgler (2006) use the first principal component of a set of economic variables as the proxy for investor sentiment.

However, the problem with these economic-based approaches is that they require an assumption about the information that those variables convey. Therefore, a significant amount of noise can be present in the final indicator (e.g. Baker and Wurgler (2006) only get 49% of total variance explained by the first principal component of the variables they consider in their sentiment indicator). Sibley et al. (2016), indeed, note that the power of the Baker and Wurgler's indicator might be related to market conditions or economic fundamentals embedded in the variables under consideration. Similarly, Stivers (2015) and Huang et al. (2015), improve Baker and Wurgler's indicator by removing fundamental components in order to reduce the noise.

Otherwise, search-based indicators of sentiment do not usually use factor analysis in their construction²⁸. Most relevant measures in this line of research (Da et al., 2011; Gao et al., 2016), focus directly on considering the search values of different terms previously selected in accordance to a pre-set sentiment bias and making composites straight from

²⁸ Dimpfl and Jank (2016), consider principal components to identify relevant variables that could be linked to sentiment for the German market, albeit aiming to explain realized volatility.

those values. Unfortunately, this methodology implies the presence of term selection bias, the lack of a proper structure and no economic meaning behind those terms finally selected for the construction of the sentiment indicator.

Therefore, the combination of principal components with searches would provide a solution of these problems by providing a factor structure to the sentiment indicator avoiding the traditional overfitting problems attached to economic-based models (Novy-Marx and Velikov, 2016). Additionally, the identification of those components that are truly relevant to isolate pure investor sentiment would help also to provide economic sense to the constituents of the sentiment indicator.

Moreover, another breakthrough of this paper is the proposal a different procedure of term selection than usually applied in the literature. Given the drawbacks related to the use of standard dictionaries, we will build on the list created by Preis et al. (2013), based on financial criteria with some additional contributions when needed. The use of a term selection with a clear financial bias should help to achieve a more precise sentiment measure. Additionally, the fact that the searches are constrained to a financial dimension guarantees not only that interest in a particular term is not biased by its interest in a different category (that could be larger), but that the common nexus among all the terms should be the agents' intentions (or sentiment) behind those searches.

Lastly, the use of global searches will scale up sentiment beyond what is the common practice in the literature: the addition of local sentiment measures (Baker et al., 2012; Gao et al., 2016). Given the global origin of the data, any measure built on it might be thought of being truly representative of global agents' intentions (Dietzel et al., 2014).

In fact, should this indicator be fully representative of global investor sentiment then it should guarantee the existence of a relationship between information diffusion and

investor intentions albeit with a higher explanatory power and economic foundation than other search-based models.

Moreover, given the globalization of information flows (and the impact on the global investors mindset), a global measure should be superior to the one based only on local information, thus invalidating the local bias prevailing in the literature.

Given all these elements, the paper is structured as follows: section 3.2 will develop the methodology for the generation of an investor sentiment index with data from digital search engines. In section 3.3, the investor sentiment index will be empirically tested against market returns defining relationships and significances and its global capabilities. In section 4, a practical application of our sentiment index through a trading strategy will be shown as well. Finally, section 5 will summarize and conclude.

3.2. Construction of a web-based investor sentiment index

The main objective of this paper is the creation of an index broadly representative of investor sentiment and based on data provided directly by agents through their use of information technologies. To achieve this target, the most important aspect is to work with data that can effectively capture expectations. The analysis takes monthly data from Google Trends (<https://www.google.com/trends>) for the period ranging from January 2004 to May 2017. This tool provides a search volume index (SVI) of queries globally or in specific geographies. These data are not provided on the number of searches but are instead scaled into an index from 0 to 100 to show the degree of popularity of a search in the whole sample, with 100 being the highest search volume for an item. Additionally, Google allows the possibility of setting search categories of the terms to be studied. Then, because this work focuses on investor sentiment, the search will be restricted to the “finance” category. This search will be performed on two geographies, US and Global.

The next step is to define the items whose SVI will be searched. Given the already mentioned limitations coming from popular dictionaries used in text analysis (Loughran and McDonald, 2011), this paper will follow Preis et al. (2013), who have already defined a comprehensive selection of words with a financial bias.

This analysis will consider 72 keywords²⁹ mostly based on that selection after removing those elements that do not show a clear definition or have insufficient queries and including some additional terms related to fixed income assets (precisely, “bankruptcy”, “yield” and “capital”) to avoid an equity bias. We will also substitute some keywords for others more representative of the market (e.g. “Dow Jones” has been exchanged for “S&P500” as to be more representative of the market benchmark- Shoven and Sialm, 2000). The exact same terms will be searched for the US and the Global geographic dimensions.

To be able to extract as much relevant information as possible, the SVI will be taken in levels instead of changes, as commonly seen in the literature (Gao et al., 2016). Additionally, a Box-Jenkins methodology will be applied to the items to obtain cleaner time series, controlled for stationarity and seasonality effects when necessary (Ngo, 2013).

3.2.1. Extracting sentiment from searches

Once proper data generation processes have been identified for each SVI, the objective is to extract information from these series that can essentially be considered as investor sentiment.

²⁹ List of terms searched on Google Trends: debt, stocks, restaurant, portfolio, inflation, housing, revenue, bankruptcy, credit, yield, unemployment, growth, investment, hedge, wedding, divorce, bonds, derivatives, profit, leverage, loss, cash, office, fine, S&P500, banking, financial crisis, happy, car, capital, finance, short sell, invest, fed, travel, expected return, gain, default, water, rich, risk, oz. of gold, success, oil, war, economy, lifestyle, greed, food, movie, ore, hold, opportunities, health, short sell, arts, culture, bubble, purchase, tourism, politics, energy, consumption, dividend, conflict, forex, home, crash, transaction, fond, work, fun.

In contrast to usual practice in search engine queries-based models (Da et al., 2015; Gao et al., 2016) that use SVI series directly to build the sentiment index, a factor analysis will be applied to the previously treated SVIs. Precisely, a principal component analysis (PCA) will be used to extract the common elements among these series (see appendix, for a detail analysis of the adequacy of this technique to the data used). Provided the type of data used (that is mostly a reflection of investor's intentionality or sentiment through information searching), the assumption behind is not that common nexus among these series should be a representation of investor sentiment (same as Baker and Wurgler (2006) do with the selection of the first principal component of their proxies for sentiment), but that there are different motivations or factors underlying that sentiment with different intensity.

In practice, this multivariate technique allows the reduction of the number of observed variables throughout the generation of a set of new variables resulting from a combination of the previous ones. As a result, these new variables just reflect the existing relations between the original variables but with no pre-set hypothesis about those relations. This is a critical issue when creating variables, for instance, from economic proxies. In fact, some of the most relevant measures of sentiment in the literature based on economic variables show a high sensitivity to the proxies used, related to the underlying assumptions behind their choice (Sibley et al., 2016). In this case, no assumption is made on the variables, and the informational content of each of them will be determined by the amount of variance incorporated into each component as a percentage of the total variance observed.

To make a proper selection of the number of components to be considered, Timmerman and Lorenzo-Seva (2011) point to three conditions: 1) the criteria should be objective, 2) the results should be interpretable, and 3) the results should be in line with theory.

Following these steps and starting with the objectivity criteria, there are several options. The first could be to apply the widely used Kaiser rule (select those components with eigenvalues greater than 1). However, contrary to frequent practice, this is not the most accurate method to identify the number of components. According to this procedure, the number of factors would show a linear relationship with the number of items used (Lloret et al., 2014). This would mean selecting 10 components from the PCA performed, and this is clearly not an acceptable choice.

Then, one additional practice to decide the number of components to be finally considered in the analysis will be to use the “scree test”. Following this technique, the number of factors would be determined by the first change in the slope of the eigenvalue seed graph. Only those components that show a significant slope before the inflection point should be considered. Although there can be some doubt about its effectiveness in those cases when there is a high number of components or several changes in slope, its validity has been widely tested (Pérez and Medrano, 2010). This test considers up to four components. However, three components will finally be selected, representing 65.6% of the total variance³⁰ (table 3.1, appendix) for the US searches (69.5% for global searches), after taking two additional criteria (clear interpretation and congruency with theory) into consideration. These three components reflect a combination of variables that can be easily interpreted (table S1, available online). In contrast, the fourth component shows a combination of variables that are difficult to be interpreted and it will be finally discarded. This classification is relevant when comparing to alternative formulations using searches (Da et al., 2015; Gao et al., 2016; Dimpfl and Jank, 2016) as these indicators come from the aggregation of terms that do not present any economic meaning beyond a positive or negative classification for sentiment. Otherwise, this methodology should help to

³⁰ Baker and Wurgler (2006) first principal component considered for the index construction, explains only 49% of the common variance among proxies.

categorize the different components of sentiment in line with the different dimensions that characterize behavioral finance theory: financial, psychological and sociological (Ricciardi and Simon, 2000). In fact, in relation with these dimensions, the first component considers variables that can be considered representative of financial expectations (PC1). The second component can be thought of as a representation of wealth expectations (PC2). The third component can be interpreted as a reflection of economic expectations (PC3).

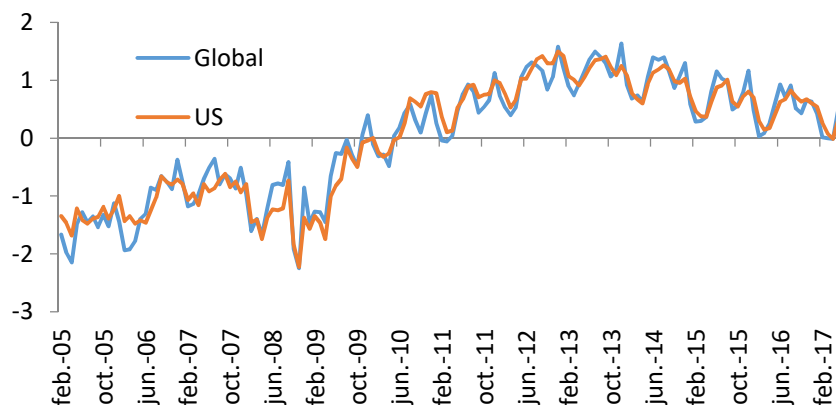
3.2.2. Sentiment index: definition and characteristics

Finally, a web-based investor sentiment index (WISI) is generated as a combination of the 3 components previously estimated (PC1, PC2 and PC3). Those constituents will be weighted in accordance with their capability to explain the total variance. This investor sentiment index has been generated for both the US ($WISI_{US}$) and global ($WISI_G$) markets (figure 3.2):

$$WISI_{US} = 0.63 * PC1 + 0.20 * PC2 + 0.17 * PC3$$

$$WISI_G = 0.66 * PC1 + 0.19 * PC2 + 0.15 * PC3$$

Figure 3.2. Web-based Investor Sentiment Index: US and Global



With respect to the main characteristics of the WISI, it presents a negative skewness (table 3.2). This result is in line with the asymmetric impact shown as well by alternative measures of sentiment present in the literature (Baker and Wurgler, 2006) or even by other elements considered as market proxies for sentiment (e.g. US Consumer Confidence Index - USCONF). It also presents a low kurtosis both compared to other measures of sentiment but also with respect to market variables.

Additionally, $WISI_{US}$ shows a significant high correlation of 0.52 at the 1% confidence level with the S&P500 and $WISI_G$ of 0.42 at 1% with MSCI World (and reaches 0.53 with MSCI ACWI). Simultaneously, the $WISI_{US}$ shows a low correlation with alternative measures of sentiment. Pearson correlation coefficients are significantly low, at 0.157 with the US Consumer Confidence index and -0.087 with the Baker and Wurgler (2006) sentiment index with a 1% confidence level. These correlations show how WISI reveals a different type of information that the one provided by alternative formulations of sentiment.

Table 3.2. Descriptive statistics

| Descriptive Statistics | | | | | | | |
|------------------------|---------|--------|--------|--------|--------|--------|--------|
| Statistics | WISI_US | WISI_G | S&P500 | MXWO | BW | USCONF | VIX |
| Mean | 0 | 0 | 0,002 | 0,001 | -0,006 | 80,094 | 19,299 |
| Median | 0,308 | 0,256 | 0,004 | 0,004 | 0,047 | 81,6 | 16,56 |
| Maximum | 1,495 | 1,636 | 0,044 | 0,045 | 0,848 | 98,5 | 59,89 |
| Minimum | -2,226 | -2,248 | -0,08 | -0,091 | -0,866 | 55,3 | 10,41 |
| Std. Dev. | 1 | 1 | 0,178 | 0,019 | 0,365 | 11,295 | 8,681 |
| Skewness | -0,329 | -0,394 | -1,011 | -1,115 | -0,467 | -0,306 | 2,079 |
| Kurtosis | 1,682 | 2,008 | 5,877 | 6,396 | 3,013 | 2,177 | 8,174 |

Sample: February 2005 - May2017, monthly intervals, given 148 observations (exception for the data for the Baker and Wurgler index. This has been taken from Wurgler's web page (<http://people.stern.nyu.edu/jwurgler>) and it is only available up to September 2015, providing 128 observations). WISI_US and WISI_G are de US and Global investor sentiment indices respectively (as they are standardized show a mean of 0 and standard deviation of 1). BW is the Baker and Wurgler sentiment index. USCONF is the Conference Board Consumer Confidence Index (uses 1985 as base year, with USCONF =100). VIX is the Chicago Board Options Exchange Market Volatility Index. S&P500 is the monthly reurn on the Standard & Poor's 500 Index (US market cap weighted stock market index). MXWO is the monthly return on the MSCI World Index (market cap weighted stock market index of +1,600 stocks worldwide).

3.3. Empirical analysis

To test the suitability of WISI as a proper estimate of investor sentiment and its capability to explain the evolution of financial assets, there is a need to consider financial markets data (financial indices and measures of volatility and market sentiment).

Price data for the different variables have been downloaded from Bloomberg with a monthly frequency. All data are price data for the different indicators and ranges for the period Jan 2005 – May 2017. The returns for stock market indices (S&P500 as US benchmark; MSCI World as developed market benchmark; MSCI ACWI as global benchmark; Ibex as Spanish benchmark) have been calculated as log stock returns:

$$R_t = \log(P_t) - \log(P_{t-1})$$

where P_t and P_{t-1} are monthly closing prices on the current and previous month, respectively.

3.3.1. Sentiment and predictability of US stock returns

Following some of the most relevant works on investor sentiment indices (Baker and Wurgler, 2006; Stambaugh et al., 2012; Sibley et al., 2016) and considering the presence of control variables to assure the real contribution of the sentiment index as a predictor of asset returns (Da et al., 2015; Mao et al., 2015), we propose the following model to be tested:

$$R_t = \alpha + \sum_{i=1}^m \beta_i R_{t-i} + \sum_{i=1}^m \gamma_i WISI_{US,t-i} + \sum_{i=1}^m \delta_i ContVab_{t-i} + \varepsilon_t \quad (1)$$

where R_t represents the return on the S&P index at time t and $WISI_{US,t-i}$ reflects the investor sentiment index at t-i (for $i=1 \dots m$) for the US market. Thus, $ContVab_{t-i}$ represents other control variables to be considered for the test.

Specifically, two control variables will be considered in the analysis. The first is market volatility as a proxy for investors' risk perception³¹. The VIX index will be taken as a reference for volatility, as it is a widely used variable in the literature (Mao et al., 2015; Eraker, 2009). The second variable to consider will be a consumer confidence index taken as a proxy for fundamentals or business cycle (Stivers, 2015). Specifically, the Consumer Confidence Index by the Conference Board (USCONF) will be considered.

The test will be run for the entire data set showing some interesting results (Model I, in table 3.3). Once considered past market returns, the lagged sentiment index and lagged control variables for the full sample, the sentiment index does not show any significant predictive capability. In fact, it can be observed that there is no clear significant variable explaining the current market return (only the third autoregressive for the market return shows some significance at the 10% confidence level).

However, as mention in the introduction, it might be a problem linked to the time framework considered for the analysis. Therefore, a Bai-Perron (2003) test for structural breaks has been applied to the stock market series. The aim of this test is to identify the existence of any break in the time series and, if so, test for its significance. A structural break has been identified in March 2009 (matching the market turning point after the market drawdown because of the 2008 financial crisis). Therefore, the analysis should be carried using the series after the break to estimate the relationship between variables (Antoshin et al., 2008).

³¹ As Bams et al. (2015, p.3) suggest, "when investors are risk averse, we expect the price of market volatility risk to be negative, because higher market volatility today can be associated with a deterioration of the future investment opportunity set. Stocks that are positively correlated with the market volatility will offer higher return when the investment opportunity set is shrinking". Equivalent results are found by Yu and Yuan (2011) and Kozhan et al. (2013).

Table 3.3. Testing local sentiment and control variables on the S&P500 return

| | Model I | Model II |
|---------------------|-------------------|-------------------|
| Constant | -0,032 (0,179) | -0,031 (0,336) |
| $WISI_{US,t-1}$ | -0,008 (0,247) | -0,019 (0,026) |
| $WISI_{US,t-2}$ | 0,008 (0,351) | 0,023 (0,055) |
| $WISI_{US,t-3}$ | 0,002 (0,744) | -0,001 (0,891) |
| R_{t-1} | 0,115 (0,438) | -0,17 (0,285) |
| R_{t-2} | 0,002 (0,990) | -0,197 (0,198) |
| R_{t-3} | 0,197 (0,062) | 0,124 (0,289) |
| VIX_{t-1} | 0 (0,722) | 0 (0,817) |
| VIX_{t-2} | 0 (0,433) | 0 (0,792) |
| VIX_{t-3} | 0 (0,819) | 0,001 (0,036) |
| $USCONF_{t-1}$ | 0 (0,785) | 0 (0,759) |
| $USCONF_{t-2}$ | 0 (0,569) | 0 (0,965) |
| $USCONF_{t-3}$ | 0 (0,215) | 0 (0,2965) |
| R^2 | 0,107 | 0,226 |
| AIC | -5,135 | -5,444 |
| Durbin-Watson stat. | 2,058 | 2,063 |

Therefore, after re-running the model for the post-break period (Model II in table 3.3) two important results come off: firstly, the explanatory capacity of the model shows enormous improvement (R^2 increases from 11% to 22%). Secondly and mostly relevant, the sentiment index shows a significant impact on future returns (first lagged coefficient of -0.019 is significant at 5% confidence level). This result is in line with existent evidence for stock price reversals after periods of high sentiment (Da et al., 2015).

At this point it is worth mentioning that there might be an issue regarding the consideration of the sentiment variables as an exogenous variable in the proposed model. Despite the above mentioned relationships between investor sentiment and stock returns there is still an ongoing discussion about the causality between these two variables (Gizelis and Chowdhury, 2016).

In order to shed some light into this point we will conduct a Granger causality test on the market return and sentiment series. This test will be performed in the context of a VAR

model, evidencing the relationship among the variables and more relevantly accounting for the simultaneity between them.

From a practical point of view the period for the analysis will be the one between March 2009 and May 2017 and the VAR will show a lag of 2 (what is consistent with the Akaike measures at different lags and with other empirical evidence³²).

The test (table 3.4) shows that the sentiment index significantly “causes” returns but not the other way around, in short periods of time. This result is consistent with a faster diffusion of information/sentiment and the greater market efficiency at the time of processing such information (Fox et al., 2016; Statman, 2018) and it also invalidates the arguments for the strong impact of returns over sentiment formation (Hoffman, 2015).

Table 3.4. Granger causality test

| Null hypothesis | Wald test | p-value |
|---------------------------------|-----------|---------|
| Sentiment does not cause return | 7,5003 | 0,0235 |
| Return does not cause sentiment | 4,3576 | 0,1132 |

Additionally, with respect to the significance of the control variables on the post-break period (Model II, table 3.3), the autoregressive market return variables do not show significant value, demonstrating that the diffusion of the non-revealed incoming information is the determining element when explaining market performance.

