



IE UNIVERSIDAD

TESIS DOCTORAL/DOCTORAL DISSERTATION

**Divulgación, utilización e interpretación de la información
financiera de los gobiernos: tres ensayos/Disclosure, use and
interpretation of the financial information of governments:
three essays**

Claudio Columbano

SEGOVIA, 2020



IE UNIVERSIDAD

TESIS DOCTORAL/DOCTORAL DISSERTATION

Divulgación, utilización e interpretación de la información financiera de los gobiernos: tres ensayos/Disclosure, use and interpretation of the financial information of governments: three essays

Claudio Columbano

Doctoral Thesis Advisor: Marco Trombetta, Ph.D.

ABSTRACT

This dissertation offers three studies aimed at shedding new light on the disclosure, use and interpretation of the financial information released by governments. The first two chapters are based on a hand-collected dataset covering the forecasts disclosed by European Union governments over the period 2001-2018. The third chapter is based on 27 interviews with sovereign rating personnel working at four credit rating agencies and on archival material. The chapters contribute novel perspectives on the peculiar financing reporting problem that confronts governments. In particular, the first chapter shows that the transparency of governmental financial disclosures responds to a cost-benefit trade-off similar to that which characterizes corporations. Accordingly, governments' disclosure choices are self-interested, but they are not always driven by opportunistic motives. The second chapter reveals that equity investors are active users of government forecasts, although they interpret them heterogeneously. In particular, the credibility of government's commitment appears to explain the emergence of differential interpretations among investors when governments issue optimistic forecasts. The third chapter documents that credit rating analysts, too, differ in their interpretation of the financial information released by governments and use their subjective judgment to make up for the shortcomings of automated valuation models. Collectively, the dissertation extends extant accounting research by providing new perspectives on the understudied financial reporting problem that characterizes governments.

RESUMEN

Esta tesis doctoral ofrece tres estudios destinados a arrojar nueva luz sobre la divulgación, utilización e interpretación de la información financiera publicada por los gobiernos. Los dos primeros capítulos se basan en un conjunto de datos recogidos por el autor que abarcan las previsiones divulgadas por los gobiernos de la Unión Europea durante el período 2001-2018. El tercer capítulo se basa en 27 entrevistas con analistas de calificación soberana que trabajan en cuatro agencias de calificación crediticia y en material de archivo. Los capítulos aportan perspectivas novedosas sobre el peculiar problema de la presentación de informes financieros que afrontan los gobiernos. En particular, el primer capítulo muestra que la transparencia de la información financiera de los gobiernos responde a una compensación de costes y beneficios similar a la que caracteriza a las empresas. En consecuencia, las decisiones de los gobiernos en materia de divulgación de información responden a intereses propios, pero no siempre están impulsadas por motivos oportunistas. El segundo capítulo revela que los inversores en capital social son usuarios activos de las previsiones gubernamentales, aunque las interpretan de manera heterogénea. En particular, la credibilidad del compromiso del gobierno parece explicar la aparición de interpretaciones diferenciales entre los inversores cuando los gobiernos emiten previsiones optimistas. En el tercer capítulo se documenta que los analistas de calificación crediticia también difieren en su interpretación de la información financiera publicada por los gobiernos y utilizan su juicio subjetivo para compensar las deficiencias de los modelos de valoración automatizados. En conjunto, la tesis amplía la investigación contable existente al proporcionar nuevas perspectivas sobre el problema de la información financiera que caracteriza a los gobiernos.

ACKNOWLEDGMENTS

This dissertation would not have been possible without the support and guidance of my advisor, Marco Trombetta. From day one, he believed in the possibility of applying a disclosure lens to analyze the peculiar financial reporting problem that governments face. He was also supportive of my willingness to engage with qualitative research methods to shed light on the behavior of one of the primary users of governments' financial information: sovereign credit rating analysts. I am thankful to Marco for the trust that he demonstrated in my ideas and for the feedback he gave me over the years.

I am also thankful to the faculty of the Department of Accounting and Management Control of IE Business School. In particular, I wish to thank Salvador Carmona, Nieves Carrera and Leopoldo Doadrio. They welcomed me to the Department in September 2015 and made me feel part of it. Thanks also to Amanda Wilford, who enthusiastically shared with me her knowledge of Management Accounting. The days in Castellón de la Plana will be missed. Thanks to Mahmoud Ezzamel, who was there from the beginning and introduced me to sociological approaches to accounting research. A final acknowledgement goes to Ileana Steccolini. She has guided me into the European public sector accounting community and welcomed me to Essex Business School. I wish to acknowledge her encouragement and support, which she showed on all occasions.

The years of doctoral work gave me the privilege to meet doctoral students and junior researchers from literally all over the world. At IE, the doctoral community of students is made of honest, curious, and respectful people that represent the best manifestation of what a university represents. The exchange of cultures, opinions and ideas took place during countless dinners, movies, and many, many drinks. Numerous friends, both old and new, both in Spain and abroad, would deserve a "thank you", too many for me to mention by name. This dissertation is the fruit of the time we spent together.

A final thanks goes to my family who rather patiently witnessed the long process of developing and writing this dissertation. Coming from a family of scientists, I can only hope that this work lives up to the standards that I have grown up seeing.

LIST OF CONTENT

Contents	
ABSTRACT	i
RESUMEN	ii
ACKNOWLEDGMENTS	iii
LIST OF CONTENT	vi
INTRODUCTION	x
INTRODUCCIÓN	xv
CHAPTER 1. WHEN DO GOVERNMENTS “GO DARK”? EVIDENCE ON GUIDANCE IN PERIODS OF UNCERTAINTY	1
1. Introduction	1
2. Related literature	6
3. Theoretical foundations and hypotheses	9
Hypothesis 1: Negative shocks to expectation precision cause a reduction in the amount of guidance	10
Hypothesis 2. The relation between negative shocks to expectation precision and the amount of guidance is moderated by the need for capital and the precision of investors’ expectations	11
4. Research design	12
4.1. <i>Models</i>	12
4.2. <i>Identification strategy</i>	3
5. Data	7
6. Results	11
6.1. <i>Validation of the “uncertainty” measure</i>	11
6.2. <i>Results of the selection model</i>	14
6.3. <i>Results of the baseline model</i>	16
6.4. <i>The role of forecast accuracy</i>	20
6.5. <i>The role of financing needs</i>	21
6.6. <i>The role of investors’ expectations</i>	29
6.7. <i>The role of opportunistic incentives</i>	30
7. Conclusions	36
CHAPTER 2. THE EQUITY INVESTORS’ RESPONSE TO GUIDANCE BY THE GOVERNMENT	37
1. Introduction	37

2. Related literature	42
3. Theoretical foundations and hypotheses	44
3.1. <i>Fiscal guidance as a disclosure event</i>	44
3.2. <i>The effect of fiscal guidance events</i>	46
Hypothesis 1a: Fiscal guidance dates are characterized by abnormal return volatility.....	47
Hypothesis 1b: Fiscal guidance dates are characterized by abnormal trading volume.....	48
3.3. <i>The effect of forecasted improvements in the budget balance</i>	48
Hypothesis 2a: The relation between forecasted budget balance improvements and abnormal return volatility does not depend on credibility.....	50
Hypothesis 2b: The relation between forecasted budget balance improvements and abnormal trading volume does not depend on credibility.....	51
3.4. <i>Operationalizing “credibility”</i>	51
4. Research design	53
4.1. <i>Event study design</i>	53
4.2. <i>Design for the causal effect model</i>	54
5. Data	60
6. Results	64
6.1. <i>Event study</i>	64
6.2. <i>The predictability of expected improvements in the fiscal outlook</i>	70
6.3. <i>Results of the selection model</i>	71
6.4. <i>Diagnostics of the selection model</i>	76
6.5. <i>Results of the baseline model</i>	80
6.6. <i>The role of credibility</i>	83
7. Conclusions	88
CHAPTER 3. THE ROLE OF JUDGMENT IN THE EVALUATION OF GOVERNMENTS: THE CASE OF SOVEREIGN RATING ANALYSTS	90
1. Introduction	90
2. Related literature	94
2.1. <i>Economic research on sovereign ratings</i>	94
2.2. <i>Sociological research on analysts</i>	96
Research Question: How does analysts’ judgment interaction with automated valuation models in the construction of sovereign ratings?.....	98

3. Conceptual framework	98
3.1. <i>Using judgment in the creation of valuation models</i>	99
3.2. <i>Using judgment to deviate from model-implied ratings</i>	99
4. Research Method	102
4.1. <i>Sampling strategy</i>	102
4.2. <i>Selection of interviewees</i>	103
4.3. <i>Other data sources</i>	103
4.4. <i>Interview protocol</i>	104
4.5. <i>Data analysis</i>	107
5. Findings	109
5.1. <i>Automated processes: the models for constructing sovereign ratings</i>	109
5.2. <i>Judgment in the creation of rating models</i>	116
5.3. <i>Judgment in the departure from rating models</i>	120
6. Discussion	126
7. Conclusions	133
CONCLUSIONS	134
CONCLUSIONES	140
APPENDIX A. THE CONSTRUCTION OF THE SCP DATASET	147
1. Retrieving the universe of Stability/Convergence Programmes	147
2. The publication dates of Stability/Convergence Programmes	152
3. Organizing the content of Stability/Convergence Programmes	157
4. Distribution of SCP publication dates	161
APPENDIX B. VALIDATION OF THE TREATMENT ASSIGNMENT MODEL OF CHAPTER 1	167
1. Validation of the propensity score assumptions	167
2. Diagnostics of the propensity score stage	173
APPENDIX C. DETAILS ON THE FILTERING PROCESS FOR THE COMPUSTAT SAMPLE USED IN CHAPTER 2	183
APPENDIX D. DETAILS ON THE CREDIT RATING INDUSTRY ANALYZED IN CHAPTER 3	185
REFERENCES	189

INTRODUCTION

This dissertation consists of three studies that aim at shedding new light on the disclosure, use and interpretation of the financial information released by governments. Compared to the study of corporate financial reporting, the literature on the external use of governmental financial information is substantially smaller (Anessi-Pessina et al., 2016; Kim et al., 2018). Studies on the determinants and consequences of publicly-available governmental financial information are partly limited by the paucity of data comparable both cross-sectionally and over time (Cicatiello et al., 2017; Khagram et al., 2013). In addition, the literature on governmental reporting takes place in different disciplines, ranging from public administration and political science to accounting, finance and economics (Zhao and Wang, 2006). Accordingly, the findings are relatively sparse, and the terminology, methods, and theories are not always consistent with each other. Thus, this dissertation is first and foremost the result of an attempt to tie together the approaches and findings of what is inherently an interdisciplinary field of study. The goal is to bring novel perspectives to the analysis of the financial reporting problem of governments. This topic is likely to become increasingly important over the coming years as governments embark on experimental policies that will likely come at a substantial cost to government finances. It is probable that we will witness a growing responsibility of governmental financial information as a pillar of effective policymaking and a tool for monitoring and coordination.

From a theoretical point of view, the chapters of this dissertation have drawn from primarily two traditions of thought: the economics-inspired disclosure theory – a branch of information economics, and the sociology-inspired stream of valuation studies (Berthoin Antal et al., 2015; Dye, 2001, 2017; Kornberger et al., 2015; Verrecchia, 2001). Disclosure theory represents the foundation of chapter 1 and chapter 2. Valuation studies represent the “domain theory” (Lukka and Vinnari, 2014) of chapter 3. Chapters 1 and 2 are related to each other by the recognition that the availability, timeliness, transparency and informativeness of financial information about the government is likely to be endogenous to

government incentives, and many of these incentives are likely to have a “political” element (Alt et al., 2006; Beyer et al., 2010; Kim et al., 2018). Accordingly, governments’ incentives to disclose or withhold information, and to do so transparently or opaquely must be taken into account before examining the consequences of information. The common thread that unifies chapters 2 and 3 is the notion of “interpretation”. It is well-known that in stylized disclosure models, the same piece of information must be interpreted in the same way by all rational agents (Kandel and Pearson, 1995; Milgrom and Stokey, 1980). Yet, both theory and empirical work recognize that markets are permeated by heterogeneous interpretation of information (Bamber et al., 1997, 1999; Beunza and Garud, 2007; Zuckerman, 1999). Chapter 2 shows that heterogeneous interpretations arise in response to optimistic forecasts by the government when their credibility is in question. Chapter 3 shows that differential interpretations are the result of shortcomings in the availability and comparability of data pertaining to governments and the uniqueness of many events that affect government finances.

The first two chapters are based on a hand-collected, unbalanced panel dataset that covers the guidance issued by all European Union (EU) governments over the period 2001-2018. Guidance is the disclosure phenomenon whereby organizational insiders issue qualitative statements or quantitative forecasts about a future outcome of interest (Hirst et al., 2008). The data I collect include the entire universe of numerical forecasts issued by governments as a result of the reporting mandate introduced in the EU in 1997 as part of the Stability and Growth Pact (SGP). The SGP is a set of two Regulations and one Resolution aimed at reinforcing the fiscal rules introduced in Europe with the 1992 Maastricht Treaty (EU, 1997a, 1997b, 1997c). Under the SGP, government guidance has both a mandatory and a voluntary component. On the one hand, the SGP mandates that governments release annual forecasts on the government debt and the budget balance (EU, 1997a, Articles 4 and 7) in reports known as Stability/Convergence Programmes (SCP). On the other hand, a set of guidelines “on the content and format” of SCPs - issued for the first time in 2001 and updated six times - establish the set of items on which governments should issue guidance

in addition to the mandatory forecasts for the debt and the budget balance. While these guidelines recommend the extent of guidance “disaggregation” (Hirst et al., 2007), its scope is ultimately decided by governments. Therefore, the SGP setting represents a rare case of mandatory guidance accompanied by voluntary disaggregation. Thanks to these institutional idiosyncrasies, I was able to construct an index that captures the extent to which governments have been willing – and able – to be transparent in their guidance. In addition, I estimate a publication date for each SCP and a measure of publication delay. This dataset is described in [Appendix A](#).

The [first chapter](#) uses the SGP dataset to measure the effect of uncertainty on the transparency of government guidance. Distinct about guidance compared to backward-looking disclosures is that one cannot assume that governments always have precise expectations (Dye, 1985; Jung and Kwon, 1988). Drawing on disclosure theory, the analysis shows that governments respond to uncertainty with a reduction rather than an increase in guidance. Therefore, governments behave *as if* periods of uncertainty represent a shock to expectation precision (Verrecchia, 1990). Hence, “going dark” in periods of uncertainty may represent a relatively benign disclosure decision by the government (Leuz et al., 2008). Interestingly, governments continue to guidance in periods of uncertainty when they need to raise capital from bond investors and bond yields are relatively high. However, their guidance is inaccurate and optimistic. Collectively, these results invite caution in drawing inferences from observed levels of transparency. Governments are transparent when they need capital, and transparency does not imply a greater accuracy of information. Indeed, the choice to be relatively opaque in periods of uncertainty does not appear to be always driven by opportunism. While the analysis cannot speak of the counterfactual behavior of governments absent the reporting mandate introduced with the Stability and Growth Pact, it seems that governments, like firms, do not require a disclosure mandate to engage in voluntary guidance to respond to capital-market needs. Indeed, the evidence points to a substantive and underexplored similarity in the guidance behavior of governments and firms.

The [second chapter](#) uses the SCP dataset by examining the information content and consensus effect (Holthausen and Verrecchia, 1990) of government guidance. The results indicate that guidance by the government is an information-rich event in equity markets. Both return volatility and trading volume are abnormally high on the day of the public release of guidance. Hence, equity investors pay attention to governmental financial information. The chapter then exploits the mandatory forecasts for the budget balance of the government to examine how investors respond to forecasted improvements in the fiscal outlook. These present a dilemma to investors: on the one hand, improvements in the fiscal outlook are sustainable and thus have persistent effects for firms' future earnings. On the other hand, improvements in the fiscal outlook are not credible because governments have incentives to show compliance with fiscal rules. (Frankel and Schreger, 2013; Gale and Orszag, 2003; Hayn, 1995). The analysis shows that investors are more concerned with credibility than sustainability. In particular, the evidence shows that announced improvements in the fiscal outlook are not informative and cause investors' opinions to diverge when the government is not credible. These findings suggest that credibility plays an important role in moderating investors' response to governments' guidance.

The [third chapter](#) uses qualitative methods to examine the valuation practices of analysts engaged with the valuation of the credit rating of a sovereign government. Prior literature indicates that sovereign ratings result from both automated valuation models and analysts' judgmental inputs, but it does not explain why judgment is needed and how it manifests in actual practice (D'Agostino and Lennkh, 2016; Ferri et al., 1999; Mora, 2006; Vernazza and Nielsen, 2015). Accordingly, prior literature suggests that rating agencies should separately disclose the rating that results from models and that which results from analysts' judgments. Drawing on 27 semi-structured interviews conducted at four global credit rating agencies between 2016 and 2019 and on archival materials, the analysis reveals the pervasiveness of analysts' judgement in all phases of the credit rating process – including the creation of valuation models. Therefore, the dichotomy between “objective” ratings derived from models and “subjective” ratings derived from analysts' judgment appears to be somewhat misplaced. The

finding that judgment is pervasive is theorized in light of an interpretation of Frank Knight's definition of uncertainty as the manifestation of an extreme classification problem (Boudreaux and Holcombe, 1989; Hoppe, 2007; Knight, 1964). Because governments seldom default, it is challenging to determine objectively which indicators best predict the default event. For the same reason, it is challenging to interpret information unequivocally. Accordingly, sovereign rating analysts confront "uncertainty" in the sense of Knight – the inability to classify individual cases into classes of cases of which outcomes are known. They predict future default risks by applying a substantial degree of judgment. They construct valuation models, too, that are only partly informed by statistical analysis. Accordingly, sovereign ratings are best understood as Knightian "estimates" rather than numerical probabilities. The chapter concludes by proposing that valuation practices in accounting and finance are likely to feature relatively more automated or judgmental processes depending on the extent to which analysts confront Knightian uncertainty – the inability to objectively classify instances.

Collectively, the dissertation extends extant accounting research by shedding light on the disclosure, use and interpretation of governmental financial information. As governments around the world embark in experimental policy initiatives that are likely to have visible effects on government finances, the financial reporting of governments will probably acquire an increasingly significant responsibility. It is hoped that the data and the analysis presented in this dissertation will contribute to this growing and important field of study.

INTRODUCCIÓN

Esta tesis doctoral consta de tres estudios que tienen como objetivo arrojar nueva luz sobre la divulgación, el uso y la interpretación de la información financiera publicada por los gobiernos. En comparación con el estudio de la información financiera de las empresas, la literatura sobre el uso externo de la información financiera gubernamental es sustancialmente menor (Anessi-Pessina et al. 2016; W. J. Kim, Plumlee y Stubben 2018). Los estudios sobre los factores determinantes y las consecuencias de la información financiera gubernamental de acceso público están limitados en parte por la escasez de datos comparables tanto transversalmente como a lo largo del tiempo (Khagram, de Renzio y Fung 2013; Cicatiello, De Simone y Gaeta 2017). Además, la bibliografía sobre la presentación de informes gubernamentales tiene lugar en diferentes disciplinas, que van desde la administración pública y las ciencias políticas hasta la contabilidad, las finanzas y la economía (Zhao y Wang 2006). Por consiguiente, los resultados son relativamente escasos y la terminología, los métodos y las teorías no siempre son coherentes entre sí. Así pues, esta tesis es ante todo el resultado de un intento de vincular los enfoques y conclusiones de lo que es inherentemente un campo de estudio interdisciplinario. El objetivo es aportar perspectivas novedosas al análisis del problema de la información financiera de los gobiernos. Es probable que este tema adquiera cada vez más importancia en los próximos años a medida que los gobiernos se embarquen en políticas experimentales que probablemente supongan un coste sustancial para las finanzas públicas. Es por lo tanto posible que asistamos a una creciente responsabilidad de la información financiera pública como pilar de la formulación de políticas eficaces y como instrumento de supervisión y coordinación.

Desde un punto de vista teórico, los capítulos de esta tesis doctoral se han inspirado principalmente en dos tradiciones de pensamiento: la teoría de la divulgación inspirada en la economía - una rama de la economía de la información - y la corriente de estudios de valoración inspirada en la sociología (Verrecchia 2001; Dye 2001, 2017; Berthoin Antal, Hutter y Stark 2015; Kornberger et al. 2015). La teoría de la divulgación representa el fundamento de

los capítulos 1 y 2. Los estudios de valoración representan la "domain theory" (Lukka y Vinnari 2014) del capítulo 3. Los capítulos 1 y 2 están relacionados entre sí por el reconocimiento de que la disponibilidad, oportunidad, transparencia y divulgación de la información financiera sobre el gobierno es probable que sea endógena a los incentivos del gobierno, y muchos de esos incentivos verisímilmente tengan un elemento "político" (Alt, Lassen y Rose 2006; Beyer et al. 2010; W. J. Kim, Plumlee y Stubben 2018). En consecuencia, los incentivos de los gobiernos para divulgar o retener información, y para hacerlo de manera transparente u opaca, deben tenerse en cuenta antes de examinar las consecuencias de la información. El hilo conductor que unifica los capítulos 2 y 3 es la noción de "interpretación". Es bien sabido que, en los modelos estilizados de divulgación, una misma información debe ser interpretada de la misma manera por todos los agentes racionales (Milgrom y Stokey 1980; Kandel y Pearson 1995). Sin embargo, tanto la teoría como los estudios empíricos reconocen que los mercados están caracterizados por una interpretación heterogénea de la información (Zuckerman 1999; Bamber, Barron y Stober 1997, 1999; Beunza y Garud 2007). El capítulo 2 muestra que las interpretaciones heterogéneas surgen en respuesta a las previsiones optimistas del gobierno cuando se pone en duda la credibilidad de los anuncios. En el capítulo 3 se muestra que las interpretaciones heterogéneas son el resultado de deficiencias en la disponibilidad y la comparabilidad de los datos relativos a los gobiernos y la singularidad de muchos acontecimientos que afectan a las finanzas públicas.

Los dos primeros capítulos se basan en un conjunto de datos de panel desequilibrado, recopilados por el autor, que abarca las previsiones económicas publicadas por todos los gobiernos de la Unión Europea (UE) durante el período 2001-2018. La previsión es el fenómeno de divulgación por el cual miembros de una organización emiten declaraciones cualitativas o cuantitativas sobre un resultado de interés futuro (Hirst, Koonce y Venkataraman 2008). Los datos que reúno incluyen todo el universo de previsiones numéricas emitidas por los gobiernos como resultado del mandato de presentación de informes introducido en la UE en 1997 como parte del Pacto de Estabilidad y Crecimiento (PEC). El

PEC es un conjunto de dos reglamentos y una resolución destinados a reforzar las normas fiscales introducidas en Europa con el Tratado de Maastricht de 1992 (EU 1997a, 1997b, 1997c). En el marco del PEC, la orientación gubernamental tiene un componente obligatorio y otro voluntario. Por un lado, el PEC ordena que los gobiernos publiquen previsiones anuales sobre la deuda pública y el saldo presupuestario (EU 1997a, artículos 4 y 7) en informes conocidos como Programas de Estabilidad/Convergencia. Por otra parte, un conjunto de directrices "sobre el contenido y el formato" de los Programas - emitidas por primera vez en 2001 y actualizadas seis veces - establece el conjunto de partidas sobre las que los gobiernos deben emitir orientaciones además de las previsiones obligatorias sobre la deuda y el saldo presupuestario. Si bien en estas directrices se recomienda el alcance de la "desagregación" de las previsiones (Hirst, Koonce y Venkataraman 2007), su alcance lo deciden en última instancia los gobiernos. Por lo tanto, el establecimiento del SGP representa un caso raro de previsión obligatoria acompañada de una desagregación voluntaria. Gracias a estas idiosincrasias institucionales, pude construir un índice que capta el grado en que los gobiernos han estado dispuestos -y son capaces- de ser transparentes en la divulgación de sus previsiones. Además, calculo una fecha de publicación para cada CPS y una medida de la demora de la publicación. Este conjunto de datos se describe en el Apéndice A.

El primer capítulo utiliza el conjunto de datos del PEC para medir el efecto de la incertidumbre en la transparencia de las previsiones gubernamentales. La diferencia entre la previsión y las revelaciones retrospectivas es que no se puede suponer que los gobiernos tengan siempre expectativas precisas (Jung y Kwon 1988; Dye 1985). Basándose en la teoría de la divulgación, el análisis muestra que los gobiernos responden a la incertidumbre con una reducción más que con un aumento de la previsión. Por lo tanto, los gobiernos se comportan como si los períodos de incertidumbre representaran un choque para la precisión de las expectativas (Verrecchia 1990). Por lo tanto, "oscurecerse" en períodos de incertidumbre puede representar una decisión de revelación relativamente benévola por parte del gobierno (Leuz, Triantis y Yue 2008). Es interesante que

los gobiernos sigan divulgando previsiones en períodos de incertidumbre cuando necesitan obtener capital de los inversores en bonos y los rendimientos de los bonos son relativamente altos. Sin embargo, sus previsiones son inexactas y optimistas. En conjunto, estos resultados invitan a la cautela a la hora de sacar conclusiones a partir de los niveles de transparencia observados. Los gobiernos son transparentes cuando necesitan capital, y la transparencia no implica una mayor exactitud de las previsiones. De hecho, la opción de ser relativamente opaco en períodos de incertidumbre no parece estar siempre impulsada por el oportunismo. Aunque el análisis no puede hablar del comportamiento hipotético de los gobiernos en ausencia del mandato de presentación de informes introducido con el PEC, parece que los gobiernos, al igual que las empresas, no exigen un mandato de divulgación para participar en las previsiones voluntarias para responder a las necesidades del mercado de capitales. De hecho, las pruebas apuntan a una similitud sustantiva y poco explorada en el comportamiento de orientación de los gobiernos y las empresas.

En el segundo capítulo se utiliza el conjunto de datos de PEC examinando el contenido de la información y el efecto de consenso (Holthausen y Verrecchia 1990) de las previsiones de los gobiernos. Los resultados indican que la divulgación de previsiones por parte de los gobiernos representa un evento rico en información en los mercados de valores. Tanto la volatilidad de los rendimientos como el volumen de negociación son anormalmente altos el día de la divulgación. Por lo tanto, los inversores en acciones prestan atención a la información financiera del gobierno. A continuación, el capítulo explota las previsiones obligatorias para el equilibrio presupuestario del gobierno a fin de examinar cómo responden los inversores a las mejoras previstas en las perspectivas fiscales. Esto plantea un dilema a los inversores: por una parte, las mejoras de las perspectivas fiscales son sostenibles y, por lo tanto, tienen efectos persistentes en las futuras ganancias de las empresas. Por otra parte, las mejoras en las perspectivas fiscales no son creíbles porque los gobiernos tienen incentivos para demostrar el cumplimiento de las normas fiscales establecidas por la UE. (Gale y Orszag 2003; Hayn 1995; Frankel y Schreger 2013). El análisis muestra que los inversores están más preocupados por la

credibilidad que por la sostenibilidad de las previsiones. En particular, las pruebas demuestran que las mejoras anunciadas en las perspectivas fiscales no son informativas y hacen que las opiniones de los inversores difieran cuando el gobierno no es creíble. Estas conclusiones sugieren que la credibilidad desempeña un papel importante en la moderación de la respuesta de los inversores a las previsiones divulgadas por los gobiernos.

En el tercer capítulo se utilizan métodos cualitativos para examinar las prácticas de valoración de los analistas que se ocupan de la valoración de la calificación crediticia de un gobierno soberano. En la bibliografía anterior se indica que las calificaciones soberanas son el resultado tanto de modelos de valoración automatizados como de las aportaciones de los analistas con criterio, pero no se explica por qué se necesita el criterio y cómo se manifiesta en la práctica real (Ferri, Liu y Stiglitz 1999; D'Agostino y Lennkh 2016; Mora 2006; Vernazza y Nielsen 2015). En consecuencia, en la literatura anterior se sugiere que las agencias de calificación deberían divulgar por separado la calificación que resulta de los modelos y la que resulta de los juicios de los analistas. Basándose en 27 entrevistas semiestructuradas realizadas en cuatro agencias de calificación crediticia mundiales entre 2016 y 2019 y en material de archivo, el análisis revela la omnipresencia del juicio de los analistas en todas las fases del proceso de calificación crediticia, incluida la creación de modelos de valoración. Por lo tanto, la dicotomía entre las calificaciones "objetivas" derivadas de los modelos y las calificaciones "subjetivas" derivadas del juicio de los analistas parece equívoca. La conclusión de que el juicio es generalizado se teoriza a la luz de una interpretación de la definición de incertidumbre de Frank Knight como la manifestación de un problema extremo de clasificación (Knight 1964; Hoppe 2007; Boudreaux y Holcombe 1989). Dado que los gobiernos rara vez incumplen sus obligaciones, resulta difícil determinar objetivamente qué indicadores predicen mejor el evento de incumplimiento. Por la misma razón, es un desafío interpretar la información de manera inequívoca. En consecuencia, los analistas de calificación soberana se enfrentan a la "incertidumbre" en el sentido de Knight: la incapacidad de clasificar los casos individuales en clases de casos cuyos resultados se conocen. Predecirán los futuros riesgos de

incumplimiento aplicando un grado sustancial de criterio. También construyen modelos de valoración que sólo se basan en parte en el análisis estadístico. En consecuencia, las valoraciones soberanas se entienden mejor como "estimaciones" *à la Knight* que como probabilidades numéricas. El capítulo concluye proponiendo que es probable que las prácticas de valoración en contabilidad y finanzas cuenten con procesos relativamente más automatizados o de juicio, dependiendo de la medida en que los analistas se enfrenten a la incertidumbre - es decir, la incapacidad de clasificar objetivamente los casos.

En conjunto, la tesis amplía la investigación contable existente al arrojar luz sobre la divulgación, el uso y la interpretación de la información financiera gubernamental. A medida que los gobiernos de todo el mundo se embarcan en iniciativas políticas experimentales que probablemente tengan efectos visibles en las finanzas públicas, la presentación de información financiera de los gobiernos adquirirá verisímilmente una responsabilidad cada vez más importante. Se espera que los datos y el análisis presentados en esta tesis doctoral contribuyan a este creciente e importante campo de estudio.

CHAPTER 1. WHEN DO GOVERNMENTS “GO DARK”? EVIDENCE ON GUIDANCE IN PERIODS OF UNCERTAINTY

1. Introduction

This chapter examines governments’ guidance choices in periods of uncertainty. Accounting scholars refer to “guidance” as the release of forward-looking information about the value of the firm (Patell, 1976). The value of guidance is well-documented in the literature and it derives from the information asymmetry that exists between managers and investors: managers privately observe signals about the future value of the firm while investors do not (Ball and Shivakumar, 2008; Beyer et al., 2010). Through guidance, managers attempt to anchor investors’ expectations around these privately observed signals with the goal of steering investment decisions and preventing the formation of over-optimistic expectations (Ajinkya and Gift, 1984; Graham et al., 2005). Consistent with information asymmetry motivating the disclosure of earnings guidance, accounting research finds that in periods of uncertainty – when information asymmetry increases – managers typically supply greater and more timely guidance (Balakrishnan et al., 2014; Billings et al., 2015). Yet, studies have also found that guidance responds to other motives. For instance, firms with poor future prospects typically reduce guidance because issuing guidance comes at a cost and the perceived benefits are limited (Leuz et al., 2008). In addition, firms sometimes issue guidance opportunistically to attract capital while exploiting investors’ inability to separate signal and noise (Kato et al., 2009; Mercer, 2004; Rogers and Stocken, 2005).

In theory, guidance by the government on the future fiscal outlook serves the similar purpose of anchoring investors’ expectations around governments’ privately observed signals about the future trajectory of the budget (Leeper, 2009). It is typically assumed that governments know more about the future fiscal outlook – partly because the government controls the evolution of future revenues and expenditures (Stiglitz, 1999). In periods of uncertainty, the demand for transparency in the *future* conduct of fiscal policy is likely to be particularly high. Therefore, it should be the case that governments issue a large amount of

guidance in periods of uncertainty. In fact, withholding guidance would leave investors at a disadvantage, as the government remains opaque about its “*intentions*, which in turn shape economic and political behavior.” (Hallerberg et al., 2009: 93–95, emphasis added). Thus, one expects governments to react to uncertainty shocks by increasing the supply of guidance on the future fiscal outlook. Yet, what if periods of uncertainty make the government, too, unable to issue accurate guidance, e.g., because governments now observe less precise signals (Penno, 1997)? Shocks to uncertainty would then give rise to a dilemma for governments: disclosing potentially *valuable* information or withholding potentially *inaccurate* information. Surprisingly enough, how governments respond to uncertainty shocks remains unknown.

This chapter exploits the SCP guidance dataset described in [Appendix A](#) to examine the causal effect of uncertainty on the amount of guidance issued by governments. To measure the amount of guidance, a peer-adjusted index of guidance intensity is calculated following prior literature (e.g., Botosan and Plumlee, 2002). The index measures the number of fiscal, economic and financial items on which a government issues a numerical, point forecast at a given forecast horizon. It thus represents an “intensity” measure of transparency (Urquiza et al., 2010). To identify the causal effect of uncertainty, a measure of uncertainty is computed as the quarter-year standard deviation in the mandatory budget balance forecasts released by governments in their SCPs.¹ As this measure is likely to be contaminated by governments’ incentives to portray a healthy fiscal outlook (Frankel, 2011; Hallerberg et al., 2009), identification is achieved by taking advantage of the institutional idiosyncrasies of the EU fiscal framework. Namely, EU fiscal rules allow to model the process that generates non-random assignment to treatment. Accordingly, selection into treatment is modeled through a policy propensity score approach in line with a growing stream of literature (Angrist et al., 2018; Angrist and Kuersteiner, 2011; Jordà and Taylor, 2016). Next, estimated treatment effects are conditioned on the inverse of the

¹ The budget balance of the government is the difference between government revenues and expenditures. It is a key metric of governments’ fiscal health in capital markets and the primary indicator used in EU supranational surveillance of fiscal performance (Mosley, 2003b; Savage, 2005).

probability of being treated (Hirano et al., 2003). Under selection-on-observables, the estimated coefficients are average treatment effects (Rosenbaum and Rubin, 1983).

The findings can be summarized as follows. While uncertainty does not have affect the overall amount of guidance issued by governments, it does reduce the number of forecasts that refer to the future evolution of the government's financial position. Guidance on the financial position of the government includes highly sensitive information of the future evolution of the debt and assets of the government – including the average maturity of bonds, the percentage that will be issued in foreign currency, the future acquisition and disposal of financial assets, as well as cash-accrual reconciliations. These items are collectively known as “stock-flow” or “debt-deficit” adjustments because they affect the financial position of the government without being recorded in the income statement. It is telling that fiscal uncertainty has an effect on these items because prior literature shows that debt-deficit adjustments are often a preferred way for governments to manage the deficit figures (Alt et al., 2014; Savage, 2005; von Hagen and Wolff, 2006). Hence, governments' guidance in periods of uncertainty appears to reflect opportunistic withholding of information.

On the other hand, governments may prefer to withhold these forecasts for “benign” reasons. The difference between benign and opportunistic motives is that the former is optimal to both insiders and investors, while the latter is optimal only to insiders. In the government literature, withholding of information is typically considered to be the result of opportunistic motives (Alesina and Perotti, 1996; Alt et al., 2006; Stiglitz, 1999). That is, governments that disclose less are considered less transparent, and less transparency is considered to be a signal of opportunism. Accordingly, prior studies find that investors reward government transparency (e.g., Glennerster and Shin, 2008). In disclosure models, however, less than full disclosure is not always the result of opportunistic motives. In addition, it is often not detrimental to investors' welfare. Indeed, Leuz, Triantis, and Yue (2008) hypothesize and find that the level of guidance chosen by firms trades off benefits and costs while also responding to insiders' preference to

preserve informational rents. Thus, it is possible that *both* benign and opportunistic motives are at play when governments choose their disclosure strategy. The study proceeds, therefore, with an analysis of heterogeneity in the causal effect of uncertainty on the amount of guidance that governments issue on the future trajectory of government debt.

The first channel examined is that guidance is withheld because it is more likely to be inaccurate in periods of uncertainty. After matching the SCP dataset with data on fiscal outcomes at the annual frequency, the results indicate that the accuracy of forecasts does not significantly diminish in periods of uncertainty. There is evidence that suggests that forecasts are, however, more optimistic in periods of uncertainty, consistent with opportunistic forecasting (Rogers and Stocken, 2005).

The second channel posits that guidance is withheld because governments prefer to avoid issuing inaccurate forecasts unless they derive an immediate benefit. The evidence is strongly consistent with this hypothesis. The results indicate that governments reduce guidance in periods of uncertainty unless they forecast a large deficit in the future. A large deficit *typically* implies that governments will need to finance it by borrowing on capital markets. Indeed, supplemental analysis confirms that governments that *do not* reduce guidance in periods of uncertainty face large rollover needs in the coming year. The results are consistent with governments issuing guidance in periods of uncertainty only when expecting to derive a benefit from it (Leuz et al., 2008).

The third channel relates governments' signals to investors' expectations. According to Verrecchia (1990), the value of guidance depends on the added value that governments' forecasts bring vis-à-vis investors' own expectations. Hence, guidance should be more valuable to investors – and more in demand – when these have noisy expectations. After matching the SCP dataset with data on secondary market government bond yield at the daily frequency, the evidence confirms that those governments that *do not* reduce guidance in periods of uncertainty face with high bond yields. Governments facing low yields “go dark” instead.

Collectively, the findings suggest that governments prefer to avoid committing to a future debt trajectory unless “forced” to do so by capital market pressures. Conversely, governments that reduce guidance in periods of uncertainty appear to do so in response to relatively low demand for information stemming from investors. Overall, these results are consistent with the notion that insiders use disclosures to reduce the “uncertainty gap” faced by investors when financing needs are most pressing and expectations are noisy (Gietzmann and Ostaszewski, 2014). The results are also consistent with Suijs’s (2007) model, in which only those firms with sufficiently uncertain prospects disclose in order to attract capital away from the risk-free asset. Accordingly, governments’ disclosure choices in periods of uncertainty appear to be driven by relatively “benign” motives in the sense of Leuz et al. (2008). At the same time, supplementary analysis examines the characteristics of governments that preserve a stable level of guidance on the future debt trajectory when faced with uncertainty shocks. Interestingly, these governments present evidence consistent with the working of opportunistic motives: their guidance is issued relatively late and it is relatively inaccurate and optimistic. This evidence confirms the intuition by Penno (1997) that in the cross-section, the intensity of guidance negatively relates to its precision. In addition, these governments are characterized by a relatively large use of debt-deficit adjustments. Hence, governments that do not reduce guidance in periods of uncertainty do so to attract capital, yet their guidance is relatively misleading (Rogers and Stocken, 2005). Overall, the evidence suggests that both benign and opportunistic rationales motivate governments’ disclosure choices in periods of uncertainty. Interestingly, by ranking governments based on the quantity of forecasts issued, one would incorrectly conclude that greater transparency is a sign of a “better” government.

The chapter contributes to the growing but limited literature on governments’ disclosure choices (Kim et al., 2018). So far, studies have emphasized the importance of electoral and fiscal incentives in portraying a rosy but ultimately untruthful picture of fiscal *outcomes* (Costello et al., 2017; Kido et al., 2012; Naughton et al., 2015; Vermeer et al., 2012; von Hagen and Wolff, 2006). Others have shown that fiscal distress, economic conditions and political dynamics

determine governments' willingness to be transparent about the budget *process* (Alt et al., 2006; Cicatiello et al., 2017; Ríos et al., 2016; Wehner and de Renzio, 2013). By focusing on the understudied phenomenon of government guidance, this chapter examines a setting where the government *may* hold imprecise view about the future fiscal outlook. The strongest contribution is to show that upon observing less than full disclosure, one should not always infer relatively sinister rationales. There are reasons for less than full disclosure to be optimal to both governments and investors. Thus, the present application of a purely disclosure lens to the analysis of governments' guidance choices expands our understanding of the government's disclosure problem in general, and of the costs and benefits of fiscal transparency in particular.

There are three additional contribution. First, this study examines the disclosure of fiscal information reported under the "statistical mode" of government accounting (Heald and Hodges, 2018) – a relatively understudied area. Second, the treatment effect approach based on inverse-probability-weighting techniques is relatively novel. It may be of interest to scholars working on the role of fiscal constraints on governments' disclosure choices (Chaney et al., 2002; Costello et al., 2017) because that these constraints effectively generate a rule for non-random assignment of units to treatment (Angrist and Kuersteiner, 2011). Finally, the empirical results on the different drivers of various facets of transparency confirm that important insights can be gained by studying the micro-aspects of disclosures indices (e.g., Copelovitch et al., 2018; de Renzio and Wehner, 2017).

The paper proceeds as follows. Section 2 reviews related literature. Section 3 discusses how disclosure theory leads to novel predictions about the voluntary amount of guidance that governments disseminate in periods of uncertainty. The research design is explained in section 4 and the data in section 5. Section 6 contains the results of the analysis and section 7 concludes.

2. Related literature

Research on the disclosure of information by governments is inherently interdisciplinary. Studies are found in accounting, political science, political

economy and public administration (Zhao and Wang, 2006). In accounting, the focus is mostly on US state governments' decision to misreport – i.e. to hide the true fiscal performance through the opportunistic use of accounting discretion.² For example, Chaney, Copley and Stone (2002) document that states in fiscal distress that are subject to a balanced budget requirement reduce funding requirement and reported pension liabilities by choosing inflated discount rates. Zhao and Wang (2006) report that states in fiscal distress and characterized by limited political competition exclude the liabilities of many state agencies from their Comprehensive Annual Financial Reports (CAFR). Kido, Petacchi and Weber (2012) emphasize the importance of the electoral cycle. They find that both the compensated absence liabilities account and the unfunded public pension liability account are manipulated by US governors ahead of elections to present healthy fiscal results as elections approach. Naughton, Petacchi and Weber (2015) highlight the role of pre-disclosure fiscal conditions. They find that governors of states facing fiscal distress under-report the public pension funding gap and are less likely to enact tax increases or expenditure cuts. Costello, Petacchi and Weber (2017) find that fiscal distress is associated with opportunistic inter-fund transfers and asset sales when governments are subject to balanced budget requirements. Two earlier studies (Cheng, 1992; Ingram, 1984) report that US states are more likely to adopt transparent reporting practices when they face strong monitoring demands by citizens, creditors, and the press, while also highlighting the role played by accounting and bureaucratic capacity. In the EU, von Hagen and Wolff (2006) document that governments achieve nominal compliance with fiscal rules by making use of financial operations that increase government debt while not being recorded in the income statement – through so-called “stock-flow” or “debt-deficit” adjustments.

Research on the forecasts issued by governments has taken place mostly outside of accounting and has had a distinct European focus, largely inspired by

² There is also a large literature on municipal and local governments' disclosure choices that is not discussed due to the different level of analysis. See Reck, Wilson, Gotlob and Lawrence (2004), and Rodríguez Bolívar, Alcaide Muñoz and López Hernández (2013) for a review. One related - municipal-level - paper that studies disclosure choices in the presence of a reporting mandate is Gore (2004).

the EU fiscal reporting framework. Specifically, this literature has studied the role of economic, fiscal, political and institutional conditions on fiscal forecast errors. For example, Hallerberg, Strauch and von Hagen (2009) find that ex-post forecast errors are more frequent during economic expansions and when states adopt highly decentralized forms of government. Pina and Venes (2011) highlight the role of political dynamics, showing that fiscal forecasts are optimistic in election years. Frankel (2011) and Frankel and Schreger (2013) find that governments are generally optimistic in their economic and fiscal forecasts, particularly at more distant forecast horizons and when they face fiscal distress. They also show that governments issue optimistic forecasts in order to show compliance with EU fiscal rules.

A last set of studies in political science and public administration investigates governments' adoption of transparent reporting practices. Alt, Lassen and Rose (2006) show that transparent budgetary practices are more likely to be adopted when governments have fiscal imbalances and face a high degree of political competition (cf. de Renzio and Wehner, 2017). Alt, Lassen and Wehner (2014) find that the adoption of transparent budgetary practices dampens the effect of the electoral and economic cycle on creative accounting, consistent with transparency dampening the political incentive to misreport (Milesi-Ferretti, 2004). Cicatiello, De Simone and Gaeta (2017) highlight the role of political determinants, showing that a more fragmented legislature and a stronger government both lead to lower transparency (cf. Ríos et al., 2016). Arapis and Reitano (2018) find that transparency increases in the size of fiscal imbalances and further document that transparency increases following economic crises.

Importantly, investors reward greater transparency. Theoretically, a commitment to adopt transparent reporting practices should be beneficial to governments as it sends a costly signal to provide markets with both good and bad news (Beyer et al., 2010). Indeed, Gelos and Wei (2005) document that investment funds allocate greater resources to transparent jurisdictions. Glennerster and Shin (2008) exploit an exogenous shock to the timeliness of governments' release of macroeconomic information and find that borrowing

costs decline by eleven percent following a signal that governments commit to greater transparency. In turn, Bernoth and Wolff (2008) show that investors react to creative accounting with wider spreads of government bonds. Bastida, Guillamón and Benito (2017) also show that interest rates are negatively associated with measures of fiscal transparency (cf. Benito et al., 2016). Copelovitch, Gandrud and Hallerberg (2018) document that borrowing costs decline for countries that release more comprehensive financial information.

3. Theoretical foundations and hypotheses

A typical assumption of extant literature on government transparency is that the government privately observes a signal about fiscal performance that it chooses whether or not to disclose. In this context, transparent reporting practices favor both the ex-ante and the ex-post role of financial information (Beyer et al., 2010): disseminating information favors monitoring while helping the private sector take informed economic decisions. Thus, transparency levels the playing field between politicians, bureaucrats, investors and the public at large (Begg, 2015; Hemming, 2013; Kopits and Craig, 1998). This rationale has led international institutions to be at the forefront of several initiatives to enhance fiscal transparency (IMF, 1998; OECD, 2002; The World Bank, 1992).³ Conversely, Alesina and Perotti (1996: 403) note the peril of opaqueness in fiscal disclosures: “Lack of transparency helps to create confusion and ambiguity on the real state of public finances, by hiding as much as possible of the current and future tax burdens, overemphasizing the benefits of spending, and underestimating the extent of current and future government liabilities.” Under this information asymmetry lens, transparency is strictly beneficial (Hood, 2006; Prat, 2006).

³ Alternatively, transparency can be beneficial because it helps the monitoring of politicians’ and bureaucrats’ actions, leading to better disciplining and screening. This view of transparency does not require that information is used by the public. The mere fact that information is publicly available induces incentive alignment between the “principal” – the public – and the “agent” – the government (Prat, 2006; Williams, 2015). Others discuss the benefits of transparency on more normative grounds (Etzioni, 2010, 2014). The debate on the benefits of transparency is reviewed elsewhere (Cicatiello et al., 2017; Heald, 2006; Hood and Heald, 2006).

Yet, the existence of information asymmetry is more often assumed than it is tested. In particular, there may be times when governments, too, do not possess information. While the assumption of perfect information endowment by governments may be descriptive when one studies the disclosure of information on *past* fiscal performance, it is less likely to be descriptive in the context of guidance about the future fiscal outlook. By definition, guidance amounts to the release of forecasts of yet unrealized outcomes. Thus, to be transparent in guidance the government must know how the future economic and fiscal outlook will evolve. While the assumption may hold in “normal” times, periods of uncertainty makes it less likely to hold. In particular, periods of uncertainty are – by definition – periods in which it is costly to acquire *precise* information that leads to accurate forecasts. In these circumstances, opportunistic withholding of information may not be the *sole* reason for less than full disclosure.

This chapter draws on disclosure theory to explain governments’ decisions to be transparent in their guidance in periods of uncertainty. Theoretically, the problem can be recast in terms of the model by Verrecchia (1990). This model is specifically concerned with managers’ incentives to disclose information as a function of the signal that managers privately observe. In the model, disclosure declines as signal precision decreases because disclosure entails a cost. When signals are of sufficiently low quality, a reduction in disclosure takes place that is in the interest of the investor – say, the shareholder – because disclosing low quality information has no value and yet entails a cost. Hence, there is a negative relation between observed *levels* – i.e. quantity - of disclosure and the amount of signal noise. Conversely, there is a positive relation between disclosure quantity and signal precision. Hence, negative shocks to the precision of expectations should lead to a decline in observed disclosure quantity. This is a first-order, supply-side effect that leads to Hypothesis 1.

Hypothesis 1: Negative shocks to expectation precision cause a reduction in the amount of guidance

On the demand side, as the precision of managers' signals declines, disclosure is less in demand unless the quality of the signals observed by the investor contains even more noise. That is, investors may still demand that managers disclose relatively imprecise information if their own expectations are sufficiently noisy. Verrecchia (1990) also noted that the conclusions of the model may be altered if one allows for (i) endogenous information acquisition, and (ii) the pricing of information risk. Regarding the former, Penno (1997) shows that in the cross-section, the quantity of information disclosed is *negatively* related to its precision if precise information is relatively costlier to acquire. Regarding the latter, Gietzmann and Ostaszewski (2014) show that when managers need capital and investors discount firm value in presence of uncertainty, information is acquired and disclosed in order to reduce the "uncertainty gap" with investors (cf. Ajinkya and Gift, 1984; Graham et al., 2005; Suijs, 2007). Disclosure quality is then negatively related to disclosure quantity because the quantity of disclosure is used to compensate for the relative imprecision of information. Both models suggest that in the cross-section, firms that disclose more information are the ones that require more capital. This leads to Hypothesis 2.

Hypothesis 2. The relation between negative shocks to expectation precision and the amount of guidance is moderated by the need for capital and the precision of investors' expectations

Overall, the model provides a simple representation that indicates testable hypotheses. Observed disclosure levels should respond to the dynamics of demand and supply that are affected by the *relative* quality of privately observed (by managers and investors) signals and managers' need for capital.

When applied to the government setting, the interesting feature of these models is that their results obtain in the presence of self-interested managers who rationally take into account investors' response to the withholding of information. Thus, the models would predict that governments do not always withhold information opportunistically to preserve informational rents. Rather,

their disclosure decisions may be “benign” in the sense that they may be aligned with the interest of investors (Leuz et al., 2008). In turn, the models allow for heterogeneous responses and are compatible with the working of opportunistic incentives. For example, it may be that *some* governments exploit periods of uncertainty to issue forecasts that are intentionally optimistic because in those periods, it is more difficult for investors to separate signal from noise (Rogers and Stocken, 2005).

Overall, the benefit of adopting a disclosure approach to study governments’ guidance decisions is that the approach allows to examine the relative role played by both benign and opportunistic motives. Both opportunistic and benign motives may drive governments’ guidance choices, and it is an empirical question which of the two motives will prevail. In turn, the disclosure approach allows to derive appropriate inferences from observed disclosure levels. The next section describes the research design used to test the hypotheses.

4. Research design

4.1. Models

The baseline model used to estimate the response of governments’ guidance to uncertainty shocks is an Ordinary Least Squares (OLS) of the form:

$$Transparency_{g,d,h} = \beta_1 Uncertainty_{g,d} + \mathbf{Controls}'\psi_{g,d} + \alpha_g + \varphi_h + \gamma_d + \epsilon_{g,d,h} \quad (1)$$

The baseline specification controls for country-specific characteristics α_g that are time-invariant as well as for possible persistence in the level of transparency. The model also includes guidance-horizon fixed effects φ_h that capture the relative difficulty of issuing guidance at longer horizons. Also, the baseline specification includes a full set of quarter-year indicators γ_d . Controlling for quarter-year indicators is necessary because SCPs tend to be published around the same quarter-year, which would expose estimates to the influence of contemporaneous unobservable shocks. $\epsilon_{g,d,h}$ is the usual error term. Except when differently stated, standard errors are always clustered two-way at the

country-date dimension.⁴ The date is the day/month/year of publication of each SCP as shown in Table 1. Clustering on the date (rather than the quarter-year) dimension is advisable because for every SCP the publication date appears as many times as the number of forecast horizons, leading to an artificial increase of observations. In addition, because each country is either treated or untreated on a given date, the country-date pair is the dimension at which treatment takes place (Abadie et al., 2017). Throughout, model (1) is referred to as the “baseline model”. If governments’ guidance intensity depends on uncertainty, the estimate of β_1 should be different from zero.

Transparency measures the amount of forecasts that each government issues in its SCP at a particular date and at a given forecast horizon. The source of this data is the SCP dataset described in [Appendix A](#). Throughout, each forecast is referred to as an “item”. Starting from the items contained in each SCP released between 2001 and 2017, each item is coded as 1 (0) if present (missing) in a given SCP. Whether an item is present or absent varies at several dimensions. Given a set of items $i = 1, \dots, N$, a government may choose to disclose guidance on item i_1 but withhold guidance on items i_2 or i_3 . Hence, guidance varies at the SCP level. In addition, given four forecast horizons $h = 0, \dots, 3$, the government may choose to release guidance on, e.g. item i_1 at horizon $h = 0$ and $h = 1$ but not at horizons $h = 2$ or $h = 3$. In addition, given a set of SCP submission years $v = 2001, \dots, 2017$, a government may release guidance on an item-horizon dyad when $v = 2001$ but withhold it when $v = 2002$. Finally, each year there are $g = 1, \dots, G$ governments in the sample. Hence, an item i appears in each SCP-horizon-year as $i_{g,h,v} \in \{0,1\}$. Next, the average frequency is computed with which item $i_{g,h,v}$ is disclosed at the same horizon and at the same SCP vintage by all $j \neq g$ governments excluding the focal government. This average is referred to as $m_{j,h,v}$. Finally, $m_{j,h,v}$ is subtracted from $i_{g,h,v}$ and the overall government-year-horizon transparency is $\sum_{j=1}^N (i_{g,v,h} - m_{j,h,v})$. Table 2 shows how *Transparency* is constructed.

⁴ Two-way clustering is implemented through the `ivreg2` routine in Stata©, available from the SSC.

Thus, *Transparency* takes the highest (lowest) value when $i_{g,v,h}$ is equal to 1 (0) and $m_{j,h,v}$ is equal to 0 (1). It maximally rewards governments that disclose a forecast that no other government discloses in the same vintage-horizon dyad, while it maximally penalizes governments that do not disclose a forecast that all other governments disclose. Because this algorithm is run for each item-vintage-horizon separately, it effectively benchmarks the item-specific transparency of a government-vintage-horizon triad to that of its peers subjected to the same reporting guidelines. Items that are published by most governments, perhaps because they are “easy” to publish contribute less to *Transparency*. In turn, items that are mandatory to disclose do not affect *Transparency* because they are published by all governments at all forecast horizons. Hence, the index captures the amount of guidance issued by governments voluntarily. The approach is similar to that used by Botosan and Plumlee (2002), who in the first part of their analysis demean their corporate disclosure index by the average disclosure level in the firm’s industry. Higher values of *Transparency* denote greater guidance intensity. The subscript d refers to the guidance date (day/month/year) in which the government publishes the focal SCP. Figure 1 plots the pooled *Transparency* score by country.

The same algorithm is run on sub-indices of transparency to capture specific facets of transparency following prior literature (Botosan and Plumlee, 2002; Bushman et al., 2004). Sub-scores are computed that reflect transparency in (a) the disclosure of forecasts on income statement items – revenues, expenditures and their components – denoted with *Transparency – Fiscal Performance* (b) balance sheet items and contingent liabilities - *Transparency – Fiscal Position*; (c) economic items – such as GDP growth, inflation, and employment rates - *Transparency – Macro*; and (d) forecasts on the assumptions underlying the fiscal plan – such as assumptions on exchange and interest rates - *Transparency – Assumptions*. Table A.4 in [Appendix A](#) contains the list of items that underlie each sub-score of *Transparency*.

The measure of uncertainty is based on the standard deviation of the expected budget balance over a horizon that includes four years beginning with the year in

which the SCP was published, i.e. t , $t+1$, $t+2$, and $t+3$. The higher the standard deviation of these forecasts, the higher the level of uncertainty, indicating that the government faces a volatile fiscal outlook. To estimate the causal effect of uncertainty, this measure is transformed into a binary variable. For each quarter-year in the sample, the median sample standard deviation of the budget balance forecasts is calculated, and a value of 1 is assigned to observations that exceed the median, and 0 otherwise. The quarter-year is a valid reference point because SCPs belonging to the same annual cycle are mostly clustered around the same quarter of each year as shown in Table 1, which contains the distribution of SCP publication dates. Hence, in each quarter-year, governments can be compared in terms of the level of uncertainty surrounding their fiscal plan. The binary measure of fiscal uncertainty is denoted with *Uncertainty*. It captures the volatility in the *expected* fiscal performance of the government.

When investigating potential heterogeneity in the effect of *Uncertainty* on *Transparency*, the following OLS is estimated:

$$Transparency_{g,d,h} = \beta_1 Uncertainty_{g,d} + \beta_2 Mechanism_{g,d} + \beta_3 Uncertainty_{g,d} * Mechanism_{g,d} + \mathbf{Controls}'\psi_{g,d} + \alpha_g + \varphi_h + \gamma_d + \eta_{g,d,h} \quad (2)$$

Where *Mechanism* is a continuous variable that captures the hypothesized mechanism behind the estimated effects. If the effect of uncertainty on transparency depends on the hypothesized mechanism, the estimates of the sum of β_1 and β_3 parameters should be statistically significant. To evaluate the strength of interaction effects and locate them in the distribution of *Mechanism*, average marginal effects (AME) of switching from *Uncertainty* = 0 to *Uncertainty* = 1 are computed over the range of *Mechanism* between its 1st and 99th percentile, while keeping other variables at their observed sample value. Throughout, I will refer to model (2) as the “mechanism model”. As with model (1), it is estimated with errors clustered two-way at the country-guidance date dimensions.

Table 1. The distribution of guidance events

Publication year	N	Publication quarter	N	Publication month	N	Publication day	N
2001	14	1	61	1	36	Monday	60
2002	12	2	225	2	13	Tuesday	71
2003	17	3	2	3	12	Wednesday	100
2004	33	4	146	4	189	Thursday	109
2005	27	<i>Total</i>	<i>434</i>	5	31	Friday	94
2006	27			6	5	<i>Total</i>	<i>434</i>
2007	31			7	0		
2008	18			8	1		
2009	18			9	1		
2010	25			10	6		
2011	27			11	58		
2012	27			12	82		
2013	25			<i>Total</i>	<i>434</i>		
2014	26						
2015	26						
2016	27						
2017	27						
2018	27						
<i>Total</i>	<i>434</i>						

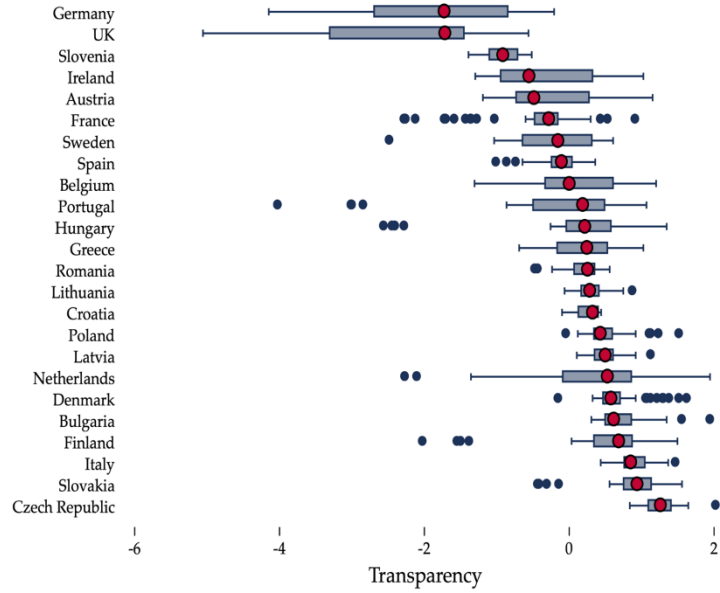
This table shows the distribution of all Stability/Convergence Programmes (SCP) in the sample by year, quarter, month and day of publication.

Table 2. Construction of the *Transparency* measure

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Country	SGP cycle	Forecast horizon	Item 1 (1=present, 0=absent)	Item 2 (1=present, 0=absent)	Mean of Item 1	Mean of Item 2	Item 1 Transparency score (=4-6)	Item 2 Transparency score (=5-7)	Transparency
A	3	0	1	1	1.00	1.00	0.00	0.00	0.00
A	3	1	1	1	0.67	1.00	0.33	0.00	0.33
A	3	2	1	1	0.33	0.67	0.67	0.33	1.00
A	3	3	0	1	0.67	0.33	-0.67	0.67	0.00
B	3	0	1	1	1.00	1.00	0.00	0.00	0.00
B	3	1	1	1	0.67	1.00	0.33	0.00	0.33
B	3	2	0	1	0.67	0.67	-0.67	0.33	-0.33
B	3	3	1	0	0.33	0.67	0.67	-0.67	0.00
C	3	0	1	1	1.00	1.00	0.00	0.00	0.00
C	3	1	0	1	1.00	1.00	-1.00	0.00	-1.00
C	3	2	0	1	0.67	0.67	-0.67	0.33	-0.33
C	3	3	0	1	0.67	0.33	-0.67	0.67	0.00
D	3	0	1	1	1.00	1.00	0.00	0.00	0.00
D	3	1	1	1	0.67	1.00	0.33	0.00	0.33
D	3	2	1	0	0.33	1.00	0.67	-1.00	-0.33
D	3	3	1	0	0.33	0.67	0.67	-0.67	0.00

This table shows how *Transparency* is calculated. For countries denoted as A, B, C, or D that release guidance in the third SGP submission cycle, each item is coded as 1 (0) if the government issues a numerical, point forecast at each forecast horizon $k=0, \dots, 3$. The item-specific mean is the average frequency with which each item is forecasted at each forecast horizon in the same submission cycle by all non-focal governments. The item-specific transparency is the difference between the item presence/absence and the item-specific mean. *Transparency* is the sum of all item-specific transparency scores at each government-forecast horizon-submission cycle triple. See Table 3 for variable definition, sources and method of construction.

Figure 1. Pooled *Transparency* scores for the governments in the sample (N=1191)



This figure plots the median and interquartile range of the pooled *Transparency* score for the governments in our sample. See Table 3 for variable definition, sources and method of construction.

4.2. Identification strategy

From an identification standpoint, the primary challenge is that *Uncertainty* is unlikely to be randomly assigned in the sample. Because *Uncertainty* is calculated based on the reported budget balance forecasts over the period $t - t + 3$, it is likely contaminated by the predictable incentives that governments have to bias these forecasts. If these reasons also correlate with *Transparency*, an OLS-estimated coefficient of *Uncertainty* would be biased. Ideally, one would want to randomly assign fiscal uncertainty shocks to government-quarter-years to then estimate the treatment effect of fiscal uncertainty on guidance intensity. Unfortunately, I am aware of no measure of fiscal uncertainty shock in either the accounting or economics literature that have the desired property of being exogenous (Fernández-Villaverde et al., 2015). Instead, a feature of the EU institutional setting can be used to *design* a scenario in which the OLS-estimates of the effects of uncertainty are unlikely to be contaminated by unobserved heterogeneity.⁵

Recall that *Uncertainty* is calculated based on the reported forecasts of the budget balance. Because the budget balance forecast is mandatory under the SGP mandate, it is always published in fiscal plans at the horizons $t, t + 1, t + 2$, and $t + 3$. This obviates the need to control for the selection mechanism that may lead to the withholding of the budget balance forecast – a problem that vexes corporate earnings guidance research (Billings et al., 2015). At the same time, governments have incentives to report a desirable budget balance forecast to show compliance with EU deficit rules. Accordingly, governments are likely to

⁵ The challenge to identification raised by non-random treatment assignment is that it violates the assumption that the potential outcome of units is independent of their treatment status. The potential, i.e. counterfactual outcome is the outcome that the treated (untreated) units would experience if untreated (treated). The independence assumption states that if the treatment status of the treated and the untreated were reshuffled, their potential outcomes would not be affected. Naturally, the counterfactual outcome is unobservable. However, if the characteristics of treated and untreated units are similar on average, i.e. if there are no differences in the moments of the distribution of characteristics in the treated and untreated subpopulations, one can estimate average treatment effect (Rubin, 1977). Pure randomization of treatment assignment achieves this *necessary* condition most often – albeit not always (Rubin, 1974). In the frequent case of non-random assignment, the propensity score solution is to *design* a sample in which the observations in the treated and untreated subpopulations are weighted differently depending on a previously estimated (inverse of) probability of treatment (Rosenbaum and Rubin, 1983).

report a positive budget balance over the medium-term. Prior literature confirms that this incentive exists and shows that it is stronger for governments in fiscal distress (Frankel and Schreger, 2013; Hallerberg et al., 2009). As a result, when one observes a large standard deviation of the medium-term budget balance, it could be due, not to uncertainty facing the government, but to the reported *plan* of balancing the budget by the end of the forecast horizon that leads to mechanically large values of the proposed measure of uncertainty.⁶

The intuition of this chapter is to use these fiscal constraints to achieve identification. Theoretically, the fiscal rules give rise to a policy rule according to which governments report an improvement in the budget balance over the horizon $t - t + 3$, particularly when they present certain characteristics such as poor fiscal performance. While this rule means that treatment is not randomly assigned, it is assigned conditional on governments' incentives to show adherence to EU fiscal rules. While the literature often expresses skepticism at the strength of these constraints (Eichengreen and Wyplosz, 1998; von Hagen and Wolff, 2006), recent evidence shows that the three percent deficit rule does act as a "magnet". The deficit rule has made governments' fiscal performance 1.5 percentage points better than in the counterfactual scenario and it has led governments to "bunch" around the 3 percent requirement (Caselli and Wingender, 2018). In the results section, evidence is presented that confirms that the deficit rule acts as a magnet with respect to reported forecasts (For evidence on the pre-crisis period, cf. Frankel and Schreger, 2013).

If the EU fiscal rules generate a predictable treatment assignment process, average treatment effects (ATE) can be computed using the inverse-probability weights (IPW) method pioneered by Hirano, Imbens, and Ridder (2003). IPW methods reweight the outcomes of treated and untreated observations to *design* a scenario in which assignment to the treatment condition is "as good as randomly assigned" (Angrist and Pischke, 2009; Rubin, 2007). In particular, IPW

⁶ Governments' willingness to plan for a balanced budget needs not come from the fiscal rules per se. It may also derive from the intertemporal balanced budget constraint that faces governments (Costello et al., 2017; Poterba, 1994). The implications for identification are unaffected.

methods achieve covariate balancing conditional on the propensity score. The propensity score is the function that relates a set of characteristics to the treatment status of each individual. If selection into treatment takes place only on observable characteristics, conditioning the treatment effect on the propensity score suffices to deliver the correct ATE (Rosenbaum and Rubin, 1983). Because the identification requirement is that treatment takes place on observables, propensity score approaches credibly address causality questions in observational studies when both theory and evidence provide robust priors on the selection process that determines treatment assignment. The policy propensity score approach is thus advisable when the researcher has strong priors as to which observables fully describe treatment assignment – usually due to institutional characteristics of the research setting. Recently, IPW methods have been used to analyze the treatment effect of fiscal and monetary policy (Angrist et al., 2018; Angrist and Kuersteiner, 2011; Jordà and Taylor, 2016). In a recent working paper, IPW methods have been used to estimate the effect of the deficit rule on government’s deficits relying on the fact that the deficit rule acts as a “catalyst” for otherwise dispersed fiscal policy choices, confirming their appeal in the EU setting (Caselli and Wingender, 2018).⁷

The EU fiscal rules give strong priors as to the process determining treatment assignment. Naturally, there are other reasons to report a forecast for the budget balance that is exceptionally high or low, most importantly expected economic conditions and politics. Thus, a model is estimated that relates a rich set of observables to *fully* describe the rule that leads a government in a given quarter-year to report forecasts for the budget balance that have a four-year standard deviation that exceeds the quarter-year-specific sample median. The model is estimated through a probit specification as $\Pr(\text{Uncertainty} = 1 | \mathbf{Controls}) =$

⁷ In turn, a benefit of the IPW approach is that is typically more efficient than conditioning the treatment effect on the full set of observable characteristics used in the propensity score stage, for two main reasons. First, it does not require the relation between observables, outcomes and treatment to be linear – an often-violated OLS restriction (Angrist and Hahn, 2004; Rubin, 1977). Second, and more importantly, conditioning on covariates does not deliver the correct ATE if their support differs significantly between treated and untreated, in which case OLS may not suffice to achieve satisfactory balancing (Jordà and Taylor, 2016).

$\Phi(-\mathbf{Controls}\beta)$ where Φ is the c.d.f of the standard normal distribution and **Controls** is the vector of covariates variables in the following model:

$$\mathbf{1}\{Uncertainty\}_{g,d} = \alpha_g + \rho_1Yields_{g,d} + \rho_2YieldVolatility_{g,d} + \rho_3GovernmentStrength_{g,d} + \rho_4PoliticalCompetition + \rho_5ElectoralPressure_{g,d} + \rho_6OutputGap_{g,d} + \rho_7FiscalGap_{g,d} + \rho_8DebtGap_{g,d} \quad (3)$$

I refer to model (3) as the “propensity score” or the “selection” model. To determine the variables to be included in it, prior literature on the determinants of reported fiscal adjustments is followed. The annualized 5-year government bond yield (*Yield*) and its annualized monthly standard deviation *YieldVolatility* is included to account for governments’ fiscal policy response to asset prices and market uncertainty (Agnello and Sousa, 2013; Borio and McCauley, 1996). The percentage of parliament seats that are loyal to the incumbent government *GovernmentStrength*, the degree of fractionalization of parliament *PoliticalCompetition* and the number of years that separate the focal year from the year of the next elections *ElectoralPressure* are included to control for political dynamics (Pina and Venes, 2011). The *OutputGap*, i.e. the difference between the actual rate of growth of GDP and the potential growth rate also enters the model because forecasts are more likely to be optimistic during booms (Hallerberg et al., 2009). Finally, the last two variables capture what are likely to be the most important explanations for government-reported, medium-term fiscal adjustment. Namely, the *FiscalGap* (*DebtGap*) is included, which is calculated as the sum (difference) between the government’s surplus (debt) and the EU fiscal limit of three (sixty) percent of GDP. Higher values of these variables denote a greater distance between the lagged fiscal performance and the medium-term objectives established by the EU fiscal surveillance framework. Fiscal and economic variables are lagged by one year compared to the SCP publication year to ensure that information is available to governments at the time of guidance. The selection model is estimated with standard errors clustered at the country dimension.

Following estimation of the propensity score model, the predicted probability that *Uncertainty* is equal to 1 is estimated. These probabilities are conditional probabilities because they are calculated conditional on the covariates, i.e. they are estimated as $p_{g,d,h} = \Pr(Uncertainty_{g,d,h} = 1 | \mathbf{Controls}_{g,d,h})$ where **Controls** include the full set of continuous and binary covariates, including the fixed effects. These probabilities are used to calculate inverse probability weights (IPW) as: $w_{g,d,h} = [(Uncertainty_{g,d,h}/p_{g,d,h}) + ((1 - Uncertainty_{g,d,h})/(1 - p_{g,d,h}))]$, with $Uncertainty \in \{0,1\}$. Note that the IPW procedure calculates a unique weight for each observation.

5. Data

Data to construct *Uncertainty* and *Transparency* come from the SCP dataset described in [Appendix A](#). The rest of the data used in the analysis consists of pre-existing economic, fiscal, financial and political conditions that are retrieved from public sources. These variables are used both in the identification strategy and to explore the heterogeneity in the treatment effect of *Uncertainty* on *Transparency*. The definition, sources and method of construction of all variables is presented in Table 3. Table 4 presents summary statistics.

Table 3. Variable definition, sources and method of construction

Variable	Construction	Source of primary data
<i>Transparency</i>	Continuous. For each item $i \in (0,1)$ disclosed in the SCP published by a government g on disclosure day d at forecast horizon k , we calculate $Transparency_{g,d,k}$ as the standardized sum of the demeaned number of items. The mean is the mean frequency with which item i is disclosed by all other governments $j \neq g$ in the same quarter-year.	Stability/Convergence Programmes
<i>Uncertainty</i>	Binary. <i>Uncertainty</i> is equal to one if <i>Uncertainty in the Fiscal Outlook</i> exceeds its quarter-year sample median	Stability/Convergence Programmes
<i>Uncertainty in the Fiscal Outlook</i>	Continuous. Equal to the standard deviation of <i>Expected Budget Balance</i> over the forecast horizon $k = 0, \dots, 3$	Stability/Convergence Programmes
<i>Expected Budget Balance</i>	Continuous. The forecast released by government for the general government net lending/net borrowing as percentage of GDP at each forecast horizon $k = 0, \dots, 3$. The general government net lending/net borrowing is measured according to the European System of Accounts (ESA)	Stability/Convergence Programmes
<i>Yields</i>	The rolling 252-day mean of the daily yield on 5-year government bonds, winsorized at the 1 st and 99 th percentile	Thomson Reuters EIKON
<i>Yields Volatility</i>	The rolling 252-day mean of the annualized monthly standard deviation of the daily first difference in the yield on 5-year government bonds. The daily first difference is the difference of the natural logarithm. Daily yields are winsorized at the 1 st and 99 th percentile	Thomson Reuters EIKON
<i>Government Strength</i>	Percentage. The percentage of seats in the legislature occupied by members that belong to political parties that support the governing coalition	Database of Political Institutions
<i>Political Competition</i>	Percentage. The probability that two randomly sampled members of the legislature belong to different political parties	Database of Political Institutions
<i>Electoral Pressure</i>	Integer. The number of years between the year of disclosure of the SCP and the year of the next elections	Database of Political Institutions
<i>Output Gap</i>	Continuous. The difference between the annual growth of real GDP and the annual growth rate of potential GDP.	AMECO
<i>Fiscal Gap</i>	Continuous. The difference between the general government net lending/net borrowing as percentage of GDP and 3 percent. The general government net lending/net borrowing is measured according to the European System of Accounts (ESA)	Eurostat
<i>Debt Gap</i>	Continuous. The difference between the general government debt as percentage of GDP and 3 percent. The general government debt is measured according to the European System of Accounts (ESA)	Eurostat
<i>Conservatism</i>	Continuous. The forecast error, defined as the difference between the general government net lending/net borrowing as percentage of GDP and the <i>Expected Budget Balance</i> . The general government net lending/net borrowing is measured according to the European System of Accounts (ESA)	Stability/Convergence Programmes and Eurostat
<i>Accuracy</i>	Continuous. The absolute value of <i>Conservatism</i>	Stability/Convergence Programmes and Eurostat
<i>Stock – Flow Adjustment</i>	Continuous. The sum of the annual first difference in the general government debt and the general government net lending/net borrowing, both as percentage of GDP. Both the general government debt and the general government net	Eurostat

	lending/net borrowing are measured according to the European System of Accounts (ESA).	
<i>GFN</i>	Continuous. The sum of negative one times the general government net lending/net borrowing as percentage of GDP and the amount of principal on government bonds maturing in one year as percentage of GDP, measured one year earlier	European Central Bank and Eurostat
<i>Redemptions</i>	Continuous. the amount of principal on government bonds maturing in one year as percentage of GDP, measured one year earlier	European Central Bank
<i>Delay</i>	Continuous. The number of days between the publication date of SCPs and the deadline for submission to the European Commission	Stability/Convergence Programmes, government sources, EU sources and Thomson Reuters
<i>Fiscal Volatility</i>	Continuous. The eight-year standard deviation of the general government net lending/net borrowing as percentage of GDP. The general government net lending/net borrowing is measured according to the European System of Accounts (ESA)	Eurostat
<i>Growth</i>	Continuous. The annual growth rate of GDP per capita in constant 2010 US dollars. The annual growth rate of GDP is the first difference in the logarithm.	World Bank
<i>Output Volatility</i>	Continuous. The eight-year standard deviation of <i>Growth</i>	World Bank
<i>EPU</i>	Continuous. The original Economic Policy Uncertainty index by Baker, Bloom, and Davis (2016), augmented with the indices created for Croatia, Greece, Ireland, Netherlands, and Sweden by the following authors: (Armelius et al., 2017; Hardouvelis et al., 2018; Kok et al., 2015; Sorić and Lolić, 2017; Zalla, 2016)	www.policyuncertainty.com

This table presents the description, method of construction and source of the variables used in the analysis.