Precisely, the lack of significance of the consumer confidence variable shows that in contrast to previous studies, the improvement of the model after the financial crisis is due neither to autocorrelation factors (Bijl et al., 2016) nor to factors linked to the business cycle (Sibley, 2016). In fact, studies using sentiment measures based on surveys have

³² Similarly, from an empirical perspective, Bijl et al. (2016), considering weekly data, see the impact of sentiment fading out after 5 weeks at most for the US market.

shown their limited explanatory capabilities and in no case predictive power (Heston and Sinha, 2016).

Similarly, VIX coefficients are not significant for lags 1 and 2, in line with the results from Granger causality test that show the largest impact of sentiment on returns on those lags. This evidence reflects the information content embedded on the WISI that overcomes the rest of variables including this one, matching the evidence by De Long et al. (1990) who find how irrational trading anticipates larger volatility.

However, VIX third autoregressive is significant at a 5% level, albeit with a low magnitude. This can reflect several factors: first, the limited time impact of sentiment information, as previously noted. Secondly, the presence of information asymmetries in the market (Sinha and Agnihotri, 2014) that are captured by the volatility index. And, thirdly, the possibility of persistence linked to structural changes in investors' mindset related to previous investment experiences (Hoffman and Post, 2015) or subjective interpretations of events (Mitroi and Oproiu, 2014; Malmendier et al., 2018).

Overall, the inclusion of control variables does not offset the explanatory capabilities of our investor sentiment indicator, acting as a reflection of non-revealed intentions by investors.

3.3.1.1. Comparison to other relevant literature

The magnitudes achieved in our model compare, for instance, to a R^2 of 1% achieved by Bijl et al. (2016), after considering a panel of 431 companies from the S&P500 and measuring the impact of weekly searches of those company names on the index return one period ahead. Similarly, Da et al. (2015) use a 30 search terms-based sentiment indicator to anticipate the return of the S&P500 and get a R^2 of 2.6% (for the 2004-2011 period) but what is more relevant, the indicator is significant only at a 10% level. Gao et al. (2016) improve previous results with a 30 positive plus 30 negative terms-based

sentiment indicator (using weekly data for the 2004-2014 period) and similar control variables but the R^2 sets at 8.8%.

An additional comparison has been made against the Baker and Wurgler index (BW) that despite being economic-based is widely accepted in the investor sentiment literature and uses a principal component methodology. As it can be seen in model III (table 3.5), the index has no significant predictive capability when estimating the equivalent equation (1) for the full sample, as it was the case for the WISI. But, as it can be seen in model IV (table 3.5), although WISI is able to show a significant negative relationship in the post-break period between lagged sentiment and future market returns, BW is not showing any significance either in this period.

Table 3.5. Impact of B&W index and control variables on the S&P500 return

| | Model III | Model IV |
|----------------|-------------------|--------------------|
| Constant | -0,04 (0,152) | -0,0135 (0,703) |
| BW_{t-1} | 0,009 (0,573) | 0,01 (0,603) |
| BW_{t-2} | -0,01 (0,651) | -0,029 (0,288) |
| BW_{t-3} | -0,005 (0,761) | 0,016 (0,411) |
| R_{t-1} | 0,113 (0,488) | -0,222 (0,236) |
| R_{t-2} | 0,04 (0,81) | -0,208 (0,252) |
| R_{t-3} | 0,197 (0,103) | 0,053 (0,0,691) |
| VIX_{t-1} | 0 (0,695) | 0 (0,802) |
| VIX_{t-2} | 0 (0,244) | 0 (0,828) |
| VIX_{t-3} | 0 (0,629) | 0 (0,289) |
| $USCONF_{t-1}$ | 0 (0,591) | 0 (0,458) |
| $USCONF_{t-2}$ | 0 (0,521) | 0 (0,734) |
| $USCONF_{t-3}$ | 0 (0,188) | 0 (0,882) |
| R^2 | 0,105 | 0,191 |
| AIC | -5,012 | -5,216 |
| Durbin-Watson | 2,052 | 2,024 |

The lack of significance of the BW index in the post-financial crisis period shows that there is some sort of information that economic indicators are not able to capture in advance compared to methods based on the use of information technologies.

3.3.2. Is globalization affecting sentiment?

Once the $WISI_{US}$ has proved effective as a measure of local sentiment, the next step is to test whether $WISI_G$ is as effective when dealing with global sentiment and markets. A priori, an index based on a global search of information should be a more efficient way to capture global sentiment than the widely seen approach of adding up sentiment for individual geographies based on local data (Baker et al., 2012; Gao et al. 2016).

Therefore, we will test the predictability of the global investor sentiment index ($WISI_G$) on the world market index (MSCI World) return (R_t) for the post-structural break period (identified in March 2009). Opposite to the US case, only lag returns are to be considered on the extended test, as there are no equivalent global control variables, or at least none that are representative enough.

$$R_t = \alpha + \sum_{i=1}^m \beta_i R_{t-i} + \sum_{i=1}^m \gamma_i WISI_{G,t-i} + \varepsilon_t \quad (2)$$

The results (Model V in table 3.6) show that sentiment is the only significant variable explaining return on the world index, validating the results already obtained on the US market and proving its validity as a reflection of global investor sentiment.

Moreover, to avoid any developed-market bias on the market index given the important weight (above 50%) of the US market on the MSCI World index, the test has also been conducted on a world index with an emerging market (EM) component, the MSCI All Countries World Index (ACWI). As Eichgreen (2006, p.160) notes, “investors tend to be imperfectly informed”, as information is difficult to obtain (even more so in markets with a lower degree of development, as is the case of EM). From a different angle, Zouaoui et

al., (2011, p.723) mention that “the impact of investor sentiment on stock market is more pronounced in countries that are culturally more prone to her like behavior, overreaction and low institutional environment”. These features can be easily found on EM, so the impact of a global investor sentiment index as a source of additional information should, *a priori*, be greater in this case.

Table 3.6. Impact of global sentiment on the MSCI World and ACWI

| | Model V | Model VI |
|---------------------|-------------------|-------------------|
| Constant | 0,009 (0,000) | 0,006 (0,438) |
| $WISI_{G,t-1}$ | -0,013 (0,038) | -0,033 (0,035) |
| $WISI_{G,t-2}$ | 0,009 (0,245) | 0,021 (0,335) |
| $WISI_{G,t-3}$ | -0,004 (0,492) | 0,012 (0,427) |
| R_{t-1} | -0,072 (0,461) | -0,127 (0,243) |
| R_{t-2} | -0,129 (0,165) | -0,127 (0,227) |
| R_{t-3} | 0,001 (0,987) | 0,065 (0,533) |
| R^2 | 0,098 | 0,105 |
| AIC | -5,22 | -3,535 |
| Durbin-Watson stat. | 1,952 | 1,974 |

The results for the test on the ACWI index return (Model VI in table 3.6) confirm the previous hypothesis for the higher intensity of sentiment in emerging markets (particularly in the post-break period, identified in November 2009). In fact, not only is the $WISI_G$ coefficient significantly high (5% level) and the impact of the coefficient on the market return larger (-0.033 vs -0.013), but the explanatory capacity of the ACWI model is greater than for the MSCI World index model (R^2 of 10.5% vs 9.8%).

Additionally, these results contradict those achieved by Gao et al. (2016), albeit with some differences in testing procedures (they consider simultaneously global and local markets sentiment indicators). They find that developed markets are more influenced by global sentiment and emerging markets by local sentiment. They explain the result as the final outcome of market integration (and standardized accounting information) whereas

our model justifies the larger impact on emerging markets due to information discovery through a growing penetration of information technologies (information effect)³³.

From a different angle, as their global sentiment indicator is a composite of six countries based on economic proxies, Baker et al. (2012) also find that superior role of global sentiment even over local market returns. Their global sentiment indicator anticipates market returns and makes local sentiment measures insignificant (although the explanatory power of the model on the countries tested for the period 1981-2006 is not high, R^2 of 1%).

In order to test the predictive capabilities of our global sentiment indicator on local markets we will run a similar exercise on the US benchmark (S&P500). Accordingly, we will test the following relation:

$$R_t = \alpha + \beta WISI_{G,t-1} + \varepsilon_t \quad (3)$$

where R_t is the return on the S&P index at time t and $WISI_{G,t-1}$ is the past global investor sentiment index. The results for the post-break period (Model VII in table 3.7) show how the global sentiment index has a significant impact on S&P returns (significance is also present even when considering lagged index values and autoregressive returns as control variables – Model VIII in table 3.7).

Thus, the explanatory power of the model improves with respect to that using the US sentiment index as the dependent variable (the R^2 increases from 3.8% with the US sentiment index to 5.4% with the global one). This result implies that contrary to the local bias prevailing in the literature, the significance of the global investor sentiment index when predicting local market returns should be at least as good as that seen by the local investor sentiment index.

³³ This would be in line as well with the literature that shows a significant relationship between internet penetration and economic growth (Amiri and Reif, 2013).

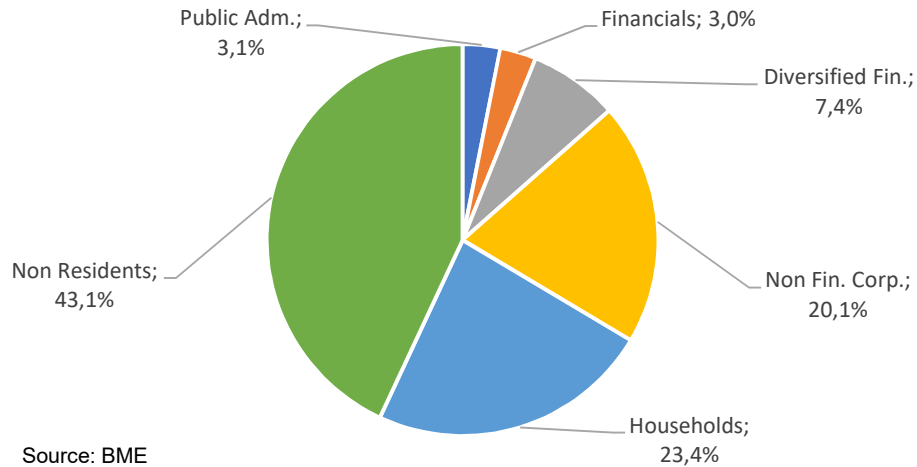
Table 3.7. Impact of global sentiment on the S&P500 return

| | Model VII | Model VIII |
|---------------------|-------------------|-------------------|
| Constant | 0,009 (0,000) | 0,009 (0,000) |
| $WISI_{G,t-1}$ | -0,006 (0,020) | -0,013 (0,021) |
| $WISI_{G,t-2}$ | | 0,008 (0,276) |
| $WISI_{G,t-3}$ | | -0,001 (0,856) |
| R_{t-1} | | -0,117 (0,218) |
| R_{t-2} | | -0,135 (0,137) |
| R_{t-3} | | -0,034 (0,713) |
| R^2 | 0,054 | 0,117 |
| AIC | -5,466 | -5,433 |
| Durbin-Watson stat. | 2,107 | 1,93 |

Despite former evidence, critique might arise on the index considered for testing the globalization of sentiment given the benchmark role of the S&P500 for global markets. Therefore, an additional test is proposed for the global sentiment index on the Spanish market benchmark: Ibex35. There is a twofold interest on this index: first, it is placed as a mid-cap index when compared to other global indices and, *a priori*, should show a lower sensitivity to global sentiment and flows attached. In fact, contrary to global indices, lagged returns have no predictive power on actual index return even for long periods and there is no structural break observed in the series.

Secondly, main participants on this index are institutional investors (only 24% of stocks were owned by households by end of 2016 - see figure 3.3). Institutional investors might be tagged as rational investors and therefore sentiment should not show a significant impact on them and subsequently on market returns.

Figure 3.3. Owners of Spanish Stocks (% market cap by end of 2016)



Therefore, the following relation will be tested:

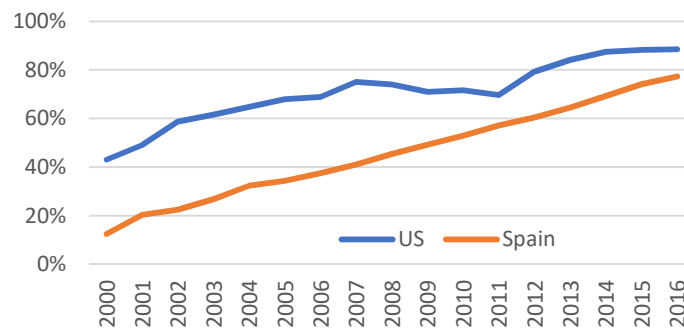
$$R_t = \alpha + \beta_i R_{t-1} + \gamma_i WISI_{G,t-1} + \varepsilon_t \quad (4)$$

where R_t is the return on the Ibex35 at time t and $WISI_{G,t-1}$ is the past global investor sentiment index.

Model 9 in table 3.8 shows the results for the period Feb2005-May 2017. These results indicate that neither the sentiment index nor the return autoregressive have significant predictive power on the index returns, as previously hypothesized.

However, before concluding that sentiment has no predictive capabilities on the Spanish market return, there is one factor that must be considered: the evolution of the penetration rate of information technologies in the Spanish market. As the matter of fact, internet penetration rates do not reach levels equivalent to those seen in the US market at the time of its structural break (70% of population) until 2014 (figure 3.4). This difference between markets is even more dramatic for the full sample, with penetration levels of 32% vs 65% in the US in 2004. Given the link between diffusion of information and sentiment, this is an element that must be considered in the analysis.

Figure 3.4. Internet penetration (% population): Spain vs US



Source: Internet Live Stats

Therefore, the model is tested again for a new sample period starting in 2014 to capture the effect of broad diffusion of information technologies in the market. Model X in table 3.8, shows that for this new sample the global sentiment index is negative (-0.0178) and significant (5% confidence level), in line with previous results for the US market and those seen in the literature (Da et al., 2015). Additionally, there is seen a larger explanatory power of the model on the index returns (R^2 of 10.8%), coming off as a key variable to anticipate market returns. These results contrast to the ones achieved by Gao et al. (2016) on the Spanish market, that showed no significance of the sentiment variable (for the period 2004-2014) and a limited explanatory power of the sentiment model (R^2 of 5.9%).

Table 3.8. Impact of global sentiment index on Ibex35

| | Model IX | Model X |
|----------------|-------------------|-------------------|
| Constant | 0,0004 (0,83) | 0,0134 (0,05) |
| $WISI_{G,t-1}$ | -0,0004 (0,85) | -0,0178 (0,04) |
| R_{t-1} | 0,0792 (0,344) | -0,1309 (0,43) |
| R^2 | 0,0065 | 0,1081 |
| AIC | -4,4903 | -4,9101 |

3.4. Trading strategy

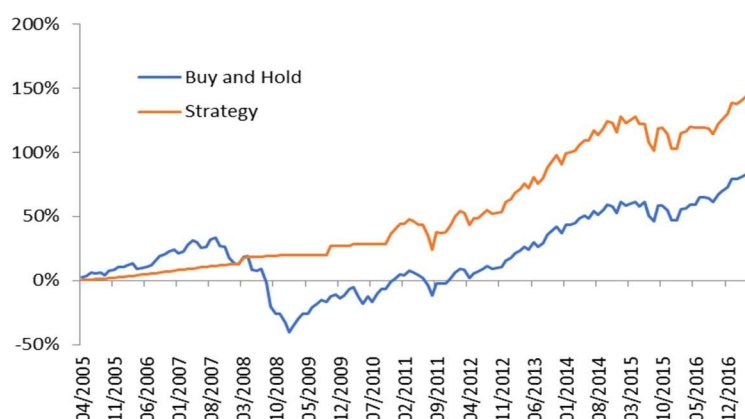
Lastly, following Preis et al. (2013) and Bijl et al. (2016) a simple trading strategy is built on the US market to assess the practical application of an investor sentiment measure as a source of return. Should the sentiment index be successful in capturing non-revealed information, then a trading strategy based on that indicator should be superior to a “buy and hold” strategy on the market index (in this case, the S&P500).

Consequently, a rule has been created in such a way that a meaningful change in the previous month’s reading of the sentiment index would trigger either a trading position on the S&P500 future or, alternatively, in a risk-free asset (3-month US T-Bill). This strategy has been conducted for the period Feb 2005-May 2017. Specifically, at the beginning of month t , a position on the 3-month T-Bill is built if $\Delta (WISI_{US,t-1}/WISI_{US,t-2}) \geq 1.5$ standard deviation (SD); otherwise, a position is taken on the S&P500 future. A 10bp transaction cost for futures trading (in line with market at the time of doing the exercise) has been considered in the return calculations.

The rationale behind this strategy is the already proved negative relationship between investor sentiment and subsequent market return. As can be seen (figure 3.5), this strategy provides a significant improvement versus just holding the market for the whole period. Precisely, the trading strategy generates an average annual return of 7.6% vs 5.2% for the buy-and-hold strategy on the S&P500.

Overall, the results support the hypothesis that a web-based sentiment index can capture some sort of information that is not present in asset prices. Although this evidence should put an end to the asymmetries between informed and uninformed investors (Fricke et al., 2014) and lead to a faster end to market anomalies, in the meanwhile it opens opportunities in the form of defining sentiment-based investment strategies that provide above market returns.

Figure 3.5. Cumulative returns of a dynamic investment strategy vs a “buy and hold” investment on the S&P500.



Source: Authors

3.5. Conclusions

This paper proposes incorporating the principal component analysis widely used in economic-based models of sentiment to search-based models. This approach has shown several advantages compared to other search-based alternatives, but one stands out: reduces noise regarding to information searches through the categorization of components. Following this procedure, only relevant information to take financial investment decisions would be considered.

Moreover, further accuracy in terms of measuring investor sentiment is achieved by using a specific financial lexicon, overcoming the problems found in the literature when using standard economic dictionaries.

The paper also shows some other interesting breakthroughs. First, it evidences how information technology penetration is a key driver for investor sentiment significance in financial markets. As a matter of fact, acceleration of internet penetration in the last decades has matched and structural change in most of global markets and supported the growing role of sentiment in markets dynamics.

Second, sentiment arises as a powerful source of information. It even captures information commonly attached to other market variables linked to sentiment as for instance implied volatility. So, it is a better proxy of non-revealed intentions/expectations.

Third, the use of global searches allows to create a global sentiment indicator representative of investor collective behavior. This produces more information than a mere addition of individual indicators and it is seen in a greater explanatory power of a global sentiment indicator when compared to a local one, invalidating the local bias rooted in the related literature.

Fourth, the sentiment indicator reveals previous undisclosed information what translates into an active contributor to market efficiency. This is seen in two ways: on the one hand, improving the explanatory power of those models focusing in predicting emerging market returns. And, on the other hand, showing how institutional investor driven markets (supposed to be essentially rational) are also impacted by sentiment and their returns linked to its evolution.

The relevance of sentiment for financial literature is evident given all these previous features. In fact, new lines of research can be opened focusing on the impact of sentiment in market efficiency. Specifically, greater information and subsequent efficiency should translate into improving risk premium estimates and gaining full control of those market anomalies that seem to proxy for some non-revealed information (e.g. momentum effect).

Appendix

PCA adequacy to SVI data.

Before applying this PCA to the SVIs two further steps need to be taken. First, the variables need to be standardized to get the components correctly estimated. If there are differences in magnitude among the variables, those with greater variance will end up dominating the first principal components and will not show real relationships.

Secondly, testing the suitability of the data for structure detection (table 1). The KMO test of sample adequacy shows a good proportion of common variance among the variables considered (values greater than 0.8 mean a good fit of the data to a factor model).

Additionally, Barlett's sphericity test indicates that the variables are correlated and suitable for this type of analysis (it rejects the null of orthogonality among the variables).

In addition to these tests, the communalities of the variables are high (table 1), evidencing how a large part of the variance of the variables is explained by the common factors, adding additional support to the method used.

Table 1. Component matrix

| Items | Component | | | | Communalities |
|-------------|-----------|---|---|---|---------------|
| | 1 | 2 | 3 | 4 | |
| home | 0,945 | | | | 0,950 |
| investment | 0,940 | | | | 0,979 |
| arts | 0,916 | | | | 0,934 |
| portfolio | 0,913 | | | | 0,951 |
| bond | 0,913 | | | | 0,978 |
| health | 0,907 | | | | 0,929 |
| oz_of_gold | -0,897 | | | | 0,940 |
| yield | 0,889 | | | | 0,932 |
| S&P500 | 0,882 | | | | 0,885 |
| housing | 0,847 | | | | 0,911 |
| purchase | 0,836 | | | | 0,862 |
| risk | 0,833 | | | | 0,910 |
| food | -0,801 | | | | 0,944 |
| derivatives | 0,800 | | | | 0,932 |
| happy | -0,794 | | | | 0,975 |
| debt | 0,792 | | | | 0,940 |
| greed | -0,788 | | | | 0,817 |

| | | | | | |
|------------------|---------|---------|---------|--------|-------|
| forex | 0,780 | | | | 0,954 |
| leverage | 0,777 | | | | 0,903 |
| ore | -0,774 | | | | 0,878 |
| cash | -0,772 | | | | 0,949 |
| tourism | 0,770 | | | | 0,709 |
| hedge | 0,753 | | | | 0,821 |
| transaction | -0,744 | | | | 0,964 |
| consumption | 0,732 | | | | 0,847 |
| office | -0,714 | | | | 0,930 |
| capital | 0,706 | | | | 0,882 |
| water | -0,696 | | | | 0,870 |
| stocks | 0,695 | | | | 0,839 |
| movies | -0,695 | | | | 0,797 |
| growth | 0,692 | | | | 0,902 |
| travel | 0,685 | | | | 0,890 |
| work | -0,628 | | | | 0,934 |
| banks | 0,612 | | | | 0,737 |
| car | | 0,863 | | | 0,932 |
| credit | | 0,703 | | | 0,803 |
| energy | | 0,678 | | | 0,871 |
| invest | | 0,676 | | | 0,838 |
| fun | | 0,658 | | | 0,835 |
| rich | | 0,630 | | | 0,783 |
| hold | | 0,614 | | | 0,866 |
| opportunities | | 0,606 | | | 0,730 |
| default | | | 0,849 | | 0,943 |
| financial_crisis | | | 0,849 | | 0,765 |
| economy | | | 0,814 | | 0,901 |
| finance | | | 0,745 | | 0,938 |
| politics | | | 0,734 | | 0,738 |
| short_sell | | | 0,712 | | 0,958 |
| loss | | | 0,620 | | 0,874 |
| inflation | | | 0,597 | | 0,816 |
| unemployment | | | | 0,796 | 0,893 |
| divorce | | | | 0,789 | 0,750 |
| % Variance exp. | 41,093% | 12,817% | 11,689% | 5,219% | |

Kaiser-Meyer-Olkin measure of sampling adequacy:

0,899

Bartlett's test of sphericity: Chi Square 18587,488.

Sig. 0

References

- Amin, R., & Ahmad, H. (2013). Does Investor Attention Matter ' S ? *Journal of Public Administration, Finance and Law*, 4, 111–125.
- Amiri, S., & Reif, B. (2013). Internet Penetration and its Correlation to Gross Domestic Product: An Analysis of the Nordic Countries. *International Journal of Business, Humanities and Technology*, 3(2), 50-60.
- Antoshin, S., Berg, A., & Souto, M. R. (2008). Testing for structural breaks in small samples. <http://www.imf.org/external/pubs/cat/longres.aspx?sk=21808>. <http://dx.doi.org/10.5089/9781451869378.001>.
- Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of Internet stock message boards. *Journal of Finance*, 59, 1259–1294. <http://dx.doi.org/10.1111/j.1540-6261.2004.00662.x>.
- Artola, C., & Galán, E. (2012). Tracking the Future on the Web: Construction of Leading Indicators Using Internet Searches. <http://dx.doi.org/10.2139/ssrn.2043056>.
- Asness, C., Moskowitz, T., & Pedersen L. H. (2013). Value and momentum everywhere. *Journal of Finance*, 68, 929-985. <http://dx.doi.org/10.1111/jofi.12021>.
- Bai, J. & Perron, P. (2003). Critical Values for Multiple Structural Changes Tests. *The Econometrics Journal*, 6(1), 72-77. <http://dx.doi.org/10.1111/1368-423X.00102>.
- Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, LXI (4), 1645–1680. <http://dx.doi.org/10.1111/j.1540-6261.2006.00885.x>.
- Baker, M., Wurgler, J., & Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of Financial Economics*, 104(2), 272–287. <http://dx.doi.org/10.1016/j.jfineco.2011.11.002>.
- Bams, D., Honavar, I., & Lehnert, T. (2015). From Time Varying Risk-Aversion to Anomalies in Market Moments' Risk Premia. <http://dx.doi.org/10.2139/ssrn.2692983>.
- Bank, M., & Brustbauer, J. (2014). Investor Sentiment in Financial Markets. <https://www.uibk.ac.at/ibf/mitarbeiter/bank/investor-sentiment-in-financial-markets.pdf>. Accessed 30 January 2018.
- Barber, B. M., & Odean, T. (2008). All that Glitters: The effect of Attention and news on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, 21(2), 785-818. <http://dx.doi.org/10.1093/rfs/hhm079>.
- Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, 45, 150–156. <http://dx.doi.org/10.1016/j.irfa.2016.03.015>.
- Choi, H. & Varian, H. (2009). Predicting the Present with Google Trends. https://static.googleusercontent.com/media/www.google.com/es//googleblogs/pdfs/google_predicting_the_present.pdf. <http://dx.doi.org/10.1111/j.1475-4932.2012.00809.x>.
- Curme, C., Preis, T., Stanley, H. E., & Moat, H. S. (2014). Quantifying the semantics of search behavior before stock market moves. *Proceedings of the National Academy of*

- Sciences of the United States of America*, 111(32), 11600–5. <http://dx.doi.org/10.1073/pnas.1324054111>.
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *Journal of Finance*, 66, 1461–1499. <http://dx.doi.org/10.1111/j.1540-6261.2011.01679.x>.
- Da, Z., Engelberg, J., & Gao, P. (2015). The Sum of All FEARS Investor Sentiment and Asset Prices. *Review of Financial Studies*, 28(1), 1–32. <http://dx.doi.org/10.1093/rfs/hhu072>.
- De Long, J.B., Shleifer, A., Summers, L. & Waldmann, R. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4), 703-738.- <http://dx.doi.org/10.1086/261703>.
- Dickinson, B., & Hu, W. (2015). Sentiment Analysis of Investor Opinions on Twitter. *Social Networking*, 04(03), 62–71. <http://dx.doi.org/10.4236/sn.2015.43008>.
- Dietzel, M. A., Braun, N., & Schäfers, W. (2014). Sentiment-Based Commercial Real Estate Forecasting with Google Search Volume Data. *Journal of Property Investment & Finance*, 32(6), 540-569.
- Dimpfl, T., & Jank, S. (2016). Can internet search queries help to predict stock market volatility? *European Financial Management*, 22(2), 171–192. <http://dx.doi.org/10.1111/eufm.12058>.
- Duch, R.M., & Kellstedt, P.M. (2011). The heterogeneity of consumer sentiment in an increasingly homogeneous global economy. *Electoral Studies*, 30(3), 399-405.
- Eichgreen, B. (2006). Can emerging markets float? Should they inflation target? In M. Vernengo (Ed.), *Monetary Integration and Dollarization: No Panacea* (pp. 149-175). Cheltenham, UK: Edward Elgar Publishing Limited.
- Eraker, B. (2009). The volatility premium. http://www.nccr-finrisk.uzh.ch/media/pdf/Eraker_23-10.pdf.
- Fang, L. H., Peress, J., & Zheng, L. (2014). Does media coverage of stocks affect mutual funds' trading and performance? *Review of Financial Studies*, 27(12), 3441-3466. <http://dx.doi.org/10.1093/rfs/hhu056>.
- Fox, E. G., Fox, M. B., & Gilson, R. J. (2016). Economic crisis and the integration of law and finance: The impact of volatility spikes. *Columbia Law Review*, 116 (2), 325-408.
- Frazzini, A. & Lamont, O.A. (2008). Dumb money: mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics*, 88(2), 299-322. <http://dx.doi.org/10.1016/j.jfineco.2007.07.001>.
- Fricke, E., Fung, S., & Goktan, M. (2014). Google search, information uncertainty, and post-earnings announcement drift. *Journal of Accounting and Finance*, 14(2), 11–27.
- Gao, Z., Ren, H., & Zhang, B. (2016). Googling Investor Sentiment around the World. http://www.cuhk.edu.hk/fin/event/symposium/SEFM_2016_paper_58.pdf.
- Ghysels, E., Plazzi, A., & Valkanov, R. (2007). Valuation in US commercial real estate. *European Financial Management*, 13, 472–497. <http://dx.doi.org/10.1111/j.1468-036X.2007.00369.x>.

- Gizelis, D., & Chowdhury, S. (2016). Investor sentiment and stock returns: evidence from the Athens Stock Exchange. https://mpra.ub.uni-muenchen.de/71243/1/MPRA_paper_71243.pdf.
- Heston, S., & Sinha, N. (2016). News versus sentiment: predicting stock returns from news stories. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2792559 / Accessed 3 March 2017. <http://dx.doi.org/10.17016/FEDS.2016.048>
- Hirshleifer, D. a., Lim, S. S., & Teoh, S. H. (2011). Limited Investor Attention and Stock Market Misreactions to Accounting Information. *Review of Asset Pricing Studies*, 1, 35-73. <http://dx.doi.org/10.1093/rapstu/rar002>.
- Hoffmann, A. O. I., & Post, T. (2015). How return and risk experiences shape investor beliefs and preferences. *Accounting and Finance*, 57(3), 759-788. <http://dx.doi.org/10.1111/acfi.12169>.
- Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies*, 28(3), 791–837. <http://dx.doi.org/10.1093/rfs/hhu080>.
- Kozhan, R., Neuberger, A., & Schneider, P. (2013). The skew risk premium in the equity index market. *Review of Financial Studies*, 26, 2174–2203. <http://dx.doi.org/10.1093/rfs/hht039>.
- Kristoufek, L. (2013). Fractal markets hypothesis and the global financial crisis: wavelet power evidence. *Scientific Reports*, 3. <http://dx.doi.org/10.1038/srep02857>.
- Lee, B. S., & Mauck, N. (2014). Information Asymmetry and the Market Response to Open Market Share Repurchases. https://mpra.ub.uni-muenchen.de/54066/1/MPRA_paper_54066.pdf. <http://dx.doi.org/10.2139/ssrn.2403377>.
- Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151–1169.
- Loughran, T., & McDonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35–65. <http://dx.doi.org/10.1111/j.1540-6261.2010.01625.x>.
- Malmendier, U., Pouzo, D., & Vanasco, V. (2018). Investor experiences and financial market dynamics. *NBER Working Paper No. 24697*.
- Mao, H., Counts, S., & Bollen, J. (2015). Quantifying the effects of online bullishness on international financial markets. <https://www.ecb.europa.eu/pub/pdf/scpsps/ecbsp9.en.pdf>.
- Mitroi, A., & Oproiu, A. (2014). Behavioral finance: new research trends, sociometrics and investor emotions. *Theoretical and Applied Economics*, 21(4), 153–166.
- Ngo, T. H. D. (2013, April). The Box-Jenkins Methodology for Time Series Models. Proceedings of the SAS Global Forum 2013 Conference, Paper 454, San Francisco, CA.
- Novy-Marx, R. & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(11), 104-47. <http://dx.doi.org/10.1093/rfs/hhv063>.

Park, H., & Sohn, W. (2013). Behavioral Finance: A Survey of the Literature and Recent Development. *Seoul Journal of Business*, 19(1), 3-42.

Pérez, E. R., & Medrano, L. (2010). Análisis factorial exploratorio: Bases conceptuales y metodológicas. *Revista Argentina de Ciencias Del Comportamiento*, 2(1889), 58–66.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, 3(1684). <http://dx.doi.org/10.1038/srep01684>.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *PloS One*, 10(9). <http://dx.doi.org/10.1371/journal.pone.0138441>.

Ricciardi, V., & Simon, H. K. (2000). What is Behavioral Finance?, *Business, Education & Technology Journal*, 2 (2), 1–9.

Shoven, J. B., & Sialm, C. (2000). The Dow Jones Industrial Average: the impact of fixing its flaws. <https://siepr.stanford.edu/research/publications/dow-jones-average-impact-fixing-its-flaws>.

Sibley, S. E., Wang, Y., Xing, Y., & Zhang, X. (2016). The information content of the sentiment index. *Journal of Banking and Finance*, 62, 164–179. <http://dx.doi.org/10.1016/j.jbankfin.2015.10.001>.

Sinha, P., & Agnihotri, S. (2014). Investigating impact of volatility persistence, market asymmetry and information inflow on volatility of stock indices using GJR-GARCH. <http://mpira.ub.uni-muenchen.de/58303>.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2), 288–302. <http://dx.doi.org/10.1016/j.jfineco.2011.12.001>.

Statman, M. (2018). Behavioral efficient markets. *The Journal of Portfolio Management*, 44(3), 76-87. <http://dx.doi.org/10.3905/jpm.2018.44.3.076>.

Stivers, A. (2015). Forecasting Returns with Fundamentals-Removed Investor Sentiment. *International Journal of Financial Studies*, 3(3), 319–341. <http://dx.doi.org/10.3390/ijfs3030319>.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62, 1139–1168. <http://dx.doi.org/10.1111/j.1540-6261.2007.01232.x>.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209–220. <http://dx.doi.org/10.1037/a0023353>.

Yook, K. C. (2010). Long-run stock performance following stock repurchases. *Quarterly Review of Economics and Finance*, 50, 323–331. <http://dx.doi.org/10.1016/j.qref.2010.03.006>.

Yu, J., & Yuan, Y. (2011). Investor sentiment and the mean–variance relation. *Journal of Financial Economics*, 100(2), 367–381. <http://dx.doi.org/10.1016/j.jfineco.2010.10.011>.

Zouaoui, M., Nouyrigat, G., & Beer, F. (2011). How does investor sentiment affect stock market crises? Evidence from panel data. *Financial Review*, 46(4), 723–747. <http://dx.doi.org/10.1111/j.1540-6288.2011.00318.x>.

4. Sense and sentiment: a behavioral approach to risk premium modelling

Juan José García Petit, Antonio Rúa Vieites, Esther Vaquero Lafuente *

Abstract

Estimates of risk premium derived from classical financial theory have consistently shown deviations from the observed levels. These limitations have been linked to the rational foundations of these theories that rely on asset prices as the main source of information. This article focuses on the need to increase the information available through the consideration of behavioral factors.

Therefore, the paper proposes an alternative methodology to estimate risk premium incorporating investor sentiment as a source of additional information.

This model is tested on the US market with the objective of obtaining a more accurate measure of risk premium than the one provided by classical financial approaches. It also offers an alternative explanation to risk-return relationship based on investor's sentiment. Finally, the use of behavioral approaches to the treatment of the risk premium will favor the control of market anomalies such as the momentum effect.

Keywords: risk premium, sentiment, valuation, beta, momentum

JEL: G40, G12

*Published online in *REFC- Spanish Journal of Finance and Accounting* on July 2, 2019.

<https://doi.org/10.1080/02102412.2019.1622067>

4.1. Introduction

Risk premium, understood as the compensation required by investors in consideration for assuming systematic risk (Gagliardini et al., 2016), is the key element to consider when valuing financial assets.

While the existing literature shows constant evolution when it comes to estimate this risk premium, it does so from multiple dimensions that goes from classical financial theory to modern behavioral approaches.

The classical financial approach has been able to establish a solid quantitative basis that facilitates a first reading of the risk component. However, the outcome might be not fully representative of the real risk premium, mostly due to the basic underlying assumptions that depart from the empirical evidence (De Long et al., 1990).

Even though *ad hoc* formulations have been made within this line of research aiming to sort this drawback and to improve the explanatory capacity of such models, the truth is that significant improvements require the need to open the paradigm to other perspectives much closer to reality (Cochrane, 2011).

In this way, behavioral finance is positioned, *a priori*, as an alternative to understand how investment processes (and the risk attached to them) work from the agent's viewpoint. Subsequently, the risk premium measure that arises from this approach should be closer to the market one (Malkiel, 2003; Lo, 2004; Kyriacou et al., 2004).

However, even considering the obvious contribution to risk premium coming out from this approach with respect to traditional formulations, it is not absent of limitations. In fact, behavioral finance cannot be considered as a full theoretical body, but a set of simultaneous explanations to the same problems analyzed by the conventional financial theory (Barberis and Thaler, 2003).

Given pros and cons, a symbiosis between both theoretical approaches might produce optimal results when estimating risk premium. Behavioral advantages of proximity to empirical evidence would be coupled by a solid methodological framework from classical financial theory.

Precisely, these capabilities are clearly complementary when facing a common goal: extracting as much information as possible (beyond the one present in prices) and get a better understanding of the dynamics of financial assets. In fact, as Thaler (2000, p.140) points out, "it seems logical that sustaining descriptive economic models in more realistic conceptions of economic agents will lead to an increase in explanatory capacity".

Therefore, any new definition of the risk premium should be based in incorporating as much information as possible in its estimation processes. In fact, Bertella et al. (2014) after considering an agent model find that the risk premium is continually underestimated because of an over-confidence bias and that an informational contribution should help to build a better measurement³⁴.

However, at this point, an additional question arises: what type of information is relevant for the determination of the risk premium? The information must have transcendence, be as broad as possible and reflect investor's sentiment³⁵.

In this line, the use of technology favors the extraction of information that can be subsequently incorporated into the processes for estimating the risk premium. As the matter of fact, the information technology highlights that prices are not reflecting all the information available (Da, Engelberg and Gao, 2011), and so that it is essential to define

³⁴ Particularly in down markets that is when the set of available information is reduced. That result would be in line with Fox et al. (2016) which show that as the available information decreases so does market efficiency.

³⁵ For instance, although there is empirical evidence of a relationship between presence in the news and asset returns (Amin and Ahmad, 2013) the presence of asymmetry is also observed (Lee and Mauck, 2014). Similarly, a repurchase of shares may indicate that these shares are undervalued but if the announcement of repurchases is made on a regular way it ceases to have an effect and profitability is no longer observed in the long term (Yook, 2010).

mechanisms of extraction of information that can flow into the market. All in all, a greater available information set should result in greater accuracy when determining the risk premium.

Against this backdrop, the objective of this article is to generate a risk premium measure with expanded information. Accordingly, a novel valuation methodology will be developed in such a way that, while keeping the structure of conventional formulations, will show greater dynamism coming from the incorporation of conditional elements and a greater information set linked to investor's sentiment.

On this basis, the effectiveness of the proposed risk premium measure and its potential advantages against classical formulations will be tested. Should these tests be successful, there would be proof of a relationship between levels of risk premium and investor sentiment. Additionally, a greater efficiency of a model with expanded information should translate into greater control of anomalies that are usually attached to classical formulations.

4.2. Determination of the risk premium through a valuation model with expanded information

The quantification of the risk premium through classical financial models has been a challenge from a methodological point of view, especially considering the presence of market anomalies contrarian to fundamental theoretical assumptions. Broadly, two alternative approaches can be highlighted in the quantification of risk:

1) Technical. This approach emphasizes the need for further development of econometric methods to overcome market anomalies. Thus, the presence of problems linked both to the robustness of the estimates and the estimation methods themselves is evident³⁶.

³⁶ For example, Wang (2003) shows an increase in the explanatory capacity of the models when considering the possibility of introducing higher moments in the estimation.