Table 4. Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Max
<u>Baseline model</u>					
<i>Transparency</i>	1191	-.03	1.06	-5.06	2.03
<i>Uncertainty</i>	1191	.48	.50	0	1
<i>Yields</i>	1191	3.23	2.14	-.47	12.51
<i>Yields Volatility</i>	1191	.17	.15	.04	.53
<i>Government Strength</i>	1191	56.15	8.28	32.08	79.87
<i>Political Competition</i>	1191	71.31	10.49	49.67	89.71
<i>Electoral Pressure</i>	1191	1.9	1.24	0	4
<i>Output Gap</i>	1191	-.48	2.66	-10.72	8.19
<i>Fiscal Gap</i>	1191	-.27	3.19	-9.9	11.7
<i>Debt Gap</i>	1191	2.32	28.26	-47.7	86.2
<u>Mechanism model</u>					
<i>Conservatism</i>	1057	-1.05	2.98	-31	3.9
<i>Accuracy</i>	1057	-1.96	2.48	-31	0
<i>Stock-Flow Adjustment</i>	1191	-1.79	3.7	-29.2	11.5
<i>GFN</i>	969	14.95	8.28	-.92	35.24
<i>Delay</i>	1165	0.25	27.43	-78	182

This table reports summary statistics of the variables used in the analysis. See Table 3 for variable definition, sources and method of construction.

6. Results

6.1. Validation of the “uncertainty” measure

The unobservable construct of interest is the noise contained in governments’ expectations about the fiscal outlook. As discussed, noise is measured through the standard deviation of the reported forecasts for the budget balance over the forecast horizon that includes year t to year $t + 3$. Then, the median standard deviation of the sample at each quarter-year pair is calculated. If the standard deviation exceeds its quarter-year median, the measure of *Uncertainty* takes value 1, and 0 otherwise. Government-quarter-years in which *Uncertainty* is equal to 1 (0) are then “treated” (“untreated”). Construct validity thus requires that high values of the standard deviation of the budget balance forecasts do capture the volatility in governments’ expectations about the fiscal outlook. This subsection shows that a continuous measure of uncertainty, the standard deviation of the budget balance forecasts (*Uncertainty in the Fiscal Outlook*) correlates with other measures of policy uncertainty, thus validating the proposed binary measure of treatment. The analysis is reported in Table 5.

Panel A reports the correlation between measures alongside significance levels. The table reassures us of the construct validity of *Uncertainty in the Fiscal Outlook*. First, compare *Uncertainty in the Fiscal Outlook* with reported data on volatility in fiscal and economic conditions. In periods of volatile fiscal and economic conditions, predicting future fiscal outcomes becomes more challenging. If *Uncertainty in the Fiscal Outlook* truly reflects forecasting difficulties, at a minimum it should correlate with lagged measures of volatility in the operating environment (Rogers and Stocken, 2005). The proxies for the volatility in the operating environment used in prior literature include the volatility of the growth rate of GDP (*Output Volatility*), the volatility in the budget balance (*Fiscal Volatility*), the Output Gap (*Output Gap*) and the growth rate of GDP (*Growth*). To calculate the volatility in lagged fiscal and economic performance, a rolling eight-year horizon is used, in line with prior research on the EU business cycle (Hallerberg et al., 2009). The measure displays expected correlation

coefficients with all its proxies. In particular, the measure is strongly and positively correlated with the volatility in the lagged fiscal balance and with the volatility in GDP growth, suggesting that predicting the future fiscal outlook is more difficult when lagged fiscal performance and economic growth have been volatile. The measure also correlates negatively with the proxies for economic conditions – lagged GDP growth and the output gap – confirming that the fiscal outlook is more uncertain when the economy is in a downturn.

Panel B compares *Uncertainty in the Fiscal Outlook* with the index of Economic Policy Uncertainty (*EPU*) (Baker et al., 2016). The *EPU* is a news-based measure of uncertainty about economic policy that has become increasingly popular in multiple disciplines (Brogaard and Detzel, 2015; Nagar et al., 2019; Pástor and Veronesi, 2013). Unfortunately, the *EPU* has not yet achieved coverage of the entire EU. The original index included data daily on the major EU economies – Germany, France, the UK, Spain and Italy. Later on, scholars adopted the *EPU* methodology to construct indices for Croatia, Greece, Ireland, Netherlands, and Sweden that are hosted on the *EPU* website. Overall, *EPU* data for these EU countries are available at monthly frequency. The dataset is then matched with the enlarged *EPU* series that includes both the original countries and the subsequent additions, resulting in 136 usable observations. The results clearly show that *EPU* is relatively persistent. The correlation between *EPU* and its first monthly lag and lead is 0.82 and 0.80, respectively. Persistence somewhat fades as the horizon lengthens, reaching about 0.65 at the second monthly lead and lag. Importantly, *Uncertainty in the Fiscal Outlook* is positively correlated with both the contemporaneous *EPU* and with its first lead and lag, with coefficients ranging between 0.17 and 0.22, significant at conventional confidence levels. These results confirm that *Uncertainty in the Fiscal Outlook* is a valid measure of uncertainty.

Table 5. Construct validity of the measure of uncertainty

Panel A: Correlation between *Uncertainty in the Fiscal Outlook* and proxies for forecasting difficulty (N=318)

Variables	(1)	(2)	(3)	(4)	(5)
(1) <i>Uncertainty</i>	1.00				
(2) <i>Fiscal Volatility</i>	0.37*	1.00			
(3) <i>Output Volatility</i>	0.15*	0.38*	1.00		
(4) <i>Output Gap</i>	-0.33*	-0.26*	-0.41*	1.00	
(5) <i>Growth</i>	-0.28*	-0.08	-0.08	0.55*	1.00

Panel B: Correlation between *Uncertainty in the Fiscal Outlook* and the *Economic Policy Uncertainty (EPU)* index (N=136)

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) <i>Uncertainty in the Fiscal Outlook</i>	1.00					
(2) <i>EPU</i>	0.22*	1.00				
(3) <i>EPU - First lag</i>	0.17*	0.82*	1.00			
(4) <i>EPU - Second lag</i>	0.08	0.66*	0.78*	1.00		
(5) <i>EPU - First lead</i>	0.20*	0.80*	0.76*	0.73*	1.00	
(6) <i>EPU - Second lead</i>	0.12	0.65*	0.69*	0.76*	0.81*	1.00

This table reports the pairwise correlation between the continuous measure of uncertainty and other proxies for economic and fiscal volatility that should affect the difficulty of forecasting. Panel A and Panel B report the pairwise correlation between the measure of fiscal uncertainty and the Economic Policy Uncertainty Index for five EU countries introduced by Baker, Bloom, and Davis (Baker et al., 2016), augmented with the indices for other selected EU countries hosted on www.policyuncertainty.com. See Table 3 for details of other contributors to the EPU and for variable definition, sources and method of construction. * denote statistical significance at the 10 percent level at least.

6.2. Results of the selection model

Table 6 reports the results of the selection model. The first four columns report the results of models that do not include country fixed effects. The coefficients always report marginal effects; hence they express the change in the probability of treatment for a one-unit change in the explanatory variable. The first two columns show that treatment assignment does not depend on the standalone effect of bond market pressure or political conditions, while the third column indicates that, as expected, governments with greater *Fiscal Gap* have a larger probability of having *Uncertainty* = 1. This probability increases by 4 percent for every 1 percentage point increase in the *Fiscal Gap*. This result holds also when estimating the full model, reported in column 5, in which the effect of *Government Strength*, too, becomes significant and negative. The negative effect indicates that governments supported by strong parliamentary majorities tend to forecast a stable fiscal outlook. Reassuringly, the inclusion of country fixed effects in column 6 does not affect the sign of these variables, but *Government Strength* loses significance in favor of the measure of *Political Competition*. In turn, in the model with country fixed effects the bond market variables become statistically significant. The estimates indicate that a one-unit increase in *Yields* (*Yields Volatility*) is associated with a 7 (74) percent increase in the probability of being treated. Note that the inclusion of country fixed effects reduces the sample by 8 observations because in one country (Malta) there are no observations with *Uncertainty* = 1. The inclusion of country fixed effects leads to a considerable improvement in the model fit as evidenced by the 8 percentage-point improvement in the Area Under the Curve (AUC). This latter specification is our best-fitting propensity score model that is used to compute IPW. The probit model confirms that prior conditions allow to predict treatment assignment. [Appendix B](#) reports extensive diagnostics of the probit model and validates the (testable) assumptions that underpin the use of IPW for computing ATE.

Table 6. Propensity score model (Marginal Effects)

	(1) Uncertainty	(2) Uncertainty	(3) Uncertainty	(4) Uncertainty	(5) Uncertainty
<i>Yields</i>	0.03 [0.03]			0.01 [0.03]	0.07* [0.03]
<i>Yields Volatility</i>	0.27 [0.33]			0.15 [0.29]	0.74* [0.32]
<i>Government Strength</i>		-0.01 [0.00]		-0.01* [0.00]	-0.01 [0.01]
<i>Political Competition</i>		-0.01 [0.00]		-0.00 [0.00]	-0.01* [0.01]
<i>Electoral Pressure</i>		-0.01 [0.01]		-0.02 [0.01]	-0.03 [0.02]
<i>Output Gap</i>			0.01 [0.01]	0.01 [0.01]	0.00 [0.01]
<i>Fiscal Gap</i>			0.04*** [0.01]	0.04** [0.01]	0.04* [0.02]
<i>Debt Gap</i>			0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
N	318	318	318	318	310
Pseudo-R2	0.01	0.03	0.07	0.09	0.23
AUC	0.5487	0.6046	0.7174	0.7261	0.8054
Country FE	No	No	No	No	Yes
Cluster	Country	Country	Country	Country	Country

This table reports estimated marginal effects of model: $\Pr(\text{Uncertainty} = 1 | \mathbf{Controls}) = \Phi(-\mathbf{Controls}\beta)$ with $\mathbf{1}\{\text{Uncertainty}\}_{g,d} = \alpha_g + \rho_1 \text{Yields}_{g,d} + \rho_2 \text{YieldVolatility}_{g,d} + \rho_3 \text{Government Strength}_{g,d} + \rho_4 \text{Political Competition}_{g,d} + \rho_5 \text{Electoral Pressure}_{g,d} + \rho_6 \text{Output Gap}_{g,d} + \rho_7 \text{Fiscal Gap}_{g,d} + \rho_8 \text{Debt Gap}_{g,d}$.

Standard errors in brackets. AUC is the Area Under the Curve. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See Table 3 for variable definition, sources and method of construction.

6.3. Results of the baseline model

This sub-section presents the results of the baseline model of the effect of *Uncertainty* on *Transparency*. Table 7 reports the estimates of the coefficient of interest across multiple specifications of the baseline model. Columns 1 and 2 report the unweighted OLS estimates of the baseline model, while columns 3 and 4 report IPW estimates. Across models, several control variables are statistically significant. A marginally significant and negative association is reported for both government strength and political competition, which indicate that governments that control a larger majority of seats in parliament and that face a more fractionalized opposition are able to withhold comparatively more information (Alt et al., 2006; Citro et al., 2019). Also, governments increase disclosures as elections approach, as indicated by the negative coefficient on *Electoral Pressure*.

In contrast, none of the estimated coefficients on *Uncertainty* are statistically significant. All models indicate a 0.1 percent reduction in *Transparency* as *Uncertainty* increases, but standard errors are too large, ranging between 0.07 and 0.08. In columns 5 and 6, two-way clustering at the country-month date and country-quarter date are experimented with, but inferences are unaffected. Lastly, column 7 presents the model results with standard errors robust only to heteroscedasticity. The negative effect of *Uncertainty* on *Transparency* is strongly significant, with standard errors dropping to about 0.04 and most control variables gaining in statistical significance. There is therefore a stark difference in statistical significance depending on the chosen error structure. On the one hand, two-way clustering at the country-quarter dimensions seem appropriate to account for the structure of the data and for treatment assignment taking place at the country-date dimension. On the other hand, two-way clustering in presence of a limited number of clusters (24 countries and 213 dates) is robust but inefficient, which perhaps leads to the inflated standard errors reported in the first six variations of the models.

Table 7. Treatment effect of uncertainty on transparency

	(1) Transparency	(2) Transparency	(3) Transparency	(4) Transparency	(5) Transparency	(6) Transparency	(7) Transparency
Uncertainty=1	-0.08 [0.08]	-0.10 [0.07]	-0.11 [0.07]	-0.11 [0.07]	-0.11 [0.07]	-0.11 [0.08]	-0.11** [0.04]
<i>Yields</i>		-0.01 [0.03]		-0.01 [0.02]	-0.01 [.]	-0.01 [0.01]	-0.01 [0.02]
<i>Yields Volatility</i>		-0.72 [6.41]		2.04 [7.20]	2.04 [7.27]	2.04 [7.99]	2.04 [3.60]
<i>Government Strength</i>		-0.01* [0.01]		-0.01* [0.01]	-0.01** [0.00]	-0.01* [0.01]	-0.01** [0.00]
<i>Political Competition</i>		-0.02 [0.01]		-0.02+ [0.01]	-0.02+ [0.01]	-0.02+ [0.01]	-0.02** [0.00]
<i>Electoral Pressure</i>		-0.04* [0.02]		-0.04 [0.02]	-0.04** [0.01]	-0.04* [0.01]	-0.04* [0.02]
<i>Output Gap</i>		-0.02 [0.02]		-0.01 [0.03]	-0.01 [0.03]	-0.01 [0.03]	-0.01 [0.01]
<i>Fiscal Gap</i>		-0.01 [0.02]		-0.02 [0.02]	-0.02 [0.01]	-0.02 [0.01]	-0.02+ [0.01]
<i>Debt Gap</i>		-0.00 [0.00]		-0.00 [0.01]	-0.00 [0.00]	-0.00 [0.00]	-0.00 [0.00]
<i>Constant</i>	0.15 [0.21]	2.26+ [1.27]	0.28+ [0.16]	2.47* [1.22]	2.47* [1.25]	2.47* [1.31]	2.47** [0.44]
N	1191	1191	1191	1191	1191	1191	1191
Adj. R2	0.70	0.71	0.81	0.82	0.82	0.82	0.82
Horizon FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Country & Date	Country & Date	Country & Date	Country & Date	Country & Month-Year	Country & Quarter-Year	Robust
Weights	No	No	IPW	IPW	IPW	IPW	IPW

This table reports OLS estimates of the model: $Transparency_{g,d,h} = \beta Uncertainty_{g,d} + \mathbf{Controls}'\psi_{g,d} + \alpha_g + \varphi_h + \gamma_d + \epsilon_{g,d,h}$

Uncertainty is binary. *Controls* is the vector of covariates from the propensity score model. IPW indicates Inverse Probability Weights. Fixed effects are estimated but unreported. Standard errors in brackets. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See Table 3 for variable definition, sources and method of construction.

The baseline model is then run on the sub-scores of *Transparency*. Recall that sub-scores are created based on the intensity with which governments issue guidance on economic items, on assumptions, and on fiscal items – separately for the future fiscal performance and the future fiscal position.

Table 8 reports the estimates of the baseline model for each transparency sub-index, first under two-way clustering (columns 1-4) and then under robust errors (columns 5-8). The estimates shed light on the reasons behind the inconsistent results of the previous analysis. *Uncertainty* does not affect governments' willingness to disclose forecasts on future economic conditions, on economic assumptions, and on the future fiscal performance - revenues, expenditures and their components. However, *Uncertainty* has a strong and statistically significant negative effect on *Transparency – Fiscal Position* that captures the amount of guidance that governments issue about the future evolution of its balance sheet items – debt, assets, contingent liabilities and their components. Shocks to uncertainty make governments reduce *Transparency – Fiscal Position* by about 0.22, which is meaningful given that the interquartile range for the overall *Transparency* index is -0.43/0.64.

The results of the baseline model indicate that uncertainty has a negative effect on the amount of guidance issued by governments on the future evolution of government debt and its components. Thus, the results somewhat support [Hypothesis 1](#). At the same time, an important insight is that the different reaction of the various facets of transparency to *Uncertainty* suggests that there is a specific rationale behind governments' release of information about the future evolution of its financial position. The items that underlie *Transparency – Fiscal Position* include a relatively compact and coherent set of items. These include the liquid financial assets of the government, the amount of bonds and loans that collectively form the government “debt” as reported in SCPs, as well as their expected maturity and amortization schedule.

Table 8. Treatment effect of uncertainty on sub-scores of Transparency

	(1) Transparency- Macro	(2) Transparency- Fiscal Performance	(3) Transparency- Fiscal Position	(4) Transparency- Assumptions	(5) Transparency- Macro	(6) Transparency- Fiscal Performance	(7) Transparency- Fiscal Position	(8) Transparency- Assumptions
Uncertainty=1	-0.09 [0.07]	0.01 [0.08]	-0.22*** [0.06]	-0.08 [0.12]	-0.09+ [0.04]	0.01 [0.04]	-0.22*** [0.03]	-0.08 [0.05]
<i>Yields</i>	0.01 [0.03]	-0.04 [0.05]	0.02 [0.04]	0.01 [0.04]	0.01 [0.03]	-0.04+ [0.02]	0.02 [0.02]	0.01 [0.02]
<i>Yields Volatility</i>	-0.03 [6.50]	-6.70 [6.22]	11.47 [7.46]	7.52 [7.22]	-0.03 [4.14]	-6.70* [3.16]	11.47*** [3.32]	7.52 [5.13]
<i>Government Strength</i>	-0.01+ [0.01]	-0.01+ [0.01]	-0.00 [0.01]	-0.00 [0.01]	-0.01** [0.00]	-0.01*** [0.00]	-0.00+ [0.00]	-0.00 [0.00]
<i>Political Competition</i>	-0.01 [0.01]	-0.03* [0.01]	-0.02 [0.02]	-0.01 [0.01]	-0.01* [0.00]	-0.03*** [0.01]	-0.02*** [0.00]	-0.01 [0.01]
<i>Electoral Pressure</i>	-0.03 [0.03]	-0.01 [0.02]	-0.02 [0.03]	-0.05+ [0.03]	-0.03+ [0.02]	-0.01 [0.02]	-0.02+ [0.01]	-0.05* [0.02]
<i>Output Gap</i>	-0.00 [0.03]	-0.04 [0.04]	0.01 [0.03]	0.01 [0.03]	-0.00 [0.01]	-0.04** [0.01]	0.01 [0.01]	0.01 [0.02]
<i>Fiscal Gap</i>	-0.01 [0.02]	-0.04+ [0.02]	-0.03+ [0.02]	0.03 [0.02]	-0.01 [0.01]	-0.04*** [0.01]	-0.03*** [0.01]	0.03* [0.01]
<i>Debt Gap</i>	0.00 [0.01]	-0.01 [0.01]	-0.00 [0.00]	-0.00 [0.01]	0.00 [0.00]	-0.01*** [0.00]	-0.00 [0.00]	-0.00 [0.00]
<i>Constant</i>	1.73 [1.33]	4.19** [1.30]	0.90 [1.44]	-0.36 [1.31]	1.73** [0.53]	4.19*** [0.49]	0.90* [0.43]	-0.36 [0.70]
N	1191	1191	1191	1191	1191	1191	1191	1191
Adj. R2	0.75	0.77	0.77	0.58	0.75	0.77	0.77	0.58
Horizon FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Country & Date	Country & Date	Country & Date	Country & Date	Robust	Robust	Robust	Robust
Weights	IPW	IPW	IPW	IPW	IPW	IPW	IPW	IPW

This table reports OLS estimates of the model: $Transparency_{g,d,h} = \beta Uncertainty_{g,d} + \mathbf{Controls}'\psi_{g,d} + \alpha_g + \varphi_h + \gamma_d + \epsilon_{g,d,h}$

Uncertainty is binary. *Transparency* is calculated separately for each item category. **Controls** is the vector of covariates from the propensity score model. IPW indicates Inverse Probability Weights. Fixed effects are estimated but unreported. Standard errors in brackets. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ See Table 3 for variable definition, sources and method of construction.

The items that contribute to *Transparency – Fiscal Position* also include the change in government debt between consecutive years and the drivers of the change. These drivers include the evolution of the deficit: other things equal, government debt in one year equals the debt in the past year, plus the deficit of the current year. However, the drivers also include accrual and revaluation adjustments as well as the acquisition and sale of financial assets, which do not pass through the income statement but directly affect debt. The limited literature on the topic indicates that governments have used these items to show compliance with the deficit rule while continuing to accumulate debt (Alt et al., 2014; e.g., Savage, 2005).⁸ Other items included in the *Transparency – Fiscal Position* include the contingent liabilities of governments, which governments currently report only seldom and with variable degrees of detail (Heald and Hodges, 2018). The fact that fiscal uncertainty leads to a reduction in the amount of guidance issued on the future fiscal position of the government raises the question of the motives behind it. The next sub-sections investigate the mechanisms behind the estimated effects.

6.4. *The role of forecast accuracy*

Governments may reduce guidance when subject to uncertainty shocks because guidance is more likely to be inaccurate. To test this possibility, *Accuracy* and *Conservatism* are regressed on *Uncertainty* and the vector of controls. *Accuracy* is negative one times the absolute value of the actual budget balance minus forecasted budget balance. Hence, the higher the value of *Accuracy*, the more accurate the forecast issued by the government. The signed value of the forecast error is denoted with *Conservatism*, which takes on positive values if the government is conservative in its guidance, and negative ones if it is optimistic. Larger levels of *Conservatism* thus indicate more conservative fiscal forecasts. If governments reduce guidance on the future financial position because forecasting is less accurate in periods of uncertainty,

⁸ Not all these “stock-flow” adjustments, as they are called in the literature, represent an example of opportunistic deficit management by governments. Some instances of asset revaluations or acquisition of financial assets are simply the result of active public debt management strategies, while others reflect the financial impact of contemporaneous market conditions (Alt et al., 2014).

there should be a relation between *Accuracy*, *Conservatism* and *Uncertainty*. Table 9 reports the results. Odd columns report results with two-way clustered standard errors, while even columns report results with standard errors robust to heteroscedasticity. Across models, there is no evidence that the accuracy of forecasts depends on the level of uncertainty. The coefficient is negative as expected, but the standard errors are too large. On the other hand, forecasts are somewhat more optimistic in periods of uncertainty when using robust errors - consistent with the intuition that uncertainty gives governments incentives to engage in opportunistic forecasting (Rogers and Stocken, 2005). Thus, some governments may reduce guidance in periods of uncertainty to avoid issuing optimistic forecasts. Others, in turn, may not reduce guidance to take advantage of reduced investors' ability to extract signals from noise.

To distinguish between opportunistic and benign rationales for guidance, the next analysis evaluates the claim that guidance may be withheld by governments in periods of uncertainty because governments prefer to avoid issuing optimistic guidance unless they can derive a financial benefit from it.

6.5. The role of financing needs

As noted in the literature on the cost and benefits of disclosure, a reduction in guidance may be "benign" if disclosure is costly and the expected benefits from disclosure are limited. For this reason, prior research finds that firms reduce voluntary disclosure when they are financially distressed and their growth prospects are anemic (Leuz et al., 2008). As noted above, for governments, too, the association between disclosure and proxies for borrowing costs is normally estimated to be positive (de Renzio and Wehner, 2017). Based on these priors, one expects a positive main effect of expected fiscal performance and transparency on the future debt trajectory. However, in periods of uncertainty governments may prefer to issue guidance only if they need to attract capital from investors. When governments foresee a healthy fiscal outlook, uncertainty may induce a decline in guidance intensity to avoid incurring a disclosure cost in the presence of a limited financial benefit. For governments expecting to be in fiscal

distress, instead, uncertainty may not lead to changes in disclosure levels as they attempt to respond to investors' informational needs (Gore, 2004).

To test the role of financing needs on guidance, the mechanism model is first run by interacting *Uncertainty* with three measures of expected fiscal performance. The first is a continuous measure, *Expected Budget Balance*, which is the budget balance as forecasted by the government. The second and third measures are binary variables that indicate whether governments have an expected budget balance that is above the EU three-percent requirement (*No Expected Excessive Deficit*) and above zero, respectively (*No Expected Deficit*).

Table 10 presents the results of the model. As anticipated, the more positive the expected fiscal performance of the government, the higher the amount of guidance provided by governments on their future debt trajectory. The coefficient, at 0.07 is substantial given a standard deviation of the *Transparency – Fiscal Position* of 0.99. The positive coefficient is also significant for the *No Expected Excessive Deficit* variable. Governments issue more guidance when their fiscal prospects are better. Turning to the role of uncertainty, a negative and significant coefficient is estimated on the interaction between *Expected Budget Balance* and *Uncertainty*. The interpretation of the interaction is that at high levels of *Expected Budget Balance*, shifting *Uncertainty* from 0 to 1 reduces guidance. Hence, governments that expect positive fiscal performance reduce the intensity of guidance when they are subject to an uncertainty shock. This finding suggests that for these governments, the benefits of disclosure do not offset the costs of committing or issuing possibly inaccurate guidance in periods of uncertainty.

Table 9. The accuracy of guidance does not decline in periods of uncertainty

	(1) Accuracy	(2) Accuracy	(3) Conservatism	(4) Conservatism
Uncertainty=1	-0.02	-0.02	-0.28	-0.28*
	[0.18]	[0.13]	[0.24]	[0.16]
<i>Yields</i>	-0.00	-0.00	-0.12	-0.12
	[0.07]	[0.07]	[0.11]	[0.09]
<i>Yields Volatility</i>	-14.27	-14.27	4.99	4.99
	[10.66]	[10.84]	[18.32]	[12.39]
<i>Government Strength</i>	0.02*	0.02*	0.01	0.01
	[0.01]	[0.01]	[0.02]	[0.01]
<i>Political Competition</i>	-0.00	-0.00	0.01	0.01
	[0.02]	[0.02]	[0.03]	[0.02]
<i>Electoral Pressure</i>	0.02	0.02	-0.00	-0.00
	[0.04]	[0.06]	[0.06]	[0.08]
<i>Output Gap</i>	0.07	0.07	0.05	0.05
	[0.06]	[0.05]	[0.08]	[0.06]
<i>Fiscal Gap</i>	-0.04	-0.04	-0.00	-0.00
	[0.04]	[0.05]	[0.04]	[0.05]
<i>Debt Gap</i>	0.03***	0.03***	0.03**	0.03**
	[0.01]	[0.01]	[0.01]	[0.01]
<i>Constant</i>	-1.57	-1.57	-1.82	-1.82
	[1.91]	[1.65]	[3.08]	[2.05]
N	1057	1057	1057	1057
Adj. R2	0.38	0.38	0.45	0.45
Horizon FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Quarter-Year FE	Yes	Yes	Yes	Yes
Cluster	Country & Date	Robust	Country & Date	Robust
Weights	IPW	IPW	IPW	IPW

This table reports OLS estimates of the model: $DV_{g,d,h} = \beta Uncertainty_{g,d} + \mathbf{Controls}'\psi_{g,d} + \alpha_g + \varphi_h + \gamma_d + \epsilon_{g,d,h}$

Uncertainty is binary. *DV* is either *Accuracy* or *Conservatism*. **Controls** is the vector of covariates from the propensity score model. IPW indicates Inverse Probability Weights. Fixed effects are estimated but unreported. Standard errors in brackets. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See Table 3 for variable definition, sources and method of construction.

Table 10. Uncertainty makes guidance decline if the expected fiscal performance is positive

	(1) Transparency-Fiscal Position	(2) Transparency-Fiscal Position	(3) Transparency-Fiscal Position
Expected Budget Balance	0.07 ⁺ [0.03]		
Uncertainty=1	-0.27 ^{***} [0.08]	-0.03 [0.09]	-0.22 ^{***} [0.06]
Uncertainty=1 # Expected Budget Balance	-0.05 ⁺ [0.03]		
No Expected Excessive Deficit=1		0.14 [0.11]	
No Expected Excessive Deficit=1 # Uncertainty=1		-0.22 ⁺ [0.09]	
No Expected Deficit=1			0.21 ⁺ [0.12]
No Expected Deficit=1 # Uncertainty=1			-0.01 [0.13]
Constant	1.11 [1.32]	0.80 [1.46]	0.93 [1.40]
N	1191	1191	1191
Adj. R2	0.77	0.77	0.77
Horizon FE	Yes	Yes	Yes
Country FE	Yes	Yes	Yes
Quarter-Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Cluster	Country & Date	Country & Date	Country & Date
Weights	IPW	IPW	IPW

OLS estimates of the model: $Transparency - Fiscal Position_{g,d,h} = \beta_1 Uncertainty_{g,d} + \beta_2 Mechanism_{g,d} + \beta_3 Uncertainty_{g,d} * Mechanism_{g,d} + \mathbf{Controls}'\psi_{g,d} + \alpha_g + \varphi_h + \gamma_d + \eta_{g,d,h}$
Uncertainty is binary. *Mechanism* is *Expected Budget Balance* in column 1, *No Expected Excessive Deficit* in column 2, and *No Expected Deficit* in column 3. **Controls** is the vector of covariates from the propensity score model. IPW indicates Inverse Probability Weights. Fixed effects are estimated but unreported. Standard errors in brackets. ⁺ $p < 0.10$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$. See Table 3 for variable definition, sources and method of construction.

Column 2 shows that the cutoff point at which governments reduce disclosure in periods of uncertainty is the three percent deficit target: the interaction term between *Uncertainty* and *No Expected Excessive Deficit* is strongly negative and significant. That is, governments that forecast a deficit that exceeds the EU target do not reduce guidance in periods of uncertainty. Instead, governments that forecast a small deficit or a surplus reduce disclosure when the fiscal outlook becomes uncertain. It is interesting to note that the cutoff point for the change in the disclosure policy is the three-percent deficit limit rather than the zero-budget balance. As reported in column 3, the interaction term between *Uncertainty* and *No Expected Deficit* is not significant. Hence, governments' disclosure choices appear to internalize the three percent deficit threshold as the meaningful cutoff point at which the perceived capital market benefit of disclosure exceeds the possible costs of inaccuracy and lack of flexibility. This evidence adds to prior studies on the role of fiscal rules as catalysts of both accounting and real actions by the government (Caselli and Wingender, 2018; Chaney et al., 2002; Costello et al., 2017; von Hagen and Wolff, 2006).

The evidence appears consistent with the intuition that governments faced with uncertainty shocks cannot "afford" to reduce guidance if they expect to have large refinancing requirements. The constant disclosure level of governments that face uncertainty shocks while expecting to run a large deficit thus suggests that governments strategically target their level of disclosure depending on their capital needs. To corroborate this result, a similar analysis is run but using a measure of financing needs that directly captures governments' immediate need for capital. The SGP data are matched with data on the amount of principal of governments' bonds as of year $t-1$ that is maturing in year t . The measure of governments' financing needs is the sum of the government deficit and the amount of maturing principal, denoted as *Government Financing Needs (GFN* in short). Compared to the *Expected Budget Balance* measure examined earlier, *GFN* has the advantage that it does not rely on governments' reported needs but on actual issue-level data. In addition, according to Gabriele et al. (2017) who propose and test the validity of *GFN*, this measure better captures governments' financing needs compared to standard deficit or debt measures because

governments with high (low) debt may have relatively low (high) financing needs depending on the maturity of the liabilities outstanding. For example, in 2014 the Greek government reported a larger amount of debt than the Italian government - at 179 percent of GDP compared to 132 percent. Greece also reported a larger deficit - 3.6 percent of GDP compared to 3 percent reported by Italy. Yet, the gross financing needs of Greece at the one-year maturity were only two-thirds of those of Italy thanks to the favorable maturity structure.⁹ Thus, governments with similar levels of debt (or deficit) do not necessarily face comparable refinancing needs.

Table 11 reports the results of the mechanism model when interacting *Uncertainty* with *GFN* in column 1. Column 2 reports the results using only the amount of principal (of any maturity) due for repayment in the disclosure year (*Redemptions*). Column 3 reports the results using a binary measure (*Large GFN*) that captures high financing needs, measured as values above the quarter-year median.

The regression table does not confirm that there is a difference in the disclosure choices of governments with large and small financing needs. Both the main effect and the interactions are statistically insignificant. It is possible that the insignificance of the estimated coefficients hides the fact that *Uncertainty* has a significant interaction effects only at specific levels of *GFN*. Figure 2 shows that this is the case. The figure plots Average Marginal Effects of switching from *Uncertainty* = 0 to *Uncertainty* = 1 over the range of *GFN*. It clearly indicates that at low levels of *GFN*, *Uncertainty* has a negative and significant effect on *Transparency*. Hence, governments that reduce disclosure in periods of uncertainty are the ones that have less need for investors' capital. On the contrary, governments that have substantial roll-over needs maintain a stable level of transparency regardless of the volatility of the fiscal outlook.

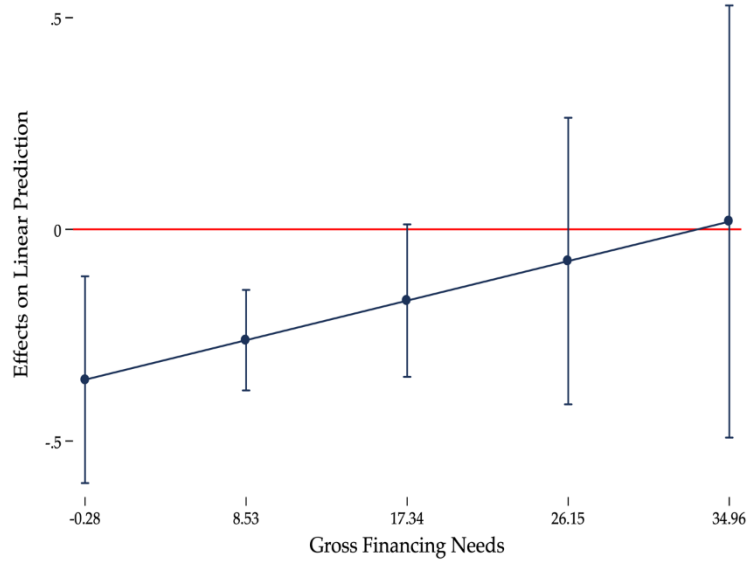
⁹ I am thankful to Carmine Gabriele for helpful discussions on the construction of the *GFN* measure.

Table 11. Uncertainty, guidance and refinancing needs

	(1) Transparency-Fiscal Position	(2) Transparency-Fiscal Position	(3) Transparency-Fiscal Position
GFN	-0.02 [0.02]		
Uncertainty=1	-0.35** [0.13]	-0.27* [0.12]	-0.19* [0.09]
Uncertainty=1 # GFN	0.01 [0.01]		
Redemptions		-0.02 [0.02]	
Uncertainty=1 # Redemptions		0.01 [0.01]	
Large GFN=1			-0.18 [0.12]
Large GFN=1 # Uncertainty=1			-0.03 [0.23]
<i>Constant</i>	1.15 [1.62]	1.00 [1.56]	0.78 [1.55]
N	969	969	969
Adj. R2	0.80	0.80	0.80
Horizon FE	Yes	Yes	Yes
Country FE	Yes	Yes	Yes
Quarter-Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Cluster	Country & Date	Country & Date	Country & Date
Weights	IPW	IPW	IPW

OLS estimates of the model: $Transparency - Fiscal Position_{g,d,h} = \beta_1 Uncertainty_{g,d} + \beta_2 Mechanism_{g,d} + \beta_3 Uncertainty_{g,d} * Mechanism_{g,d} + \mathbf{Controls}'\psi_{g,d} + \alpha_g + \varphi_h + \gamma_d + \eta_{g,d,h}$
Uncertainty is binary. *Mechanism* is *GFN* in column 1, *Redemptions* in column 2, and in column 3 it is *Large GFN*. **Controls** is the vector of covariates from the propensity score model. IPW indicates Inverse Probability Weights. Fixed effects are estimated but unreported. Standard errors in brackets. * $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See Table 3 for variable definition, sources and method of construction.

Figure 2. Uncertainty causes a reduction in guidance only at low levels of debt refinancing needs



This figure shows the average marginal effects (AME) of *Uncertainty* on *Transparency – Fiscal Position*. AME are calculated based on small increments of the moderating variable shown on the horizontal axis. All other variables are kept at their sample values. Bars indicate 90 percent confidence bands. See Table 3 for variable definition, sources and method of construction.

6.6. *The role of investors' expectations*

So far, the evidence is consistent with government *not* reducing guidance when the need to attract capital from investors is pressing. Other governments “go dark” instead. This can be considered a rational response to a positive disclosure cost in presence of limited disclosure benefits. Nevertheless, investors may still benefit from guidance about the future evolution of the debt if their own expectations about the fiscal outlook are noisier than those of the government. To examine this question, governments' response to uncertainty shocks is evaluated at varying levels of investors' own uncertainty about the future evolution of the fiscal outlook. Two continuous measure of investors' expectations are computed based on secondary market data.¹⁰ The first measure is the annualized standard deviation of the daily secondary market yield on government bonds with 5-year maturity (*Yields Volatility*). Following Borio and McCauley (1996), it is calculated as the 252-days moving average of the annualized value of the monthly standard deviation of the change in daily (logarithmic) yields. A binary measure (*High Volatility*) is also created that take value 1 if the *Yields Volatility* is above its quarter-year median, and 0 otherwise. The second continuous and binary measures are computed in the same way, but now focusing on the yield itself (*Yields* and *High Yields*, respectively). If governments' disclosure reduction is benign, the reduction in *Transparency* should be concentrated in periods when both the level and the volatility of government bond yields are relatively low.

Table 12 reports the results. The negative main effect of *Uncertainty* is once more negative and strongly significant. The effect of the standalone interaction term varies is consistently positive, and it is strongly significant for the indicator variable that captures *High Yields*. The result indicates that governments reduce guidance in periods of uncertainty unless the level of bond yields is high. The

¹⁰ Secondary market data are preferable over macro-fiscal data because both their geography and frequency match that of the treatment variable. In addition, they closely capture investors' views on the fiscal risk of the government, as opposed to reported fiscal risk. Compared to government interest payments, in particular, secondary market bond yields approximate the yield that governments pay at the upcoming bond auctions and thus represent immediate fiscal stress as opposed to the inherited borrowing costs (Ballard-Rosa et al., 2019).

same inference can be obtained by observing the AME plotted in Figure 3. The figure once more indicates the substantial difference in the reaction to *Uncertainty* depending on capital market conditions. When the volatility of investors' expectations is low, governments reduce guidance in response to fiscal uncertainty shocks. Once investors' expectations become volatile, however, governments respond with greater guidance.