2) Informational. This approach shows that the problem lies in the limitations at the time of incorporating all potential available information into the model. It is not a matter of identifying factors that provide solutions to specific anomalies, but of looking for processes that reveal all the available information, especially when quantifying the magnitude of the risk premium. That is to say, we face models that condition the results to the existing information at all times³⁷.

This informational approach has led to some improvement in the explanatory capability of the asset dynamics and is compatible with traditional valuation methodologies (Hansen and Richard, 1987). In fact, the consideration of conditional information factors has even allowed to capture some of the market anomalies linked to traditional valuation approaches³⁸.

The main problem of the traditional valuation models does not arise from the introduction of additional information, but from that being narrowly done and without relaxing theoretical assumptions.

In fact, proposals for improvements in conventional finance focus on the treatment of the assumption of efficiency in the selection of portfolios. However, not much attention is paid to other relevant assumptions of classical financial theory and in particular two of them: a) investors have the same expectations about expected returns in a context of fully efficient markets and without transaction costs; b) even assuming homogeneity of expectations, these are conditioned according to the information available at each moment and, therefore, should be considered dynamically and not as a constant (Stambaugh, 1982; Hansson and Hördall, 1998; Best and Byrne, 2001).

³⁷ In any case, it does not mean incorporating new information, but processing the data in order to obtain additional information. There is no discussion about market efficiency and the fact that prices reflect all the available information.

³⁸For example, Lettau and Ludvigson (1999) manage to capture the value effect by introducing additional risk factors in the model using market data (the dividend/price ratio, the long-short term returns differential and a consumption/wealth ratio).

Any attempt to develop a valuation model to overcome these limitations involves, first, the introduction of variability in those components that capture the market risk premium. This implies the variability both in the market variance and in the covariance between the financial assets and the market reference³⁹.

Secondly, the treatment of the assumption of market efficiency and the available information set. The assumption of market efficiency, and that asset prices reflect all the available information, must be relaxed.

The incorporation of informational components in the valuation model, that is, conditioning the model to a broader information set⁴⁰ (introducing greater dynamism), implies the need to reformulate the evolution of the covariance between the assets and the market variable as a defining element of the risk premium (Bollerslev, 1987; Ng, 1991; De Santis y Gerard, 1997; Hanson y Hördal, 1998).

In this way, the methodological objective focuses, on the one hand, on limiting any substantial loss of information in the data extraction process. On the other hand, in adding superior information (not included in the asset prices) about agents' expectations.

4.2.1. Implementation of a sentimental asset pricing model

The model to be developed from which the market risk premium component will be estimated is conceptually based on the work of Morelli (2003) for the British market. Our implementation to this model will focus on two aspects: 1) the incorporation of investor sentiment as a conditional information factor, and 2) the process of estimating variances and conditional covariances.

³⁹ Fama and French (1989) show that the risk premium is not constant but varies throughout the asset cycle.

⁴⁰ Either by generating processes that help to reveal all the information implicit in the prices or considering additional explanatory factors

On the formulation of a traditional valuation model (precisely the *Capital Asset Pricing Model* - CAPM⁴¹) and following the notation of Morelli (2003), informational conditionality can be introduced on the valuation model, in such a way that:

$$E(r_{it}^{\phi}) = \beta_i^{\phi} E(r_{Mt}^{\phi}) \quad (1)$$

where,

$$\beta_i^{\phi} = cov(r_{it}r_{Mt}/\phi_{t-1})/var(r_{Mt}/\phi_{t-1}) \quad (2)$$

$E(/\phi_{t-1})$, indicates the conditioned expectation to the set of all information available ϕ at time t-1. r_{it} is the excess return of asset i with respect to the risk free rate at time t and r_{Mt} is the market excess return at time t.

The introduction of this informational conditionality would allow a more precise asset valuation, as well as a better estimation of the risk premium. Therefore, the discussion around the introduction of conditional informational factors focuses on the definition and scope of the information set.

In this sense, this work presents a double innovation. First, an investor sentiment indicator will be considered as a source of additional information (Ho and Hung, 2009), overcoming the limitations of using only asset prices (proper of traditional valuation models). This investor sentiment indicator (WISI) is based on data coming from agents' use of information technologies (see appendix for a brief explanation). This indicator probes to be more representative of sentiment than commonly used sentiment indicators based on economic factors (Baker and Wurgler, 2006, Antoniou, Doukas and Subrahmanyam, 2016).

⁴¹ In its traditional formulation, $E(R_i) = R_f + \beta_{iM}[E(R_M) - R_f]$ CAPM reflects how the risk premium of an asset is proportional to its systemic risk and, therefore, this factor can be quantified by a cross sectional regression type $E(R_i) = \alpha_0 + \lambda_1\beta_{iM} + \varepsilon_i$

Therefore, a regression of the returns of each one of the assets and market series will be run on the lagged investor sentiment index (WISI):

$$r_{i,t} = c + bWISI_{G,t-1} + \tau_{i,t} \quad (3)$$

As a result, the informational component is isolated from the of price component reflecting its contribution to valuation. Then, the residuals ($\tau_{i,t}$) of this regression will be taken for the estimation of variances and conditional covariances.

Following Morelli (2003) the assumption that the series follow an AR (1) process will be kept. This will generate conditional returns such as:

$$\tau_{i,t} = E(\tau_{i,t}/\psi_{t-1}) + \varepsilon_{i,t} \quad (4)$$

Where ψ_{t-1} represents the available information set obtained with the application of the previous process on the returns series. Although the information set ψ_{t-1} is still less than \emptyset (perfect information) it is higher than the one obtained not considering the presence of sentiment information.

This broader information set also helps to relax one of the assumptions of traditional valuation models: the presence of homogeneous expectations. Although investor sentiment reflects global aggregate expectations, these are derived from data generated individually and, therefore, heterogeneous in their origin.

The second innovation of this work focuses on the determination of the variability of the risk premium through the estimation of variances and conditional covariances.

In the case of Morelli (2003), the estimate of conditional variances and covariances is done through the application of ARCH processes on the residuals obtained from the autoregressive processes of the returns series such as:

$$\varepsilon_{it}\varepsilon_{Mt} = E(\varepsilon_{it}\varepsilon_{Mt}/\psi_{t-1}) + \eta_{it} \quad (5)$$

and,

$$\varepsilon_{Mt}^2 = E(\varepsilon_{Mt}^2/\psi_{t-1}) + \eta_{Mt} \quad (6)$$

The first of the sequences coincides with the conditional covariance between r_{it} r_{Mt} and the second one, coincides with conditional variance of r_{Mt} . However, this approach to the calculation of conditional variances is not optimal for the extraction of all available information.

Accordingly, the estimation of the conditional market variance $E(\varepsilon_{M,t}^2)$ will be carried by applying a GARCH (1,1) process⁴² and the estimation of conditional covariance between assets and market $E(\varepsilon_{i,t}\varepsilon_{M,t})$ will be run through a DCC-GARCH (1,1) process that solves the limitations of Morelli's work (2003)⁴³.

This DCC (dynamic conditional correlation) focuses on the conditional correlations between different financial assets (Engle, 2002). They are characterized by the fact that the variances and conditional covariances do not depend only on the delays of each of them for each variable, but there are also cross-effects between the variables considered (Saiti et al., 2014).

The process is implemented following a procedure in two steps:

- 1) The suitable GARCH (q,p), $h_t = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$, is estimated for each of the variables considered, as shown in previous section.
- 2) Residuals from the previous step are standardized ($S_{i,t} = \frac{\varepsilon_{i,t}}{\sqrt{h_{i,t}}}$) and they are used as an input for the estimate of conditional correlations in a DCC model (example for two assets):

$$(C_{i,j,t}) C_{i,j,t} = (1 - \phi_1 - \phi_2)\bar{C} + \phi_1 S_{i,t-1} S_{j,t-1} + \phi_2 C_{i,j,t-1}$$

⁴² The GARCH (1,1) formulation usually provides the best fit in most cases (Bollerslev et al., 1992; Chou, 1988).

⁴³ Morelli (2003) estimates the conditional covariance by multiplying the individual residuals and then modeling through an ARCH process. This procedure is assuming that the conditional correlation is constant (Bollerslev, 1990) which, in turn, assumes independence from conditional variances and does not consider changing behaviors (Saiti et al., 2014).

ϕ_1 and ϕ_2 capture the effects of previous shocks and dynamics on past correlations. It captures the dynamics of the conditional correlations and allows obtaining the dynamic conditional covariances that will be used in the valuation model.

In short, this methodology allows a dynamic monitoring of the risk premium, with a much more accurate estimate thereof, before the consideration of an information set that exceeds the information currently available in asset prices and a dynamic estimate of the relationships between assets and market.

As a result, we will obtain a sentiment conditional asset pricing model such as:

$$E(r_{it}^{\Psi}) = \beta_i^{\Psi} E(r_{Mt}^{\Psi}) \quad (7)$$

Where,

$$\beta_i^{\Psi} = E(\varepsilon_{i,t} \varepsilon_{M,t} / \psi_{t-1}) / E(\varepsilon_{M,t}^2 / \psi_{t-1}) \quad (8)$$

This valuation model conditioned on sentimental information has several potential advantages compared to the formulations of conventional financial theory (even in its most elaborated factor versions):

- 1) It presents a much more realistic picture of the level of risk premium for each asset at each moment of time. Meanwhile, traditional models only reflect average exposure levels regarding market risk.
- 2) It shows that available information is the key to an accurate quantification of the risk premium. The sentiment component is integrated into the traditional formulation, explaining the divergences in the risk-return relationship with respect to empirical observation that would be linked to exposure to investor sentiment.
- 3) This model should be more efficient given the broader information set, and therefore would make it possible to deal with market anomalies linked to traditional formulations.

4.3. Empirical contrast of a conditional sentimental valuation model

Once the methodology for estimating the valuation model has been defined, the next step is its empirical testing. However, before that we will proceed to introduce the variables to be considered. Additionally, a descriptive analysis of those variable will be done in order to verify their suitability and modify if necessary.

The research focuses on the US market during the period between March 2009 and May 2017, in the form of monthly observations⁴⁴. The time frame and the frequency are in line with those used for the estimation of the sentiment indicator.

Regarding the specific market data and the approach of the work, this will focus on a sector classification of the American market (Johnk and Soydemir, 2015) that moves away from the traditional formulations that primarily focus on the analysis of portfolios built according to different criteria (Fama and McBeth, 1973)⁴⁵.

In relation to the origin of the data, the direct price series for the different sectors have been obtained from Bloomberg, taking the first level of the GICS (Global Industry Classification Standard) classification of the S&P index (table 4.1). The market index considered as benchmark is the S&P500.

⁴⁴ The start date is derived from the presence of a structural change in the US market observed in March 2009 through the application of a Bai-Perron test.

⁴⁵ As pointed out by different literature the consideration of portfolios when performing this type of analysis is not more efficient than the use of single stocks as they present issues linked to thin trading (Dimson and Marsh, 1984), changing composition (Soufian, 2001), size issues (Avramov and Chordia, 2006), or limited factor structures (Lewellen et al., 2010).

Table 4.1. Sectors of the S & P500

| Variables | Bloomberg ticker |
|------------------------|-------------------------|
| Cyclical consumption | S5COND_INDEX |
| Staples | S5CONS_INDEX |
| Energy | S5ENRS_INDEX |
| Financials | S5FINL_INDEX |
| Healthcare | S5HLTH_INDEX |
| Industrials | S5INDU_INDEX |
| Information technology | S5INFT_INDEX |
| Materials | S5MATR_INDEX |
| Telecommunications | S5TELS_INDEX |
| Utilities | S5UTIL_INDEX |
| S&P500 | SPX_INDEX |

Returns data are calculated monthly as follows:

$$r_t = (P_t - P_{t-1})/P_{t-1}$$

r_t reflects returns in the month t , P_t price in the month t y P_{t-1} price in the previous month.

Finally, note that as a risk-free instrument it has been decided to take a three-month T-Bill, with monthly return data (to make them compatible with sector returns) also obtained from Bloomberg.

4.3.1. Descriptive analysis of the series

Table 4.2 shows the main statistics for the sectors and market series. This first descriptive approach attempts to identify the presence or absence of normality factors in the series (a question of relevance given that classical valuation models rest on the assumption of normality). Then, the asymmetry and kurtosis of the distributions are collected altogether with a specific test for normality⁴⁶for each of the series.

⁴⁶ The Jarque-Bera test considers normality as null hypothesis.

Table 4.2. Descriptive statistics

| | S&P500 | Materials | Utilities | Information technologies | Industrials | Health | Financials | Energy | Cyclical consumption | Staples | Telecoms |
|-------------|-----------|----------------|-----------|--------------------------|-----------------|-----------|-----------------|-----------|----------------------|-----------|-----------|
| Mean | 0.012747 | 0.012299 | 0.008285 | 0.016347 | 0.014900 | 0.012736 | 0.015188 | 0.005283 | 0.017866 | 0.010696 | 0.005822 |
| Median | 0.014759 | 0.005867 | 0.010212 | 0.018420 | 0.011894 | 0.011730 | 0.018137 | 0.006411 | 0.017748 | 0.010449 | -0.000359 |
| Max | 0.107723 | 0.175996 | 0.079163 | 0.121114 | 0.177236 | 0.090075 | 0.221746 | 0.169516 | 0.185312 | 0.061592 | 0.093194 |
| min | -0.081976 | -0.165876 | -0.095770 | -0.084514 | -0.098267 | -0.080478 | -0.115554 | -0.125819 | -0.098028 | -0.059996 | -0.093248 |
| Std. Dev. | 0.037055 | 0.058084 | 0.035588 | 0.046298 | 0.047947 | 0.035286 | 0.057521 | 0.053808 | 0.045571 | 0.029073 | 0.039882 |
| Assymmetry | -0.067139 | 0.241884 | -0.540136 | 0.011585 | 0.327790 | -0.284783 | 0.568788 | -0.018192 | 0.388185 | -0.199525 | 0.019957 |
| Kurtosis | 3.140454 | 3,82605 | 3,1316 | 2,879024 | 3,907105 | 3,169883 | 4,434816 | 3,183553 | 4,17187 | 2,276989 | 2,48323 |
| Jarque-Bera | 0.155753 | 3,78013 | 4,885269 | 0.062585 | 5,167074 | 1,457221 | 1,38302 | 0.144438 | 8,15112 | 2,81319 | 1,108158 |
| P-value | 0.925079 | 0.151062 | 0.086932 | 0.969192 | 0.075506 | 0.482579 | 0.000993 | 0.930327 | 0.016983 | 0.244976 | 0.574601 |

Sample: Mar2009-May2017

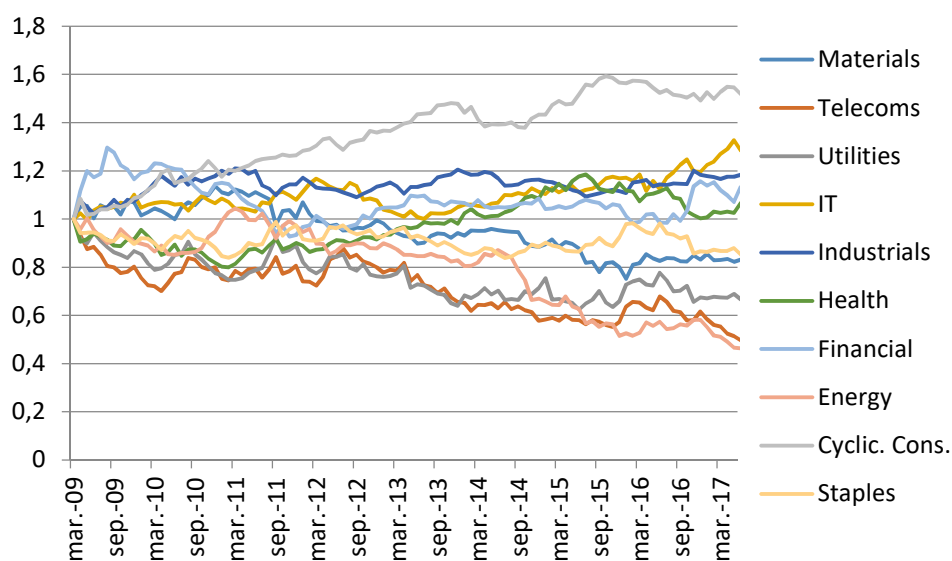
Based on the observation of these moments and the test considered, it can be concluded that for this period most of the series of sector returns can be assumed to be normal, except for: utilities, industrials, financials and cyclical consumption.

This data also highlights the fact that in addition to the market series only those sectors that present a more defensive nature show on average a negative asymmetry (this is a greater probability of generating negative returns): utilities, health, energy and staples. However, it must be noticed that this period reflects positive returns on average and high sentiment⁴⁷.

These evidences are also reflected from the point of view of the evolution of the cumulative returns of the different sectors in relation to the market. In fact, these defensive sectors are among those that present lowest returns for the whole period (figure 4.1).

⁴⁷ In moments of high sentiment, the price of the kurtosis is positive, while sustained exposure to asymmetry risk leads to lower expected returns (Bams et al., 2015).

Figure 4.1. Cumulative sector returns vs S&P500



Source: Authors

Additionally, and prior to test the conditional sentimental valuation model and given the relevance of the informational component, the sensitivity of the different sectors to the global sentiment will be also tested. Therefore, an impulse-response (IR) analysis is performed in an autoregressive vector model (VAR) on the sector returns and the global sentiment index.

The VAR models have their origin in the work of Sims (1980) by setting a linear model with n-variables in which each variable is explained not only by its own delays but by the delays of the rest of the variables. The impulse-response analysis reflects the response in the current and future values of a variable when facing a unitary shock in the value of another variable (Stock and Watson, 2002).

Appendix 2 shows the results of the impulse-response analysis between the global sentiment variable and the different sectors. From this preliminary analysis we obtain that: 1) there is an impact to a greater or lesser extent between sentiment and most sectors (apart from the health sector where it is neutral). 2) The relationship tends to fade from

the third period onwards. 3) The impact is higher in the first period (particularly in cyclical consumption and materials) and negative (except for utilities which is positive, in line with the defensive nature of the sector).

After all this evidence, two initial hypotheses are to be tested: first, sectors with higher betas will present a greater sensibility to sentiment.

Second, since the sentiment provides additional information to be distinguished from the systematic risk, the betas of these sectors must be lower than those obtained in those formulations not considering the sentiment component (and the risk premium factor will be also smaller). This is, it must be a more accurate measure of systematic risk (Gagliardini et al., 2016).

4.3.2. Modeling of sentimental conditional betas

In this section the estimate of the extended betas is carried out by applying the methodology previously mentioned. In this respect, the first step goes through the incorporation of the sentimental component to the returns series (table 4.3), through the estimation of equation (3).

The results of this regression show that for some sectors there does not seem to be a significant relationship with the global sentiment indicator. This result goes hand in hand with the evidence derived from the previous impulse-response analysis in which the reaction was limited (or even contradictory, as in utilities) in relation to those sectors (mainly defensive ones with low sensitivity to sentiment).

Tabla 4.3: Sentiment- return relationship parameters

| Sectors | c | b |
|----------------|----------|----------|
| Materials | 0,023* | -0,021** |
| Telecoms | 0,009 | -0,006 |
| Utilities | 0,009*** | -0,003 |
| IT | 0,023* | 0,013*** |
| Industrials | 0,024* | -0,017** |
| Health | 0,013* | -0,001 |
| Financials | 0,026* | -0,021** |
| | | - |
| Energy | 0,014*** | 0,016*** |
| Cyclical cons. | 0,026* | -0,016** |
| Staples | 0,015* | -0,008 |

*significant 1%

**significant 5%

***significant 10%

With these residuals of the previous relationship, we proceed to the estimation of the conditional variances and covariances in which we incorporate an autoregressive function of one period in the evolution of those residuals (table 4.4), as shown in equation (4).

Table 4.4: AR (1) parameters

| | γ | ψ |
|----------------|----------|----------|
| Materials | 0,003 | -0,178** |
| Telecoms | -0,001 | -0,093 |
| Utilities | 0 | -0,241** |
| IT | 0,001 | -0,203** |
| | | - |
| Industrials | 0 | 0,179*** |
| | | - |
| Health | 0,002 | 0,196*** |
| Financials | 0 | 0,0482 |
| Energy | 0 | -0,029 |
| Cyclical cons. | 0 | -0,226* |
| | | - |
| Staples | 0,000 | 0,185*** |
| S&P500 | 0 | -0,143 |

*significant 1%

**significant 5%

***significant 10%

On the residuals $\varepsilon_{i,t}$ the conditional variances of the different reference series are obtained (table 4.5) responding to a GARCH model (1.1):

$$h_t = \omega + \alpha_i \varepsilon_{t-1}^2 + \beta_i h_{t-1}$$

Table 4.5: GARCH (1,1) parameters

| | ω | α | β |
|----------------|----------|----------|---------|
| Materials | 0 | 0,333** | 0,616* |
| Telecoms | 0 | 0,073*** | 0,927* |
| Utilities | 0 | 0,076 | 0,796* |
| IT | 0 | 0,19** | 0,729* |
| Industrials | 0 | 0,283* | 0,410* |
| Health | 0 | 0,069 | 0,459 |
| Financials | 0 | 0,084*** | 0,848* |
| Energy | 0,002 | 0,193 | 0,024 |
| Cyclical cons. | 0 | 0,325** | 0,325 |
| Staples | 0,000 | 0,056 | 0,307 |
| S&P500 | 0,000 | 0,187** | 0,535** |

*significant 1%

**significant 5%

***significant 10%

Finally, the conditional covariances derived from the DCC-GARCH (1,1) model introduced in the previous chapter will be obtained (table 4.6):

$$C_{i,j,t} = (1 - \phi_1 - \phi_2)\bar{C} + \phi_1 S_{i,t-1} S_{j,t-1} + \phi_2 C_{i,j,t-1}$$

Once the variance and conditional covariances of the returns adjusted for sentiment are obtained, we have the informational components to implement the sentimental conditional risk premium model.

At the time of introducing conditionality, the estimation procedure will be adapted in two phases, used by Fama and McBeth (1973). First, the betas are estimated. Given the methodology proposed, betas are obtained from their decomposition into variances and conditional covariances, according to the equation (8).

Table 4.6: DCC (1,1) parameters

| | ϕ_1 | ϕ_2 |
|----------------|----------|----------|
| Materials | 0,009 | 0,944 |
| Telecoms | 0,031 | 0,999* |
| Utilities | 0,02 | 0,816 |
| IT | 0,149*** | 0,579* |
| Industrials | 0,082* | 0,902* |
| Health | 0,023 | 0,927* |
| Financials | 0,077 | 0,201 |
| Energy | 0,176** | 0,531** |
| Cyclical cons. | 0,077 | 0,201 |
| Staples | 0,031 | 0,766* |

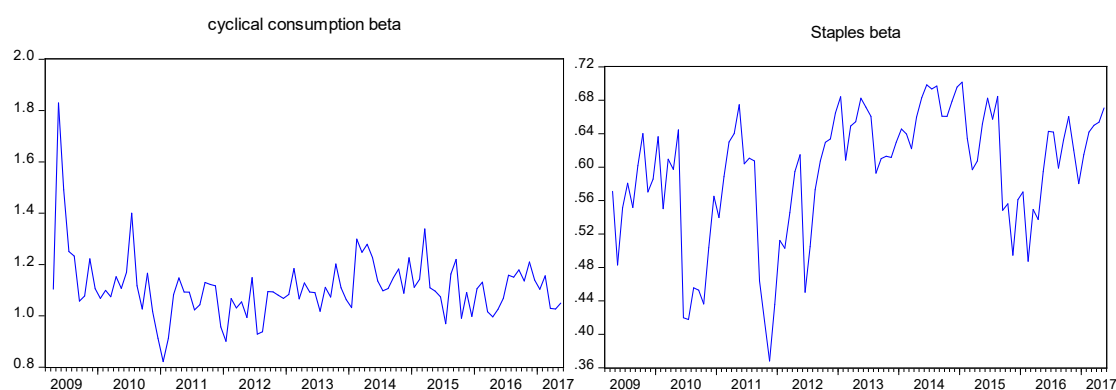
*significant 1%

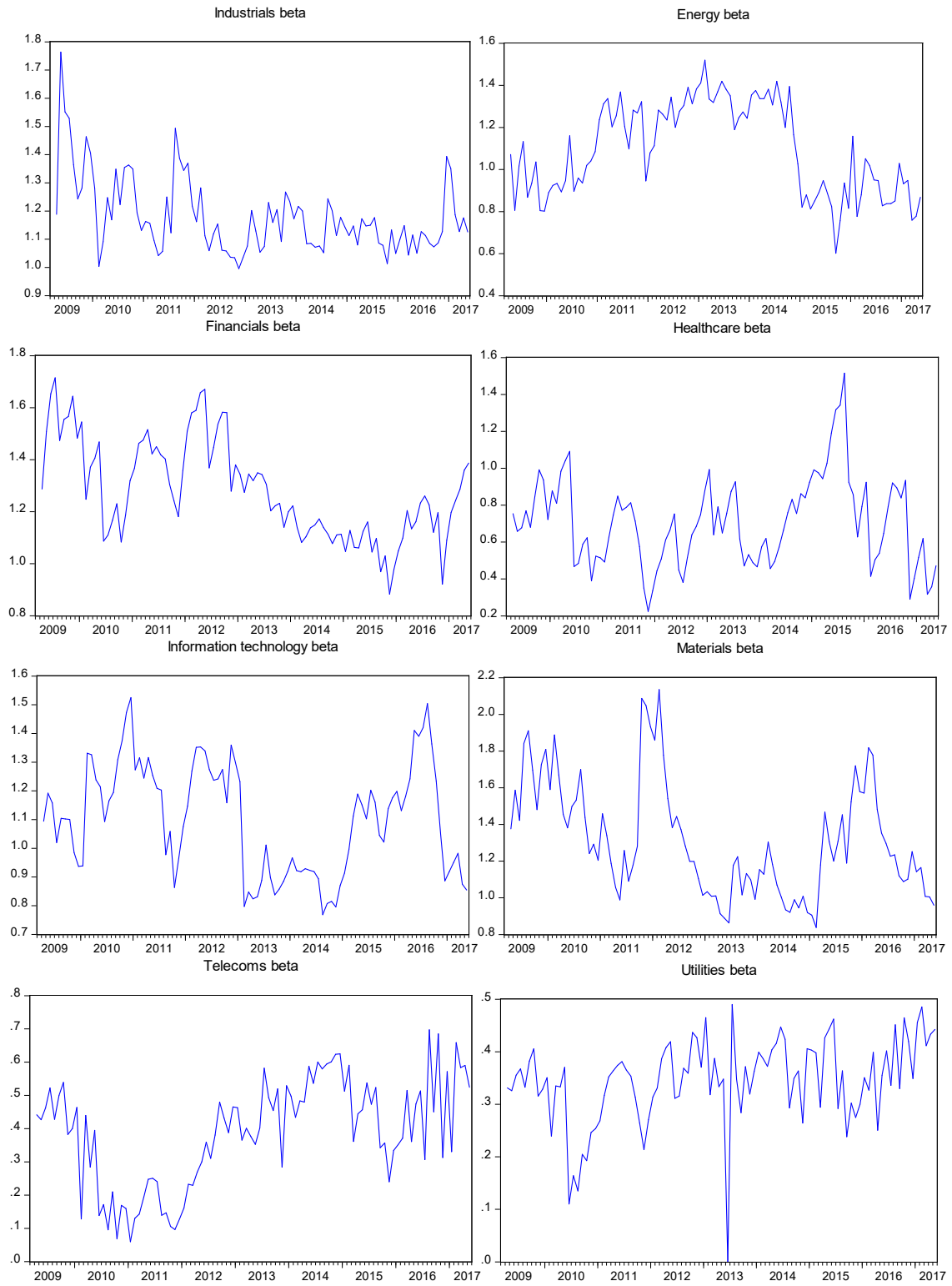
**significant 5%

***significant 10%

With the conditional valuation model and unlike the classic formulation (CAPM), the beta is estimated dynamically, for each month over the period of the sample, discarding those periods incorporated in the different estimation processes. It finally results in a total of 98 estimates of sentimental conditional beta (April 2009-May 2017) for each sector (figure 4.2).

Figure 4.2. Sentimental conditional beta





Source: Authors

4.3.2.1. Empirical evidence from sentimental betas

The first evidence that emerges when observing these betas is that the highest beta sectors match those with higher sensitivity to global sentiment. This result is in line with the first hypothesis raised⁴⁸.

Secondly, it is observed that in those periods of an upward trend in sentiment, the highest beta sectors show a less pronounced beta. On the contrary, in periods in which the sentiment is less positive, betas are accentuated (Antoniou et al., 2016).

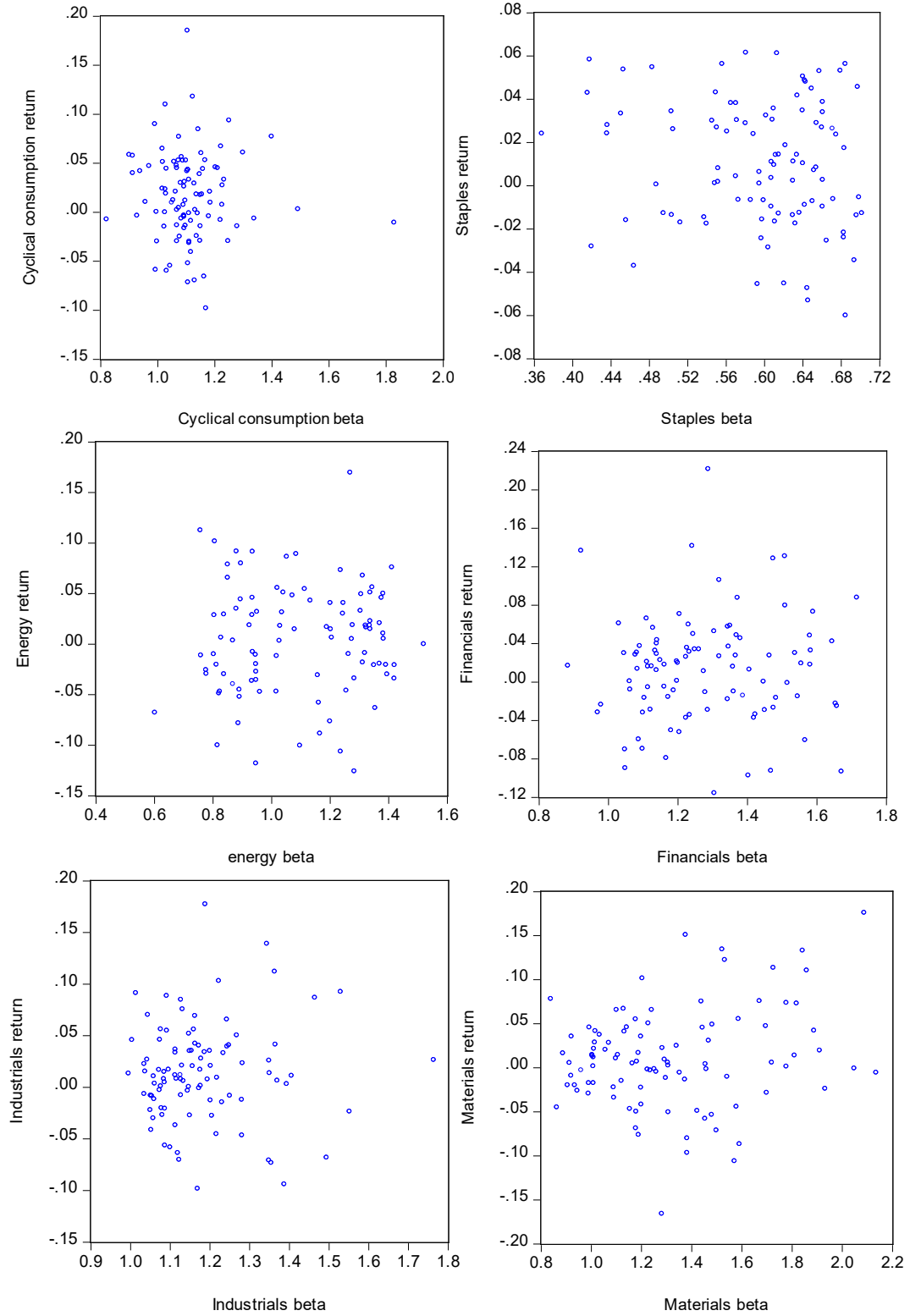
Additionally, if a relationship between beta and sector return is considered (figure 4.3) other interesting results can be noticed:

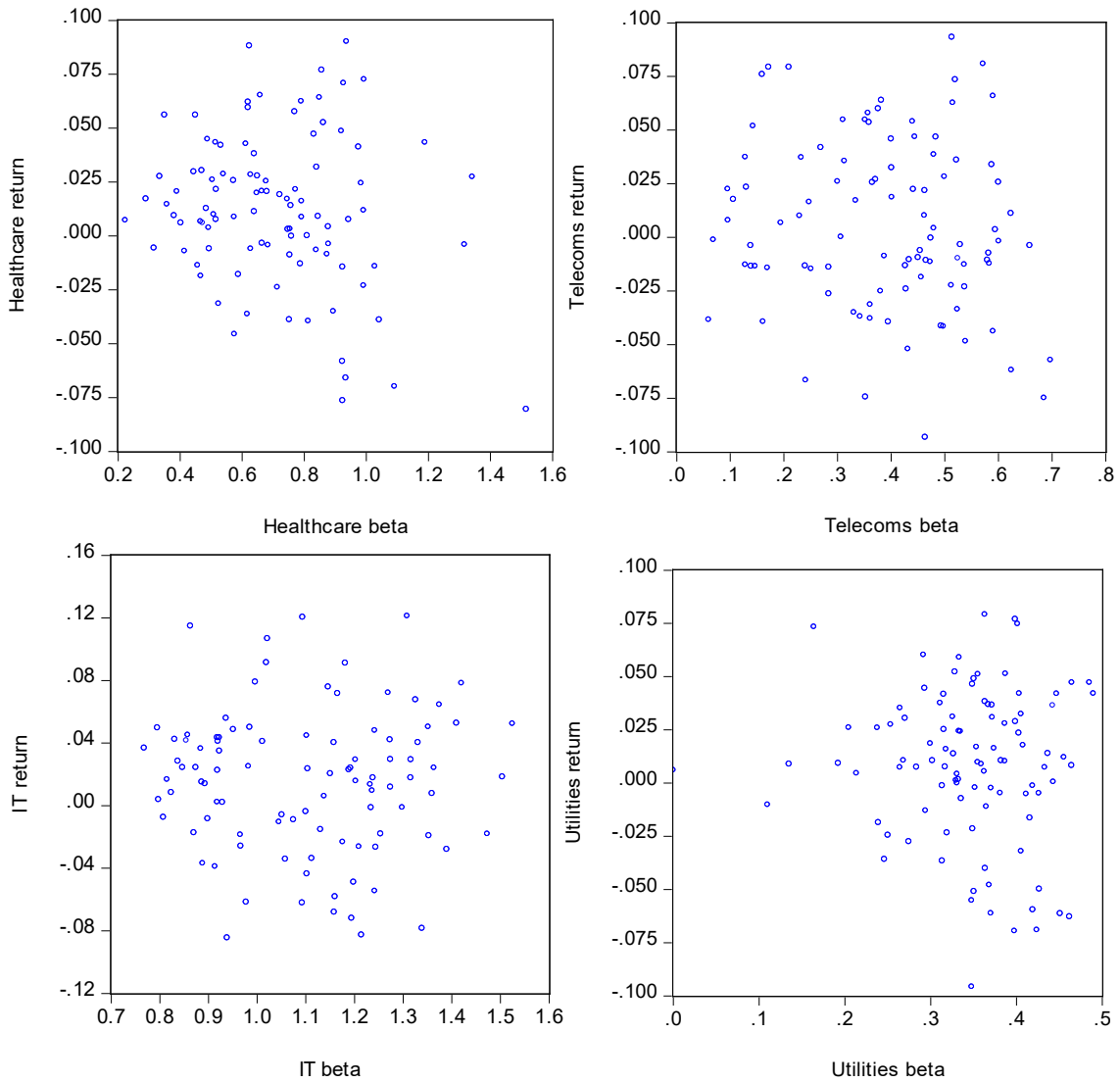
- 1) The higher the beta the greater the dispersion of returns. This is particularly remarkable in those sectors with a reduced sensitivity to global sentiment.
- 2) There is a linear relationship between beta and return⁴⁹, with a higher probability of positive returns to beta increases in those sectors with higher sensitivity to sentiment. This relationship exceeds the evidence found using traditional models that usually show a neutral relationship between return and beta (Roll y Ross, 1994).

⁴⁸ Glushkov (2006) finds, from a stockholder point of view, that the increase in irrational intermediation in a stock increases the correlation of this stock with the sentiment factor and leads to a higher beta.

⁴⁹ Also seen in Ben Sita (2018) or Tang and Shum (2003), although these last ones showed that the sign of the relationship changed from positive to negative when market moved from uptrend into downtrend.

Figure 4.3. Sector returns vs sentimental conditional betas





Source: Authors

Lastly, in order to test the positive contribution of this methodology with respect to the formulations derived from the classical theory an additional test is performed. There is a comparison between the beta obtained through the traditional CAPM model (unconditional) and the conditional sentimental beta (considering the arithmetic mean of all the betas to obtain a beta of the whole period). Even the conditional beta is calculated without considering the sentimental element to verify that this factor provides additional information (table 4.7).

Table 4.7. Comparison of betas for the period April 2009-May2017

| Sectors | Unconditional beta | Conditional beta | Sentimental conditional beta |
|------------------------|---------------------------|-------------------------|-------------------------------------|
| Materials | 1,351 | 1,32 | 1,317 |
| Telecoms | 0,477 | 0,43 | 0,388 |
| Utilities | 0,341 | 0,32 | 0,344 |
| Information technology | 1,135 | 1,129 | 1,103 |
| Industrials | 1,205 | 1,33 | 1,18 |
| Healthcare | 0,756 | 0,64 | 0,708 |
| Financials | 1,337 | 1,31 | 1,28 |
| Energy | 1,024 | 1,55 | 1,095 |
| Cyclical consumption | 1,153 | 1,126 | 1,111 |
| Staples | 0,588 | 0,57 | 0,595 |

As it can be seen in the results the unconditional beta is higher in most cases (except for the energy sector in both conditional versions, staples in the sentimental version and industrials in the simple conditional). In turn, sectors that present a greater sensibility to sentiment show smaller betas in the conditional sentimental version as opposed to the simple conditional model.

This result confirms the second hypothesis. The introduction of sentimental information contributes to debug systematic risk in the form of smaller betas compared to non-sentimental versions.

However, it remains to be checked that the risk premium factor is also smaller in those sectors that are more sensitive to the sentiment component, something that makes up the second phase of the estimation procedure.