Overall, the results are consistent with the theory that when investors have uncertain priors over future outcomes, disclosure serves the purpose of alleviating an "uncertainty gap" (Gietzmann and Ostaszewski, 2014). The results can also be interpreted in light of Suijs's (2007) model, in which only the managers of firms with sufficiently uncertain prospects disclose so as to attract capital away from the risk-free asset. These analysis provide support for [Hypothesis 2](#).

6.7. The role of opportunistic incentives

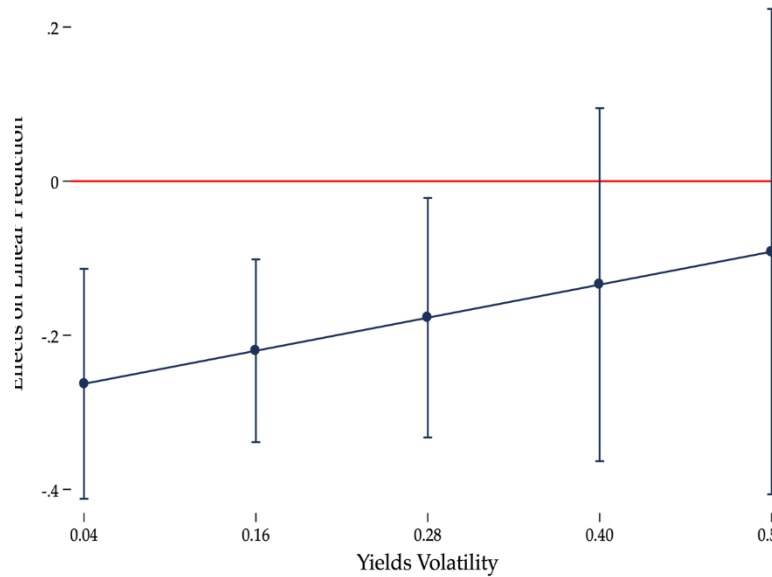
So far, the evidence has established that governments that expect to rely on creditors do not reduce guidance when faced with uncertainty shocks. The evidence also indicates that governments reduce guidance in response to uncertainty only when investors' expectations are relatively precise. An immediate follow-up question is whether the guidance issued by governments in need for capital has any value. In fact, periods of fiscal uncertainty and volatile investors' expectations may give governments the incentive to issue optimistic to take advantage of investors' inability to formulate precise fiscal forecasts (Rogers and Stocken, 2005). This incentive may make governments choose to continue to issue guidance in periods of uncertainty, only to release forecasts that are inaccurate and optimistic and blame the inaccuracy on the uncertainty surrounding the fiscal outlook (Hood, 2007). If this explanation is correct, there should be cross-sectional variation in the kinds of governments that do not reduce guidance when subject to fiscal uncertainty.

Table 12. Uncertainty, guidance and investors' expectations

	(1) Transparency- Fiscal Position	(2) Transparency- Fiscal Position	(3) Transparency- Fiscal Position	(4) Transparency- Fiscal Position
Uncertainty=1	-0.28**	-0.40**	-0.27**	-0.38***
	[0.09]	[0.15]	[0.09]	[0.06]
Uncertainty=1 # Yields Volatility	0.35			
	[0.43]			
Uncertainty=1 # Yields		0.06		
		[0.04]		
High Volatility=1			-0.04	
			[0.08]	
High Volatility=1 # Uncertainty=1			0.12	
			[0.13]	
High Yields=1				-0.22
				[0.16]
High Yields=1 # Uncertainty=1				0.35**
				[0.14]
Constant	0.87	1.19	1.60	1.17
	[1.42]	[1.42]	[1.50]	[1.41]
N	1191	1191	1191	1191
Adj. R2	0.77	0.77	0.76	0.77
Horizon FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Quarter-Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Cluster	Country & Date	Country & Date	Country & Date	Country & Date
Weights	IPW	IPW	IPW	IPW

OLS estimates of the model: $Transparency - Fiscal Position_{g,d,h} = \beta_1 Uncertainty_{g,d} + \beta_2 Mechanism_{g,d} + \beta_3 Uncertainty_{g,d} * Mechanism_{g,d} + \mathbf{Controls}'\psi_{g,d} + \alpha_g + \varphi_h + \gamma_d + \eta_{g,d,h}$
Uncertainty is binary. *Mechanism* is *GFN* in column 1, *Redemptions* in column 2, and in column 3 it is *Large GFN*. **Controls** is the vector of covariates from the propensity score model. IPW indicates Inverse Probability Weights. Fixed effects are estimated but unreported. Standard errors in brackets. * $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See Table 3 for variable definition, sources and method of construction.

Figure 3. Uncertainty causes a reduction in guidance only if government bond yields are low



This figure shows the average marginal effects (AME) of *Uncertainty* on *Transparency – Fiscal Position*. AME are calculated based on small increments of the moderating variable shown on the horizontal axis. All other variables are kept at their sample values. Bars indicate 90 percent confidence bands. See Table 3 for variable definition, sources and method of construction.

To test this explanation, four indicators of probable opportunistic disclosures by governments are examined. These are: (i) the ex-post forecast accuracy of governments that in periods of uncertainty do not reduce transparency; (ii) their level of conservatism; (iii) the level of debt-deficit (also known as “stock-flow”) adjustments that, as discussed above, are often considered a way for governments to manage the reported deficit, and (iv) the delay in the publication of guidance – a proxy for governance quality (Andersen et al., 2014). *Stock – Flow Adjustments* is equal to the annual change in debt plus the reported deficit, which if positive signals possible misreporting. Finally, *Delay* is the number of days between the publication date and the deadline for submitting the SCP to the European Commission. *Accuracy* and *Conservatism* are as defined above.

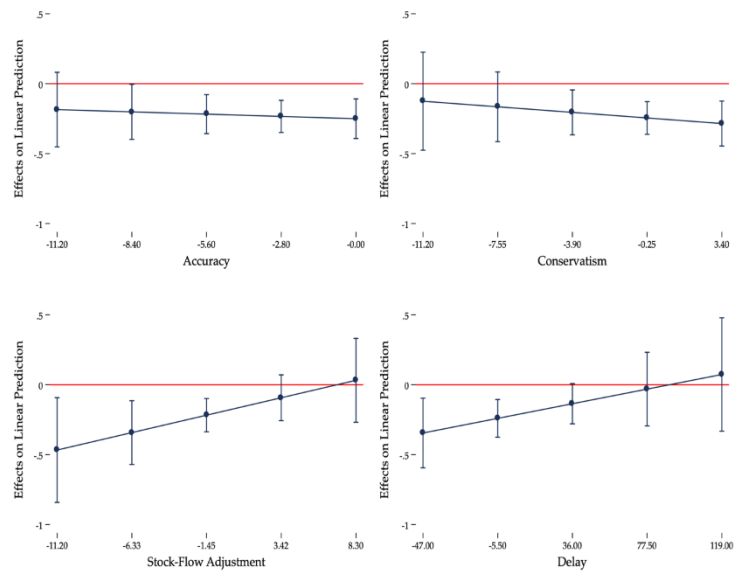
Results are reported in Table 13. *Uncertainty* continues to have a negative effect on transparency but the interaction terms are not significant. Figure 4 plots the AME to locate the effects of *Uncertainty* on *Transparency – Fiscal Position* for levels of opportunistic incentives. The figures are concordant in indicating that governments that do not reduce *Transparency* in periods of *Uncertainty* are characterized by attributes that suggest possible opportunism in reporting. Guidance by these governments is on average less accurate and more optimistic than that of governments who respond to *Uncertainty* with less guidance. The estimates indicate that governments that guide in periods of fiscal uncertainty make budget balance forecasts that exceed actual outcomes by over 4 percent of GDP. The optimistic bias of governments that guide in periods of uncertainty is consistent with uncertainty giving incentives for opportunistic forecasting (Rogers and Stocken, 2005). In addition, these governments always report a negative *Stock – Flow Adjustment* in the year prior to the SCP disclosure, consistent with this indicator capturing deficit management techniques to some extent (von Hagen and Wolff, 2006). Finally, governments that guide in periods of uncertainty publish their SCP with a delay of over a month, which is consistent with these governments attempting to preserve informational rents (Alesina and Perotti, 1996; Alt et al., 2006).

Table 13. Uncertainty, transparency and opportunistic incentives

	(1) Transparency- Fiscal Position	(2) Transparency- Fiscal Position	(3) Transparency- Fiscal Position	(4) Transparency- Fiscal Position
Accuracy	0.00			
	[0.01]			
Uncertainty=1	-0.25***	-0.25***	-0.18**	-0.23***
	[0.07]	[0.06]	[0.06]	[0.07]
Uncertainty=1 # Accuracy	-0.01			
	[0.02]			
Conservatism		0.01		
		[0.01]		
Uncertainty=1 # Conservatism		-0.01		
		[0.02]		
Stock-Flow Adjustment			-0.02	
			[0.01]	
Uncertainty=1 # Stock-Flow Adjustment			0.03	
			[0.02]	
Delay				0.00
				[0.00]
Uncertainty=1 # Delay				0.00
				[0.00]
Constant	1.32	1.33	0.97	1.01
	[1.36]	[1.37]	[1.45]	[1.40]
N	1057	1057	1191	1165
Adj. R2	0.77	0.77	0.77	0.77
Horizon FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Quarter-Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Cluster	Country & Date	Country & Date	Country & Date	Country & Date
Weights	IPW	IPW	IPW	IPW

OLS estimates of the model: $Transparency - Fiscal Position_{g,d,h} = \beta_1 Uncertainty_{g,d} + \beta_2 Mechanism_{g,d} + \beta_3 Uncertainty_{g,d} * Mechanism_{g,d} + \mathbf{Controls}'\psi_{g,d} + \alpha_g + \varphi_h + \gamma_d + \eta_{g,d,h}$
Uncertainty is binary. *Mechanism* is *Accuracy* in column 1, *Conservatism* in column 2, *Stock – Flow Adjustments* column 3, and *Delay* in column 4. **Controls** is the vector of covariates from the propensity score model. IPW indicates Inverse Probability Weights. Fixed effects are estimated but unreported. Standard errors in brackets. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See Table 3 for variable definition, sources and method of construction.

Figure 4. *Uncertainty, Transparency and opportunistic incentives*



This figure shows the average marginal effects (AME) of *Uncertainty* on *Transparency – Fiscal Position*. AME are calculated based on small increments of the moderating variable shown on the horizontal axis. All other variables are kept at their sample values. Bars indicate 90 percent confidence bands. See Table 3 for variable definition, sources and method of construction.

7. Conclusions

This chapter has analyzed if and how an increase in the uncertainty that surrounds governments' expectations of the future fiscal outlook affects the transparency of guidance that governments disclose to the public. The findings can be summarized as follows. Governments' guidance choices are highly responsive to bond market informational needs. Governments reduce guidance on the future evolution of debt and assets in periods of uncertainty only if they do not face immediate refinancing needs. Also, the reduction of guidance in periods of uncertainty takes place only when investors have relatively precise expectations as to how to price in governments' fiscal risk. Conversely, when investors have highly uncertain expectations governments do not reduce guidance in response to uncertainty shocks. At the same time, the value-added of guidance issued by transparent governments in periods of uncertainty is questionable: governments that do not reduce the amount of guidance in periods of uncertainty issue forecasts that turn out to be inaccurate and optimistic ex-post. In addition, these governments publish their guidance with substantial delay, and they present evidence of opportunistic reporting. Overall, the evidence suggests that periods of uncertainty give the government the incentive to use guidance somewhat opportunistically. Conversely, periods of fiscal uncertainty give the opportunity to reduce disclosure for relatively "benign" reasons (Leuz et al., 2008) to governments that are not highly reliant on capital markets.

CHAPTER 2. THE EQUITY INVESTORS' RESPONSE TO GUIDANCE BY THE GOVERNMENT

1. Introduction

Understanding the costs and benefits of disclosure regulations is an important goal of accounting research (Bushman and Landsman, 2010; Leuz and Wysocki, 2016). It is typically the government that imposes disclosure rules on firms, most often in light of their alleged beneficial outcomes - to transparency, cost of capital, liquidity or competition. Yet, what happens when the government faces a regime of mandatory disclosure? This chapter draws on the dataset of SCP guidance described in [Appendix A](#) to analyze the information content and consensus effect of mandatory government guidance on the future evolution of the budget. In this chapter, it is particularly important to note that a stated purpose of guidance under the Stability and Growth Pact (SGP) was to facilitate investors' understanding of governments' medium-term plans for the budget: the SGP regulations state that "member states would also be required to make their programmes public in the interests of transparency and so as to increase public understanding and *market scrutiny*" (EC, 1996: 10, emphasis added; Hallerberg et al., 2009). Accordingly, the SGP introduced the mandatory disclosure of fiscal plans alongside mandatory guidance of selected elements of the fiscal outlook – the budget balance and the level of debt. In addition, as discussed in [Chapter 1](#), the SGP allowed for voluntary disclosure of a large set of accompanying forecasts. Here, I draw on the SGP mandate to evaluate the extent of market scrutiny of guidance by the government by focusing on the equity markets response to the forecast for the budget balance. The main methodological innovation over prior work is that the potential outcome framework is adopted to estimate the *causal* effect of guidance (Angrist, 2010; Rubin, 1974). The results presented in this chapter also represent the first cross-sectional, longitudinal study of government guidance, covering 23 governments over the period 2001-2017.

As in much work on the information content of announcements, the equity market response to the *release* of disclosures – a binary event – is separated from the response to its *content* (e.g., Ricci, 2015). Standard methods are

adopted in the first part of the analysis. The value of disclosure to equity investors is estimated through an event study that contrasts event-day return volatility and trading volume (Chae, 2005; Landsman and Maydew, 2002) with the counterfactual outcome represented by the absence of guidance (Kothari and Warner, 2007). Then, the value of disclosed forecasts is examined by computing the average causal effect of treatment. Treatment is defined as a day in which government announces that the budget balance – the difference between revenues and expenditures – will improve compared to its last reported value, similar to Kato, Skinner and Kunimura (2009).

Although the SGP mandate results in a large number of forecasts disclosed by governments, the focus on this chapter is on the forecast for the budget balance, for two reasons. First, it is salient to government. The budget balance is a metric that attracts great attention in capital markets (Mosley, 2003b). In the bond market in particular, forecasts of the budget balance affect interest rates and are closely monitored by rating agencies (e.g., Moody's, 2016). In addition, the budget balance has a close connection to the annual change in government debt (Alt et al., 2014; von Hagen and Wolff, 2006). This connection makes the evolution of the budget balance an indicator of the probability of future tax increases because government debt cannot increase indefinitely (Gan and Qiu, 2019; von Hagen and Wolff, 2006). Thus, this metric should be intuitively important to equity investors. Second, the issuance of budget balance forecasts is mandatory under the SGP mandate (EU, 1997a). As a result, budget balance forecasts are always available in the dataset, obviating the need to formulate a selection mechanism (Rogers and Van Buskirk, 2013). Naturally, the reported forecast will depend on governments' incentives - in particular the need for highly indebted governments to show compliance with EU fiscal rules. The approach chosen is to exploit the predictability of reported budget balance forecasts to *design* (Rubin, 2007) a sample in which the probability of treatment is independent of the potential outcome. This approach is inspired by recent work in fiscal economics, which estimates the effectiveness of fiscal rules in the European Union (EU) by reweighting observations by the inverse of their treatment probability (Angrist and Kuersteiner, 2011; Caselli and Wingender,

2018). A similar approach has been used recently to estimate the effect of austerity policies (Jordà and Taylor, 2016). In this chapter, the approach is adopted to estimate the information content and consensus effect of fiscal guidance in equity markets.

The theoretical mechanism investigated in this chapter is the role of guidance credibility. As noted by Mercer (2004), the mere issuance of guidance is unlikely to attract the attention of investors unless managers are able to persuade the market that guidance is truthful. Indeed, while guidance should be valuable to investors as a tool to reduce information asymmetry about the *future* value of the firm (Ajinkya and Gift, 1984), it is only ex-post verifiable, raising the possibility that forecasts are optimistically biased ex-ante (Rogers and Stocken, 2005). Interestingly, studies on the characteristics of governments' guidance indicate that fiscal forecasts are on average optimistic, particularly in the long term (cf. Alesina and Perotti, 1996; Frankel, 2011). The average bias in governments' forecasts suggests that equity investors may react to guidance only when it is credible, raising the question of what characteristics of guidance increase its information content. Yet, no study to my knowledge has examined the role of credibility in the context of fiscal guidance. In this chapter, two aspects of credibility will be tested: the credibility of information - a function of the past level of optimism and the extent of disaggregation of guidance (Hirst et al., 2008) - and the credibility of the government – a function of political dynamics (Alt et al., 2006).

The findings can be summarized as follows. First, the analysis indicates that guidance is a valuable event to equity investors. On average, I document a 1.2 percent spike in abnormal return volatility on the guidance date and a 9.9 percent increase in abnormal trading volume. The event is preceded by several days characterized by abnormal trading activity, consistent with information acquisition prior to a scheduled announcement (Kim and Verrecchia, 1991a). In turn, trading volume continues to be abnormally high for three days following the event even in the absence of price changes, consistent with guidance spurring opinion divergence across investors (Kandel and Pearson, 1995).

Second, the analysis shows that announcements of short-term improvements in the budget balance are associated with 0.43 percent higher return volatility in the five days surrounding the date of guidance, but only when the government has been conservative in its last forecast. In turn, abnormal trading volume is 0.53 percentage points higher in the presence of forecasts of a short-term improvement in the budget balance.

Third, the driver of opinion divergence across investors appears to be the credibility of guidance. The results indicate that trading volume is 2.2 percent higher in response to forecasted improvement in the budget balance when the government was optimistic in its last forecasts, and about 0.6 percent higher when the government is politically weak. Collectively, the results suggest that fiscal forecasts *do not* reduce market uncertainty unless the government is credible. Hence, governments can take actions to anchor investors' expectations around their reported forecast. Surprisingly, trading volume is higher in response to forecasted budget balance innovations when the government issues disaggregated guidance, suggesting that disaggregation may not increase the credibility of forecasts.

The primary contribution of this paper is to test the claim that fiscal forecasts reduce information asymmetries in markets (Alt et al., 2006; Brown et al., 2017; Kopits and Craig, 1998; Stiglitz, 1999). It is well-known that the desired reduction in information asymmetries has prompted international institutions to become vocal promoters of transparency in the fiscal and budgetary domain (IMF, 1998; Mehrpouya and Salles-Djelic, 2019; Mosley, 2003a; OECD, 2002; The World Bank, 1992). Yet, as Heald (2006: 71) puts it: "It is vitally important that claims to transparency are tested rather than allowed to go unchallenged". In this context, this chapter expands our understanding of the actual consequences of transparency in government guidance. The evidence complements extant studies that focus on the signaling role of transparency in capital markets (Bastida et al., 2017; Benito et al., 2016; Glennerster and Shin, 2008). It also complements research on the value of fiscal forecasts in bond markets (Laubach, 2009; Thorbecke, 1993; Wachtel and Young, 1987). With the exception of

Bafundi, Columbano and Carrera (2018), who analyze the case study of Spanish SGP disclosures, no study that I am aware of tests the role of fiscal forecasts in equity markets. In addition, prior work does not examine trading volume and it does not attempt to estimate causal effects.

The broader contribution of this paper is to extend the presently limited research on the consequences of central governments' disclosure choices (Kim et al., 2018). This subfield of accounting research is attracting growing interest, as mounting levels of public debt and deficit put the financial reporting "lever" of fiscal policy at the forefront (Naughton and Spamann, 2015). This chapter complements other efforts in the study of governments' use of accounting and disclosure tools, e.g. the opportunistic management of GDP figures (Lyu et al., 2018), the electoral cycle in (mis)reporting the fiscal position (Kido et al., 2012), and the effect of fiscal stress on accounting discretion (Chaney et al., 2002; Costello et al., 2017; Naughton et al., 2015; von Hagen and Wolff, 2006). The documented evidence about the conditions that make fiscal disclosures an effective channel to anchor expectations points to a novel and important responsibility for the financial disclosures of governments.

Finally, this chapter contributes to the literature on disclosure regulation that has mostly focused on the effects of *corporate* disclosure mandates (Badertscher et al., 2013; Leuz and Wysocki, 2016; Shroff et al., 2014). The findings may assist in assessing the success of the SGP reporting mandate. They complement the work of others who have examined the track record of governments in adhering to SGP requirements related to, e.g. the accuracy and bias of forecasts (Frankel and Schreger, 2013; Hallerberg et al., 2009; Leblond, 2006; Merola and Pérez, 2013).

The rest of the chapter is structured as follows. Section 2 reviews related literature. Section 3 discusses the theoretical foundations and hypothesis. Section 4 explains the research design. Section 5 presents the data and section 6 the results. Section 7 concludes.

2. Related literature

The primary related literature investigates the use of macroeconomic-and policy-level information in the valuation of equity. One study particularly close to the analysis presented in this chapter is Bonsall, Bozanic and Fischer (2013) who show that firms' stock returns respond to the guidance issued by firms that are highly exposed to macroeconomic conditions. Also related is Li, Richardson and Tuna (2014) who find that macroeconomic forecasts are relevant to forecasting firm profitability in countries where firms generate part of their revenues. Recently, Nagar, Schoenfeld and Wellman (2019) show that the earnings response coefficients (bid-ask spreads) of firms decline (increase) in periods of policy uncertainty. Finally, Wisniewski and Moro (2014) document abnormal market returns when EU politicians announce their long-term policy priorities. These papers demonstrate that equity investors revise their beliefs about firm-specific profitability based on, amongst other inputs, macroeconomic and policy information (Konchitchki, 2016). This is consistent with firm-level price changes responding to systematic factors impounded in aggregate economic and financial information (Anilowski et al., 2007; Ball et al., 2009). None of these papers, however, investigate government-released guidance about the expected fiscal outlook.

The second related literature investigates the value of earnings guidance released by corporate managers. Since Patell (1976), research has shown that guidance is informative to investors, and more so than earnings announcement (Ball and Shivakumar, 2008; Beyer et al., 2010). The value of guidance rests on its ability to reduce information asymmetry about "the parameters of the distribution of firms' *future* cash flows and earnings" (Billings et al., 2015: 161, emphasis added). Most related to this paper is the study by Kato, Skinner and Kunimura (2009) who find that quasi-mandatory Japanese earnings forecasts are generally optimistic, that optimism in forecasting is persistent, and that forecasts are generally informative. Importantly, the authors document that the information content of a forecast decreases in its past level of optimism, consistent with investors discounting signals that are relatively inaccurate (Verrecchia, 2001). Other studies on the benefits of guidance include Billings, Jennings and Lev

(2015) who show that volatility declines following guidance, particularly for firms with a proven track record in providing guidance (Balakrishnan et al., 2014). Also related to this chapter are studies on the information content of disaggregated guidance. Disaggregation occurs when guidance on earnings is accompanied by forecasts of individual items – e.g., revenues and expenses (Hirst et al., 2008). Extant literature generally finds that information content increases in the extent of guidance disaggregation, suggesting that disaggregation increases the credibility of forecasts (Hirst et al., 2007; Lansford et al., 2013; Merkley et al., 2013). All these studies take place in the corporate setting.

At the central government level, Bafundi et al. (2018) conduct an analysis of stock returns, absolute returns, trading volume and return volatility in the days surrounding SGP guidance by the Spanish government. They find evidence of a price reaction to guidance and an erratic trend in abnormal trading volume, which is abnormally high around the event and sharply declines immediately following guidance. However, their analysis represents a one-country case study, it does not examine the role of divergence of opinions across investors and it estimates associations rather than causal effects. Other studies of the capital market effect of fiscal guidance examine changes in government bond yields following fiscal announcements (Wachtel and Young, 1987). A relatively recent study is Laubach (2009) who finds that an increase in expected deficits is associated with an increase in interest rates, particularly at the longer end of the yield curve (cf. Thorbecke, 1993). The evidence is not conclusive, however. In the Euro area, McMenamin, Breen, and Muñoz-Portillo (2015) find that a tightening of the fiscal stance is associated with negative yields. Turning to the equity market, Darrat (1988, 1990) examines stock market movements in Canada around the release of realized deficit data and finds that larger deficits are associated with negative returns. Ardagna (2009) studies annual stock returns before and after large changes in fiscal performance in 16 OECD countries and also finds a negative association. Agnello and Sousa (2013) adopt a VAR-based approach and document similar evidence (However, see Tavares and Valkanov, 2003). This chapter differs from these articles in several ways. First, earlier studies measure the equity market effect of *realized* fiscal outcomes rather than guidance. Second,

they examine a different outcome - *signed* stock returns - rather than return volatility and trading volume. Third, they examine associations over long event windows – months, quarters or years, while this chapter uses daily data that are less likely to be affected by the possible arrival of confounding news (Kothari and Warner, 2007). Finally, they do not attempt to estimate causal effects.

3. Theoretical foundations and hypotheses

3.1. Fiscal guidance as a disclosure event

The goal of this chapter is to estimate the equity investors' response to guidance issued by the government on its budget. The “response” under study is the combination of abnormal return volatility and trading volume on the days immediately surrounding the guidance event. To understand under which conditions this response should arise and the form that it is likely to take, I recast this research question as a standard disclosure problem in the “association-based” tradition (Verrecchia, 2001). Association-based disclosure models connect information to investors' actions by assuming that investors use information rationally to predict firms' future value but can only do so noisily because the former is only an imperfect indicator of the latter. Borrowing the notation from Verrecchia (2001: 105), consider a typical model for a generic firm i and day d , for convenience.

$$p_{i,d} - p_{i,d-1} = \frac{n_i}{h_i + n_i} * (y_{i,d} - m_i) \quad (1)$$

Where p indicates the stock price and $p_{i,d} - p_{i,d-1}$ is the change in stock prices for firm i between two consecutive days. The larger this difference, the larger the predicted price change. If one is interested in the information content of announcements rather than in the sign of the price reaction, $p_{i,d} - p_{i,d-1}$ is expressed as absolute difference $|p_{i,d} - p_{i,d-1}|$. When taking logarithms of p , this difference is equal to the daily stock return. The unsigned return (or its square) is the measure of information content most used in prior literature (Beaver, 1968; Landsman and Maydew, 2002).

An announcement arrives on day d that contains information in the form of a number $y_{i,d}$ that expresses, for instance, the forecast of firm i future earnings u_i . Investors' expectation about u_i is, e.g. its unconditional mean m_i . If information is a noisy measure of firms' earnings, i.e. if $y_i = u_i + e_i$, then e_i is a stochastic term that measures the noise contained in y_i and e.g., $e_i \sim N(\mu_i, v_i)$. Then, $y_{i,d} - m_i$ is the extent to which the forecast issued on day d deviates from expectations and it is positively associated with announcement-day returns. This first-order effect is moderated by two firm-specific parameters: n_i is the precision of the information disclosed, i.e. the inverse of the signal variance v_i so that $n_i = 1/v_i$; h_i is the inverse of the variance in the firm's performance. Equation (1) thus states that the equity market response to information y disclosed on day d is moderated by the relative precision n of information: more precise information amplifies the content of news.

What is the role of trading volume in this model? There are two possible characterizations of trading volume. In the first one (Holthausen and Verrecchia, 1990; Kim and Verrecchia, 1991b), investors differ in the precision (i.e. the "quality") of their own private information. In the second one (Kandel and Pearson, 1995) investors differ in the way they interpret the correlation between the information disclosed $y_{i,d}$ and firm value. In the first case, trading continues to be a function of absolute price changes $|p_{i,d} - p_{i,d-1}|$. In the second case, trading arises even in the absence of price changes. This is because of diverging interpretations of announcements. For instance, given the relation between information and underlying firm value $y_i = u_i + e_i$ two investors may have different opinions about the mean μ_i of the stochastic term e_i . Following an announcement, the difference in investors beliefs is $\mu_{i,1} - \mu_{i,2}$ and trading occurs even in the absence of price changes (Kandel and Pearson, 1995). Adjusting trading volume for the observed price changes thus allows to recover the percentage of trading volume that is due to opinion divergence (cf. Garfinkel, 2009; Garfinkel and Sokobin, 2006).

It is important to note the role of subscript i . The subscript denotes "information about firm i " rather than "information disclosed by firm i ". Sometimes, as in the

case of earnings guidance, the two coincide. At times, however, information arrives that affects expectations about firms' future earnings that is not disclosed by the firm itself: these are cases of information spillover, as discussed in [section 2](#). Within this class of models, investors' response to governments' guidance on the budget exists if guidance is an indicator of the underlying value of the firm, i.e. if governments' expected budgetary developments belong to the class of signals y_i . Instead, if guidance on the budget does not provide valuable information about u_i , we should not expect returns volatility on the guidance announcement date. At the same time, if guidance is correlated with firm value for at least some investors, then even in the absence of price changes trading volume should increase due to opinion divergence. The arguments in favor and against these hypotheses are reviewed in what follows.

3.2. The effect of fiscal guidance events

For equity investors to be interested in fiscal guidance, fiscal forecasts must alter investors' beliefs about "the parameters of the distribution of firms' future cash flows and earnings" (Billings et al., 2015: 161). This is the case if expected changes in the budget balance (noisily) predict firms' future profitability, i.e. if guidance on the budget $y_{i,d}$ is an indicator of firms' future earnings u_i . Based on prior literature, this is likely. Choices on taxes and spending directly affect earnings through the impact of taxes on corporate income and capital gains, subsidies, and contracts (Belo et al., 2013). Indirectly, they affect earnings by modifying consumers' and employees' preferences for income, saving, and consumption, as well as investment returns (Belo and Yu, 2013; Giavazzi et al., 2000). Hemming (2013) and Ramey (2011) review the impact of fiscal policy on the economy.

Guidance is then likely to have information content insofar as it represents a forward-looking indicator of future firms' earnings. For guidance to be informative, however, it must also be the case that governments know more than investors about their own future budgetary choices (Alt et al., 2006; Brown et al., 2017). In terms of equation (1), this means that $y_{i,d} \neq m_i$ on average. There is evidence that this is the case. Paredes, Pérez, and Perez-Quiros (2015) show that SCP

guidance is incrementally informative of future fiscal outcomes over real-time data, particularly when fiscal policy changes radically. Merola and Pérez (2013) also show that supranational institutions that issue fiscal forecasts cater to the information set of the government when making their own predictions. Hence, evidence suggests that at least a portion of guidance should contain private information of the government. Thus, guidance should result in a decrease of information asymmetry *between* the government and the average equity investor, as stated in [Hypothesis 1a](#):

Hypothesis 1a: Fiscal guidance dates are characterized by abnormal return volatility

Contrary to this hypothesis is the theory that fiscal choices are irrelevant to firms' long-term earnings - the so-called "Ricardian" approach to deficits (Barro, 1974, 1989). The Ricardian approach states that rational agents offset any announcements of a deficit or surplus increase because they anticipate that governments will have to reverse this decision in the future to maintain a balanced budget. As the government announces, e.g., a future deficit increase in the form of higher transfers to households, firms' expected earnings do not increase because the additional household income will be saved rather than consumed. Under the Ricardian hypothesis, $y_{i,d}$ is not an indicator of u_i . The Ricardian approach may thus be seen as the null version of [Hypothesis 1a](#) (Tavares and Valkanov, 2003).

In practice, it is possible and likely that investors hold different views as to the effect of fiscal guidance on firms' earnings. In terms of equation (1), μ_i may be different in the opinion of different investors. Indeed, the very sign of μ_i may be different. Some investors may interpret announcements of budget balance improvements as being positive for investments, e.g. because of their effect on interest rates (Thorbecke, 1993). Others may interpret them as being negative, e.g. because of their effect on aggregate demand (Mosley and Singer, 2008). Therefore, it is not straightforward to predict whether the effect of fiscal guidance is positive or negative on average (Ellahie and Ricco, 2017). If investors differ in

their assessments of the implications of fiscal outcomes on firms' future earnings, trading volume should accompany the release of guidance. [Hypothesis 1b](#) is thus stated as follows:

Hypothesis 1b: Fiscal guidance dates are characterized by abnormal trading volume

3.3. The effect of forecasted improvements in the budget balance

The next research question is, what is the information content of guidance when governments forecast an improvement in the fiscal outlook? There are two competing mechanisms that can drive equity investors' response to forecasted improvements, namely persistence and credibility, which lead to opposite predictions.

On the one hand, improvements in the budget balance, if implemented, can be sustained indefinitely because a budget surplus is sustainable in the long term. In contrast, a deterioration in the budget balance can only be temporary as it leads to explosive debt dynamics (Gale and Orszag, 2003). Under the "persistence" channel, announcements of future budget balance improvements should represent a more persistent indicator of firms' future earnings and, therefore, carry greater information content than announced deteriorations. In terms of equation (1), y is a stronger proxy for u when $y > 0$ than when $y \leq 0$. This argument is analogous to that made in the corporate setting, that the disclosures of loss-making firms have relatively low information content because losses are not persistent (Hayn, 1995).

On the other hand, an announced improvement in the future budget balance should be less informative to equity investors than a forecasted deterioration because it is less likely to be credible. Two elements support this claim. First, EU fiscal rules encourage governments to portray an improvement in the fiscal outlook, meaning that from the point of view of equity investors, governments are likely to announce a budget balance improvement regardless of the signal that

the government privately observes. This means that y will tend to be biased towards positive values, on average. In turn, the extent of bias is likely to be predictable by investors based on pre-disclosure fiscal, economic and political conditions. In particular, Frankel and Schreger (2013) show that the average forecasted budget balance innovation is positive in their pre-crisis sample of EU governments, particularly for governments in fiscal stress.¹¹ This means that y is not exogenous and therefore not randomly assigned in the sample. In turn, the predictability of the bias in y implies that investors are likely to have an expectation m that is conditional on the drivers of the bias. As a result, the difference $y - m$ should be smaller when $y > 0$, as should the corresponding information content. Second, because budget balance improvements amount to either a cut in expenditures or an increase in taxes, they are less electorally popular than announcements of budget balance deterioration (e.g., Poterba, 1994). As noted by Alesina and Perotti (1996: 403), budget balance forecasts can be used strategically so that announced fiscal actions are “systematically postponed to year 2, 3, or 4 of the plan and never implemented”. Therefore, even though announcements of future budget balance improvements have more persistent implications for firms’ earnings, they are less likely to be implemented, making y a noisier proxy of u when $y > 0$. Under the “credibility” explanation, promises of future budget balance improvements are less credible than forecasts of future budget balance deteriorations. Hence, they should have limited information content. If the “credibility” explanation is correct, governments may attempt to restore the credibility of forecasted budget balance improvements by modifying the moderating parameter $\frac{n_i}{h_i+n_i}$ and, in particular, the precision n of the y signal. Precise forecasts amplify the effect of $y - m$, particularly when $y > 0$.

The interaction between the first-and second-order effects imply that an analysis of the information content of fiscal guidance must examine return volatility under conditions of high and low credibility. This is because under persistence, one expects return volatility to be higher when $y > 0$, *particularly* when guidance is credible. Instead, under credibility, one expects return volatility

¹¹ Below, I confirm and extend this result.

to be lower when $y > 0$, *unless* guidance is credible. Accordingly, the strategy adopted in this chapter is to distinguish between the credibility and the persistence explanation by empirically measuring credibility and then estimating the causal effect of forecasted budget balance improvements at both low and high levels of credibility.

In turn, the causal effect of forecasted budget balance improvements on trading volume depends on the source of opinion divergence across investors. Under the persistence explanation, investors' opinions should be less dispersed when the government announces an improvement rather than a deterioration in the budget balance. As underlined by Gale and Orszag, when governments announce a deterioration in the budget balance, "individuals would not know how the deterioration in the long-term budget outlook associated with the [deficit-increasing] provisions will ultimately be resolved. *The key point is that uncertainty is not eliminated, and may well be increased*, by enacting legislation that is clearly unsustainable." (Gale and Orszag, 2003: 470, emphasis added). Because budget balance improvements are persistent, they remove uncertainty as to the actions that governments will have to take to restore a balanced budget in the future. Under persistence, trading volume should therefore be lower when $y > 0$ compared to $y \leq 0$. Instead, under the "credibility" channel investors should trade when governments announce budget balance improvements only when these are not credible. This is because investors differ in their opinion about whether announcements will translate into policies. Therefore, under persistence, one expects trading volume to be higher when $y < 0$, *particularly* when guidance is credible. Instead, under credibility, one expects trading volume to be higher when $y > 0$, *only* when guidance is not credible. The hypotheses of this subsection are thus expressed in null form.

Hypothesis 2a: The relation between forecasted budget balance improvements and abnormal return volatility does not depend on credibility

Hypothesis 2b: The relation between forecasted budget balance improvements and abnormal trading volume does not depend on credibility

3.4. Operationalizing “credibility”

Three aspects of credibility will be examined. First, the role of the government’s reputation as an optimistic forecaster will be considered. While guidance is ex-ante unverifiable, it is ex-post verifiable as soon as outcome data become available. It is widely accepted, therefore, that the past accuracy of a forecaster matters to investors when choosing the extent of their response to guidance (Mercer, 2004). The corporate literature provides robust evidence that guidance issued by firms with a prior history of inaccuracy loses in information content (Kato et al., 2009; Rogers and Stocken, 2005). The literature on governments’ forecasts does not examine the moderating role of forecast accuracy. Nevertheless, prior studies show that governments have incentives to be optimistic in their budget balance forecasts – particularly when subject to fiscal rules (Frankel, 2011; Frankel and Schreger, 2013; Hallerberg et al., 2009). This facet of credibility is denoted with *Conservatism*, which is the extent to which the actual budget balance innovation exceeded the forecast issued by the government in its prior SCP. I concentrate on the most recent budget balance forecast in line with Kato et al. (2009). Then, government-years in which the previous change in the budget balance exceeded or fell short of the forecasted change are denoted with *Conservative* and *Optimistic*, respectively.