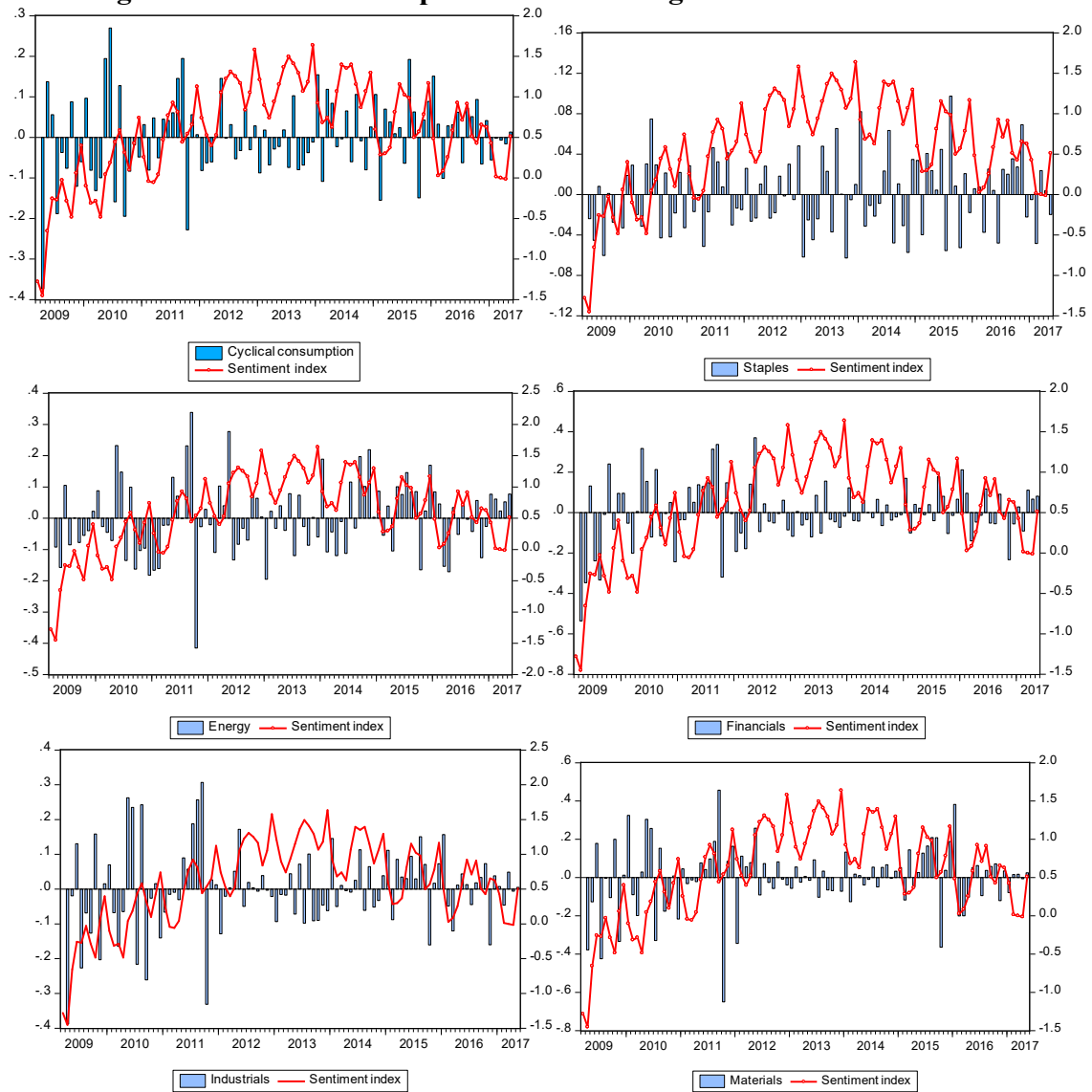
4.3.3. Assessing the sentimental risk premium factor

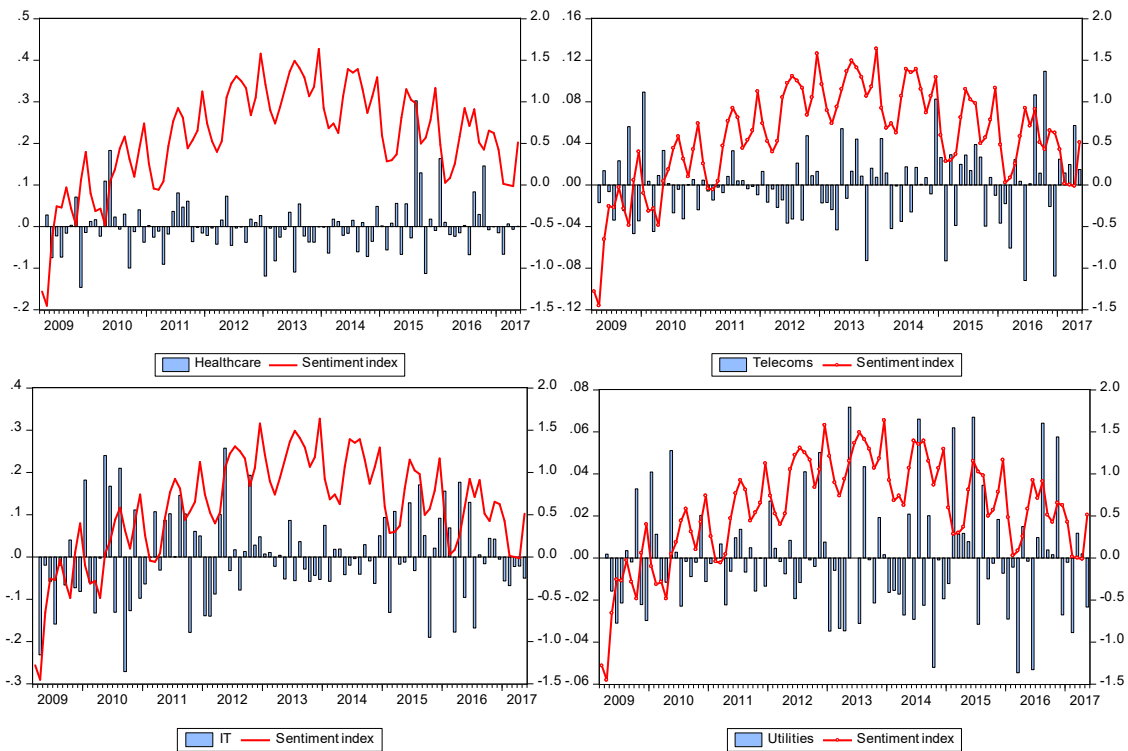
Next step is to determine whether the risk premium factor λ_i contributes significantly to explain the excess returns of assets (r_{it}). Specifically, we proceed to estimate the equation:

$$r_{it} = \lambda_i \beta_{Mi} + \varepsilon_i \quad (9)$$

Then, 98 estimates of risk Premium are obtained for each sector, one for each month of the sample (figure 4.4).

Figure 4.4. Sectorial risk premium factor vs global sentiment indicator





Source: Authors

Several significant results can be highlighted:

1. In relation to the theoretical risk-return relationship, risk premium shows that this relationship is not stable through time (Aragó and Matallin, 2002, Alonso and Restoy, 1995). This modeling allows a much more active risk monitoring as well as the determination of the sign of its contribution to sector return in a given period.
2. Additionally, a positive and significant relationship is observed on average between risk premium and excess return (table 4.8). Although reduced in both approaches, risk premium is lower in the sentimental conditional model, reinforcing the informational role of the sentiment and its impact when defining more precisely the risk premium.
3. Finally, there is an alternative explanation to the relationship between sector returns and risk premium. Unlike usual explanations linked to market volatility (Alcalá et al., 1993; Morelli, 2003; De las Heras and Nave, 2002) or

macroeconomic conditions (Ferson and Harvey, 1999; Vassalou, 2004) it can be attributed to the investor sentiment. Therefore, in those stages of sustained increase in sentiment, the return-risk premium ratio becomes positive and in those phases of correction of sentiment the sign of the relation reverses.

Table 4.8: Unconditional vs conditional risk premium factor

| Sectors | Unconditional | Conditional |
|----------------|---------------|----------------|
| Materials | 0,009* | 0,009* |
| Telecoms | 0,012 | 0,007 |
| Utilities | 0,024* | 0,021* |
| IT | 0,014** | 0,013** |
| Industrials | 0,012** | 0,012** |
| Healthcare | 0,017** | 0,013** |
| Financials | 0,011** | 0,010** |
| Energy | 0,005 | 0,005 |
| Cyclical cons. | 0,015** | 0,014** |
| Staples | 0,018** | 0,016** |

*significant 5%

**significant 1%

4.4. Sentiment and market anomalies: the momentum effect.

The incorporation of informational factors linked to investor sentiment into classic valuation models increases their explanatory capacity, and favors a dynamic perspective closer to the empirical evidence of the risk premium.

However, despite this contribution, we still need to test the model's capabilities when dealing with market anomalies. Although some of the most known anomalies have been captured by including sentiment as a factor into classical models⁵⁰, the momentum effect has not been fully captured yet. This section will focus on the formal treatment of this anomaly by our model.

⁵⁰ Stambaugh et al. (2014) after combining Baker and Wurgler (2006) sentiment index and the Fama and French model (1993) are able to demonstrate that the sentiment captures 11 market anomalies on a non-spurious form.

Much of the problems attached to this anomaly may be due to the use of explanatory factors linked to the asset or macro cycles and the absence of a clear relationship between these and the momentum effect itself⁵¹.

In this respect, the hypothesis to be tested is that investor sentiment would capture the momentum effect when it comes to explaining the asset excess return. If so, this would imply that momentum proxies for elements linked to the investor sentiment⁵².

Testing this hypothesis with sectors is more challenging than the traditional use of portfolios based on the trend of single stock returns (Wang, 2003). The consideration of sectors is more demanding as a single sector gathers stocks with higher and lower returns (Moskovitz and Grinblat, 1999; Lewellen, 2002).

Anyhow, the Jegadeesh and Titman (1993) methodology will be used in the construction of a positive momentum strategy based on the evolution of sector returns.

In accordance with this methodology, the path of asset returns for 6 months (training period) will be taken as a reference to select a sector along with a maintenance period of another 6 months. As a result, we will build a portfolio made of the three sectors with the highest return at every time. This portfolio shows an average monthly return of around 1% (13% per annum, close to the 12% returns found by Jegadeesh and Titman (1993) for a momentum strategy in the US market).

In order to test the direct effect of the momentum factor when explaining the sector excess returns, the following regression is estimated:

$$r_{i,t} = C + \alpha_1 M_t + \varepsilon_t \quad (10)$$

⁵¹ For example, Hurst et al. (2017) taking a century of history of the US market, they study the behavior of momentum strategies finding out that they show good behavior in different macro environments (even in 8 of the 10 crises identified since 1880). Additionally, Zhang (2005) even points out that the risk premium of the momentum factor is pro-cyclical.

⁵² In fact, some authors link the momentum effect to specific elements linked to sentiment such as herding behavior (Moskovitz y Grinblatt, 1999).

Where $r_{i,t}$ corresponds to the excess of returns of i sector in period t and, $M_{t,,}$ responds to the returns of the momentum portfolio for the period t.

The results (table 4.9) show how the momentum component is highly significant for all sectors.

Table 4.9. Sector excess returns vs. momentum factor

| Sector | C | Momentum |
|-------------|------|----------|
| Materials | 0,00 | 0,84* |
| Telecoms | 0,00 | 0,39* |
| Utilities | 0,00 | 0,33* |
| IT | 0,01 | 0,60* |
| Industrials | 0,01 | 0,73* |
| Healthcare | 0,01 | 0,44* |
| Financials | 0,01 | 0,78* |
| Energy | 0,00 | 0,68* |
| Cyc.cons. | 0,01 | 0,67* |
| Staples | 0,01 | 0,36* |

*significant 1%

**significant 5%

Given these premises, next step implies testing whether the sentimental risk premium ($\lambda_{i,t}$) is able to capture that momentum effect when explaining excess returns:

$$r_{i,t} = C + \alpha_1 M_t + \alpha_2 \lambda_{i,t} + \varepsilon_t \quad (11)$$

Once the sentimental risk premium factor is included, two important results (table 4.10) must be noted: on the one hand, those sectors that are more sensitive to sentiment (in which risk premiums are highly significant) cancel out the significance of the momentum factor (apart from cyclical consumption and information technologies, albeit the significance levels for the momentum factor decline).

On the other hand, even in those cases where the momentum factor is still highly significant (mostly in sectors that are not sensitive to sentiment) the size of the component

decreases substantially. This result evidences how the risk factor is able to capture part of the effect even in sectors less sensitive to sentiment.

Table 4.10. Sector excess return vs. momentum factor and risk premium

| Sector | C | Risk premium | Momentum |
|-------------|------|--------------|----------|
| Materials | 0,01 | -0,33* | 0,01 |
| Telecoms | 0,00 | -1,01* | 0,08* |
| Utilities | 0,01 | -1,29* | 0,03** |
| IT | 0,01 | -0,42* | 0,03** |
| Industrials | 0,01 | -0,40* | 0,00 |
| Healthcare | 0,01 | -0,48* | 0,09* |
| Financials | 0,01 | -0,37* | 0,03 |
| Energy | 0,00 | -0,45* | 0,02 |
| Cyc.cons. | 0,02 | -0,43* | 0,04** |
| Staples | 0,01 | -0,79* | 0,03* |

*significant 1%

**significant 5%

All these findings point to one conclusion and arise an additional hypothesis. The conclusion: there is supportive evidence to incorporate the sentiment component into valuation models as a source of information beyond asset prices, partially confirming our former hypothesis.

As for the new hypothesis: investor sentiment might deal with market anomalies (eg. momentum effect) as these can be thought a result of undisclosed information.

In order to test this hypothesis we will run the following relation:

$$M_t = C + \alpha_1 WISI_{G,t-1} + \varepsilon_t \quad (12)$$

Where $WISI_{G,t-1}$ is the investor sentiment indicator delayed one time period.

Table 4.11. Momentum factor and global sentiment

| | Coefficient | T-statistic | p-value |
|----------|-------------|-------------|---------|
| C | 0,019 | 2,902 | 0,000 |
| WISI(-1) | -0,015 | -1,929 | 0,05 |

The results (table 4.11) clearly confirms the hypothesis. The sentiment indicator significantly advances the behavior of the momentum factor or, equivalently, the momentum effect reacts to sentiment specific information⁵³.

4.5. Conclusions

This paper proposes an alternative methodology to estimate risk premium that challenges traditional financial approaches. In fact, it considers investor sentiment as a source of superior information to market prices. As a result of this additional informational contribution several conclusions can be highlighted: first, the introduction of a changing beta over time overcomes the limitations of traditional valuation models that measure risk premium for the whole period. Therefore, the sentimental valuation model is more efficient than traditional formulations.

Secondly, the incorporation of investor sentiment contributes to further refining the risk premium estimate, even better than the conditional versions of traditional valuation models that do not take sentiment into consideration.

Thirdly, and for the full sample, the positive risk-return relationship is confirmed since a more precise measure of systematic risk is found. This does not mean that within the sample period there are not moments in which the relationship is negative, but it provides a justification as to why in these periods high beta sectors do not have high returns. The explanation would be linked to the evolution of investor sentiment and sector sensitivity to it⁵⁴(especially, in times of reversal of global investor sentiment).

⁵³ Overcoming formulations supported by macroeconomic variables that are unable to explain this momentum effect (Griffin et al., 2003).

⁵⁴ In line with the work of Bams et al. (2015) from the perspective of higher moments in the distribution of returns.

Fourthly, an additional contribution to traditional explanations in the literature (mostly related to volatility factors) focuses on the role of investor sentiment as the key determinant of the risk-return relationship.

Finally, the incorporation of sentimental information allows to control some of the anomalies not explained from the stand of classical financial theory. Precisely, it is found that the momentum effect reacts to sentiment information and ceases to be significant once investor sentiment is explicitly incorporated into the risk premium estimation.

Appendix 1. A web-based investor sentiment index (WISI)

This index assumes that sentiment is a function of information. Prices are not revealing all the information needed to capture investor expectations (Da et al., 2011) and therefore a sentiment measure built on information beyond prices should be more efficient way to capture those expectations. This non-revealed information can be captured by making an active use of information technologies. Digging into data coming out from these sources can be highly productive at the time of reaching conclusions on collective behavior as they are commonly accessible to global investor community at the same time (Dietzel et al. 2015; Fricke et al., 2014).

Specifically, this measure will rely on the use of search engine queries. This approach searches for the existence of a relationship between number of queries of specific terms and the subsequent reaction of economic and financial variables.

Google has become the most prominent tool used here (Artola and Galán, 2012; Choi and Varian, 2009; Dimpl and Jank, 2016; Da, Engelberg, and Gao, 2011, 2015; Preis et al., 2013). The relevance of this tool is evident considering that according to Statcounter Global Statistics (www.gsstatcounter.com), Google holds the 91.5% of search engine market share worldwide as of October 2017.

The sentiment index will be created by using monthly data from Google Trends (<https://www.google.com/trends>) for the period ranging from January 2004 to May 2017. This tool provides a Search Volume Index (SVI) scaled from 0 to 100 to show the degree of popularity of a search into the whole sample being 100 the highest search volume achieved for an item. Additionally, items will be search within the “finance” category with a global scope.

Regarding to the items to be searched, Loughran & Mcdonald (2011) show that popular dictionaries of terms are not the best instruments to be used. For instance, according to

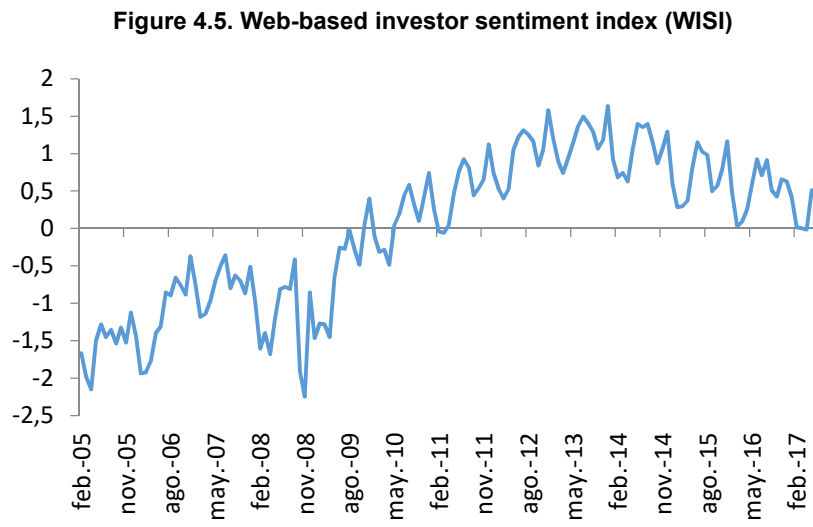
their study, three quarters of word counts in 10-k filings based on Harvard dictionary are typically not negative in a financial context. To avoid these biases this paper will follow Preis et al. (2013) who have defined a comprehensive selection of words with a financial tilt⁵⁵. Some fine-tuning will be applied to that list by removing those elements that do not present a clear meaning or have insufficient queries and including some additional terms related to fixed income assets (precisely, “bankruptcy”, “yield” and “capital”) to avoid an equity bias.

Next step is to make a proper data treatment that in the case of time series will mean to look for the proper ARIMA-SARIMA specification consistent with stationary series. Once proper data generation processes have been identified for each SVI the objective is the extraction of information from those series that can be considered essentially as investor sentiment. Then, a principal component analysis (PCA) methodology will be used to extract the common elements among those series. Provided the type of items searched and their categorization the assumption behind is that the common nexus among those series should be a representation of investor sentiment.