Second, the extent to which budget balance forecasts are disaggregated is considered. As noted in the literature review, disaggregation increases the credibility of forecasts by making guidance more precise (Hirst et al., 2007). In particular, disaggregated guidance provides a costly signal that expectations are precise. This phenomenon occurs for two reasons. First, disaggregation allows for the ex-post verification of a larger number of items, thus enhancing investors’ monitoring ability. Second, disaggregation makes it harder for managers to manage earnings. Indeed, corporate studies find that information content is higher for disaggregated than aggregated guidance (Hirst et al., 2007; Lansford

et al., 2013; Merkley et al., 2013). In the government setting, I am aware of no studies that examine the moderating role of disaggregation. However, I build on the corporate literature to treat disaggregation as a measure of credibility that increases the information content of guidance. As described in [section 3 of Appendix A](#), the extent of disaggregation is computed as the number of forecasts issued by governments in their SCPs as a percentage of the total number of forecasts recommended by the EU guidelines. I define *Aggregated (Disaggregated)* guidance that which exceeds (falls short) of the sample median.

Third, I measure the extent to which governments can credibly commit that the forecasted budget balance improvement will translate into actual outcomes. This aspect of credibility is a function of political negotiations between the government and the legislature. A large literature in political economy analyzes the consequences of political bargains on fiscal outcomes (e.g., Alesina et al., 1998). The consensus is that the greater the conflict between the executive and the legislative branches of government, the lower the probability of budget balance improvements (Alesina and Perotti, 1996; Alt and Lowry, 1994; Hallerberg et al., 2009; Poterba, 1994). For example, Alt and Lowry (1994) find that US states in which the governor belong to a party different from that which controls the legislature are more likely to run deficits. Poterba (1994) similarly finds that divided states are less likely to tighten the budget. Hence, the credibility of announcing an improvement in the budget balance should be lower under weak governments and governments that face a competitive political landscape. I define a *Strong Government* as one that commands more than the median share of parties in parliament based on the sample distribution. A *Weak Government* is defined symmetrically. Similarly, a *Cohesive Parliament* is one where the probabilities that two deputies picked at random from the parties in parliament belong to different parties is less than the sample median. A *Divided Parliament* is defined symmetrically. These measures are common in the political science literature as a proxy for the ability of governments to implement policies in a timely and comprehensive fashion (e.g., Alt and Lassen, 2006; Wehner and de Renzio, 2013).

4. Research design

4.1. Event study design

To test [Hypotheses 1a](#) and [Hypotheses 1b](#), the event study method is adopted (Kothari and Warner, 2007). From firm-level data, abnormal return volatility (*ABVAR*) and abnormal trading volume (*ABT*) are computed. To compute *ABVAR*, I follow Landsman and Maydew (2002) and first estimate abnormal returns as the prediction error of a one-factor market model where firm-level daily returns are regressed on the equal-weighted average daily return of all firms in a country. Defining with $r_{i,c,d}$ the return of firm i in country c on day d , the market model is estimated on the trading days $d - 100$ to $d - 11$ as:

$$r_{i,c,d} = \alpha_{i,c} + \beta_{1i,c} * r_c + e_{i,c,d}$$

The abnormal return *AR* is the prediction error of this regression in the event window that includes the days $d - 10$ to $d + 10$:

$$AR_{i,c,d} = r_{i,c,d} - (\hat{\alpha}_{i,c} + \hat{\beta}_{1i,c} * r_c)$$

The abnormal return volatility *ABVAR* for each firm and day is the variance of the abnormal return in the event window, divided by its variance in the estimation period S^2 :

$$ABVAR_{i,c,d} = \frac{AR_{i,c,d}^2}{S_{i,c}^2} \quad (2)$$

ABVAR is standardized so that it has mean 0 rather than 1 in the sample.

To measure abnormal turnover, the method proposed by Chae (2005) is followed. Then, abnormal turnover (*ABT*) is computed as the difference between the proportion of shares traded on the event day and the mean proportion during the estimation window. The estimation-window proportion of shares traded captures liquidity trading and is therefore uncorrelated with trading due to information arrival. Denoting with $V_{i,c,d}$ the number of the shares of firm i in country c traded on day d and with $Sh_{i,c,d}$ the number of firms' shares outstanding on that day, *ABT* is:

$$ABT_{i,c,d} = \ln\left(\frac{V_{i,c,d}}{Sh_{i,c,d}}\right) - \frac{1}{D} \sum_{d=-11}^{-D} [\ln\left(\frac{V_{i,c,d}}{Sh_{i,c,d}}\right)] \quad (3)$$

With D being the number of days in the estimation window. As shown by the subscripts, in the event study the unit of analysis is the firm-day and multiple firms are nested in the same country. Thus, the statistical significance of $ABVAR$ and ABT is evaluated through an intercept-only OLS regression that allows to cluster standard errors at the two-way country-and guidance-date dimension, which are the dimensions within which the sample firms are nested (Abadie et al., 2017; Angrist and Pischke, 2009).

4.2. Design for the causal effect model

To test [Hypothesis 2a](#) and [Hypothesis 2b](#) in a causal sense, the forecasted budget balance improvement (henceforth: $FBBI$) is computed in line with Kato et al. (2009) as the difference between the budget balance forecast that government of country c discloses on day d of year t at the horizon $t + k$ (with $k = 1,2,3$) and the budget balance at year $t - 1$.

$$FBBI_{c,d}^{t+k} \equiv Budget\ Balance\ Forecast_{c,d}^{t+k} - Budget\ Balance_{c,t-1}$$

If above (below) zero, $FBBI$ indicates that the government expects its fiscal performance to improve (deteriorate) between $t - 1$ and t .

In an association study, event-day $ABVAR$ and ABT would be regressed on $FBBI$. However, the resulting estimates would be not causal effects because $FBBI$ is most likely not randomly assigned to the observations in the sample. As discussed, EU fiscal rules ensure that governments with certain characteristics – such as poor fiscal performance - almost certainly *report* that the budget balance will improve over the forecast horizon. If governments with weak fiscal performance have a greater probability of reporting positive values of $FBBI$, then its OLS-estimated coefficient would be biased. The ideal strategy to deal with this problem would be to observe the mean budget balance improvement expected by investors ahead of guidance and assume that the discrepancy between investors' expectations and governments' forecast is randomly assigned. This

strategy would amount to measure the parameter m in equation (1) and use the difference $y - m$ as a continuous measure of treatment using an OLS estimator. Unfortunately, investors' expectations are unobservable. In the corporate literature, it is common to use analysts' earnings forecasts as a proxy for expectations. My scanning of available sources indicates that this strategy is not viable in the SGP setting, largely because of under-coverage of many countries by analysts. As an alternative, I could use the fiscal forecasts released by either the European Commission (EC) or the International Monetary Fund (IMF) to proxy for investors' expectations. However, neither the EC nor the IMF are representative of the market and their own forecasts are biased (Merola and Pérez, 2013). In addition, both institutions itself cater to the information set of the government when issuing fiscal forecasts (ibid.). This is because the government has private information that these international institutions do not have. As a result, it is unclear that the fiscal forecast produced by international institutions represents the average expectation of the market. Hence, it is unclear that the difference between EC forecasts and governments' forecasts would represent the average surprise component $y - m$ in investors' average expectations.

Instead of attempting to replicate markets' expectations, I exploit the predictability of governments' fiscal forecasts to my advantage. I follow the approach proposed by Caselli and Wingender (2018) and utilize the EU fiscal rules to *design* (Rubin, 2007) a treatment assignment mechanism based on the endogenous and observable propensity of governments to report a positive *FBBI*. The estimator adopted to compute Average Treatment Effect (ATE) is the Augmented Inverse Probability Weighted (AIPW) estimator recently used by Jordà and Taylor (2016) to estimate average treatment effects while controlling for selection into treatment. The AIPW estimator is a doubly-robust estimator that delivers the correct ATE in presence of misspecification of either the treatment or the outcome model (Jordà and Taylor, 2016; Robins et al., 1994) and it takes the following form:

$$\widehat{ATE} = \frac{1}{n} \sum_t \left\{ \begin{aligned} & \left[\frac{Treatment * \gamma_t}{\hat{p}_t} - \frac{(1-Treatment) * \gamma_t}{(1-\hat{p}_t)} \right] - \\ & \left(\frac{Treatment - \hat{p}_t}{\hat{p}_t(1-\hat{p}_t)} \right) * \\ & [(1 - \hat{p}_t) \pi_1(\mathbf{X}_t, \hat{\theta}_1) + \hat{p}_t \pi_0(\mathbf{X}_t, \hat{\theta}_0)] \end{aligned} \right\} \quad (4)$$

The AIPW estimator of the ATE has three terms.

The first term computes the mean difference in the outcome γ between treated ($Treatment = 1$) and untreated ($Treatment = 0$) observations based on the estimated inverse probability of treatment \hat{p} . This is a standard IPW estimator based on differences in reweighted means (Hernán and Robins, 2020; Hirano et al., 2003). To calculate \hat{p} , a probit model is implemented that defines assignment to treatment to be a function of a vector \mathbf{X} of observable characteristics, i.e. $\hat{p} = \Pr(Treatment = 1 | \mathbf{X})$. Then, inverse probability weights (IPW) are calculated as $\frac{Treatment}{\hat{p}} + \frac{(1-Treatment)}{(1-\hat{p})}$.

The second term is a multiplicative adjustment term that ensures that IPW sum up to 1 and that stabilizes extreme weights, delivering a gain in efficiency and avoiding the need to truncate extreme weights when predicted probabilities become extremely high (Austin and Stuart, 2015; Lunceford and Davidian, 2004).

The third term is the OLS-predicted outcome π_1 for the treated group conditional on the IPW-weighted estimates $\hat{\theta}_1$ obtained by regressing y over the vector of variables \mathbf{X} in the subpopulation of the treated, less the equivalent predicted outcome π_0 in the subpopulation of the untreated. Regressing the outcome separately in the two subpopulations allows for the possibility that the effect of treatment differs for observations with certain characteristics that also correlate with their observed treatment status (e.g., Bonsall et al., 2013).

The advantage of the AIPW estimator is that if selection takes place on observables and all observations have a positive and similar probability of being treated regardless of their actual treatment status, the reweighted sample achieves covariate balancing and allows to compute the correct ATE.

Like all IPW estimators, AIPW are most effective in the presence of binary treatments (Hernán and Robins, 2020). However, *FBBI* is a continuous variable. Therefore, extensive diagnostics are conducted to examine which level of *FBBI* approximates the conditions that allow me to interpret my AIPW-estimated ATE as the true causal effects. There are two conditions that must be respected, conditional exchangeability and positivity (Angrist et al., 2018; Hernán and Robins, 2020). Conditional exchangeability states that if the characteristics \mathbf{X} that determine whether a unit is treated and that correlate with outcomes are observable, then after conditioning the effect of treatment on \mathbf{X} , the difference in outcomes between treated and untreated is a causal effect. The condition is also known as “selection-on-observables”. Positivity requires that the probability of treatment is positive for all levels \mathbf{x} of all variables contained in \mathbf{X} . The positivity condition implies that observations in the sample cannot have a zero or unit probability of (not) receiving treatment at all levels of all variables contained in \mathbf{X} . Equivalently, instead of conditioning the effect of treatment on \mathbf{X} , the researcher can condition on a function $\pi(\mathbf{X})$, known as a propensity score (Rosenbaum and Rubin, 1983). The propensity score is most often modeled as a logit or probit, after which conditional probabilities of treatment \hat{p} are predicted for each observation. Selection on observables is then met if outcomes are independent of treatment conditional on the propensity score, and if all predicted probabilities exceed zero, i.e. $0 < \hat{p} < 1$. If these conditions hold, causal effects are computed by weighting each observation by the inverse of its probability of being treated (Angrist et al., 2018; Angrist and Kuersteiner, 2011; Rubin, 1977).

I experiment with two *Treatment* definitions defined as *Positive* and *Large* that are based on cutoffs defined by the sample distribution of *FBBI* at each forecast horizon $t + k$ at four forecast horizon ($k = 0,1,2,3$), for a total of $2 \times 4 = 8$ possible treatment conditions. The eight candidate treatment are all binary variables that take value 1 when *FBBI* exceeds a cutoff point, and they are 0 otherwise. The first measures of *Treatment* is $Positive_{c,d}^{t+k}$, which takes value 1 when $FBBI_{c,d}^{t+k}$ is above zero, and 0 otherwise. $Positive_{c,d}^{t+k}$ thus denotes as treated all countries c that on day d forecast a budget balance innovation above

0 at horizon $t + k$. Observations are untreated otherwise. The second dose of treatment is $Large_{c,d}^{t+k}$, which takes value 1 when $FBBI_{c,d}^{t+k}$ exceeds or is equal to the 50th percentile of its sample distribution, and 0 otherwise. As a result, the same observation may be, e.g. in treatment condition $Treatment = 1$ when $Treatment \equiv Positive_{c,d}^{t+1}$, but untreated when $Treatment \equiv Positive_{c,d}^{t+2}$, depending on the relative amount of $FBBI_{c,d}^{t+1}$ compared to $FBBI_{c,d}^{t+2}$.

To predict assignment into the *Positive* and *Large* conditions, the probability is modeled that governments issue a forecast that reveals that $Treatment^{t+k} = 1$ conditional on a comprehensive set of pre-disclosure economic, fiscal and political conditions.¹²

The selection of variables is informed by prior literature. The past level of debt and budget balance are included because of the pressure for governments in fiscal stress to show compliance with fiscal rules (Costello et al., 2017; Frankel and Schreger, 2013). The level of GDP growth and its volatility are included because the future deficit depends on economic dynamics through its impact on tax receipts and subsidies (Hallerberg et al., 2009). The volatility in the budget balance is included because volatility gives incentives to forecast opportunistically (Rogers and Stocken, 2005). Similar to Costello et al. (2017), the annual amount of disbursement from the EU to the country is included to capture the importance of dependence on the EU. An indicator for euro area membership is also included, as research has found that euro adopters perceive deficit limits as relatively more binding (Frankel and Schreger, 2013). Next, I control for political dynamics that prior literature argues should correlate with governments' incentives to announce a certain level of the surplus or deficit. Capital markets are generally assumed averse to deficit spending because this increases inflation and thus interest rates, which reduce the value of equity (Frieden, 1988, 1991, 2016; Mosley and Singer, 2008). Thus, the annual

¹² Note that to compute a consistent ATE it is not necessary that governments actually deliver the promised budget balance improvements - there is, however, evidence that this is the case (Caselli and Wingender, 2018; European Fiscal Board, 2019). Observations are treated or untreated solely based on the *reported* forecasts. Therefore, all that is required is that selection into the reported forecasts is based on the observables included in the prediction model.

percentage of listed shares owned by households as percentage of GDP is included to proxy for citizens' relative preferences for deficit spending. Next, measures of political dynamics that correlate with governments' preferences for, and ability to, implement prudent budgetary choices enter the model. These include a measure of government strength that is based on the percentage of members of parliament that belong to parties loyal to the government and a measure of political competition based on the probability that two members of parliament chosen at random belong to different parties (Wehner and de Renzio, 2013). These measures capture the strength of constraints on governments that, as noted in the hypothesis development section, affect the probability of announcing an improvement in the deficit. To capture citizens' demands for larger deficits stemming from the electoral cycle, a variable is included that measures the number of years that separate year t from the year of the next election (Brender and Drazen, 2008; Kido et al., 2012). Finally, the lagged value of $Treatment_{c,d}^{t+k}$ is included to control for the persistent nature of budgetary adjustment (Jordà and Taylor, 2016). The probit model takes the form:

$$\begin{aligned} \Pr (Treatment_{c,d}^{t+k} = 1) = & \beta_1 Debt\ Gap_{c,t-1} + \beta_2 Budget\ Gap_{c,t-1} + \\ & \beta_3 Unemployment_{c,t-1} + \beta_4 Growth_{c,t-1} + \beta_5 Output\ Gap_{c,t-1} + \\ & \beta_6 Budget\ Volatility_{c,t-1} + \beta_7 Output\ Volatility_{c,t-1} + \beta_8 FDI_{c,t-1} + \\ & \beta_9 Trade_{c,t-1} + \beta_{10} EU\ Funds_{c,t-1} + \beta_{11} Stock\ Market\ Exposure_{c,t-1} + \\ & \beta_{12} Euro_{c,t-1} + \beta_{13} Government\ Strength_{c,t} + \beta_{14} Political\ Competition_{c,t} + \\ & \beta_{15} Electoral\ Pressure_{c,t} + \beta_{16} Treatment_{c,t-1}^{t+k} + \varepsilon_{c,d}^t \quad (5) \end{aligned}$$

In estimating this model, standard errors are clustered at the country dimension.

Finally, when computing the ATE based on the AIPW estimator in (4), I control for the vector X of variables in model (5) plus two additional variables. I control for the revision in $FBBI$ between the focal guidance event and the previous event to account for the possibility that forecast revisions affect market reactions (Kato et al., 2009). I denote this variable with *Revision*. In addition, I control for the contemporaneous innovation in the forecast for economic growth, denoted with

FGRI for “Forecasted Growth Innovation”. The contemporaneous growth forecast may correlate with both *FBBI* and investors’ response because *FBBI* is reported by governments in percentage of GDP and is thus affected by *FGRI* (Frankel and Schreger, 2013). When the dependent variable is *ABT*, I further control for contemporaneous *ABVAR* and for *Bid – ask spread* to control for the effect of information content on trading decisions and for the effect of information asymmetry following disclosure (Chae, 2005; Kim and Verrecchia, 1994). Compustat Global does not provide bid-ask prices. Therefore, the calculation of the daily *Bid – ask spread* follows the procedure outlined in Corwin and Schultz (2012) that exploits high and low daily stock prices to estimate bid-ask quotes.¹³ When computing the ATE at high and low levels of credibility, I implement the ATE estimator separately for bins defined by the sample median of the credibility measure, as in Jordà and Taylor (2016).

5. Data

Data on the forecasts issued by EU governments in their SCP and on the dates of guidance and are hand-collected from public sources, government sources and the press. The data collection procedure is described in [Appendix A](#). The capital market variables are calculated from daily firm-level data sourced from Compustat Global-Security Daily files through WRDS. Details on the filtering procedure are in [Appendix C](#). The variables that measure pre-treatment characteristics and the treatment effect modifiers are retrieved from both public sources and the SCPs. The definition, sources and method of construction of all variables used in the analysis is described in Table 1. Summary statistics are presented in Table 2.

¹³ I am thankful to Shane Corwin for helpful discussions on the construction of this measure.

Table 1. Variables definition

Variable	Description	Construction	Source of data
<i>CABVAR</i> _{<i>d+k</i>}	Continuous. Cumulative abnormal return volatility on event day <i>d ± k</i> trading days	The sum of the residuals of a regression of firm-specific returns on country-level returns. The estimation window of 100 trading days prior to the event day <i>d</i> . <i>k</i> = 0,1,2,3,4. If <i>k</i> = 0, <i>CABVAR</i> is the event day abnormal return volatility.	Compustat Global
<i>CABT</i> _{<i>d+k</i>}	Continuous. Cumulative abnormal log turnover on event day <i>d ± k</i> trading days	The sum of the daily natural logarithm of the ratio of shares traded over shares outstanding, less its average value in the estimation window. The estimation window contains 100 trading days prior to the event day <i>d</i> . <i>k</i> = 0,1,2,3,4. If <i>k</i> = 0, <i>CABT</i> is the event day abnormal log turnover.	Compustat Global
<i>Positive</i> ^{<i>t+k</i>}	Integer. Announcement of a future budget balance improvement	1 if the forecasted budget balance improvement between year <i>t</i> and <i>t + k</i> is positive, 0 otherwise. <i>k</i> = 0,1,2,3.	Stability/Convergence Programmes
<i>Large</i> ^{<i>t+k</i>}	Integer. Announcement of a future budget large balance improvement	1 if the forecasted budget balance improvement between year <i>t</i> and <i>t + k</i> is above the sample median, 0 otherwise. <i>k</i> = 0,1,2,3.	Stability/Convergence Programmes
<i>FGR</i> ^{<i>t+k</i>}	Continuous. Forecasted change in the growth rate	The difference between the forecasted rate of GDP growth at <i>t + k</i> and <i>t - 1</i> . <i>k</i> = 0,1,2,3	Stability/Convergence Programmes
<i>Revision</i>	Continuous. Revision in the forecasted budget balance innovation	The difference between the forecasted budget balance innovation between two consecutive years	Stability/Convergence Programmes
<i>Bid – ask spread</i>	Continuous. Bid-ask spread	Based on the formula in Corwin and Schultz (2012)	Compustat Global
<i>Debt</i>	Continuous. General government consolidated gross debt divided by GDP at current market prices	One-year lag of the value	Eurostat
<i>Budget Balance</i>	Continuous. General government net lending/net borrowing divided by GDP at current market prices	One-year lag of the value	Eurostat
<i>Growth</i>	Continuous. Growth rate of GDP per capita adjusted for purchasing power	100*the one-year lag of the value	World Bank
<i>Fiscal Volatility</i>	Continuous. Volatility in the budget balance	One-year lag of the eight-year rolling standard deviation of <i>Budget Balance</i>	Eurostat
<i>Output Volatility</i>	Continuous. Volatility in the growth rate of the economy	One-year lag of the eight-year rolling standard deviation of <i>Growth</i>	Eurostat
<i>EU Funds</i>	Continuous. Dependence on the European Union (EU)	100*the lag of the value of payments disbursed by the European Union to the country divided by GDP at current market prices	European Commission
<i>Stock market exposure</i>	Continuous. Value of listed shares owned by citizens	The value of listed shares owned by households and non-profit institutions serving households divided by GDP at current market prices	Eurostat

<i>Euro adopter</i>	Integer. The country adopts the euro as their national currency	1 if yes, 0 otherwise	European Central Bank
<i>Government Strength</i>	Continuous. The extent of support of the government by political parties represented in the legislature	Variable <i>MAJ</i> in the original source of data is the percentage of seats held by parties loyal to the cabinet	Database of Political Institutions, 2017 vintage
<i>Political Competition</i>	Continuous. The extent of fragmentation in the political parties represented in the legislature	Variable <i>FRAC</i> in the original source of data is the probabilities that two deputies picked at random from parties represented in the legislature belong to different parties	Database of Political Institutions, 2017 vintage
<i>Electoral Pressure</i>	Count. The number of years left to the next election	1*the variable <i>YRCURNT</i> in the original source of data, so that higher values denote greater pressure	Database of Political Institutions, 2017 vintage
<i>Conservative</i>	Integer. The government was conservative in its last forecast	1 if the difference between actual and forecasted <i>Budget Balance</i> was greater than zero in the last forecast, 0 otherwise	Stability/Convergence Programmes
<i>Optimistic</i>	Integer. The government was optimistic in its last forecast	1- <i>Conservative</i>	Stability/Convergence Programmes
<i>High Transparency</i>	Integer. The government is transparent in the current guidance	1 if the share of forecasts issued in the current Stability/Convergence Programme exceeds the sample median, 0 otherwise	Stability/Convergence Programmes
<i>Low Transparency</i>	Integer. The government is not transparent in the current guidance	1- <i>High Transparency</i>	Stability/Convergence Programmes
<i>Strong Government</i>	Integer. The government commands a strong majority in the legislature	1 if the parties supporting the government hold more than the sample average number of seats in parliament, 0 otherwise	Database of Political Institutions, 2017 vintage
<i>Weak Government</i>	Integer. The government does not command a strong majority in the legislature	1- <i>Strong Government</i>	Database of Political Institutions, 2017 vintage
<i>Cohesive Parliament</i>	Integer. The government faces a cohesive parliament	1 if <i>Political Competition</i> is lower or equal to its sample mean, 0 otherwise	Database of Political Institutions, 2017 vintage
<i>Divided Parliament</i>	Integer. The government faces a fractionalized parliament	1- <i>Cohesive Parliament</i>	Database of Political Institutions, 2017 vintage

This table presents the definition, method of construction and source of the variables used in the analysis.

Table 2. Summary statistics

	N	Mean	St. Dev	Median	Min	Max
<i>CABVAR_d</i>	251	0	.17	-.03	-.24	1.5
<i>CABVAR_{d+1}</i>	248	.02	.3	-.04	-.47	1.96
<i>CABVAR_{d+2}</i>	244	.03	.39	-.07	-.53	2.31
<i>CABVAR_{d+3}</i>	214	.05	.52	-.08	-.71	2.69
<i>CABVAR_{d+4}</i>	201	.08	.67	-.08	-.83	4.38
<i>CABT_d</i>	251	.02	.39	0	-1.57	1.47
<i>CABT_{d+1}</i>	248	.06	.68	-.01	-1.77	2.36
<i>CABT_{d+2}</i>	244	.11	1	.08	-2.52	3.07
<i>CABT_{d+3}</i>	214	.23	1.34	.11	-2.67	4.41
<i>CABT_{d+4}</i>	201	.34	1.7	.11	-3.68	6.31
<i>Positive^t</i>	251	.57	.5	1	0	1
<i>Positive^{t+1}</i>	251	.65	.48	1	0	1
<i>Positive^{t+2}</i>	250	.73	.44	1	0	1
<i>Positive^{t+3}</i>	218	.82	.38	1	0	1
<i>Large^t</i>	251	.2	.4	0	0	1
<i>Large^{t+1}</i>	251	.32	.47	0	0	1
<i>Large^{t+2}</i>	250	.46	.5	0	0	1
<i>Large^{t+3}</i>	218	.53	.5	1	0	1
<i>FGRI^{t+k}</i>	250	0.32	2.71	0.2	-11.1	16.6
<i>Revision</i>	251	.16	3.15	0	-17.6	26.9
<i>Bid – ask spread</i>	240	.01	.04	0	0	.52
<i>Debt</i>	251	65.21	30.19	63.6	3.5	132.7
<i>Budget Balance</i>	251	-2.81	3.78	-2.6	-32.4	4.9
<i>Growth</i>	251	1.6	3.33	1.7	-18	11.9
<i>Fiscal Volatility</i>	251	2.3	1.74	1.81	.56	14.22
<i>Output Volatility</i>	251	2.66	1.91	2.2	.47	10.57
<i>EU Funds</i>	251	.79	1.14	.21	0	5.36
<i>Stock market exposure</i>	251	8.37	6	7	0	34.1
<i>Euro adopter</i>	251	.42	.49	0	0	1
<i>Government Strength</i>	251	56.58	7.94	55.04	36.08	79.87
<i>Political Competition</i>	251	71.57	10.51	73.08	49.67	89.71
<i>Electoral Pressure</i>	251	-.2	1.33	-2	-4	0
<i>Conservative</i>	251	.57	.5	1	0	1
<i>Optimistic</i>	251	.43	.5	0	0	1
<i>High Transparency</i>	251	.48	.5	0	0	1
<i>Low Transparency</i>	251	.52	.5	1	0	1
<i>Strong Government</i>	251	.45	.5	0	0	1
<i>Weak Government</i>	251	.55	.5	1	0	1
<i>Cohesive Parliament</i>	251	.44	.5	0	0	1
<i>Divided Parliament</i>	251	.56	.5	1	0	1

This table reports summary statistics of the variables used in the analysis. Variables are defined in Table 1.

6. Results

6.1. Event study

Table 3 and Figure 1 report the results of the test of [Hypothesis 1a](#) and [Hypothesis 1b](#). In the table, panel A lists the coefficients and standard errors of the intercept-only regression of daily abnormal return volatility and turnover. The analysis indicates that the guidance date is characterized by both the highest return volatility and highest trading volume, at 1.2 and 9 percentage points above expectations, respectively. The remaining days of the event window are mostly characterized by a level of return volatility statistically indistinguishable from zero, with the exception of the days $d - 9$ and $d - 6$, which are marginally significant and negative. Trading volume is also abnormally high on all days preceding the event and it remains positive and marginally significant until $d + 3$, after which it returns to normal levels. The high level of turnover in the days preceding guidance is indicative of differential pre-disclosure beliefs driven by heterogeneous private information (Kim and Verrecchia 1991a, 1991b). In turn, abnormal trading on the days immediately following the event – days $d + 1$ through $d + 3$ - takes place even absent price changes. This latter finding suggests that for at least two days following fiscal guidance, investors may be trading on different interpretations of the implications of fiscal guidance for firms' future earnings (Garfinkel and Sokobin, 2006; Kandel and Pearson, 1995). This possibility is explored in the tests of [Hypothesis 2a](#) and [Hypothesis 2b](#) in the next section.

Two robustness tests are conducted. First, the definition of “event date” is modified to include different sets of days surrounding the event day d . This in line with prior research (Białkowski et al., 2008; Ricci, 2015). It also takes into account the estimated measurement error in the event days as explained in [section 2 of Appendix A](#). In Panel B of Table 3, each row denotes the length of the event window. The results confirm that $ABVAR$ is statistically significant only on the event day, while ABT is statistically significant (and positive) on all days that surround the event window chosen. Hence, guidance appears to have strong information content on the day of its disclosure to the public. The effect is sharp,

and it is not preceded by days with abnormal returns that may suggest privileged access to leaked budget guidance material. In turn, trading activity is abnormally high before, during and immediately after the event.

Next, a placebo test inspired by Chae (2005) is run to establish that the capital market measures are not biased in the sample. Placebo values of abnormal return volatility and turnover are estimated on a sample of 100, randomly selected days that exclude the true event dates. Because these dates are selected at random, the placebo estimates of cumulative return volatility and turnover should be statistically indistinguishable from zero. Results of the placebo test are presented in Figure 2. Return volatility and turnover are cumulated between the first day and the last day of the event window, i.e. $d - 10$ to $d + 10$. The top charts report cumulative values on the actual guidance days, while the bottom panels refer to the randomly selected dates. The results clearly indicate that the estimates in Table 3 are not picking up measurement bias or random noise in either *ABVAR* or *ABT*.

Collectively, the results suggest that fiscal guidance reduces information asymmetry *between* governments and investors. However, the findings of abnormal trading volume indicate that following guidance, traders' expectations do not immediately converge. The finding that fiscal guidance does not eliminate uncertainty is consistent with recent evidence on the effect of Covid-19 fiscal announcements on government bond yields (Ettmeier et al., 2020). It also qualifies the earlier findings by Bafundi et al. (2018). The authors found abnormal trading volume on both the day of release of SCPs by the Spanish government and the preceding trading day. However, they documented a sharp drop in trading volume on day $d + 1$, followed by a return of trading volume towards positive levels in the remaining days. While the findings presented here are consistent with theirs in the pre-announcement period, their evidence of an abnormal post-announcement trading volume appear idiosyncratic in light of my cross-country results. In particular, the results raise the possibility that post-announcement trading activity is explained by the content of the information

disclosed by governments and its credibility. The next set of analysis explores this possibility.

Table 3. The release of government guidance is associated with abnormal return volatility and trading activity

Panel A: Daily value

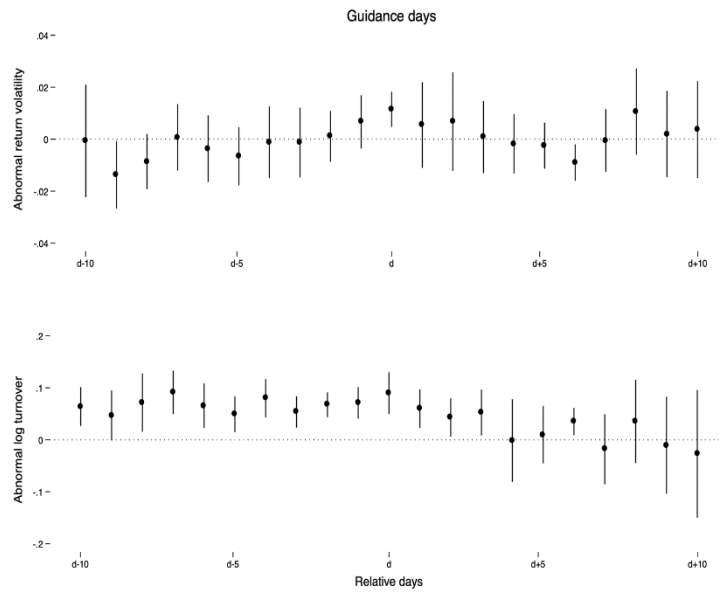
Days to event	ABVAR			ABT		
	Beta	s.e.	N	Beta	s.e.	N
-10	-0.001	[0.013]	44834	0.064**	[0.023]	44568
-9	-0.014 ⁺	[0.008]	43646	0.047	[0.029]	43389
-8	-0.009	[0.006]	43004	0.071 [*]	[0.034]	42755
-7	0.001	[0.008]	44716	0.091***	[0.025]	44470
-6	-0.004	[0.008]	46102	0.066 [*]	[0.026]	45844
-5	-0.007	[0.007]	45344	0.049 [*]	[0.021]	45144
-4	-0.001	[0.008]	45250	0.080***	[0.022]	45013
-3	-0.001	[0.008]	45235	0.054**	[0.018]	44992
-2	0.001	[0.006]	45631	0.067***	[0.014]	45405
-1	0.007	[0.006]	46565	0.071***	[0.018]	46314
0	0.012**	[0.004]	46888	0.090***	[0.024]	46651
1	0.005	[0.010]	43165	0.060**	[0.023]	42870
2	0.007	[0.012]	41257	0.043 ⁺	[0.022]	40985
3	0.001	[0.008]	43322	0.052 ⁺	[0.027]	43062
4	-0.002	[0.007]	45594	-0.001	[0.048]	45334
5	-0.002	[0.005]	42789	0.010	[0.033]	42497
6	-0.009 [*]	[0.004]	42321	0.035 [*]	[0.016]	42078
7	-0.001	[0.007]	41693	-0.018	[0.041]	41434
8	0.011	[0.010]	42682	0.035	[0.049]	42425
9	0.002	[0.010]	43999	-0.011	[0.057]	43742
10	0.004	[0.011]	43295	-0.027	[0.075]	42980

Panel B: Cumulative values

Event window	CABVAR			CABT		
	Beta	s.e.	N	Beta	s.e.	N
<i>d</i>	0.012**	[0.004]	46888	0.090***	[0.024]	46651
<i>d</i> - 1/ <i>d</i> + 1	0.024	[0.018]	42328	0.270***	[0.054]	41650
<i>d</i> - 2/ <i>d</i> + 2	0.039	[0.034]	37917	0.516***	[0.080]	36961
<i>d</i> - 3/ <i>d</i> + 3	0.035	[0.046]	35658	0.747***	[0.099]	34475
<i>d</i> - 3/ <i>d</i> + 1	0.030	[0.026]	40789	0.454***	[0.076]	39787
<i>d</i> - 2/ <i>d</i> + 1	0.028	[0.021]	41144	0.370***	[0.065]	40301
<i>d</i> - 1/ <i>d</i> + 2	0.035	[0.030]	39100	0.396***	[0.063]	38291
<i>d</i> - 1/ <i>d</i> + 3	0.037	[0.039]	37196	0.511***	[0.074]	36272

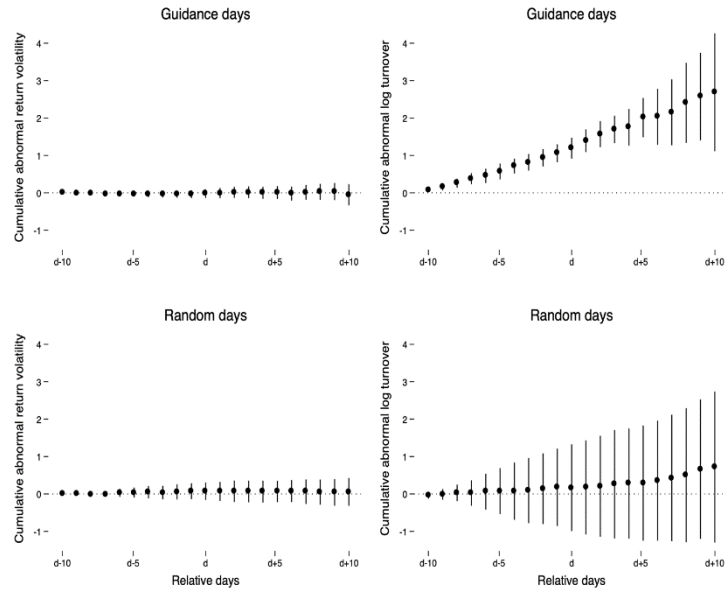
This table reports daily abnormal return volatility (*ABVAR*) and daily abnormal trading volume (*ABT*) in the twenty-one days surrounding the event day *d* in which the government issues guidance (Panel A) and cumulative abnormal return volatility (*CABVAR*) and abnormal trading volume (*CABT*) over the days indicated in the event window (Panel B). Standard errors clustered two-way at the country and guidance date level are in brackets. ⁺ $p < 0.10$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$. Variables are defined in Table 1.

Figure 1. Guidance dates are associated with abnormal return volatility and trading activity



This figure plots daily abnormal return volatility (*ABVAR*) and abnormal daily turnover (in logarithms) (*ABT*) in the twenty-one trading days surrounding the release of a Stability/Convergence Programme (SCP), denoted with *d*. Point estimates are the coefficients on an intercept-only OLS regression of each variable with standard errors clustered two-way at the country-guidance date level. Bars represent 90 percent confidence intervals. Variables are defined in Table 1.

Figure 2. Cumulative abnormal return volatility and trading activity around actual and placebo dates



This figure plots cumulative abnormal return volatility and turnover (in logarithms) in the twenty-one trading days surrounding the release of a Stability/Convergence Programme (SCP), denoted with d . Point estimates are the coefficients on an intercept-only OLS regression of each cumulated variable with standard errors clustered two-way at the country-guidance date level. Bars represent 90 percent confidence intervals. In the top panels, the actual dates of release of SCPs are used. In the bottom panels, a random sample of 100 dates is used as placebo. Variables are defined in Table 1.

6.2. The predictability of expected improvements in the fiscal outlook

The next set of tests shows that budget balance forecasts on average portray an improvement in the fiscal outlook and are optimistic. These tests serve to provide support for the assumption that budget balance improvements are predictably positive and optimistic. These tests are based on earlier work on SCP forecasts that is, however, limited to the pre-crisis period (Frankel and Schreger, 2013).

Table 4 compares forecasted and actual budget balance improvements. The results are estimated through an intercept-only OLS regression with standard errors clustered at the country level. The number of observations depends on the forecast horizon, and in some cases the outturn data necessary to estimate fiscal outcomes are not available. The table clearly reveals the predictable pattern in government fiscal forecasts. Consider the forecasted change in the budget balance $FBBI$. The table shows that the average government expects $FBBI$ to improve at all horizons. As the forecast horizons lengthens, so does the magnitude of the forecasted improvement. For example, governments on average forecast an improvement of 0.36 percent of GDP at horizon t , compared to an improvement of 1.9 percent of GDP at horizon $t + 3$. All forecasted improvements are statistically different from zero, confirming that governments on average report that the budget balance will improve over the forecast horizon. Now consider the actual change in the budget balance BBI . The results show that BBI is always positive, indicating that governments not only announce, but also take real actions that make the budget balance improve over time. However, the actual improvement is not monotonically increasing in the forecast horizon and it becomes statistically indistinguishable from zero at horizon $t + 3$. Thus, governments forecast that the budget balance will improve more than it actually does. To confirm the statistical significance of this result, consider *Conservative* – the difference between BBI and $FBBI$. The table shows that the forecasted and actual budget balance are statistically indistinguishable only at the short-term horizons t and $t + 1$. Instead, forecasts on average fall short of actual results at horizons $t + 2$ and $t + 3$. As forecasts on average fall short of outcomes, SGP

forecasts are on average optimistically biased, and increasingly so as the forecast horizon lengthens.

6.3. Results of the selection model

The next set of tests reports the results of predicting budget balance improvements. A probit model is used to predict budget balance improvements at the four forecast horizons t through $t + 3$. The tests serve to confirm that budget balance improvements are predictably issued by governments with certain characteristics. Table 5 presents the results of the model when the improvement is measured as a positive change in the forecasted budget balance, i.e. *Positive* = 1. Table 6 presents the results when improvement is measured as a large positive change in the forecasted budget balance, i.e. *Large* = 1.

Across models, evidence on the drivers of forecasted budget balance improvements is robust and consistent. The treatment predictors behave largely as expected. The most important predictor is the lagged budget balance, which enters with negative sign: as governments past performance improves, the probability of a *Positive* change in the forecasted budget balance declines significantly at all forecast horizons, consistent with governments in healthy financial conditions having a lesser need to show compliance with EU deficit rules. It is also consistent with governments using reporting discretion so as to present a healthy fiscal picture in times of fiscal stress (Brender and Drazen, 2008; Kido et al., 2012; Naughton et al., 2015). A second consistent predictor of governments' probability of announcing budget balance improvements is the percentage of equity shares held by households. The estimates confirm that as households become more exposed to equity markets, governments become more averse to deficit spending as it increases interest rates and reduces equity values.

Table 4. Government guidance portrays an optimistic improvement in the fiscal outlook

Model	Dependent variable	Coefficient	s.e.	N
(1)	<i>FBBI t</i>	0.36 [*]	[0.16]	251
(2)	<i>FBBI t+1</i>	0.82 ^{**}	[0.22]	251
(3)	<i>FBBI t+2</i>	1.44 ^{***}	[0.26]	250
(4)	<i>FBBI t+3</i>	1.89 ^{***}	[0.30]	219
(5)	<i>BBI t</i>	0.42 ^{***}	[0.10]	251
(6)	<i>BBI t+1</i>	0.72 ^{***}	[0.14]	236
(7)	<i>BBI t+2</i>	0.64 ^{**}	[0.17]	216
(8)	<i>BBI t+3</i>	0.30	[0.25]	166
(9)	<i>Conservative t</i>	0.06	[0.12]	251
(10)	<i>Conservative t+1</i>	-0.11	[0.18]	236
(11)	<i>Conservative t+2</i>	-0.88 ^{**}	[0.25]	215
(12)	<i>Conservative t+3</i>	-1.66 ^{***}	[0.34]	165

This table reports the estimated coefficient of an intercept-only regression of forecasted budget balance innovations (*FBBI*), actual budget balance innovations (*BBI*), and *Conservative* 0, 1, 2, and 3 years ahead of year *t*. Standard errors clustered two-way at the country and guidance date level are in brackets. ^{*} $p < 0.10$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$. Variables are defined in Table 1.

The statistical significance of other predictors varies with the forecast horizon. Higher levels of government debt and economic growth make it more likely that governments forecast a positive change in the budget balance at horizons t through $t + 3$. Political competition also increases the probability of forecasted budget balance improvement, but only at forecast horizons t and $t + 1$. Dependence on EU assistance marginally predict *Positive*, but only at the shortest horizon. Finally, there is some sparse evidence that growth volatility and the lagged treatment condition increase the probability of treatment. The evidence also indicates that predictability increases in the forecast horizon, so that the least predictable forecasts are the short-term ones (Frankel and Schreger, 2013). The results presented in Table 6 about the predictability of *Large* forecasted budget balance improvements portray a largely similar picture. In particular, the lagged budget balance continues to be a predictor of treatment across all forecast horizons. All models have a fairly high predictive accuracy, with the Area Under the Curve (AUC) ranging between 85 and 97 percent. Hence, announcements of budget balance improvements are largely predictable based on a limited set of observables. This finding is consistent with the intuition that treatment assignment in the context of EU deficit rules follows a predictable pattern.

Table 5. Expected improvements in the fiscal outlook are predictable

	(1) <i>Positive^t</i>	(2) <i>Positive^{t+1}</i>	(3) <i>Positive^{t+2}</i>	(4) <i>Positive^{t+3}</i>
<i>Debt</i>	0.01* [0.00]	0.01+ [0.01]	0.02** [0.01]	0.02 [0.01]
<i>Budget Balance</i>	-0.35*** [0.05]	-0.53*** [0.08]	-0.53*** [0.07]	-0.60*** [0.09]
<i>Growth</i>	0.13* [0.05]	0.15** [0.05]	0.14** [0.05]	-0.05 [0.08]
<i>Fiscal Volatility</i>	-0.00 [0.08]	-0.14 [0.11]	-0.15 [0.14]	-0.22 [0.17]
<i>Output Volatility</i>	0.02 [0.09]	0.18* [0.07]	0.31* [0.12]	0.17 [0.11]
<i>EU Funds</i>	0.24+ [0.13]	0.11 [0.13]	0.07 [0.18]	0.31 [0.20]
<i>Stock market exposure</i>	0.04* [0.02]	0.05* [0.02]	0.06* [0.02]	0.07+ [0.04]
<i>Euro adopter</i>	-0.15 [0.25]	-0.26 [0.24]	0.10 [0.35]	-0.17 [0.50]
<i>Government Strength</i>	0.01 [0.01]	-0.00 [0.02]	-0.00 [0.01]	-0.03+ [0.02]
<i>Political Competition</i>	0.03*** [0.01]	0.03** [0.01]	0.02 [0.02]	0.03 [0.03]
<i>Electoral Pressure</i>	0.05 [0.09]	-0.03 [0.08]	-0.12 [0.10]	-0.18* [0.09]
<i>Lagged Positive^t</i>	0.16 [0.26]			
<i>Lagged Positive^{t+1}</i>		0.18 [0.21]		
<i>Lagged Positive^{t+2}</i>			0.46+ [0.28]	
<i>Lagged Positive^{t+3}</i>				0.54 [0.40]
<i>Constant</i>	-4.65*** [1.40]	-4.32* [1.77]	-4.44* [1.96]	-2.42 [2.91]
N	219	222	217	173
Pseudo-R2	0.31	0.45	0.53	0.64
AUC	0.85	0.91	0.94	0.96
Cluster	Country	Country	Country	Country

This table reports the results of the probit model for the dependent variable indicated below the model number. AUC is "Area Under the Curve". Standard errors clustered at the country level are in brackets. * $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Variables are defined in Table 1.

Table 6. Expected, large improvements in the fiscal outlook are predictable

	(1) <i>Large^t</i>	(2) <i>Large^{t+1}</i>	(3) <i>Large^{t+2}</i>	(4) <i>Large^{t+3}</i>
<i>Debt</i>	-0.00 [0.01]	0.02** [0.01]	0.02** [0.01]	0.02** [0.01]
<i>Budget Balance</i>	-0.39*** [0.09]	-1.07*** [0.16]	-0.68*** [0.09]	-0.62*** [0.11]
<i>Growth</i>	0.14+ [0.07]	0.22** [0.06]	0.11+ [0.06]	0.19+ [0.09]
<i>Fiscal Volatility</i>	0.08 [0.11]	-0.34*** [0.09]	-0.35*** [0.09]	-0.16 [0.10]
<i>Output Volatility</i>	0.01 [0.08]	0.63*** [0.15]	0.39*** [0.10]	0.22+ [0.13]
<i>EU Funds</i>	0.20+ [0.11]	0.12 [0.15]	-0.06 [0.17]	-0.13 [0.21]
<i>Stock market exposure</i>	0.07** [0.03]	0.11** [0.04]	0.03 [0.03]	0.03 [0.02]
<i>Euro adopter</i>	-0.24 [0.44]	0.12 [0.35]	0.22 [0.31]	0.03 [0.30]
<i>Government Strength</i>	0.00 [0.02]	-0.04 [0.02]	-0.02 [0.02]	-0.04** [0.02]
<i>Political Competition</i>	-0.01 [0.02]	0.02+ [0.01]	0.02+ [0.01]	0.01 [0.02]
<i>Electoral Pressure</i>	-0.18+ [0.10]	-0.15 [0.10]	0.14+ [0.07]	0.05 [0.08]
<i>Lagged Large^t</i>	0.38+ [0.22]			
<i>Lagged Large^{t+1}</i>		0.08 [0.40]		
<i>Lagged Large^{t+2}</i>			0.83** [0.28]	
<i>Lagged Large^{t+3}</i>				0.65+ [0.35]
<i>Constant</i>	-2.92 [1.84]	-7.87*** [1.52]	-4.13*** [1.16]	-2.15 [1.85]
N	233	233	232	177
Pseudo-R2	0.43	0.68	0.59	0.60
AUC	0.91	0.97	0.94	0.95
Cluster	Country	Country	Country	Country

This table reports the results of the probit model for the dependent variable indicated below the model number. AUC is "Area Under the Curve". Standard errors clustered at the country level are in brackets. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Variables are defined in Table 1.

6.4. Diagnostics of the selection model

I now examine the distribution of predicted probabilities to examine the extent to which they are evenly distributed across treated and untreated observations. The distribution of empirical probabilities is an important diagnostic when using inverse-probability weights for two reasons. First, if there are extreme probabilities on either side of the distribution, some observations obtain extreme weights in the calculation of treatment effects because the weights are calculated as $w_{c,d} = \frac{Treatment_{c,d}}{\hat{p}_{c,d}} + \left(\frac{1-Treatment_{c,d}}{1-\hat{p}_{c,d}}\right)$. This problem can be dealt with by truncating extreme weights or by applying a stabilization factor (Jordà and Taylor, 2016). More problematic is the second aspect of highly skewed empirical distributions. Because identification requires positivity, it must be that all observations in the sample have positive probability of being either treated or untreated conditional on the estimated propensity score. Ideally, one would want these probabilities to be uniformly distributed in both the treated and untreated subpopulation. In observational studies, a uniform distribution is unlikely to arise. Yet, the empirical distribution can be inspected to establish whether there is sufficient overlap in the distribution of \hat{p} for treated and untreated observations.

Table 7 reports summary statistics of the predicted probabilities. Both the average and the median predicted probabilities increase substantially at the forecast horizon lengthens for all doses of treatment. This finding indicates that as the forecast horizon lengthens, an increasingly large number of governments expects the fiscal outlook to improve. For example, the average (median) predicted probability that $Positive = 1$ is 0.63 (0.66) at $t + 0$, compared to 0.82 (0.99) at $t + 3$. For $Large = 1$, the average (median) predicted probability is 0.21 (0.09) at horizon $t + 0$, rising to 0.67 at horizon $t + 3$. Figure 3 plots these distributions for the twelve definitions of treatment and the two treatment conditions $Treatment = 1,0$. Each graph shows the empirical propensity scores, first for the untreated and then for the treated observations using bins with length equal to 0.025. Consider the top-left chart of Panel A. The estimated probabilities for $Positive^t$ are relatively uniformly distributed when $Treatment = 0$. While there are relatively few observations with high probabilities of treatment, the distribution

is relatively uniform within levels of \hat{p} . Similarly, there are enough observations with $Treatment = 1$ for approximately all predicted probabilities, although there is a somewhat excess mass of observations with extremely high treatment probabilities – above 90 percent. Instead, consider the top-left chart of the second panel, when treatment is defined as *Large*^t. The distribution of predicted probabilities is visibly different between treatment conditions, with an extremely high number of untreated observations that present very low estimated treatment probabilities, and with only two observations for any level of $\hat{p} > 0.6$. The distribution of the treated observations is relatively uniform, but there are very few observations (21 percent of the total) that are predicted to be untreated. The figure also shows that as the horizon that defines treatment changes from $k = 0$ to $k = 1, 2, 3$ the estimated propensity scores become exceedingly skewed towards extreme values and the extent of overlap between the distribution of predicted probabilities of treated and untreated observations falls markedly. Based on this analysis, IPW methods can be employed to compute average treatment effects only for *Positive* doses of treatment and only at horizon $k = 0$. Other definitions of treatment fail to respect the positivity assumption because there is not sufficient overlap in the distribution of predicted probabilities of treatment across treated and untreated observations.

In summary, treated (untreated) observations are too (un)likely to be treated and their outcomes cannot be contrasted for most definitions of treatment based on forecasted budget balance innovations. Hence, in the causal model I only compute ATE defining treatment as a forecasted short-term budget balance improvements, i.e. *Positive*^t.

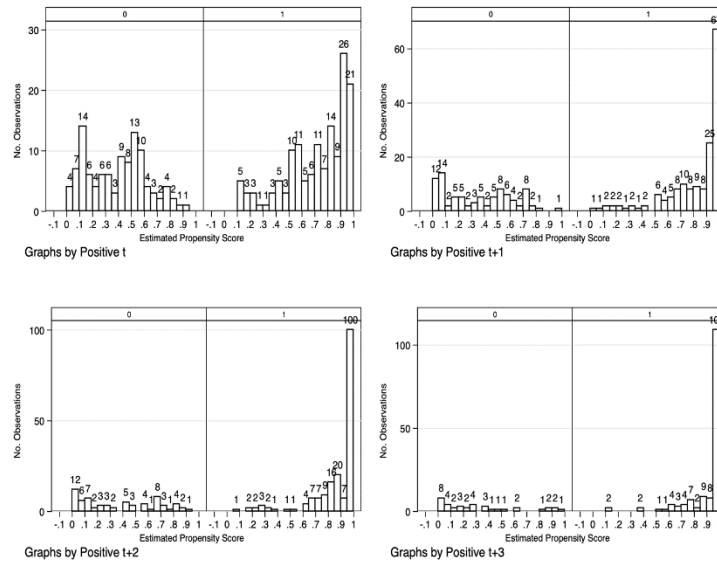
Table 7. Summary statistics of predicted probabilities of treatment

Variable	N	Mean	St. dev	Median	Min	Max
Pr(<i>Positive</i> ^{t+0})	233	.63	.29	.66	.01	1
Pr(<i>Positive</i> ^{t+1})	229	.69	.32	.8	0	1
Pr(<i>Positive</i> ^{t+2})	224	.77	.32	.91	0	1
Pr(<i>Positive</i> ^{t+3})	194	.82	.31	.99	0	1
Pr(<i>Large</i> ^{t+0})	233	.21	.27	.09	0	1
Pr(<i>Large</i> ^{t+1})	233	.36	.4	.1	0	1
Pr(<i>Large</i> ^{t+2})	233	.48	.4	.41	0	1
Pr(<i>Large</i> ^{t+3})	197	.57	.39	.67	0	1

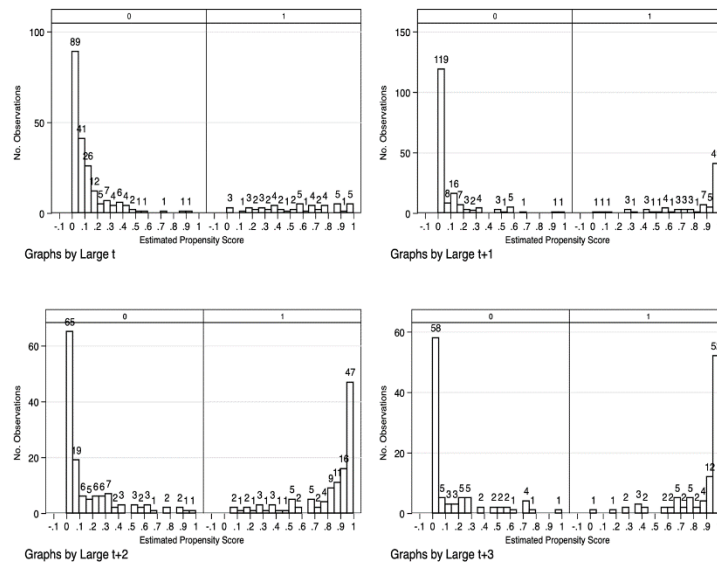
This table reports summary statistics of the predicted probabilities \hat{p} of observed treatment status. The treatment status is shown in parentheses. Predicted probabilities are computed as $\hat{p} = \Pr(\text{Treatment}^{t+k} = 1 | \mathbf{X})$, where \mathbf{X} is the vector of control variables of the probit model.

Figure 3. Distribution of predicted empirical probabilities by treatment status (1 for treated, 0 for untreated)

Panel A: *Treatment* \equiv *Positive*



Panel B: *Treatment* \equiv *Large*



This figure plots the empirical probabilities that *Treatment* \equiv *Positive* (Panel A) and *Treatment* \equiv *Large* (Panel B) at the forecast horizon indicated below each graph. The empirical probabilities are estimated through the probit model. Variables are defined in Table 1.

6.5. Results of the baseline model

This sub-section presents the result of testing [Hypothesis 2a](#) and [Hypothesis 2b](#). Table 8 reports estimates of the treatment model in the announcement period. Panel A reports Average Treatment Effects (ATE) computed as the contrasts in cumulative return volatility as governments shift from the untreated to the treated condition based on the AIPW estimator in (4). Panel B reports the same analysis, but for cumulative (log) turnover. As in Chae (2005), the announcement period includes the five trading days between $d - 2$ and $d + 2$. Thus, the accumulation period begins at $d - 2$ (column 1) and it ends at $d + 2$ (column 5). Cumulative abnormal return volatility and trading volume are denoted with $CABVAR$ and $CABT$, respectively.

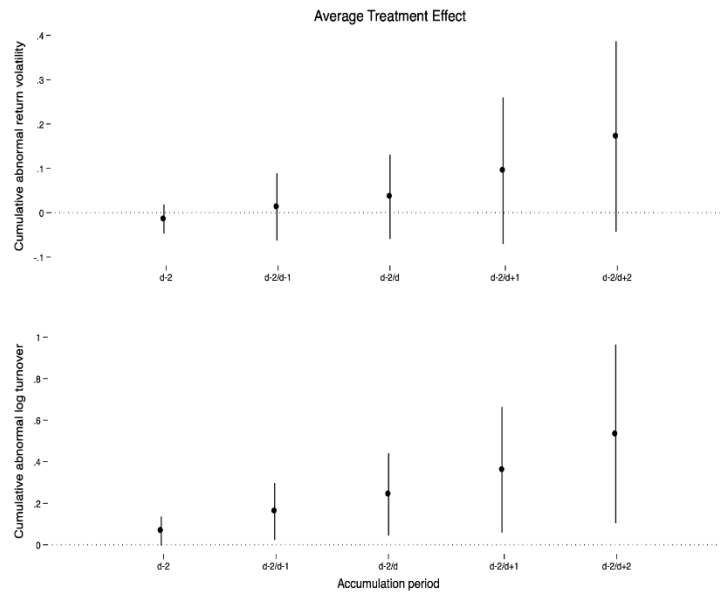
The results in Panel A show that forecasting a short-term improvement as opposed to deterioration in the fiscal outlook does not increase the volatility of stock returns. Over the announcement period $d - 2/d + 2$, cumulative abnormal return volatility is positive – at 17.2 percent – but standard errors are too large. The top chart of Figure 4 visualizes these estimates throughout the announcement period, revealing a simultaneous increase in both $CABVAR$ and its standard error. Hence, equity investors attribute equal information content to announcements that the budget balance will improve or deteriorate in the short term. Panel B reports ATE for cumulative log turnover. In stark contrast with the results on cumulative return volatility, trading volume is significantly higher in the treated condition. Trading volume is not statistically different from zero on day $d - 2$, becoming marginally significant on day $d - 1$. From the event day onwards, cumulative abnormal trading volume is about 24.2 percent higher for treated than untreated units. This difference exceeds 53 percent at the end of the announcement period. The bottom chart of Figure 4 visualizes this trend.

Table 8. Expected improvements in the fiscal outlook cause more trading activity than expected deteriorations

Panel A: Cumulative abnormal return volatility (<i>CABVAR</i>)					
	(1)	(2)	(3)	(4)	(5)
	<i>d-2</i>	<i>d-2/d-1</i>	<i>d-2/d</i>	<i>d-2/d+1</i>	<i>d-2/d+2</i>
ATE	-0.014	0.013	0.036	0.095	0.172
s.e.	[0.019]	[0.044]	[0.055]	[0.096]	[0.125]
N	249	246	242	213	200
Panel B: Cumulative abnormal log turnover (<i>CABT</i>)					
	(1)	(2)	(3)	(4)	(5)
<i>Accumulation period</i>	<i>d-2</i>	<i>d-2/d-1</i>	<i>d-2/d</i>	<i>d-2/d+1</i>	<i>d-2/d+2</i>
ATE	0.066	0.160 ⁺	0.242 [*]	0.361 ⁺	0.534 [*]
s.e.	[0.041]	[0.080]	[0.115]	[0.176]	[0.250]
N	239	236	232	203	190

This table reports average treatment effects (ATE) as the contrast between *CABVAR* (Panel A) and *CABT* (Panel B) for treated and untreated observations over the event days indicated below the model number. ATE are estimated through the estimator in (4). Treated observations are those for which $Positive^t = 1$. The outcome model is omitted for presentation purpose. Standard errors clustered at the country level are in brackets. ⁺ $p < 0.10$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$. Variables are defined in Table 1.

Figure 4. Expected improvements in the fiscal outlook cause more trading activity than expected deteriorations



This figure plots average treatment effects (ATE) of events in which the government announces that the budget balance will improve between year $t - 1$ and year t . ATE are estimated through the estimator in (4). The outcomes are abnormal return volatility and abnormal trading volume (in logarithms) in the top and bottom graph, respectively. Outcomes are cumulated from two days before to the event day d to two days after. Bars represent 90 percent confidence intervals. Variables are defined in Table 1.

These findings do not support the hypothesis that investors respond to forecasted budget balance improvements because of uncertainty over their sustainability – the “persistence” explanation. If investors traded in response to sustainability concerns, the treatment effect in Panel B would be negative. Instead, the evidence suggests that investors disagree over the probability that guidance will translate into outcomes – i.e. they may disagree about its “credibility”. The “persistence” explanation is also not supported by the result that investors do not attribute greater information content to forecasted short-term improvements rather than deteriorations in the budget balance. Should investors find forecasts that portray a budget balance improvement informative because of their persistence, the treatment effect in Panel A should be positive. Instead, under the credibility explanation, the treatment effect can be positive only if guidance is credible.

6.6. The role of credibility

The next set of analyses explores the validity of the credibility explanation for these results. Table 9 reports ATE for bins defined by the level of credibility of the reported forecast. Once more, Panel A (B) reports the cumulative ATE when the outcome variable is return volatility (turnover). In the first set of results, the bins are defined as government-years in which the previous forecasted budget balance innovation exceeded or fell short of the actual innovation, meaning that the previous forecast was *Conservative* or *Optimistic*, respectively. In the second set of results, the bins are defined as government-years in which the current forecasted budget balance innovation is either *Aggregated* or *Disaggregated*.

Panel A reveals that cumulative return volatility is significantly higher – and relatively large in magnitude – when governments that have been conservative in the past SCP announce that the fiscal performance will improve in the short run. This result implies that announcements of future improvements in the budget balance are more informative than announcements of future deteriorations, but only when the government is credible. In turn, equity investors do not similarly reward optimistic governments regardless of the content – positive or negative -

of their guidance. Hence, equity markets' response to fiscal guidance does not depend on the information disclosed when the government has been optimistic in the past, whereas it does depend on it when the government has been conservative. This evidence supports the "credibility" channel. In turn, the results in Panel B indicate that optimism plays a major role in explaining abnormal trading activity in response to fiscal guidance. The evidence indicates that trading activity is larger both statistically and economically already two days before the announcement of positive guidance. More importantly, trading volume increases by over 50 percentage points between $d - 2$, $d - 1$ and the announcement day d . During the two days following the announcement, trading volume increases a further 60 percentage points in each day. Hence, investors trade in response to forecasted budget balance improvements, but only when the government has not been credible in the past. Turning to the role of disaggregation, I document that forecasted improvements generate more trade than forecasted deteriorations when guidance is disaggregated. This effect begins on the event day and it becomes large by the end of the announcement period. This finding does not support the "credibility" channel because disaggregated forecasts should be more credible than aggregated ones. I find, however, that this is not the case.

Finally, Table 10 examines the role of credibility of "implementing" the forecasted budget balance improvement. I now estimate treatment effects for bins defined by the strength of the incumbent government. Panel A documents a lack of significance of credibility as a moderator of treatment on return volatility. Instead, the results in Panel B provide additional evidence that trading volume responds to forecasted improvements in the budget balance because of concerns about credibility. Cumulative abnormal volume is about 30 percent higher when a weak coalition announces an improvement as opposed to a deterioration (or a zero change) in the budget balance, and about 38 percent higher when the government confronts a divided parliament.

Table 9. The causal effect of expected improvements in the fiscal outlook depends on the credibility of guidance

Panel A: Cumulative abnormal return volatility (<i>CABVAR</i>)					
<i>Condition</i>	(1) <i>d-2</i>	(2) <i>d-2/d-1</i>	(3) <i>d-2/d</i>	(4) <i>d-2/d+1</i>	(5) <i>d-2/d+2</i>
<i>Conservative</i>	-0.04 [0.04]	0.04 [0.07]	0.10 [0.10]	0.17 [0.14]	0.43* [0.20]
<i>Optimistic</i>	-0.00 [0.02]	0.06 [0.06]	0.02 [0.08]	0.03 [0.13]	-0.01 [0.17]
N	249	246	242	213	200
<i>Condition</i>	(1) <i>d-2</i>	(2) <i>d-2/d-1</i>	(3) <i>d-2/d</i>	(4) <i>d-2/d+1</i>	(5) <i>d-2/d+2</i>
<i>Disaggregated</i>	-0.01 [0.04]	0.09 [0.10]	0.08 [0.11]	0.22 [0.16]	0.29 [0.20]
<i>Aggregated</i>	0.01 [0.01]	0.01 [0.03]	0.00 [0.06]	0.04 [0.10]	0.12 [0.14]
N	249	246	242	213	200
Panel B: Cumulative abnormal log turnover (<i>CABT</i>)					
<i>Condition</i>	(1) <i>d-2</i>	(2) <i>d-2/d-1</i>	(3) <i>d-2/d</i>	(4) <i>d-2/d+1</i>	(5) <i>d-2/d+2</i>
<i>Conservative</i>	-0.08 [0.14]	-0.09 [0.23]	0.12 [0.22]	0.07 [0.34]	-0.19 [0.60]
<i>Optimistic</i>	0.39** [0.13]	0.34 [0.22]	0.91** [0.30]	1.50** [0.43]	2.15** [0.65]
N	239	236	232	203	190
<i>Condition</i>	(1) <i>d-2</i>	(2) <i>d-2/d-1</i>	(3) <i>d-2/d</i>	(4) <i>d-2/d+1</i>	(5) <i>d-2/d+2</i>
<i>Disaggregated</i>	0.10 [0.10]	0.24 [0.16]	0.34* [0.18]	0.73* [0.26]	0.96** [0.32]
<i>Aggregated</i>	-0.01 [0.12]	0.08 [0.16]	0.14 [0.22]	0.15 [0.30]	0.21 [0.44]
N	239	236	232	203	190

This table reports average treatment effects (ATE) as the contrast between *CABVAR* (Panel A) and *CABT* (Panel B) for treated and untreated observations over the event days indicated below the model number for bins defined by the condition indicated in the first column. ATE are estimated through the estimator in (4). Treated observations are those for which $Positive^t = 1$. The outcome model is omitted for presentation purpose. Standard errors clustered at the country level are in brackets. * $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Variables are defined in Table 1.

Collectively, the results suggest that the reason that leads investors to engage in abnormal amounts of trading activity when the government announces a short-term improvement in the budget balance is that investors are uncertain as to the ability of governments to implement the promised adjustment. Uncertainty increases for weak governments and for governments that have been optimistic in their past forecasts. These findings are incompatible with the notion that trading responds to the uncertainty caused by announcements of an increased debt accumulation. Instead, the findings broadly support the hypothesis that investors respond to forecasted short-term improvements in the budget balance only when these are credible. However, the findings indicate that the extent of disaggregation of forecasts does not increase the credibility of forecasts. I am not aware of studies that relate forecast disaggregation to trading volume in the corporate literature, and this is a surprising finding that may be interesting to investigate in future research.

Recall that these effects are estimated controlling for the contemporaneous return volatility and bid-ask spread. In addition, note that in none of the days of the announcement period, return volatility is significantly different from zero. Hence, trading volume takes place for reasons unrelated to both information asymmetry across traders (as measured by the bid ask spread) and information content (as measured by return volatility). Indeed, virtually identical results obtain when replacing the measure of abnormal trading volume with the measure of opinion divergence (standardized unexplained volume) in Garfinkel (2009; cf. Garfinkel and Sokobin, 2006).

Table 10. The causal effect of expected improvements in the fiscal outlook depends on the probability of implementation

Panel A: Cumulative abnormal return volatility (*CABVAR*)

<i>Condition</i>	(1) <i>d-2</i>	(2) <i>d-2/d-1</i>	(3) <i>d-2/d</i>	(4) <i>d-2/d+1</i>	(5) <i>d-2/d+2</i>
<i>Strong Government</i>	-0.04 [0.04]	0.02 [0.10]	0.09 [0.13]	0.33 [0.20]	0.36 [0.24]
<i>Weak Government</i>	0.01 [0.02]	0.09* [0.04]	0.10 [0.06]	0.08 [0.10]	0.11 [0.14]
Observations	249	246	242	213	200

<i>Condition</i>	(1) <i>d-2</i>	(2) <i>d-2/d-1</i>	(3) <i>d-2/d</i>	(4) <i>d-2/d+1</i>	(5) <i>d-2/d+2</i>
<i>Cohesive Parliament</i>	0.01 [0.09]	0.03 [0.12]	0.14 [0.13]	0.17 [0.22]	0.39 [0.27]
<i>Fragmented Parliament</i>	0.04 [0.05]	0.03 [0.06]	0.01 [0.08]	0.12 [0.12]	0.15 [0.14]
Observations	249	246	242	213	200

Panel B: Cumulative abnormal log turnover (*CABT*)

<i>Condition</i>	(1) <i>d-2</i>	(2) <i>d-2/d-1</i>	(3) <i>d-2/d</i>	(4) <i>d-2/d+1</i>	(5) <i>d-2/d+2</i>
<i>Strong Government</i>	0.04 [0.08]	0.16 [0.15]	0.29 [0.32]	0.38 [0.40]	0.80 [0.64]
<i>Weak Government</i>	0.13 [0.07]	0.33* [0.15]	0.33* [0.16]	0.42+ [0.24]	0.57+ [0.32]
Observations	239	236	232	203	190

<i>Condition</i>	(1) <i>d-2</i>	(2) <i>d-2/d-1</i>	(3) <i>d-2/d</i>	(4) <i>d-2/d+1</i>	(5) <i>d-2/d+2</i>
<i>Cohesive Parliament</i>	0.07 [0.13]	0.25 [0.24]	0.12 [0.21]	0.15 [0.40]	0.40 [0.50]
<i>Fragmented Parliament</i>	0.07 [0.06]	0.22* [0.08]	0.39* [0.16]	0.42+ [0.22]	0.59+ [0.30]
Observations	239	236	232	203	190

This table reports average treatment effects (ATE) as the contrast between *CABVAR* (Panel A) and *CABT* (Panel B) for treated and untreated observations over the event days indicated below the model number for bins defined by the condition indicated in the first column. ATE are estimated through the estimator in (4). Treated observations are those for which $Positive^t = 1$. The outcome model is omitted for presentation purpose. Standard errors clustered at the country level are in brackets. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Variables are defined in Table 1.

A direct comparison of the results with earlier literature can be attempted, but only tentatively. Prior work on budget balance forecasts generally finds that announcements of budget balance deteriorations are associated with an increase in interest rates – consistent with crowding out effects (Thorbecke, 1993). Hence, one might expect an increased return volatility in the treated condition. However, earlier studies use a continuous as opposed to binary measure of budget balance changes. While this choice has the benefit of utilizing the entire information that the variable contains (Alesina, Favero et al., 2015), it is exposed to the risk of extrapolation if large innovations are unevenly distributed between treated and untreated units. A similar point is made by Jordà and Taylor (2016) in their study of the treatment effect of actual (as opposed to forecasted) budget balance improvements. As it has been in [section 6.4](#), this risk is material. In addition, earlier studies do not consider the role of credibility. With respect to the documented effects on trading volume, with the exception of Bafundi et al. (2018), I am not aware of comparable studies on the effect of fiscal guidance on trading volume. The results presented here may explain their finding of a positive post-announcement trading activity in the Spanish market. In the present sample, the Spanish government is in the treated condition more than 80 percent of the times.

7. Conclusions

This chapter has analyzed the equity markets response to guidance issued by the government on the future trajectory of the budget balance – the difference between government revenues and expenditures. While the results indicate that guidance by the government is an information-rich event in equity markets, the analysis shows the important role played by the credibility of guidance as a moderator of investors' response. Indeed, the findings indicate that as governments announce that the budget balance will improve in the short term, investors remain generally skeptical. In particular, investors interpret announcements of short-term improvements in the budget balance heterogeneously. The degree of heterogeneity in investors' response appears to be a function of the credibility of governments' forecasts. In particular, the evidence shows that investors' opinions diverge more when the government has

been optimistic in its last forecast and when it holds a weak political mandate. These findings suggest that credibility plays an important role in moderating investors' response to governments' guidance. Surprisingly, the results indicate that credibility is negatively affected by the extent of disaggregation of guidance.

Future research may investigate why it is the case that in the government setting, disaggregation appears to make guidance less credible. It would be consistent with the results of [Chapter 1](#) that investors discount the value of disaggregated forecasts because the intensity of guidance may be a substitute for its accuracy. In addition, it would be important to examine the relation between bond market investors' response and equity investors' response. If it is the case that trading in secondary bond markets is affected by guidance on the budget, then part of the trading volume response documented in this analysis may be driven by portfolio rebalancing rather than opinion divergence. At the same time, by controlling for the level of information asymmetry in markets, the level of liquidity trading and the portion of trading that responds to information content, the analysis presented in this chapter alleviates concerns over alternative explanations for the results.

CHAPTER 3. THE ROLE OF JUDGMENT IN THE EVALUATION OF GOVERNMENTS: THE CASE OF SOVEREIGN RATING ANALYSTS

1. Introduction

Why is it the case that the Republic of Singapore is given a Aaa credit rating by Moody's, just like the Federal Republic of Germany? And how can it be that, thanks to this rating, the tiny city-state ranks higher than France whose rating is, in turn, the same as that of Kuwait? Besides, why was the rating on the UK downgraded by S&P in the aftermath of the Brexit vote, but that of Spain was unaffected by the political unrest in Catalonia?¹⁴

This chapter adopts qualitative methods to examine the making of sovereign ratings - the credit ratings of sovereign governments.¹⁵ Credit ratings belong to the large, growing and somewhat discredited family of financial indicators. According to Farlow (2015: 228), they are “[p]erhaps the best-known, and most ubiquitous” among them. Indeed, government ratings are everywhere: there are now over 100 governments rated by the three largest credit rating agencies. In contrast, in the 1970s Moody's only rated five governments, S&P rated seven, and Fitch rated none (Gaillard, 2012). Governments issue about 40% of the global stock of bonds and the value of those that are rated approaches \$50 trillion (D'Agostino and Lennkh, 2016; S&P, 2015). Sovereign ratings correlate with governments' debt, financing costs and interest rates, making them a compact measure of governments' financial performance, position and prospects (Afonso

¹⁴ All ratings are as of 26 June 2018. Source: company websites.

¹⁵ Sovereign ratings apply to debt instruments traded in capital markets that are issued by governments and held by private-sector investors. The domain of application of a sovereign rating thus excludes the debt held, or the credit lines granted, by official-sector institutions such as the International Monetary Fund and the World Bank, as well as the debt held by foreign governments. Also, sovereign ratings also do not cover the obligations of a government towards employees, suppliers, and pension beneficiaries (S&P, 2014a). Regulation 462/2013 amending Regulation 1060/2009 defines sovereign ratings in a somewhat broader sense, including the ratings applied to regional and local authorities of the State. In this paper, the focus is on the rating of government entities' creditworthiness at the highest level, that is a federal or central government (EU, 2009b, 2013).

et al., 2011; Cantor and Packer, 1996). They influence private sector borrowing costs, credit ratings and asset prices, as well as the economy at large (Borensztein et al., 2013; Chen et al., 2016; Corsetti et al., 2013; Ferreira and Gama, 2007; Gande and Parsley, 2005). The recent financial crisis was a painful reminder of the systemic importance of sovereign ratings.