These components can be attached to different elements being part of explanations of sentiment in the existent literature. Precisely, three components will be finally selected such as the first one, considers variables that can be representative of financial expectations (PC1). The second component can be thought as a representation of wealth expectations (PC2). And, the third component can be interpreted as a reflection of economic expectations (PC3). The combination of those items will provide the global

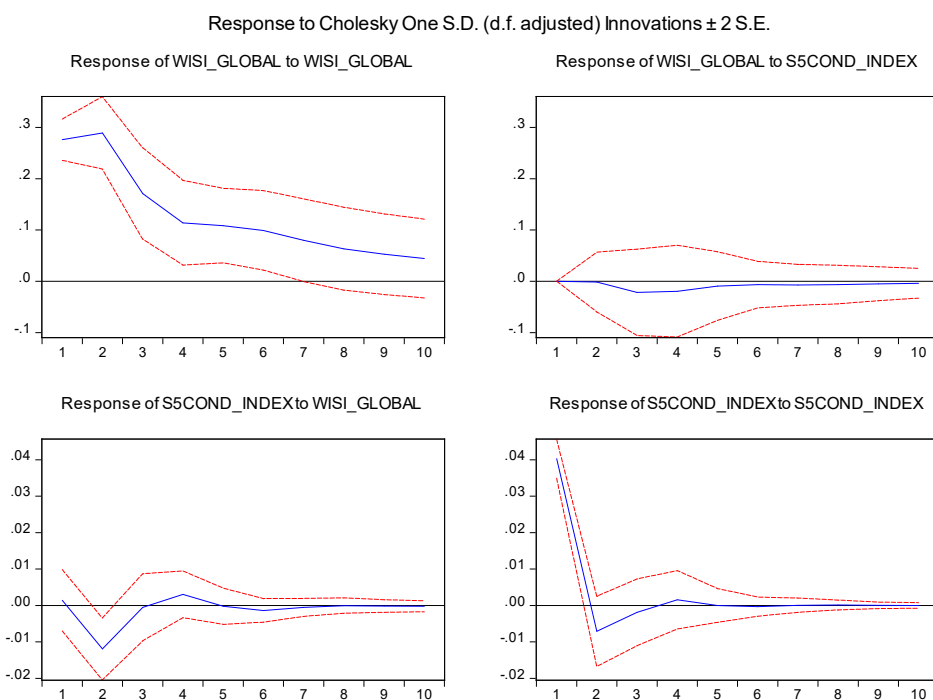
⁵⁵ List of the terms searched on Google Trends: debt, stocks, restaurant, portfolio, inflation, housing, revenue, bankruptcy, credit, yield, unemployment, growth, investment, hedge, wedding, divorce, bonds, derivatives, profit, leverage, loss, cash, office, fine, S&P500, banking, financial crisis, happy, car, capital, finance, short sell, invest, fed, travel, expected return, gain, default, water, rich, risk, oz. of gold, success, oil, war, economy, lifestyle, greed, food, movie, ore, hold, opportunities, health, short sell, arts, culture, bubble, purchase, tourism, politics, energy, consumption, dividend, conflict, forex, home, crash, transaction, fond, work, fun.

investor sentiment index (WISI) that will be used in the conditional valuation model (figure 4.5).

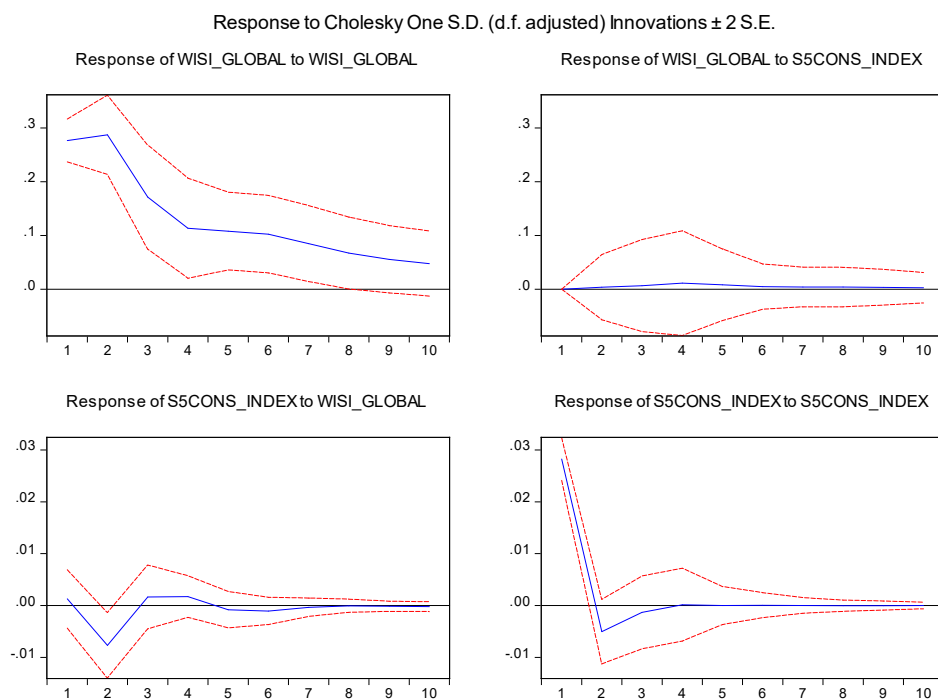


Appendix 2. Impulse-response analysis

1. Cyclical consumption.

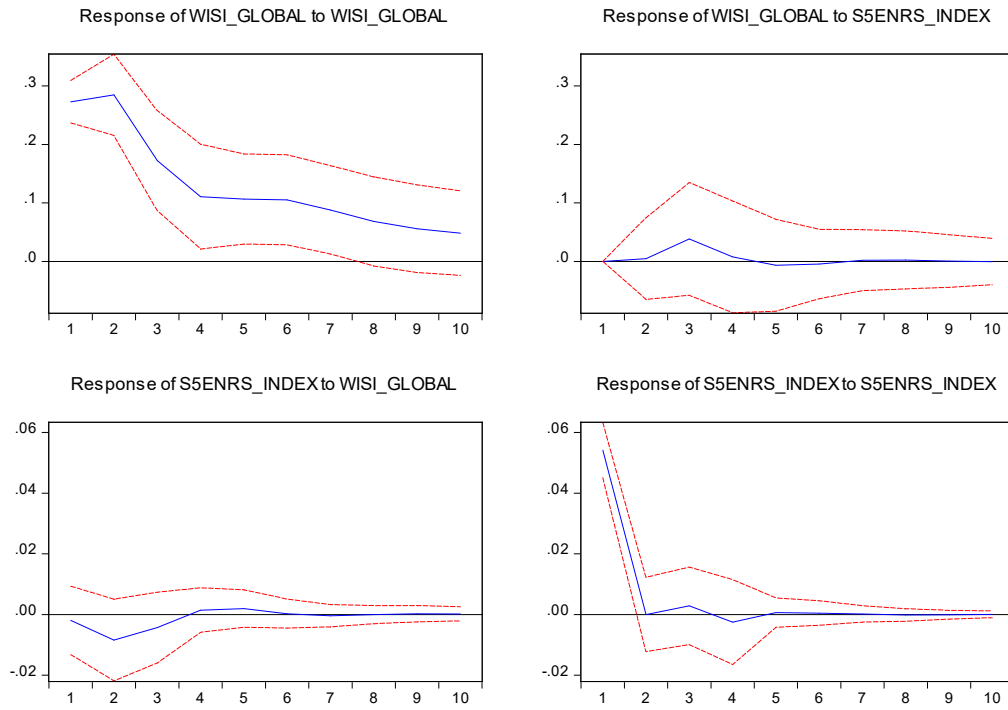


2. Staples.



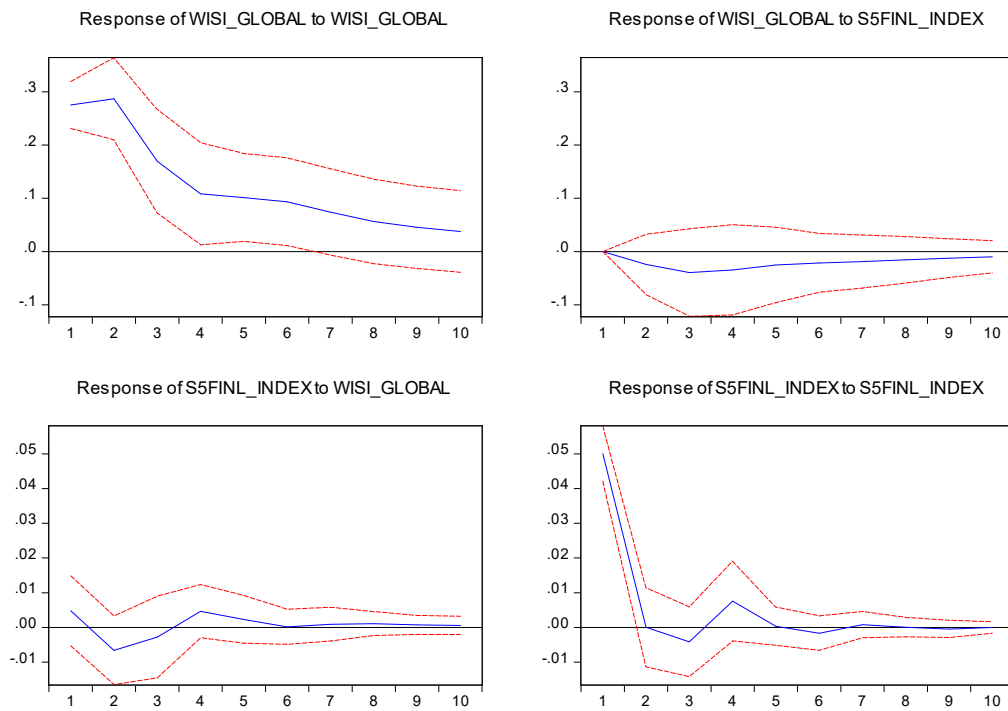
3. Energy.

Response to Cholesky One S.D. (d.f. adjusted) Innovations ± 2 S.E.



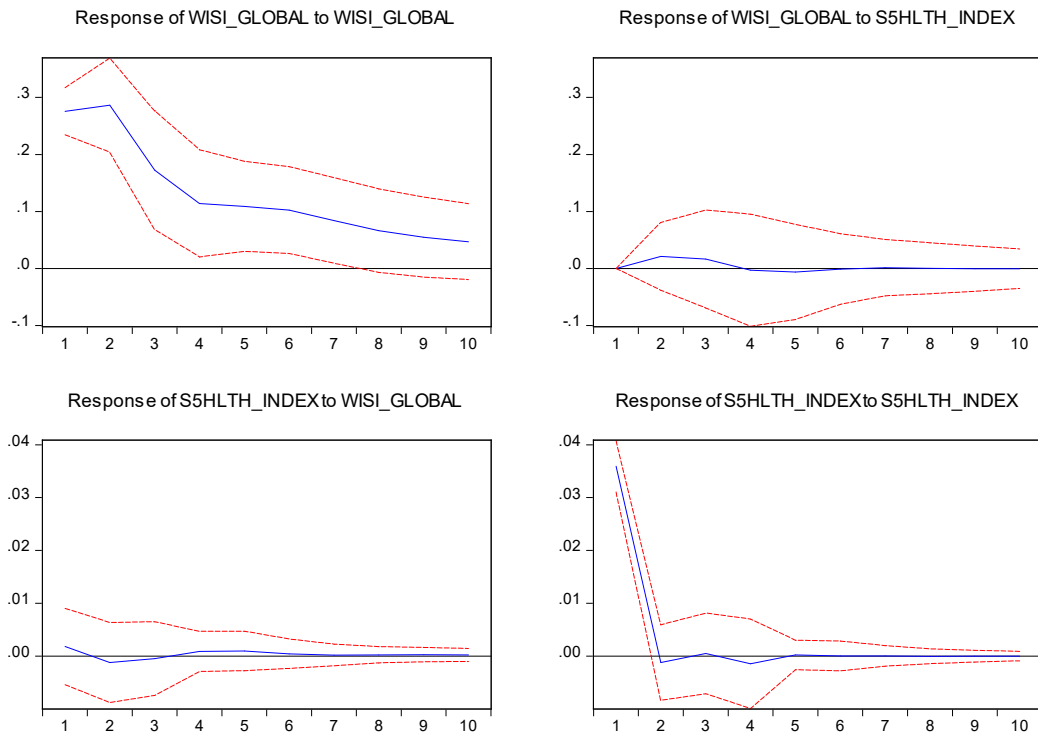
4. Financials.

Response to Cholesky One S.D. (d.f. adjusted) Innovations ± 2 S.E.



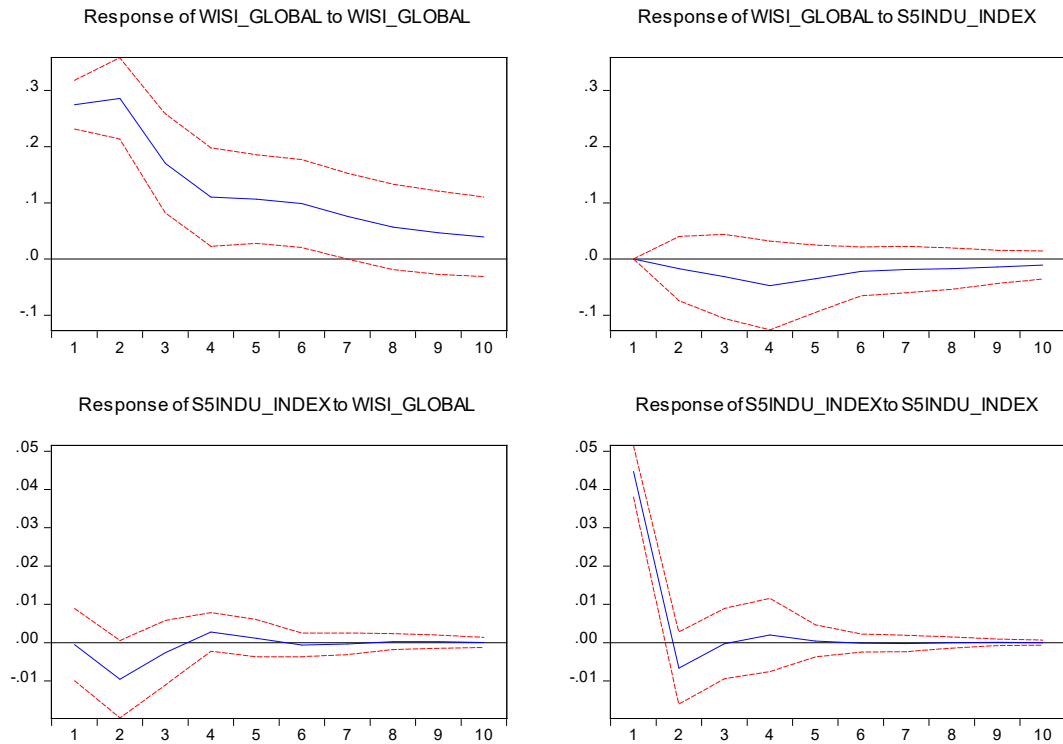
5. Healthcare.

Response to Cholesky One S.D. (d.f. adjusted) Innovations \pm 2 S.E.



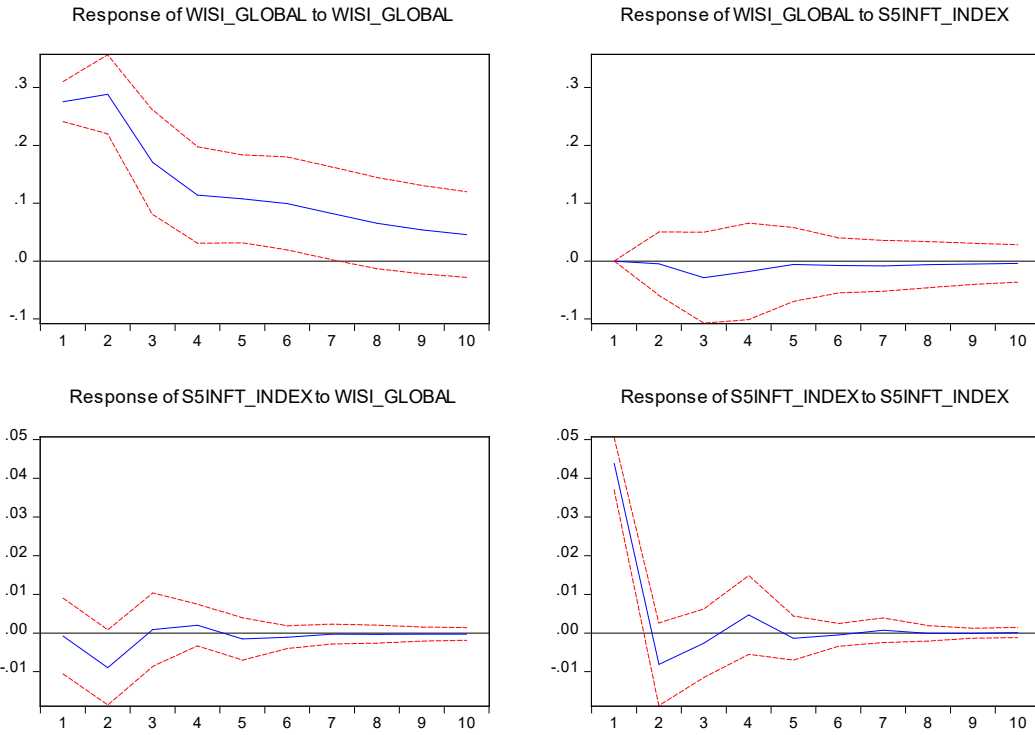
6. Industrials.

Response to Cholesky One S.D. (d.f. adjusted) Innovations \pm 2 S.E.



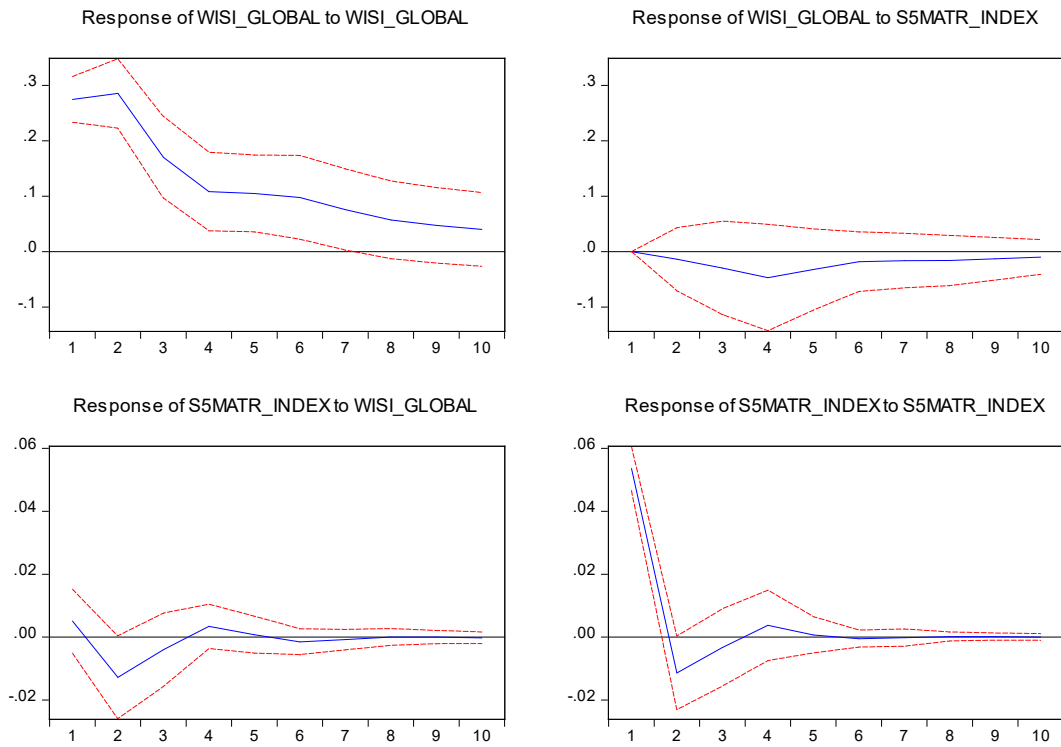
7. Information technology.

Response to Cholesky One S.D. (d.f. adjusted) Innovations ± 2 S.E.