Yet, we know very little about how these ratings come about. As Bruner and Abdelal (2005: 198) note, there is “lack of awareness of what rating agencies actually do”. Prior to the financial crisis, research on sovereign ratings largely focused upon emerging markets (Gande and Parsley, 2005). After some European countries experienced rating downgrades, especially Greece in 2010, scholarly attention shifted to understanding what leads a credit rating agency (CRA) to change its assessment of governments’ creditworthiness. An interesting avenue pursued by economics research has been to develop models to disentangle the relative importance of “hard data” vis-à-vis analysts’ subjective judgements in the making of ratings (Afonso et al., 2011; D’Agostino and Lennkh, 2016; Ferri et al., 1999; Mora, 2006; Vernazza and Nielsen, 2015). While these articles indicate that analysts’ valuation practices involve a judgmental component, a comprehensive examination of the role of judgment in the making of sovereign ratings is presently lacking. This is an important limitation because the judgmental component of sovereign ratings is often considered to be a manifestation of CRAs’ incentive to issue procyclical and often inflated ratings to please paying issuers (Bolton et al., 2012; Mathis et al., 2009). Accordingly, several scholars proposed to publish two separate ratings – one that expresses the result of automated rating model, and one that expresses analysts’ judgment (D’Agostino and Lennkh, 2016; Jollineau et al., 2014). In its 2008 review of the CRA industry, the Securities and Exchange Commission (SEC) lamented the lack of transparent disclosure of “factors that were considered in developing a particular rating” and that “[t]he rationale for deviations from the model or out of model adjustments was not always documented” (SEC, 2008b: 19). Accordingly the SEC proposed that CRAs “keep a record of the rationale for any material difference between the credit rating implied by the model and the final credit rating issued.” (ibid. 20, cf. SEC, 2008a). This proposal assumed that the rating

generated by models are void of judgment, and that the application of judgment derives from opportunistic behavior. The material presented here will shed some light on the practical feasibility of these proposals.

Drawing on 27 interviews conducted at four global credit rating agencies (CRAs) and on archival materials, this chapter examines the role of analysts' judgment in the making of sovereign ratings. Building on prior conceptual frameworks on the work of equity analysts (Abhayawansa et al., 2015), the material indicates that analysts' judgment permeates all stages of the sovereign rating process. Judgment is deployed by analysts as they select and neglect indicators that feed their automated rating models. Judgment is also deployed by analysts as they recommend a credit rating that deviates from that indicated by models. Judgment also manifests itself in analysts' interpretation of information and indicators, even if these are gathered in quantitative form and from trusted sources. In an attempt to theorize from the material, the chapter examines the rationale that analysts most commonly mention in connection with the application of judgment: the small sample of sovereign default renders statistical modes of inferences unfeasible. Accordingly, information must be interpreted, and sovereign ratings are the result of qualitative inferences. Automated rating models are not used to construct accurate ratings. Rather, their role is for ratings to be constructed following a consistent and transparent process. For ratings to be accurate, judgment is necessary.

In the discussion section, this finding will be theorized in terms of Frank Knight's discussion of the concepts of risk and uncertainty (Knight, 1964). It will be shown that these concepts are labels that Knight adopts to describe a rather more fundamental valuation problem confronting all forward-looking actors such as rating analysts. As analysts attempt to infer whether a certain piece of information indicates that the risk of sovereign default has changed, they confront the challenge that sovereigns hardly default. Therefore, it is challenging to determine whether information points to a higher or lower default risk because the small sample of sovereign default precludes drawing this inference statistically – and therefore, objectively. This implies that each government - and

each event that it goes through - must be analyzed on a case-by-case basis. But, in Knightian terms, a case-by-case analysis is one in which a case is considered to be a relatively “unique” instance. When a case is unique, it does not belong – by definition – to a wider class of mostly similar cases. Then, statistical methods of inference cannot be applied because statistics only applies to classes of cases and never to individual instances (Guseva and Rona-Tas, 2001; Hoppe, 2007). Accordingly, exact numerical probabilities of default cannot be calculated. Instead, sovereign ratings represent “estimates” in the Knightian sense: “an estimate or intuitive judgment is somewhat like a probability judgment”. However, it is different from numerical probabilities because it rests on “*no valid basis of any kind for classifying instances.*” (Knight, 1964: 223–224, original emphasis). The point bears emphasis. Numerical, exact probabilities can be inferred from data when history provides a sufficiently large sample of occurrences, such that any “new” case can be classified into a category within that sample. Then, probabilities can be calculated and expressed numerically. Instead, when cases are unique, only judgmental estimates can be expressed that, while not being entirely arbitrary, do not rest on robust empirical evidence. The possibility of assigning probabilities or making estimates is what distinguishes risk and uncertainty: “risk” and “uncertainty” denote not so much the possibility of attributing probabilities to future events but rather the possibility of classifying cases into categories. In the case of sovereign ratings, analysts mostly confront uncertainty and ratings themselves express estimates rather than probabilities. Overall, the findings of this inquiry highlight that judgment is a pervasive feature of sovereign credit ratings and that – somewhat contrary to what was claimed by Besedovski (2018) - sovereign rating analysts remain generally distrustful of automated models and statistical modes of inference. In addition, unlike Carruthers (2013), the findings suggest that sovereign rating analysts do not “transform” uncertainty into risk. Rather, they seem to be aware that ratings express, at best, judgmental estimates. The benefit of theorizing valuation practices using Knight is that it allows to state the testable propositions that: (i) valuation practices in accounting and finance always begin with the attempt to assign individual instances to broader classes; and (ii) valuation practices are

more automated when this activity of classification is easier and more judgmental when it is relatively more challenging. The sovereign rating industry thus represents a “paradigmatic case” (Patton, 2002) of a valuation challenge that – it will be proposed - confronts all valuation practices in accounting and finance.

The chapter proceeds as follows. Section 2 reviews related literature. Section 3 introduces the conceptual framework that guided the data collection and analysis. Section 4 discusses the sources of data and the method of analysis. Section 5 outlines the findings. Section 6 discusses how the findings enriches prior literature and states two propositions. Section 7 concludes. [Appendix D](#) describes the credit rating industry.

2. Related literature

By studying the valuation practices of sovereign rating analysts, this chapter connects with the sociological literature that studies the construction of quantitative financial indicators, of which ratings are a prime example (Farlow, 2015; Rottenburg et al., 2015). Prior studies use qualitative methods to investigate analysts’ valuation practices but focus primarily on sell-side equity analysts. It was not until recently that qualitative work on sovereign ratings began to be conducted. In addition, there is a substantial body of research that uses econometric methods to dissect the role of judgment in the making of sovereign ratings. This literature is reviewed first, before discussing related sociological studies.

2.1. Economic research on sovereign ratings

Economic studies of sovereign ratings typically adopt a functionalist standpoint: the role of CRAs is to alleviate information asymmetry between lenders and borrowers by specializing in the production of an indicator of default probability (Cantor and Packer, 1994; White, 2010). Because of the arms-length nature of capital market transactions (Rajan, 1992), this role is particularly valuable to bond investors.

In an early study of sovereign ratings determinants, Cantor and Packer (1996) document the importance of eight explanatory variables: Gross Domestic

Product (GDP) per capita, GDP growth, the rate of inflation, the budget balance, the net income flows vis-à-vis the rest of the world, the amount of debt owned by foreign residents, the level of economic development and the history of default (cf. Mosley, 2003b). More recently, Afonso, Gomes and Rother (2011) adopt a sophisticated econometric specification to separate short-and long-term determinants of sovereign ratings. They show that in the short-term, rating changes are explained mostly by economic and financial results. In the long-term, however, the quality of a nation's institutions, the history of default and the extent of dependence on financial flows from abroad become relatively more important.

Interestingly, rating agencies disagree on the level of sovereign ratings. Hill, Brooks and Faff (2010) report that instances of disagreement for sovereigns rated by more than one agency are more common than instances of agreement. In addition, disagreement is visibly more pronounced in the case of sovereign as opposed to corporate ratings. They also show that in a sample of countries rated by the three larger agencies between 1990 and 2006, the numerical variables that best explain the variance in S&P ratings differ from those that explain the variance at Moody's and Fitch, suggesting that agencies differ in their judgment on how best to measure creditworthiness.

A number of studies attempt to separate the relative roles of judgement and automated models in the final rating assessment. Ferri, Liu and Stiglitz (1999) conceptualize sovereign ratings as the weighted sum of two components: an "objective" rating that results from the application of a rating model, and a "subjective" rating that results from the application of analysts' judgement. The weights reflect the relative importance of quantitative and judgmental inputs in the final rating assessment. The authors show that prior to the East Asian crisis, the model-implied rating of East Asian sovereign governments was systematically lower than the observed rating, suggesting that judgment was used by CRAs to inflate ratings. Following the crisis, CRAs downgraded issuers too aggressively compared to the model results, suggesting that judgment was used conservatively to restore reputational capital. Mora (2006) confirms that judgment led to inflated ratings prior to the crisis, while showing that the observed

procyclical behavior of ratings was due to model misspecification. In turn, her article documented substantial inertia in sovereign ratings. Gärtner, Griesbach and Jung (2011) adopt a similar approach to examine CRAs' behavior during the euro area sovereign debt crisis. They show that countries affected by the crisis were unduly penalized by the application of CRAs' judgment. Vernazza and Nielsen (2015) also follow this method and documented that while the model-implied rating has predictive ability, subjective adjustments do not. In fact, the application of judgment biases ratings in the wrong direction. Finally, D'Agostino and Lennkh (2016) re-engineer the model-implied credit rating of Moody's through a replication of its methodology. They find that while some countries systematically benefit from the exercise of judgment by Moody's analysts, others are typically penalized. The authors also attempt to explain the reasons behind the judgmental component of ratings and document that judgment is relatively persistent.

While econometric studies reveal that sovereign ratings are the result of both automated models and judgment, they do not attempt to investigate how judgment emerges or why it is needed. In addition, these studies hint at the fact that judgment may be used in some circumstances, but do not explain what these circumstances are. The reason is largely methodological. While quantitative studies are able to answer the question of "what" drives sovereign ratings, they are lesser able to provide insights over "how" ratings are constructed in practice. In contrast, sociological research focuses on "how" questions through qualitative methods. The next sub-section elaborates on this literature.

2.2. Sociological research on analysts

The sociological literature on analysts mostly focuses on explaining why analysts often disagree with each other – a manifestation of subjective judgment. Beunza and Garud (2007) argue that disagreement arises because different analysts interpret, construct, and frame their valuation objects – e.g., stocks - as belonging to either the same sector or to different ones (cf. Giorgi and Weber, 2015). Depending on the result of this framing, analysts may issue a buy/sell recommendation that has little to do with existing public or private information and

much to do with the sector to which analysts have framed the firm as belonging to. Blomberg (2016) finds that analysts' recommendations on a stock differ depending on the way in which analysts conceive of their work. While some analysts conceive their work as being based on purely quantitative analyses, others emphasize the importance of embeddedness in the social world of markets. In an earlier study of sell-side analysts, Johansson (2007) also identified the importance of relational capital to analysts' reception *and* provision of value-added, contextual information. Relatedly, Spence et al. (2019) delve deeper into analysts' embeddedness in financial markets and document that analysts use equity reports instrumentally, not to generate valuable information but rather to gain entry into more profitable businesses. Imam and Spence (2016) ask why analysts are listened at all by equity investors, given that their forecasts are notoriously prone to bias and error. They find that analysts' value lies in the contextual information that accompany forecasts rather than in the forecasts themselves.

Related articles on credit scoring systems in finance suggest that a strong tension may exist between judgment and quantitative models. For example, Poon (2009) shows that in the consumer finance industry, credit scoring practices have become increasingly dependent on automated credit scoring models. The automated scoring system found the support of US regulatory authorities because of its "mechanical objectivity" (Porter, 1994): credit guidelines were no longer subject to interpretation with potentially discriminatory consequences as lenders were forced to apply "a non-manipulable credit factor" (Fourcade and Healy, 2013; Poon, 2009). Somewhat similarly, Pollock and D'Adderio document that the methodology for ranking IT firms by industry consulting firm Gartner attempts to discourage the application of analysts' judgment: "In order to attempt to remove the idea of bias and partiality from the ranking, individual analysts were now increasingly circumscribed by a new material and organizational reality (increasingly explicit assessment criteria, a methodology team scrutinizing their work, the need to provide explicit evidence for choices, a spreadsheet that plotted dots, etc.)" (Pollock and D'Adderio, 2012: 575; Porter, 1994; Power, 1997).

These studies go a long way towards explaining the way in which analysts use information and to some extent unveil the role of judgments in the process of, e.g. stock recommendation and in the making of rankings and credit scoring systems. Yet, these studies do not examine sovereign rating analysts. Only recently have scholars begun to explore the making of sovereign ratings. Penét (2015) studies the rating reports written by analysts at Moody's during the Greek sovereign debt crisis. He claims that these reports acted as "figuring devices" through which Moody's articulated and communicated what analysts perceived as the relevant risks and the pressing uncertainties, while underlying that they were intertwined with a broader contextual, legal and political factors (cf. Mallard and Pénet, 2013). His study is based on archival material, however, and could not benefit from analysts' own narrative of the process underlying rating practices. Recently, Besedovski (2018) describes the changing epistemology of ratings – once more focusing on Moody's. She proposes that Moody's has become increasingly "financialized" over the last forty years insofar as the calculative practices embraced by its rating analysts have shifted from judgment to automated models. While this shift has affected mostly the structured finance departments of Moody's, the author documents a growing trend towards a quantitative construction of sovereign ratings. Yet, the article does not delve deeper into the relative importance and rationale of sovereign analysts' judgment in the rating process. Hence, the research question of this chapter is:

Research Question: How does analysts' judgment interaction with automated valuation models in the construction of sovereign ratings?

3. Conceptual framework

This section discusses the framework that was adopted to analyze the tension between sovereign rating analysts' judgment and automated valuation models. This framework is presented in Figure 1. It is informed by prior literature that documents the coexistence of automated valuation models and analysts' judgment. Accordingly, the field of CRAs was entered with the following themes

in mind.

3.1. Using judgment in the creation of valuation models

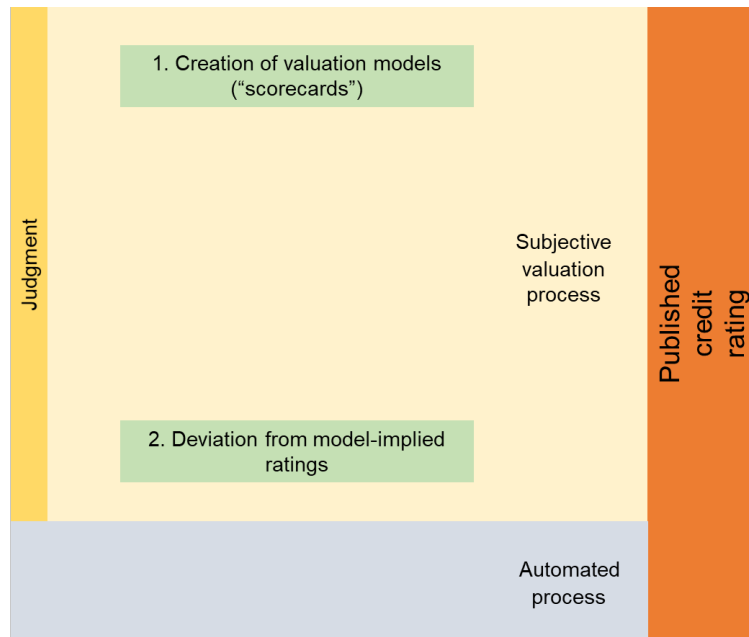
This theme explores the possibility that analysts use judgment in the creation of rating models. As noted in Section 2, economic research often decomposes observed sovereign ratings into an objective and subjective component. The subjective component is estimated as the residual variance left unexplained by an econometric model. The model-implied rating is considered to be “objective” because the model variables are not manipulated by rating analysts. However, sociological research finds that (equity) analysts often use different analytical frames in their valuation practices. Thus, it is possible that in the sovereign rating industry, too, analysts’ judgment affects the creation of valuation models. A possible manifestation of judgment in model building is that for variables to enter a model, they must be selected as relevant drivers of creditworthiness. For example, Mehrpouya and Samiolo (2016) documents substantial judgment in the choice of the indicators that underpin ranking models that measure the progress that pharmaceutical companies make towards socially responsible goals. On the basis of which criteria do sovereign rating analysts choose rating indicators at the expense of others? And, how is the predictive ability of these indicators established? In addition, economic research posits that ratings are the weighted average of “objective” ratings derived from models and “subjective” ratings informed by analysts’ judgment. To what extent does judgement play a role in the weighting of indicators, and why do analysts select weights that depart from those suggested by the sovereign rating model? These questions speak to the role of judgment in the construction of “objective” valuation models.

3.2. Using judgment to deviate from model-implied ratings

This theme examines the extent to which analysts trust the result of automated valuation models. As noted by D’Agostino and Lennkh (2016: 22), “a non-trivial element of sovereign risk for wealthier economies in distress remains unexplained by the scorecard.” The “scorecard” is the name given across the industry to the set of indicators that feed automated valuation models and to the weighting function that ties the indicators together. CRAs are open about the

possibility that ratings deviate from scorecards. Moody's note that "the rating factors in the scorecard do not constitute an exhaustive treatment of all the considerations that are important for sovereign bond ratings" and that "[r]atings may also reflect circumstances in which the weighting of a particular factor will be substantially different from the weighting suggested by the scorecard." (Moody's, 2016: 28). But, what are these "circumstances" and "considerations", and why is the scorecard unable to capture them fully? Further, how do analysts justify the application of judgment in the deviation from the model results? These questions shed light on the role of judgment in constructing a more "accurate" credit rating.

Figure 1. Conceptual framework of analysts' uses of judgment in the construction of sovereign credit ratings



This figure visualizes the Conceptual Framework used in the data collection and analysis.

4. Research Method

4.1. Sampling strategy

The first phase of a qualitative research design is the selection of the case to be studied (Flyvbjerg, 2006). This phase of the research design is particularly critical to qualitative inquiry because of the resources that this method typically requires and because of the possibility that repeated access to interviewees will not be possible. It is therefore crucial for the selected case to provide the greatest information content with respect to the research question. In this context, representative sampling is generally agreed not to be the most effective sampling strategy (Flick, 2009; Flyvbjerg, 2006). Purposeful sampling (Patton, 2002) was adopted instead so as to obtain an “information-rich” case. An information-rich case is defined as one “from which one can learn a great deal about issues of central importance to the purpose of the research” (Patton, 2002). As the research question is to investigate sovereign rating analysts’ practices, analysts at sovereign departments of CRAs were the natural candidate.

With a few rating agencies dominating the credit rating industry in the EU, access to four CRAs was obtained. To preserve confidentiality, these CRAs will be referred to as CRAa, CRAb, CRAc, and CRAd. A fifth CRA declined to participate. All CRAs had their headquarters of European operations in London at the time interviews were conducted. Two of them can be characterized as “large”, while two can be characterized as “small” based on their market shares at end-2019 (ESMA, 2019). All agencies issue rating on issuers belonging to all asset classes – sovereigns, financial and nonfinancial corporations, and structured finance. This fact ensures that the chosen CRAs have extensive coverage and understanding of the rating industry (ESMA, 2019).

A total of 27 interviews were conducted. The “sample size” was determined by theoretical saturation. According to Glaser and Strauss, theoretical saturation is reached when “no additional data are being found whereby the sociologist can develop properties of the category” (Glaser and Strauss, 1967: 61). Theoretically important categories were partly elaborated in advance of interviews based on the conceptual framework introduced in the previous section. In turn, the novel

and relatively unexplored context made it probable that interviews would reveal important themes and concepts not captured by the conceptual framework. Accordingly, interviews were conducted in several waves over the period 2016-2019 in order to permit iteration between the research question and the interview material (Flick, 2009). Each interview lasted on average 71 minutes. Except for five interviews with sovereign analysts at CRAb – during which only hand notes were permitted - all interviews were tape recorded and transcribed. Note that the number of interviews is comparable with prior work on analysts (e.g., Abhayawansa et al., 2015; Barker et al., 2012; Blomberg, 2016; Johansson, 2007).

4.2. Selection of interviewees

The second phase of the qualitative research design is the selection of interviewees. The majority of interviewees were purposefully chosen to be analysts who have (or had) primary responsibility over a portfolio of sovereign governments and voting rights during rating committees. The analysts in charge of the portfolio being the ones who propose a credit rating and the rating committee being the forum where these valuations are discussed, debated and voted on, the selection of the interviewees fits the research question. Senior staff without involvement in analytical matters but with involvement in the development or validation of methodologies, as well as with regulatory affairs, was also interviewed. These interviews allowed to gain familiarity with the (now regulated) industry of credit ratings, to learn about its evolution to the present day, and to obtain a non-analyst view of the expectations that CRAs place on analysts. Because interviewees could assist in answering the research question, they are “experts” (Flick, 2009; Meuser and Nagel, 2009). Table 1 summarizes the interviewees’ profile.

4.3. Other data sources

Additional sources of data included archival material. This consisted of fourteen documents released by CRAs, including rating methodologies, rating criteria, default studies and explanatory documents on the rating process. The methodologies consulted were those current at end-2016, but earlier

methodologies were consulted to identify significant changes. Rating reports and press releases issued by all major CRAs as well as news coverage were also consulted in order to gain familiarity with the industry. In addition, twenty-seven documents issued by EU institutions were consulted. These included legislation that delineates public expectations concerning the activities of CRAs with specific focus on the transparency and accountability of the process of sovereign ratings, the use of ratings by market participants, the penalties for deviations from regulations, ESMA's guidelines regarding the validation of rating methodologies, and responses by CRAs to ESMA on the appropriateness of regulatory guidelines.¹⁶ Table 2 lists the archival material.

4.4. Interview protocol

A semi-structured approach to the interviews was adopted. The semi-structured approach allows interviewees' views to shape researchers' understanding of the phenomenon under analysis, particularly when the field is relatively understudied (Barker et al., 2012). The black box of rating agencies is such that terms, expressions, and themes relevant to the daily work of analysts often required elaboration during the interview. The flexibility of the semi-structured approach is to permit such elaborations through the development of a flexible interview protocol. At the same time, the protocol gives structure to interviews conducted under time constraints and with little chance of repeated access to the field, which guarantees that the major themes are covered in each interview (Qu and Dumay, 2011). Finally, the interview protocol serves to show the experts advance preparation as these are more likely to be insightful if the researcher is seen as competent (Flick, 2009; Meuser and Nagel, 2009).

¹⁶ ESMA is the "European Securities and Markets Authority" and it is the regulator of CRAs in the EU. See [Appendix D](#).

Table 1. Interviewees' Profile

Interview ID	CRA ID	Position held (Analyst/Other)	No of interviews
1&4	A	Analyst	2
2	A	Analyst & Other	1
3	A	Analyst	1
5	A	Analyst	1
6	A	Analyst	1
7	A	Analyst	1
8	A	Analyst	1
9	A	Analyst	1
10	A	Other	1
11	A	Other	1
1	B	Other	2
1	B	Other	2
2	B	Analyst	2
3	B	Analyst	1
4	B	Other	1
5	B	Analyst	1
1	C	Analyst	1
2	C	Former Analyst	1
1	D	Analyst	1
2	D	Analyst	1
3	D	Analyst	1
4	D	Analyst	1
5	D	Analyst	1

This table shows the interviewees' profile and the number of interviews held at each agency and with each interviewee.

Table 2. List of archival material

#	Author	Title
1	EU	Regulation (EC) No 1060/2009 of the European Parliament and of the Council
2	EC	Public Consultation on Credit Rating Agencies
3	EU	Regulation (EU) No 1095/2010 of the European Parliament and of the Council
4	EU	Regulation (EU) No 513/2011 of the European Parliament and of the Council Proposal for a Regulation of the European Parliament and of the Council amending
5	EC	Regulation (EC) No 1060/2009 on credit rating agencies
6	EC	Commission Delegated Regulation No 447/2012
7	EU	Regulation (EU) No 462/2013 of the European Parliament and of the Council Sovereign ratings investigation: ESMA's assessment of governance, conflicts of interest, resourcing adequacy and confidentiality controls
8	ESMA	Sovereign Rating Criteria
9	Fitch	Guide to Credit Rating Essentials
10	S&P	How We Rate Sovereign
11	S&P	2014 Annual Sovereign Default Study and Rating Transitions
12	S&P	Rating Sovereign Governments
13	DBRS	On the validation and review of Credit Rating Agencies' methodologies
14	ESMA	Sovereign Bond Ratings
15	Moody's	Implementation of the Regulation (EU) No 462/2013 on Credit Rating Agencies
16	ESMA	What are Rating Criteria
17	S&P	Principles of Credit Rating
18	S&P	Sovereign Rating Methodology
19	S&P	Study on the Feasibility of Alternatives to Credit Ratings
20	EC	EU supervision of credit rating agencies - well established but not yet fully effective
21	ECA	Study on the State of the Credit Rating Market
22	EC	Discussion Paper on the validation and review of Credit Rating Agencies' methodologies
23	ESMA	Comments on ESMA Discussion Paper on the validation and review of credit rating agencies' methodologies
24	Moody's	2015 Annual Sovereign Default Study and Rating Transitions
25	S&P	Guidelines on the validation and review of Credit Rating Agencies' methodologies
26	ESMA	On the Discussion Paper on the validation and review of Credit Rating Agencies' methodologies
27	ESMA	Sovereign Rating Criteria
28	Fitch	Proposed Refinements to Sovereign Bond Ratings Methodology
29	Moody's	S&P Global Ratings Definition
30	S&P	Comments on ESMA Consultation Paper on the Guidelines on the validation and review of credit rating agencies' methodologies
31	Moody's	Rating Sovereign Governments
32	DBRS	Guidelines on the validation and review of Credit Rating Agencies' methodologies
33	ESMA	Competition and choice in the credit rating industry Market share calculation required by Article 8d of Regulation 1060/2009 on Credit Rating Agencies as amended
34	ESMA	Sovereign Bond Ratings
35	Moody's	Box 3 Eurosystem publishes more detailed criteria for accepting rating agencies
36	ECB	Reply to ESMA Discussion Paper on the validation and review of Credit Rating Agencies methodologies
37	DBRS	Reply to ESMA Discussion Paper on the validation and review of Credit Rating Agencies methodologies
38	Fitch	Reply to ESMA Discussion Paper on the validation and review of Credit Rating Agencies methodologies
39	Moody's	Reply to ESMA Consultation Paper on the validation and review of Credit Rating Agencies methodologies
40	S&P	Reply to ESMA Consultation Paper on the validation and review of Credit Rating Agencies methodologies
41	Fitch	Reply to ESMA Consultation Paper on the validation and review of Credit Rating Agencies methodologies

This table lists the type and source of archival material consulted.

Before the first set of interviews, a pilot interview was conducted that is not included in the final sample. As the next interviews took place, the attempt was made to gain familiarity with analysts' work. The first set of questions asked the interviewee to outline his/her educational and professional background. The second set of questions, following Blomberg (2016), asked interviewees to narrate their typical working day. The goal was to elucidate how analysts conceive of their own work and thus prioritize analysts' "own lived experiences" (ibid., 284). Typically, the next set of questions followed up on a comment made by analysts in their description of the job. Unlike Blomberg (2016), however, the questions did not seek to identify different "kinds" of analysts. Rather, the concern was with valuation practices and the role of judgment in ratings. In particular, the questions focused on how analysts rationalize the proper place for judgment in the rating analysis. Often, the questions led analysts to tell concrete examples of circumstances in which the application of judgment led to a deviation from the rating model to achieve greater accuracy. A more "analytical" set of questions (Kreiner and Mouritsen, 2005) asked analysts to hypothesize counterfactual scenarios in which judgment would not be necessary. This last set of question forms the basis of the theorizing of the empirical material discussed in the last section.

4.5. Data analysis

Similar to Imam and Spence (2016), the initial data analysis method chosen was to highlight themes that were common to the interviews. In the first stage of interviews, the theorizing was left relatively unspecified while the conceptual framework provided indication as to some theoretically-important themes derived from prior literature on analysts' valuation practices. These were:

1. *Automated valuation processes*
2. *Creation of valuation models*
3. *Deviation from model-implied ratings*
4. *Published credit rating*

Excerpts from the interview transcripts were assigned one of the codes of the list. In addition, excerpts from transcripts that did not fall into the above list of

codes and that were frequently mentioned by interviewees were coded separately. As the transcripts were analyzed, some of these excerpts were assigned codes based on concepts derived from prior literature on commensuration and ranking practices. These codes were:

5. *Analysts choose what is valuable.* Interviewees commonly referred to the “theory” and “evidence” that inspires the choice of the indicators underpinning rating models. This theme connects to the literature on valuation and classification that indicates that the selection of valuation inputs includes a normative moment: “Forms of quantification never simply reflect the world. Instead, they are the product of a series of interpretive decisions about what to quantify, how to categorize, and how to label it.” (Kornberger et al., 2015; Rottenburg et al., 2015: 11)
6. *Administrative replication.* Interviewees often referred to the importance of ratings being “consistent”. Building on Power (2004), administrative replication captures the notion that automated valuation exercises can be replicated in different contexts and by different people. They thus have “consistency”.
7. *The accuracy of mechanical objectivity.* A common theme raised by interviewees related to the accuracy of credit ratings produced by automated valuation models. This theme relates to Porter (1994), who suggests that measurement systems based on “mechanical objectivity” are often preferred although they are less accurate: “There is a strong incentive to prefer readily standardizable measures to highly accurate ones, where these ideals are in conflict” (ibid., 391).

Additional themes that often emerged from the material are: the intertwining of indicators (Schultz et al., 2001); commensuration and comparability (Espeland and Stevens, 1998); perception and intuition (Daston and Galison, 2010); accounting numbers, intelligence and information gathering (Knorr Cetina, 2011); indicators as tools; categorization, encoding and statistical associations (Desrosieres, 1998, 2001); uncertainty (Knight, 1964), trust in models (Porter, 1995); and small sample.

5. Findings

5.1. Automated processes: the models for constructing sovereign ratings

References to the automated valuation process that underlies a sovereign rating are found both in the archival material consulted and in the interview transcripts. Within the archival material, the process is outlined in the rating methodologies that are regularly published by CRAs. Methodologies go by slightly different names in the industry. S&P speaks of the “Sovereign Rating Methodology” (S&P, 2014a). Moody’s refers to it as “Rating Methodology: Sovereign Bond Ratings” (Moody’s, 2016), while Fitch speaks of “Sovereign Rating Criteria” (Fitch, 2016b). The purpose of methodologies is to illustrate: (a) the analytical pillars that CRAs consider to be related to future probabilities of default of governments; and (b) the weights attached to each pillar. These two elements are commonly referred to as a “scorecard”.

The analytical pillars contained in the scorecard reflect some broadly-defined dimensions that, according to CRAs, affect the willingness and ability that a government will default in the future. Typically, these pillars include aspects of the governments’ own financial performance and position, macroeconomic and financial prospects and institutional/political dynamics. The analysis of the archival material of the major CRAs indicates that sovereign ratings are based on four to six dimensions:

- The quality, predictability, and transparency of institutional and political structures
 - The structure and the prospects of the economy
 - The competitiveness of the economy vis-à-vis the rest of the world
 - The past and the future financial conditions of the government
 - The degree of monetary and exchange-rate flexibility
 - The exposure of the economy to exogenous shocks

The heading given to each pillar varies across CRAs: for example, S&P speaks of the combination of 1 and 2 above as the “institutional and economic profile of a country” which reflects the CRA’s assessment “of the resilience of a country’s

economy, the strength and stability of its civil institutions, and the effectiveness of its policymaking” (S&P, 2014a); it echoes the pillar known at Moody’s as “economic resiliency”, which combines the assessment of the country’s economic and institutional strength into one dimension (Moody’s, 2016). At Fitch, “macroeconomic performance, policies and prospects” is the dimension that summarizes the volatility of the growth rate of the economy as well as its inflation dynamics, being very close to DBRS’s notion of “economic structure and performance” (DBRS, 2016; Fitch, 2016b). These pillars are the headings under which CRAs select a set of mostly quantitative indicators to “measure” the different facets of default risk. The indicators chosen by the four largest agencies operating in the EU are listed in Table 3.

As one interviewee stated, many of the inputs underpinning the sovereign rating methodology are:

“Macro. All macro. Yeah. Then there’re other aspects. There’s the banking sector. There’s the monetary, liquidity, and all these. They’re all factual. They’re just simply the numbers.” (CRAa.1).

Many indicators are “just simply the numbers”, reflecting analysts’ understanding of the automated, mechanical stage of this moment of the valuation process. These numbers are then “commensurated” (Espeland and Stevens, 1998) to generate a score. In turn, the scores are commensurated through a system of weights.

Table 3. The models for rating sovereign governments

Rating pillar	Indicator(s) used
S&P	
1. Degree of effectiveness, stability and predictability of policy-making process	n.a.
2. Income levels, growth prospects and economic diversity and volatility	(1) GDP per capita; (2) average real growth rate of GDP per capita; (3) share of GDP accounted for by the major economic sector
3. External liquidity, external debt	(1) Net external debt/Current Account Receipts; (2) Gross External Financing Needs/Current Account Receipts plus usable reserves; (3) Percentage of global foreign currency reserves denominated in the country's currency
4. Fiscal performance	(1) Change in gross government debt as percentage of GDP; (2) Percentage of liquid assets available for withdrawal; (3) Interest expenditures as percentage of revenues; (4) Net government debt as percentage of GDP; (5) Contingent liabilities (banks' assets as percentage of GDP)
5. Credibility and effectiveness of monetary policy, development of financial and capital markets	(1) Degree of independence of central bank and clarity over its goals; (2) Wide array of monetary instruments; (3) Ability to act as lender of last resort; (4) Low inflation; (5) Large amount of deposits in banks and of bond instruments traded in capital markets
Moody's	
1. Growth and scale of the economy	(1) Real GDP growth; (2) Volatility of real GDP growth; (3) Nominal GDP; (4) GDP per capita; (5) Indicator for credit boom; (6) Competitiveness index
2. Institutional framework and effectiveness	(1) Indices of government effectiveness, rule of law, and corruption; (2) Inflation level and volatility; (3) History of default
3. Debt burden	(1) Government debt as percentage of GDP; (2) Trend in government debt; (3) Share of government debt denominated in foreign currency; (4) Other public sector debt; (5) Public sector assets
4. Political risk, liquidity risk, banking sector risk	(1) Current account balance plus FDI as share of GDP; (2) International Investment Position as share of GDP; (3) Domestic political risk and geopolitical risk; (4) Government liquidity risk and market stress indicators; (5) Contingent liabilities (Size of banking system)
Fitch	
1. Governance quality, wealth and flexibility of the economy	(1) World Bank governance indicators; (2) GDP per capita; (3) Share of world GDP; (4) Years since the last default; (5) Amount of credit in the economy
2. Macroeconomic performance and policy credibility	(1) Real GDP growth volatility; (2) Inflation rate; (3) Real GDP growth
3. Government financial performance and position	(1) Government debt as percentage of GDP; (2) Government interest expenses as percentage of revenues; (3) Government fiscal balance as percentage of GDP; (4) Share of government debt denominated in foreign currency
4. Financial performance and position vis-à-vis non-resident lenders	(1) Sovereign foreign assets as percentage of GDP; (2) Foreign exchange reserves; (3) Current Account Balance plus FDI as percentage of GDP; (4) Reserve currency indicator; (5) Commodity dependence
DBRS	
1. Overall fiscal performance	(1) Fiscal balance as percentage of GDP
2. Debt stock and maturity structure	(1) Government gross debt; (2) Average maturity of public sector debt; (3) Change in debt associated with adverse shock
3. Economic growth and productivity	(1) Growth in real GDP per capita; (2) Level of GDP per capita in US\$; (3) standard deviation of GDP growth rate; (4) Size of the economy and diversification of trade
4. Policy credibility	(1) Inflation; (2) Total deposits held by households and firms; (3) Change in government bond yields during slowdowns
5. External imbalances	(1) Current account balance as percentage of GDP; (2) Net International Investment Position as percentage of GDP; (3) Exchange rate flexibility; (4) Reserve currency status
6. Institutional environment	(1) World Bank indices of Voice and Accountability and Rule of Law

This table lists the indicators used in the automated model of the four major credit rating agencies operating in the European Union as described in their methodologies (DBRS, 2016; Fitch, 2016b; Moody's, 2016; S&P, 2014a).

Figure 2 portrays the way in which at S&P, the scores of two pillars are combined to obtain an indicative rating (S&P, 2014a). The first score (Institutional and economic profile) represents S&P's assessment of the ability of national institutions and the effectiveness of policy-making to promote sustainable public finances, combined with the country's economic prospects. The second score (Flexibility and performance profile) is a summary measure of the country competitiveness vis-à-vis the rest of the world, the quality of its public finances, and the effectiveness of monetary policy in the country. The figure shows that scores on individual pillars are combined to obtain an indicative rating. Through the commensuration of these scores, rating models construct equivalent values. At S&P, for example, a score ranging from 1 to 1.7 on the Flexibility and Performance Profile can be combined to a score of 2.5 on the Institutional and Economic profile to generate an indicative rating of "aa+". Now suppose that the Flexibility and Performance Profile score drops to the range 1.8-2.2; then, an "aa+" indicative rating can still be generated if the Institutional and Economic profile obtains a score of 2. At the same time, compensation is not always allowed. For example, an "extremely weak" institutional and economic profile at S&P cannot be combined with *any* assessment of flexibility and performance, resulting in "N/A". Thus, the weighting process appears to represent not only a commensuration of different dimension. It also seems to create a hierarchy of values between the building blocks. Weak scores on one dimension are not always related balanced by scores on other dimensions.

According to interviewees, this fact highlights a relatively higher importance of some pillars compared to others.