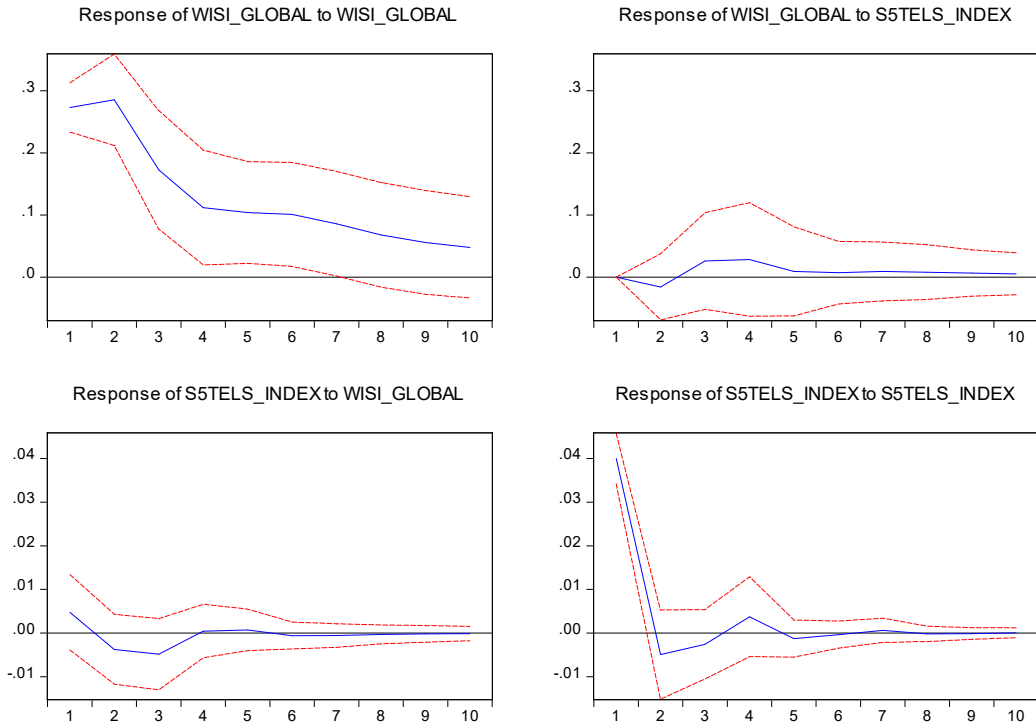
8. Materials.

Response to Cholesky One S.D. (d.f. adjusted) Innovations ± 2 S.E.



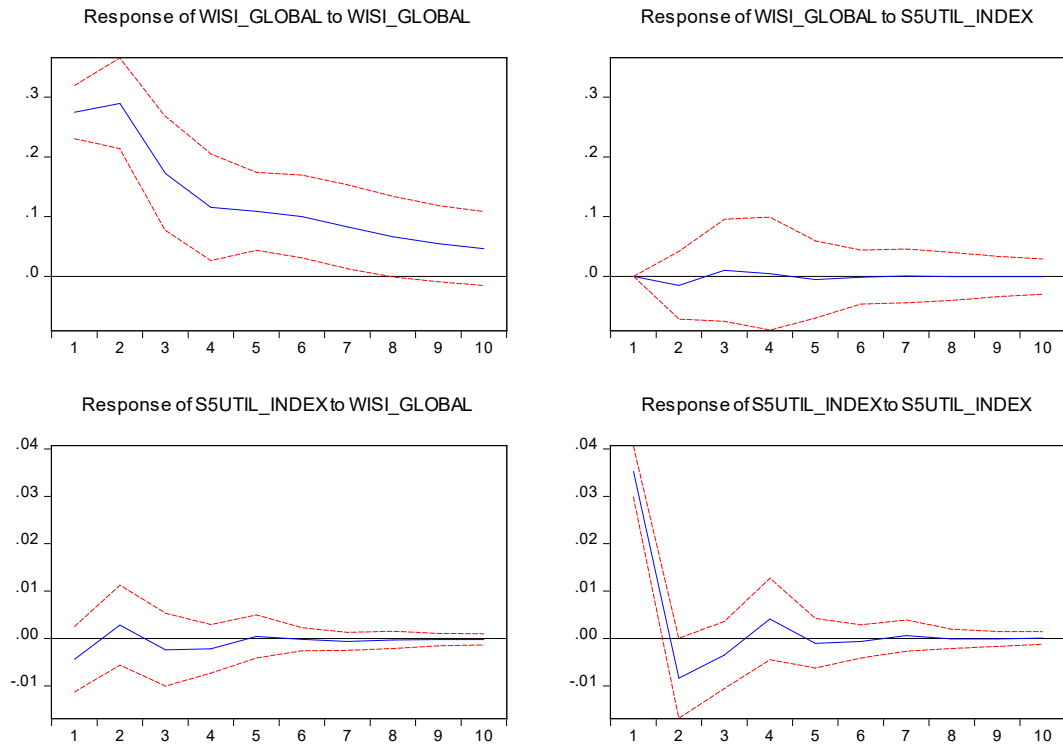
9. Telecommunications.

Response to Cholesky One S.D. (d.f. adjusted) Innovations ± 2 S.E.



10. Utilities.

Response to Cholesky One S.D. (d.f. adjusted) Innovations ± 2 S.E.



References

- Alcalá et al. (1993). Prima de riesgo y volatilidad en el mercado de valores español. *Revista de Economía Aplicada*, 1,3, 95-117.
- Alonso, F., & Restoy, F. (1995). La remuneración de la volatilidad en el mercado español de renta variable. Banco de España, documento de trabajo 9508, 1-49.
- Amin, R., & Ahmad, H. (2013). Does Investor Attention Matter ' S ? *Journal of Public Administration, Finance and Law*, (4), 111–125.
- Antoniou, C., Doukas, J. A., & Subrahmanyam, A. (2016). Investor Sentiment, Beta, and the Cost of Equity Capital. *Management Science*, 62,2, 347-367.
- Aragó, V., & Matallín, J.C. (2002). Relación entre la volatilidad y la prima de riesgo. Evidencia para el mercado español. <https://xiforofinanzas.ua.es/programa/trabajos/1023.pdf>.
- Avramov, D., & Chordia, T. (2006). Predicting stock returns. *Journal of Financial Economics*, 82,2,387-415.
- Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, LXI(4), 1645–1680.
- Bams, D., Honavar, I., & Lehnert, T. (2015). From Time Varying Risk-Aversion to Anomalies in Market Moments' Risk Premia. <http://dx.doi.org/10.2139/ssrn.2692983>.
- Barberis, N., & Thaler, R.H., (2003). A survey of behavioral finance. In Constantinides, G., Harris, M., & Stulz, R. (Eds.), *Handbook of the Economics of Finance* (pp.1052-1090), Amsterdam: North-Holland.
- Ben Sita, B. (2018). Estimating the beta-return relationship by considering the sign and the magnitude of daily returns. *Quarterly Review of Economics and Finance*, 67, 28-35.
- Bertella, M.A, Pires, F.R., & Feng, L. (2014). Confidence and the stock market: an agent based approach. *PLoSone* 9(1): e83488. <https://doi.org/10.1371/journal.pone.0083488>.
- Best, P & Byrne, A. (2001). Measuring the equity risk premium. *Journal of Asset Management*, 1, 3, 245-256.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of Economics and Statistics*, 69(3), 542-547.
- Bollerslev, T. (1990). Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model. *The Review of Economics and Statistics*, 72,3,498-505.
- Bollerslev, T., R.Y. Chou, & K.F. Kroner (1992). ARCH modelling in finance. *Journal of Econometrics* 52, 5-59.

- Chou, R. (1988), "Volatility persistence and stock valuations: some empirical evidence using GARCH", *Journal of applied econometrics*, 3, 279-294.
- Cochrane, J. H. (2011). Presidential Address: Discount Rates. *Journal of Finance*, 66(4), 1047–1108.
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *Journal of Finance*, 66, 1461–1499.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98, 4, 703-738.
- De Santis, G., & Gerard, B. (1997). International asset pricing and portfolio diversification with time varying risk. *Journal of Finance*, 52, 1881-1912.
- De las Heras, A., & Nave, J. (2003). Análisis no paramétrico de estacionalidad en los rendimientos de la deuda pública española. *Revista de Economía Financiera*, 1, 65-79.
- Dimson, E., & Marsh, P. (1984). The impact of the small firm effect on event stocks and the performance of published UK stock recommendations. *Journal of Financial Economics*, 17, 113-142.
- Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20(3), 339-350.
- Fama, E., & McBeth, J. (1973). Risk, return and equilibrium: empirical tests. *Journal of Political Economy*, 71, 607-636.
- Fama, E., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25, 1, 23-49.
- Fama, E., & French, K. (1993). Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics*, 33, 3-56.
- Ferson, W. E., & Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. *Journal of Finance*, 54(4), 1325-1360.
- Fox, E. G., Fox, M. B., & Gilson, R. J. (2016). Economic crisis and the integration of law and finance: The impact of volatility spikes. *Columbia Law Review*, 116, 2.
- Gagliardini, P., Ossola, E., & Scaillet, O. (2016). Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets. *Econometrica*, 84(3), 985–1046.
- Glushkov, D. (2006). Sentiment Beta. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=862444.
- Griffin, J. M., Ji, X., & Martin, J. S. (2003). Momentum Investing and Business Cycle Risk: Evidence from Pole to Pole. *The Journal of Finance*, 58, 6, 2515-2547.

- Hansen, L., & Richard, S. (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica*, 55, 587-613.
- Hansson, B., & Hördahl, P. (1998). Testing the conditional CAPM using multivariate GARCH-M. *Applied Financial Economics*, 8, 377-388.
- Ho, C., & Hung, C. H. (2009). Investor sentiment as conditioning information in asset pricing. *Journal of Banking and Finance*, 33(5), 892-903.
- Hurst, B., Ooi, Y. H., & Pedersen, L. H. (2017). A Century of Evidence on Trend-Following Investing. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2993026.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: implication for stock market efficiency. *Journal of Finance*, 48, 65-92.
- Johnk, D. W., & Soydemir, G. (2015). Time-Varying Market Price of Risk and Investor Sentiment: Evidence from a Multivariate GARCH Model. *Journal of Behavioral Finance*, 16(2), 105-119.
- Kyriacou, K., Madsen, J., & Mas, B. (2004). The equity premium. https://www.researchgate.net/publication/4924220_The_equity_premium.
- Lee, B. S., & Mauck, N. (2014). Information Asymmetry and the Market Response to Open Market Share Repurchases. https://mpa.ub.uni-muenchen.de/54066/1/MPRA_paper_54066.pdf.
<http://dx.doi.org/10.2139/ssrn.2403377>.
- Lettau, M., & Ludvigson, S. (2001). Consumption, aggregate wealth and expected stock returns. *The Journal of Finance*, 56(3), 815-849.
- Lewellen, J. (2002). Momentum and Autocorrelation in Stock Returns. *Review of Financial Studies*, 15, 2, 533-564.
- Lewellen, J., Nagel, S., & Shanken, J. (2010). A Skeptical Appraisal of Asset-Pricing Tests A Skeptical Appraisal of Asset-Pricing Tests. *Journal of Financial Economics*, 96(2), 175-194.
- Lo, A. W. (2004). The Adaptive Markets Hypothesis. *The Journal of Portfolio Management*, 30(5), 15–29.
- Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17(1), 59-82.
- Morelli, D. (2003). Capital asset pricing model on UK securities using ARCH. *Applied Financial Economics*, 13(3), 211–223.
- Moskowitz, T. J., & Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance*, 54(4), 1249-1290.

- Ng, L. (1991). Tests of the CAPM with time varying covariances: a multivariate GARCH approach. *Journal of Finance*, 46, 1507-1521.
- Roll, R., & Ross, S. A. (1994). On the Cross-sectional Relation between Expected Returns and Betas. *The Journal of Finance*, 49(1), 101-121.
- Saiti, B., Bacha, O. I., & Masih, M. (2014). The diversification benefits from Islamic investment during the financial turmoil: The case for the US-based equity investors. *Borsa Istanbul Review*, 14(4), 196-211.
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1-48.
- Soufian, N. (2001). Empirical content of CAPM and APT across time. <https://e-space.mmu.ac.uk/1462>.
- Stambaugh, R. (1982). On the exclusion of assets from tests of the two parameter model. *Journal of Financial Economics*, 10, 235-268.
- Stambaugh, R. F., Yu, J., & Yuan, Y. (2014). The long of it: Odds that investor sentiment spuriously predicts anomaly returns. *Journal of Financial Economics*, 114(3), 613–619.
- Stock, J. H., & Watson, M. W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97(460), 1167–1179.
- Tang, G. Y. N., & Shum, W. C. (2003). The conditional relationship between beta and returns: Recent evidence from international stock markets. *International Business Review*, 12(1), 109-126.
- Thaler, R. H. (2000). From Homo Economicus to Homo Sapiens. *Journal of Economic Perspectives*, 14(1), 133–141.
- Vassalou Maria, X. Y. (2004). Default Risk in Equity Returns. *The Journal of Finance*, 59(2), 831-868.
- Wang, K. (2003). Asset pricing with conditioning information: a new test. *The Journal of Finance*, 58, 161-196.
- Yook, K. C. (2010). Long-run stock performance following stock repurchases. *Quarterly Review of Economics and Finance*, 50, 323–331.
- Zhang, L. (2005). The value premium. *Journal of Finance*, 60(1), 67-103.

5. Conclusions, limitations and future research

5.1. Conclusions

This thesis contributes to the existing literature on the study of the risk premium by setting a symbiosis between the different theoretical approaches to it. In this regard, starting from a modeling typical of the classical financial theory, specific elements to behavioral finance are incorporated in order to highlight the presence of undisclosed information about investors' preferences. The incorporation of this information into the study translates into a more accurate definition of the stock market's risk premium.

This achievement has gone through the overcoming of two intermediate objectives already outlined in the introduction chapter:

1. The establishment of a quantitative measure of investor sentiment.

Based on the analysis of the existing literature on the investor sentiment, a novel definition of investor sentiment is elaborated. This definition entitles the development of a conceptual framework founded on the processing of information as the key factor for the configuration of a sentiment measure.

Following this interpretation of sentiment, a methodology is developed underpinned with the use of information technologies as the source of data extraction on the agents' intentions and that, treated using factorial methods, allows the creation of a broadly representative measure of sentiment.

This approach surpasses the results obtained from the use of direct methods of sentiment estimation, such as those based on surveys (which, despite showing a significant contemporary explanatory capacity, miss the predictive capabilities of the one developed here). It also surpasses the most popular formulations based on the use of economic information (e.g., Baker and Wurgler, 2006), or even those based on the use of web

searches (e.g., Gao et al., 2016) but lacking an economic explanation behind the results obtained and the relationship with the dynamics of financial assets.

Overall, this thesis concludes that it is possible to create a broadly representative measure of investment sentiment, grounded on the search for information. The use of aggregate information from web searches, in combination with the previously mentioned methodology, makes it possible to identify and quantify the sentiment component by creating an indicator free of assumptions and representative of collective behavior.

The results obtained when testing the significance of the sentiment indicator in explaining the evolution of financial assets, have shown the relevance of sentiment not only as an explanatory but as a predictive factor for the performance of those assets. This role has been prominent particularly over the last decade, in line with the strong global penetration of information technologies.

Additionally, setting a link between sentiment and information has made it possible to identify a process of globalization in this sentiment supported by the transmission of information flows and the high penetration and global accessibility of information technologies.

2. The elaboration of a sentimental risk premium measure.

The present work highlights the informative limitations of the classical theories in determining stock market risk premium levels, even using alternative processes to extract as much information as possible from the market. Therefore, the incorporation of undisclosed information about agents' intentions contributes to a greater precision in the estimation of that risk premium.

In this line, a methodology is developed that, starting from classical foundations (e.g., Morelli's (2003) conditional capital asset modelling), implements the information set by incorporating the estimated investor sentiment.

The consideration of the sentimental component helps both to isolate more clearly the market systematic risk and to dynamically monitor the exposure to it. Then, a definition for a dynamic risk premium much closer to reality is finally achieved.

Moreover, all this contributes to improve modelling and market efficiency as it incorporates additional information to the one present in asset prices and it is done in an increasingly faster way as the processes of global dissemination of information also increase their speed.

The empirical contrast of the model also provides interesting results that make an additional contribution to the existing literature. Firstly, it is observed that sectors with higher beta match those with higher sensitivity to global sentiment.

Secondly, considering the linear relationship between beta and return, there is a greater probability of positive returns on beta increases in those sectors with a greater sensitivity to sentiment, contrary to the traditional literature that has shown a neutral relation between profitability and beta.

In addition, once the systematic risk component is isolated once the information component is considered and, for the whole period, the positive risk-return relationship is confirmed. This result does not mean that within the sample there are no periods in which the relationship might be negative, but it provides a proper explanation to it linked to the evolution of investor sentiment. A different explanation to traditional literature on the subject that links that relationship mainly to volatility factors.

In short, this work highlights the improvement achieved in the determination of the risk premium derived from the introduction of a sentimental component into the classic valuation models or equivalently the superior results that can be obtained from mixed formulation that incorporate behavioral elements and sound classical methodologies.

5.2. Limitations and future research

The results obtained in the work are positive and promising on the success of the methodologies used. However, it is also true that the proposals are not exempt from certain limitations that could limit their scope.

Among these limitations, the most relevant are:

1. Within the formulation of the sentiment-information theoretical relationship, one of the key pieces is the diffusion of information and its globalization. Although this relationship is evident in the dynamics of penetration of information technologies and the increase in their accessibility, the truth is that there is no quantification of the speed at which the information is disseminated (which it contributes to the homogenization of expectations). In this regard, the application of methodologies from other fields might be relevant, such as the application of search methods to epidemiology and how information flows can anticipate the evolution of disease transmission (Ginsberg et al., 2009; Fenichel et al., 2011; Towers et al., 2015).
2. An additional aspect for improvement comes from the methodology applied to the estimation of the conditional sentimental risk premium and precisely to the estimation of the conditional covariance and its assumed generalization for all sectors. Probably, in some sectors less sensitive to sentiment it would be possible to consider a methodology that captures the information in an alternative way (starting with a more precise definition of the underlying data generating processes), contributing, for instance, to a greater control of the momentum anomaly that it is not fully explained for those sectors.
3. Another element to consider would be a better definition of the items used in the analysis. The term selection considered in this thesis (based on the work of Preis et al., 2013) presents the limitation of having an Anglo-Saxon bias. This fact is a limitation

in the face of the growing weight of emerging markets (eg. China) as well as from other developed markets (eg. Germany) in which local language is the most relevant. Likewise, we must consider the dynamism also of the items that might also vary over time.

4. Finally, an additional research dimension is opened within the ethical-regulatory scope insofar as the presence of a quasi-monopoly in the search engines can raise doubts about a risk of conditioning or even manipulation (Hirshleifer, 2015) on agent's behavior derived from the handling of information.

References

- Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, *LXI* (4), 1645–1680.
- Fenichel, E. P., Castillo-chavez, C., Ceddia, M. G., Chowell, G., & Gonzalez, P. A. (2011). Adaptive human behavior in epidemiological models. <https://www.pnas.org/cgi/doi/10.1073/pnas.1011250108>.
- Gao, Z., Ren, H., & Zhang, B. (2016). Googling Investor Sentiment around the World. http://www.cuhk.edu.hk/fin/event/symposium/SEFM_2016_paper_58.pdf.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014.
- Hirshleifer, D. (2015). Behavioral Finance. *Annual Review of Financial Economics*, *7*(1), 133–159.
- Morelli, D. (2003). Capital asset pricing model on UK securities using ARCH. *Applied Financial Economics*, *13*(3), 211–223.
- Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, *3*(1684), 1-6.
- Towers, S., Afzal, S., Bernal, G., Bliss, N., Brown, S., Espinoza, B., & Castillo-Chavez, C. (2015). Mass Media and the Contagion of Fear: The Case of Ebola in America. *PloS One*, *10*(6), 1-13.