"From my career I know there are certain statistics that sort of have been good predictors, for example private sector credit growth, as the sort of fundamental predictor of potential crisis." (CRAa.1)

"Usually when you have a sovereign default, it's preceded by a balance of payments crisis. I don't actually think that there's been a single default of a surplus country." (CRAc.1)

Figure 2. Commensuration of analytical rating pillars at S&P

Table 2 Indicative Rating Levels From The Combination Of The Institutional And Economic Profile With The Flexibility And Performance Profile												
Institutional and economic profile												
Flexibility and performance profile	Category	Superior	Extremely strong	Very strong	Strong	Moderately strong	Intermediate	Moderately weak	Weak	Very weak	Extremely weak	Poor
Category	Assessment	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Extremely strong	1 to 1.7	aaa	aaa	aaa	aa+	aa	a+	a	a-	bbb+	N/A	N/A
Very strong	1.8 to 2.2	aaa	aaa	aa+	aa	aa-	a	a-	bbb+	bbb	bb+	bb-
Strong	2.3 to 2.7	aaa	aa+	aa	aa-	a	a-	bbb+	bbb	bb+	bb	b+
Moderately strong	2.8 to 3.2	aa+	aa	aa-	a+	a-	bbb	bbb-	bb+	bb	bb-	b+
Intermediate	3.3 to 3.7	aa	aa-	a+	a	bbb+	bbb-	bb+	bb	bb-	b+	b
Moderately weak	3.8 to 4.2	aa-	a+	a	bbb+	bbb	bb+	bb	bb-	b+	b	b
Weak	4.3 to 4.7	a	a-	bbb+	bbb	bb+	bb	bb-	b+	b	b-	b-
Very weak	4.8 to 5.2	N/A	bbb	bbb-	bb+	bb	bb-	b+	b	b	b-	b-
Extremely weak	5.3 to 6	N/A	bb+	bb	bb-	b+	b	b	b-	b-	'b-' and below	'b-' and below

Assigning 'CCC+', 'CCC', 'CCC-', and 'CC' ratings is based on "Criteria For Assigning 'CCC+', 'CCC', 'CCC-', And 'CC' Ratings," Oct. 1, 2012.

This figure shows how commensuration of analytical pillars (building blocks) takes place at Standard and Poor's. Source is the methodology for rating sovereign governments (S&P, 2014a).

The analytical pillars, the numerical indicators and the weights are combined to generate a model-implied ratings that represent CRAs' attempt to establish a certain degree of consistency in thinking about credit risk. Analysts are expected to adhere to the results of the scorecard as much as possible. According to the archival material consulted, S&P states that analyst-adjusted ratings should differ by indicative ratings by no more than one notch, save for exceptional circumstances (S&P, 2014a). Fitch states that the "qualitative overlay" should be limited to at most three notches (Fitch, 2016b).

Interviewer: How much do rating analysts at your CRA rely on the models, like the methodologies, for sovereign rating?

Interviewee: How much we rely on the methodology?

Interviewer: Yes.

Interviewee: Entirely. (CRAc.1)

The role of scorecards is thus to ensure that the rating process is automated to an extent and based on "objective" criteria – such as a transparent list of indicators. An interviewee explained what this entails.

"Investors around the world want [CRAb] to have an opinion on the risk of a bond, the risk that someone will not repay, and they want us to have a similar framework of thinking about this and a similar meaning to our rating whether it's a bond issued in China, the US, wherever." (CRAb.1).

Through methodologies, CRAs attempt to instill the mentality of "administrative replication" conceptualized by Power (2004). That ratings must be globally consistent and based on shared frameworks of thinking is an aspect not brought to light by prior research. The material indicates that the "administrative replication" of ratings represents an important element of CRAs' public discourse. Independence and uniformity are crucial because credit ratings are issued by the CRA rather than by the individual analyst. Indeed, analysts' subjective idiosyncrasies is downplayed in favor of a global, "shared" view of credit risk. The goal as much as possible is to avoid analysts to have an undue influence on the process by which ratings are produced. The same interviewee explained that, over the years, the construction of methodologies has shifted from being the result of an individual decision-making process to a CRA-level one, tracing an

interesting parallel to the same dynamic of institutionalization documented by Pollock and D'Adderio (2012) in an altogether different ranking organization:

“Originally the analyst used to be the lead on the methodology and then you’d have a body that would then review it and make sure that it made sense. We were more fragmented in our methodologies. We now have a (...) team but that’s not just one person sitting in a black box sort of coming out with a great thought. Then they bring in the analytical teams and then there’s a decision-making process.” (CRAb.1)

At the same time, the quote speaks of CRAs issuing an “opinion” on default risk. The choice of the term is not idiosyncratic to the interviewee. For example, S&P clarifies that ratings express “the agency’s opinion about the ability and *willingness* of an issuer, such as a corporation or state or city government, to meet its financial obligations in full and on time.” (S&P, 2014b: 2, emphasis added). Similar definitions are found across the industry (Moody’s, 2020; e.g., Scope Ratings, 2019b). Indeed, White (2010) notes that “opinion” is a favorite word in the CRAs industry. More importantly perhaps, EU regulations establish the nature of ratings to that of “opinions” in Article 3 of the regulation on credit rating agencies (EU, 2009b, known as ‘CRA I’): “‘credit rating’ means an opinion regarding the creditworthiness of an entity, a debt or financial obligation, debt security, preferred share or other financial instrument, or of an issuer of such a debt or financial obligation, debt security, preferred share or other financial instrument, issued using and established and defined ranking system of rating categories.”. Yet, if ratings are opinions, then their epistemic standing differs from that of objective “truths”. Indeed, if ratings are opinions, they cannot be “objective opinions”. Instead, they must have a *subjective* component. An interviewee tried to explain.

“The fundamental – the Gordian knot of the rating agencies – is that rating agencies issue subjective opinions. They essentially operate in a very kind of small niche of macro-economics and so the same way macroeconomists are completely incapable of telling you what GDP growth is going to be next year, rating agencies are not capable of getting it systematically right on who’s going to default and who’s not going to default.” (CRAa.11)

This quote connects the nature of ratings as opinions with the ability to accurately predict default: ratings are opinions because they may be inaccurate

and thus do not represent statements of facts. Yet, ratings must be consistent, and subjectivity is downplayed by CRAs' strive towards consistency. Behind the veil of what appeared to be an unproblematic description of what ratings are, a tension begins to appear between the characterization of ratings as "subjective opinions" and the ambition/pretense of an objective and "shared" epistemology of ratings based on consistency, which almost hints at credit ratings as a "view from nowhere" (Nagel, 1986; Samiolo, 2012). Indeed, these quotes demonstrate the inherent tension between objectivity and subjectivity in the construction of credit ratings. Analysts estimate the risk that a government will not repay its obligations because this is what investors expect them to do. This risk, however, is calculated and framed as a globally consistent measure. Risk assessments are framed as opinions – yet they must derive from a shared way of thinking about risk and they must be independent of the site of their production. This gives rise to a tension between objectivity and subjectivity – between CRAs' global aspirations and analysts' individual judgements. In what follows, this tension is explored.

5.2. Judgment in the creation of rating models

At stake in the selection of pillars, indicators and weight – i.e. in the construction of the scorecard - is the possibility of knowing and concretely, valuing a government. In this sense, the construction of methodologies sets in motion a valuation practice (Dewey, 1943; Muniesa, 2011). As pillars and measures are included in methodologies, others are excluded. In selecting dimensions and measures, the methodologies of CRAs construct the "valuable" government as one that possesses certain traits, perhaps at the expense of others. This is an exercise that can be fiercely contested. For instance, in 2014 the Italian Supreme Audit Institution notified S&P of its intention to begin a €234 billion lawsuit against the agency because it "never in its ratings pointed to Italy's history, art or landscape which, are universally regarded as the basis of its economic strength". S&P discharged the allegation as "frivolous and without

merit”.¹⁷ How do analysts select the pillars that underpin their automated rating models?

“We all have views on what would be more effective at predicting the future, and you know, and well...being flippant about it, you know, you could say that the Chief Executive of a company that wears a red tie consistently, is a danger for credit risk. [Laughter] That could be a valid methodology. I think we’d have a few problems from a regulatory point of view. So, it is about saying, ‘Look. Out of all this information’, you know, we’re not auditors of information, we’re not users of all the information out there, but we have got to say, ‘If we’re in this sector trying to predict the risk that an investor’s going to lose money on investing in this bond, what are the issues that we see as the fundamental drivers?’” (CRAb.1)

The above quote shows that pillars are chosen insofar as they are “indicators” (Porter, 2015) of credit risk. Consider the role that institutions play in determining credit risk. According to the archival material, CRAs consider institutional quality as an indicator of lower default risk. S&P, for instance, values “highly effective” policy-making because it “enables governments to address periods of economic distress”, and penalizes ineffective policy-making as a credit-negative, representing one of the “factors that can pose risks to institutional stability and, in turn, lead to substantial policy shifts” (S&P, 2014a). Similarly, Fitch states, “Governance indicators seek the capture (...) the effectiveness of governments and institutions in managing economic activity and absorbing adverse shocks.” (Fitch, 2016a: 12). In this quote, we see that institutional quality is mobilized as an attribute of a government. It can be high or low: if it is high, the probability of “absorbing adverse shocks” in the future is equally high and the risk of default is correspondingly low: institutional quality reduces the probability of default by giving the government a greater ability to react to shocks that might or might not take place. Conversely, low institutional quality weakens the prospects of absorbing external shocks, which increases the probability of default. How is the existence of this relation assessed?

“[In] a country that performs weakly on some governance indicators or has poor or weak institutions there just might be a less efficient deployment of

¹⁷ See “Italy accuses S&P of not getting ‘la dolce vita’” (4 February 2014) by Stephen Foley and Guy Dinmore on ft.com. Retrievable at <https://www.ft.com/content/6ed9649e-8dab-11e3-bbe7-00144feab7de>. (Last access: 2 February 2020)

resources, bot capital and labor. And so, those that are in charge of the resources—their motivation might be more parochial (...) That means they'll build a bridge somewhere and nowhere or they'll build a whole town sometimes and now that town's empty. And that comes with a cost, and that's why you see the debt stock in aggregate—not necessarily in the public sector—but total debt has increased so much and it's just inefficient. So, I would say you see great inefficiencies when there are some questions around institutional strengths.” (CRAa.7)

Having weak institutions is costly to the government finances. It is costly today, as resources are mobilized “to build a bridge somewhere”. It will be costly tomorrow, as “that town’s now empty”. While the financial reporting system of governments might not recognize these probable outlays as liabilities today, they will likely contribute to larger deficits at some point in the future.¹⁸ But now consider the case of country corruption. In CRAs’ methodologies, high levels of corruption are considered as an indicator of lower institutional quality and greater default risk (e.g., Moody’s, 2016; Scope Ratings, 2019a). Yet, there are those who have differing views.

“So, corruption. Yeah, it’s a problem and it can hamper growth etc., or rule of law or, you know, [it] avoids accountability and all this kind of stuff. It’s all good things, you’re going to have that, but is it really determining directly the ability and the willingness of a government to pay? In some cases, in others it does not. Now an example I’d like to give here is Romania under Ceausescu. I mean we didn’t have the government indicators then, but they probably would have been very, very bad. (...) But the man, however mad he was, he said: ‘I want to be debt-free. I want to pay back the debt. It may be a matter of people starve for it – so be it.’ So, you have to draw the balance here somewhere between what you think is positively relevant for the question at hand and what do you normatively or ethically want to see for society to represent.” (CRAc.1)

It is not so much that corruption is not “a problem”. It certainly is. Yet, it does not always affect default risk in the mind of analysts – “in some cases” it does, in others “it does not”. These differences are clearly borne out by an exercise of judgement by analysts.

Interviewer: “But it’s normally quite a fine line, isn’t it, to distinguish between what is qualitative and where judgement is?”

Interviewee: “Yeah, because when you do the scorecard, you are judging which variables and which timeframe.” (CRAa.9)

¹⁸ In practice, accounting systems recognize these liabilities in a variety of ways. See Heald and Hodges (2018).

It would be incorrect to assume that models do not include a judgmental component. Rather, the evidence presented here indicates that the “automated” models that lead to what the economics literature considers to be the “objective”, model-implied ratings result from the application of judgment. It is instructive to see how CRAs defend the selection of indicators. Consider how Moody’s justifies the inclusion of economic indicators in its rating model.

“In terms of the historical importance of Economic Strength and the correlation with credit stress, our research on 29 sovereign defaults between 1997 and 2012 indicates that long-term economic stagnation was the principal underlying cause of default in 10% of cases and a contributing factor in many others.” (Moody’s, 2016: 8)

Moody’s attempts to identify the fundamental drivers of creditworthiness through statistical analysis. In this way, a government that faces “long-term economic stagnation” can be assigned to the category of low-quality borrowers. Yet, Moody’s acknowledges that between 1997 and 2012 only 29 defaults took place, a limited sample. What do analysts make of the statistical justification of the inclusion of certain indicators in methodologies? An interviewee remarked that while CRAa does carry out statistical tests, it refuses to rely entirely on them in the case of governments.

“We’ve taken inputs from a variety of sources, looked at it [sic]. We’ve done statistical tests as much as we can within the limited pool of sovereign defaults. We’ve used a range of different tests to try to conclude what are the best metrics that you can use and I think what we’ve found is that there are some that do a pretty good job and get you, say, 80-90 percent of the way there, but inevitably there are, at times, accounting differences between different countries which cannot be fully reconciled, and in other cases just the indicators themselves work for 80 percent of your countries but not the other 20 percent.” (CRAa.3)

Statistical tests are conducted “as much as we can”. The pool of sovereign governments that have defaulted in recent years is limited. Sometimes, the indicators do not “work”. The interviewees were then asked for counterfactual examples: what would be needed for sovereign analysts to embrace more quantitative methods of choosing rating indicators?

“If you’re dealing with a retail bank, at least at a given point in time in a given economic environment, you’ve got so many people you can start doing a lot of

statistical work on it and get a good idea of markers of better or worse performance. That's where it doesn't apply to rating agencies: they don't deal in thousands and thousands and thousands of little credits; they deal in a relatively finite number and the problem is going through time.” (CRAa.11)

The limited availability of default episodes is a challenge against embracing a fully statistically grounded culture of inference (Mikes, 2011). Interviewees indicate that CRAs' confidence in such methods may increase, *if only* analysts could run such analysis on a larger set of government defaults. The same interviewee attempted to qualify the epistemic qualities of rating models.

“We know that finance is essentially the result of the interaction of human beings; that most processes that result from interaction of human beings don't follow steady...sort of you know Gaussian distributions and things like that; that the world's very unlikely to be some linear combination of nicely observable parameters. So, we all know that these kinds of score cards and things like that are very simplistic representations in reality: [they are] models in the sense of what a model is, which is a simplification of reality which is only useful insofar as it serves a purpose. But, no one's pretending that's – I mean unless you're doing physics, but here where in a social science effectively – no one's pretending that it represents some sort of rules of nature.” (CRAa.11)

No analyst pretends that models represent “some sort of rules of nature”. But then, to what extent do analysts trust the ratings that result from the mechanical application of the CRA model? The next sub-section is devoted to examining this question.

5.3. Judgment in the departure from rating models

“As you have probably seen, we have this quantitative approach which we call a scorecard and you receive first of all outcome of this work, which places the sovereign in a certain area across all the sovereigns. So, you have kind of a starting point when you start with the analysis, and you can then dig deeper into the analysis.” (CRAAd.3)

According to interviewees, there is a sense in which the automated valuation is respected by analysts. For example, interviewees state that financial numbers and other indicators are included in the scorecards without any adjustment by the analyst. Analysts' adjustments take place only *after* the model has generated its results.

Interviewee: "The data goes into the machine cleanly."

Interviewer: "Just as is?"

Interviewee: "As is. So, in the model China's unemployment rate is 4.3, just as it was for the last half decade."

Interviewer: "Even if highly suspicious, the figures go in as they are?"

Interviewee: It's in and then if we think that they're suspicious we'll make an adjustment ex-post. (CRAa.7)

"The scorecard provides an indicative factor score. The rating I recommend is the scorecard rating plus my professional judgment. Then, it is subjected to scrutiny in the Rating Committee meeting to produce the final rating that we publish. (CRAb.5)

The outcome of the model is an indicative rating – a "starting point" - which is only the beginning of analysts' activity. The indicative rating generated through the mechanical objectivity of a scorecard is only provisional. It is "the infrastructure on which judgment would rest." (Daston and Galison, 2010: 329). The question is why analysts would ever change the results of the model. One interviewee explained what this entail.

Interviewer: "So you could put all the figures, quantitative, qualitative converted into numbers, into the score card and say it's, I don't know, BBB for the sake of argument. And then you say, 'I'm not going to recommend this. I'm going to recommend BBB+?"

Interviewee: "That's right."

Interviewer: "And why would the analyst change the rating suggested by the score card?"

Interviewee: "Well because, going back to this thing again, for certain sovereigns, although this scorecard has been strongly back-tested, no scorecard can give a perfect indication of how important certain of these factors are to a specific sovereign. So the lead analyst may think for example, looking at this, that institutions in a country and rule of law are so critically important for the wellbeing of that country and the health of that country's trade and investment, but even if the scorecard, based on all the numbers, shows a lower grade, the strength of these institutions and the ability of these institutions to react to any stress implies that, really, perhaps we should have given more weight to that." (CRAa.1)

Analysts deviate from models because models cannot take into account the idiosyncrasies of a government. By aiming for consistency in the rating, models may overweight or underweight the importance of certain indicators. But, analysts do not remain passive. Questions then focused on which areas of the analysis are most often affected by judgment. According to interviewees, a favorite domain

is politics. Politics matter to sovereign rating analysts because the history of sovereign defaults indicates that governments often renege on their financial obligations even in the presence of adequate financial capacity to remain current on debt payments. Indeed, Gaillard (2012) finds that ever since the first manual on “The Credit of Foreign Governments” issued in 1918, Moody’s included the willingness to repay the debt, known as the good faith of the government, as one of the criteria used to rate a sovereign. After a century, the approach is unchanged and an important pillar of creditworthiness remains the assessment of “whether the country’s institutional features are conducive to supporting its ability and willingness to repay its debt.” (Moody’s, 2016: 13). Similarly, at S&P, “the assessment of a governmental obligor focuses on the potential that the obligor might default even when it has the resources to meet its financial commitments.” (S&P, 2011).¹⁹ An analyst confirmed.

*“The default of a sovereign is always a political decision. Always, in the end.”
(CRAc.1)*

If politics matter so much to credit risk in the sovereign space, how do CRAs quantify the “potential” that default decisions will occur? The archival material indicates that at S&P, political scores are “primarily” informed by qualitative analysis (S&P, 2014a). DBRS lists a set of numerical variables that attempt to measure the political environment, yet the agency emphasizes the role of qualitative analysis: “[p]olicy intentions, political and social cohesion and willingness to uphold the sanctity of contracts are difficult to determine by statistics alone, and often require judgement calls to formulate a fair and balanced sovereign credit opinion.” (DBRS, 2016: 11)

“How can you get political [assessments] statistically? I mean...yes, you can get opinion polls on elections. You can get surveys of people’s likes and dislikes, but you really need the human element.” (CRAa.1)

Automated models do not suffice in accurately measuring political dynamics: “you need the human element”. But, precisely the human element would disappear if analysts blindly embraced the results of automated models. An

¹⁹ This theme is developed more fully in Gulati and Scott (2013).

approach to valuation that seeks objective measurement would privilege hard numbers and models. Yet, this approach would fail to capture an important dimension of creditworthiness, i.e. governments' intentions to remain current on their obligations. Hence the defense of judgment in the analysis of politics.

"Just because a country has the capacity on paper - given these assumptions and given these indicators - it does not mean that they have the willingness to do it. And that's why... this is the dismal science staring us in the face! Because if Argentina, even though they have, you know, 5% of GDP in debt and they're sitting on, you know, 30% GDP in reserves, they don't want to pay, they're not going to do it. And so, they default, even though the framework that you would otherwise use has them looking pretty good." (CRAa.7)

Note how this interviewee combines the quantitative and the qualitative, while noticing that the former alone would lead to an inaccurate classification. That is, a government that quantitatively appears to be "pretty good" and thus deserving of a high rating, is really one that should be classified as risky. The interviewee questions the appropriateness of mechanical classifications based on a purely numerical exercise.

Another analyst explained that judgement affects not only the measurement of unquantifiable phenomena, such as political dynamics. It also affects the use that analysts make of numerical information.

Interviewer: "Is it descriptive of the work of an analyst to say that quantitative information is simply taken for what it is, or is there a role for professional judgment also in dealing with quantitative information?"

Interviewee: "There's a certainly a role because again, I always say every model is wrong and some are hopefully useful, but as you know the limits of a quantitative...to understand the limits of a quantitative model you need to have a sense of what the variable tells you or does not tell you." (CRAAd.4)

According to this interviewee, analysts' skepticism of models reflects the awareness that models have an inadequate representational capacity. Rather, they are instrumental to the development of an approximate - albeit provisional - classification. Indeed, the quotes above indicate that the indicative rating generated by automated models is not taken at face value: analysts "understand the limits of a quantitative model" and apply judgment to adjust the indicative

rating and *more accurately* assign a rating to a government. Judgement is necessary to grasp patterns that models are incapable of identifying. An analyst shared the experience of adjusting the indicative rating score.

Interviewee: “[the government] made a really hard choice and I thought that both of those sections were being unfairly critical in the standard model because it was showing lots of volatility. But I think the volatility in part was a function of the government making a hard choice, and so we gave them a notch above what the model would say. So that’s where the human element comes into play. I think that’s a pretty good example.”

Interviewer: “So this is where your expert judgement comes in? To say the scorecard end-result is not right?”

Interviewee: “Right. It’s sort of...you’re interpreting how...what the data is communicating to you through the model. We thought it works in most case, but in this case it is actually penalizing [the country] where I think they should be getting a benefit because their deficit didn’t widen that much and their debt to GDP ratio didn’t increase that much. Because instead of increasing spending as a way of smoothing out consumption, they did the opposite.” (CRAa.7)

In this context, the automated model was penalizing a sovereign for having engaged in austerity measures that contributed to a sharp drop in economic growth and “lots of volatility”. Yet, the drop happened for the “right” reasons, according to this analyst. Judgment – the ability to interpret “what the data is communicating to you *through* the model” eventually reversed the conclusions of the model. Another interviewee explained the role of judgment in a downgrade.

“The headline numbers were looking good, or good enough at least, and the deficits were coming down. They were below 3 percent, and there’s this obsession with Maastricht in Europe and the Middle East specifically. And so, why did we downgrade? The reason is that we thought the policy is going into a dead end. So, what they [the government authorities] were doing is they were artificially gaming the European Accounting Standards by Eurostat by securitizing future flows, which means that you have more money now, but you have less money in the future. Now, this is what I call a loan.” (CRAc.1)

The complementary nature of judgment vis-à-vis models thus becomes a key part of analysts’ epistemology. Sometimes, the ability to interpret becomes critical when numbers are simply missing and when events are highly contextual.

I can give you a concrete example that just came to my mind and then it will be clear. For example, what does the sacking of the central bank governor in

Turkey means for the sovereign rating of Turkey? Of course, there is no data or indicator to quantify this, right? There, of course, the analyst has to assess these events, basically by trying to judge whether...what impact this single event, the sacking of the governor of the central bank, could have on the independence of the central bank, which is key to the sovereign rating. (CRAd.1)

Other times, judgment is necessary because governments go through unprecedented events.

"If we were to speak about [country name] before the elections or the UK post-Brexit, the interpretation of the political situation is much more tricky [sic]. So, when we dig into qualitative there is much more perception and the comparability aspect might not kick in because it's really a unique event. So that's the difficulty obviously: when you speak about the UK that goes through a referendum, that decides to leave the [European] Union, again it's all unprecedented." (CRAa.5)

To sovereign analysts, ratings are at times constructed based on "a sample of one or fewer" (March et al., 1991). The interviewees indicate that confidence in models would increase, *if only* analysts could run such analysis on a larger set of governments.

"If you had a hundred countries, you could say, 'Well eighty of them will go down that route and the other twenty go the other route', then we can do this. Then you can do some sort of weighting, but you only have one. So, I think you really have to come down and say, 'Put your head in your hands. This is the likely outcome.'" (CRAc.1)

The situation that prevents analysts to fully trust the model results is that numbers must be interpreted and contextualized. In addition, models cannot be fully trusted because in an attempt to gain consistency, they discard valuable information. Sometimes, information is not available, or not in qualitative forms. Other times, information is new because events are without precedents. The interviewees also indicate that, *if* the number of cases on which to run a model was larger, a statistical mode of inference could be acceptable. Except, it is not. Thus, the work of analyst acquires an interpretive character.

"Sometimes, people can read fifty things, but the most important pure information can be gleaned out of three documents and a sense of something. Three documents and then some sort of sense of political events and experience of how to interpret that political event." (CRAa.1)

The skill required of analysts is the development of “a sense of something”. This “sense” is the ability to interpret events, numbers, and indicators – to put them in context.

“I think the indicators are tools that we use to craft the narrative and sometimes it’s not necessarily the outcome of the indicator but rather what that indicator is telling you.” (CRAa.7)

Interpretation is most obviously needed when peering into the future, as analysts attempt to formulate forecasts and projections.

“There’s always a degree of interpretation as far as the forecasts are concerned because the forecast that we base our decision on, and of course the forecast is subjective, you know? We cannot be certain. So, the analytical judgement is actually relevant here but it’s not an algorithm where you just put in data and then you turn the handle and the rating comes out the other end.” (CRAc.1)

Yet, interpretation is also needed when confronting economic and financial indicators that, if simply inserted into the model, would lead to inaccurate inferences. Then, analysts attempt to interpret what is behind the number. The complementary nature of judgment vis-à-vis models thus becomes a key part of analysts’ epistemology.

The quantitative model could capture let’s say... if you have a fiscal adjustment the quantitative model will capture an improvement because the fiscal balance is going down. If I then look at the budget and I see public investment cuts, education cuts, but no growth consolidation measures, then obviously I should adjust this qualitatively and say okay, the model is telling me that something is getting better but at the same time qualitatively I know that the measures that are being introduced will adversely affect the growth prospects of this economy. That’s when this interaction comes in and that’s why I do think a thorough analysis needs to look at both. (CRA d.4)

6. Discussion

The remarks shared by analysts during these interviews lead to the insights that judgment is an integral component of the work of sovereign ratings. Judgment is exercised in the creation of automated rating models, in the choice to deviate from model-implied ratings, and in the interpretation of numerical and qualitative information. The reason for this form of “quantitative skepticism” (Mikes, 2011) is two-fold. First, governments seldom default. As a result, it is

challenging to establish what indicators best serve the purpose of alerting CRAs to the dimensions that are relevant to assessing default risk. Judgment is therefore unavoidable because the “sample” of sovereign defaults is too small to pretend that rating models reflect the working of economic laws or of a theory of sovereign defaults. Second, data must be interpreted because the same data may indicate that the risk of default has increased or decreased in light of available evidence. Accordingly, the role of rating models is primarily to ensure consistency in the application of judgment. Conversely, the role of models is not to eliminate judgment altogether, as models are themselves the product of analysts’ judgment.

I theorize these findings by referring to the distinction between “risk” and “uncertainty” proposed by the Chicago economist Frank Knight in his 1921 book “Risk, Uncertainty and Profit (Knight, 1964 [1921]). Knight’s distinction between risk and uncertainty has been interpreted in a variety of ways and used in several different contexts. Not all these interpretations can coexist. Indeed, Knight is considered an eclectic writer, which presents “peculiar difficulties to critics and interpreters” (McKinney, 1977: 1438). For example, LeRoy and Singell (1987) speak of outcomes that are insurable as those that allow for the operation of the law of large numbers, and uninsurable outcomes those that do not. The first outcomes raise (insurable) risks to actors, while the second ones raise uncertainty. According to Khalil (1997), uncertainty describes situations in which an event can evolve a priori into any of several, potentially plausible outcomes. Power (2004) relates the distinction between risk and uncertainty to the distinction between the measurable and the unmeasurable. Mousavi and Gigerenzer (2014) define “risk” the situation in which both outcomes and probabilities are known and “uncertainty” that where all outcomes but not all probabilities are known. They further define “fundamental uncertainty” the situation characterized by imperfect knowledge of both outcomes and probabilities. In accounting and finance, the most popular distinction between risk and uncertainty does have to do with the possibility of assigning probabilities. For example, Knights and Vurdubakis (1993) consider risk the situations that allow for probability calculations, and uncertainty the situations that prevent them (cf.

Karpik, 2010; Thomsen and Skærbæk, 2018). Froud (2003) defines risk as a situation in which numerical probabilities can be assigned, and uncertainty the situation when this cannot be done. Knorr-Cetina (2011), too, interprets the distinction in terms of whether probabilities can or cannot be assigned, adding that “[a]nalysts, we may say, are in the business of transforming uncertainty into risk.” (ibid., 433).

In this section, the distinction between risk and uncertainty is discussed in light of a separate line of scholarship that emerged in Austrian economic circles (Boudreaux and Holcombe, 1989; Foss, 1993; Hoppe, 2007; Klein, 2009). These scholars have argued that underpinning the distinction between risk and uncertainty in Knight is a problem of classification. Risk allows for attributing numerical probabilities to an event because the event that actors face is similar to other events of which we know the outcome. As a result, when there are many events similar to the one at hand, it is possible to use statistical methods of inferences to derive a numerical probability. In contrast, in periods of uncertainty it is not possible to attribute numerical probabilities to events because the event in question is not sufficiently similar to other events of which we know the outcome. Therefore, the more limited the number of events of which we know the outcome, the less likely is that we will be able to apply statistical techniques to these events. In these cases, we cannot attribute a numerical probability to events, but only make “estimates”. Knight is explicit about the connection between probabilities, estimates and the ability to classify events within a class of similar cases.

“an estimate or intuitive judgment is somewhat like a probability judgment, but very different from either type of probability judgement already described” [because it rests on] “no valid basis of any kind for classifying instances.” (Knight, 1964: 223–224)

Thus, Knight draws a connection between the ability to classify instances and the kind of predictions that we can make about them. Recall how one interviewee mentioned that “If you had a hundred countries, you could say, ‘Well eighty of them will go down that route and the other twenty go the other route’ (...) then you can do some sort of weighting, but you only have one”. This quotes

emphasized the challenge of attributing numerical values to future outcomes in the presence of unique cases. Another interviewee mentioned that “If you’re dealing with a retail bank, at least at a given point in time in a given economic environment, you’ve got so many people you can start doing a lot of statistical work on it and get a good idea of markers of better or worse performance. That’s where it doesn’t apply to rating agencies: they don’t deal in thousands and thousands and thousands of little credits; they deal in a relatively finite number and the problem is going through time.” This quote points to an essential problem confronting sovereign rating analysts, i.e. the limited sample that would underpin any statistical calculation. Indeed, recall the interviewee who said that the firm had done “statistical tests as much as we can within the limited pool of sovereign defaults”. It is the challenge of attempting to generalize from events that are rare and unique that gives rise to uncertainty in the sense of Knight.

“The practical difference between the two categories, risk and uncertainty, is that in the former the distribution of the outcome in a group of instances is known (either through calculation a priori or from statistics of past experience), while in the case of uncertainty this is not true, the reason being in general that it is impossible to form a group of instances, because the situation dealt with is in a high degree unique.” (Knight, 1964: 233)

The distinction between risk and uncertainty thus refers to the ability of actors to classify units into classes. According to this last quote, the extreme circumstance that prevents the assignment of numerical probabilities through statistical methods occurs when a situation is entirely unique and unprecedented. Defining a situation as unprecedented amounts to stating that it shares no properties with other situations. When an individual instance does not have properties that are common to a wider set, we do not know what its outcome will be. Then, actors who face unique or novel situations face a classification problem; they do not know in which class to place the case.

“The essential and outstanding fact is that the “instance” in question is so entirely unique that there are no others or not a sufficient number to make it possible to tabulate enough like it to form a basis for any inference of value about any real probability in the case we are interested in.” (Knight, 1964: 226)

Not knowing which categories best accommodate the case at hand, the language of statistical probabilities ceases to be useful, because that language

speaks of properties that are present in classes of instances, and when instances form part of no homogenous class, which property they possess is unknown (Hoppe, 2007). When cases are singular and known to be so, speaking of them in terms of probabilities becomes meaningless. When there is only one case, one cannot make any *probabilistic* prediction. Recall how interviewees mentioned the challenge posed by unprecedented events, stating “when you speak about the UK that goes through a referendum, that decides to leave the [European] Union, again it’s all unprecedented”. When there is only one case, one cannot make any probabilistic prediction. Similarly, sovereign rating analysts often confront uncertainty in this Knightian sense: they cannot entirely rely on past experience to assign a government to a rating category because the history of sovereign defaults does not provide a sufficiently robust sample of cases that allows to *automatically* make an exact, numerical prediction about the future. Similarly, information must be interpreted because the limited sample of sovereign defaults significantly limits the reliability of statistical models that predict default based on a limited set of publicly available indicators. Rather, analysts make “estimates”. Estimates neither emerge out of statistical knowledge of the behavior of classes of similar cases nor are they best understood as a priori predictions. Instead, they are estimates in a Knightian sense. They reflect sovereign rating analysts’ awareness that “cases are unique and classes correspondingly *unreal*” (Knight, 1946: 105, emphasis added).

Thus, in the sovereign rating space, “analysts confront (...) the imperative to decide with a significant but limited knowledge.” (Beunza and Garud, 2007: 13; Callon and Muniesa, 2005). Reading this problem through Knight clarifies exactly what is “imperative” and what is “limited” in this kind of knowledge. Imperative is the need to assign a rating. Sovereign rating analysts must arrive at a decision over which category of creditworthiness—the AAA, AA, BB+, etc. of the rating table—best describes each government. This is what users expect of CRAs. The classification expectation clashes, however, with the reality they often encounter in their daily work: the history of sovereign default is limited, and information does not have an unambiguous interpretation as a result. The basis for evaluation is a limited case-by-case analysis that, while informed by prior experience and

comparison with other situations, often remains geared towards unique evaluations of unique situations that are not subsumed by the rigidity of automated calculative techniques. Rather, judgment surrounds the entire rating process – the creation of valuation models, the deviation from model-implied ratings, and the interpretation of information– as the pragmatic response to an inherent classification problem. Accordingly, sovereign rating analysts issue Knightian “estimates” rather than “probabilities” and face Knightian “uncertainty” rather than “risk”.

The aspect of this chapter that seems to have the broadest implications for other studies of valuation practices in accounting and finance is that valuations that are forward-looking – such as credit ratings – always require finding a solution to a classification problem. It is a classification problem that has an epistemic trait insofar as it unavoidably involves forming a view of what each case – a firm, a stock, a government - is most similar. If the case to be valued is similar to other cases, prior knowledge of the outcome of those cases can be used to formulate a probabilistic prediction for the case at hand. Else, probabilistic predictions cannot be made, and one will have to turn to qualitative estimates that have no statistical basis and no probabilistic character. In practice, cases will be never completely unique and never completely homogenous, so that the classification problem amounts to establishing or constructing such similarity and difference. The key point, however, is that such construction is the essence of forward-looking valuation such as the construction of ratings. The classification problem described here does not affect those organizations that rank organizations based on past performance, the focus of much prior literature (Jeacle and Carter, 2011; Kornberger and Carter, 2010; Mehrpouya and Samiolo, 2016; Schultz et al., 2001). When performance has already occurred, there is no uncertainty over the similarity between past, present and future. There is also no ambiguity over what information implies for future performance, because future performance is not what is being valued and ranked. Instead, the insight that forward-looking practices that are concerned with making inferences about the future confront a classification problem is, I propose, applicable to other forward-looking valuation sites, where it will vary in degree but not in kind. Specifically, I

propose that the degree to which valuation practices will embrace quantitative as opposed to qualitative modes of inference will depend on the strength of the challenges to classification as perceived by analysts. Therefore, the first moment of a valuation process will be geared towards assessing the belonging of cases to classes. This is stated as the first proposition:

Proposition 1. The first moment of a valuation process is devoted to assessing if a case belongs to a class

The evidence presented here also qualifies the conclusion reached by prior work - that ratings *always* result from the assemblage of large samples of borrowers and the adoption of quantitative, model-based valuation techniques. For example, Mehrpouya and Samiolo (2016) argue that the difference between the ranking practice they examine in the context of the creation of the Access to Medicine Index and credit ratings is that while the former “is not based on formal statistical analysis”, the latter “seek to predict behaviour based on the law of large numbers and relatively “mechanized” statistical approaches” (Mehrpouya and Samiolo, 2016: 563 footnote 31). In a similar vein, Fourcade and Healy (2013: 563) claim that over the past decades “credit rating moved to probabilistic predictions based on statistical analysis of historical population data.” The evidence presented here shows that the automation of valuation practices depends on a problem of classification that may affect certain rating sites more than others. In the case of sovereign rating analysts, model-based valuation has not replaced analysts’ judgment. Indeed, analysts’ judgement shapes both the construction and the use of valuation models. The evidence also qualifies prior studies that suggest that the making of rankings neglects information that is available for some but not all entities, thus excluding dimensions that are either unquantifiable or idiosyncratic (Espeland and Sauder, 2007; Espeland and Stevens, 1998). The “Knightian” discussion of the making of sovereign ratings qualifies this notion by positing that information that makes units comparable reduces uniqueness and allows to more easily attribute a unit to a common class, whereas idiosyncratic traits make this process relatively more difficult. Information that is unquantifiable, idiosyncratic and incomparable continues to be

part of the valuation process but now has to be interpreted. In the case of sovereign ratings, the predominance of idiosyncratic qualities makes it more likely that analysts will resort to estimates as opposed to statistically grounded modes of inference. Hence, the second proposition:

Proposition 2. Valuation processes are less likely to be automated when valuation objects are relatively more unique

7. Conclusions

This chapter has used qualitative methods to examine the valuation practices of analysts engaged with the valuation of the credit rating of a sovereign government. Drawing on 27 semi-structured interviews conducted at four global credit rating agencies between 2016 and 2019 and on archival materials, the analysis has revealed the pervasiveness of analysts' judgement in all phases of the credit rating process – from the creation of valuation models to the interpretation of public information on the performance and prospects of the government. The analysis indicates that sovereign rating analysts work in the conditions of uncertainty rather than risk in the sense of Knight (1964) because information does not have unambiguous interpretation. In turn, this is due to the fact that the sample of sovereign defaults is too small to rely on the past to determine which indicators represent a reliable predictor of default. As a result, sovereign ratings are best understood as estimates rather than probabilities. The chapter proposes that understanding forward-looking valuation practices in accounting and finance as resulting from the perceived (by analysts) uniqueness of individual cases allows to predict which valuation sites are more likely to adopt automated as opposed to judgmental approaches to valuation.

CONCLUSIONS

This dissertation has provided novel perspectives on the financial reporting problem of governments. Through the use of both quantitative and qualitative studies and the creation of an original dataset, questions at the intersection of accounting, political economy, sociology and finance have been explored.

The first chapter shows that governments' guidance choices in periods of uncertainty are a function of the informational needs of bond market investors. Governments reduce guidance on the future evolution of debt and assets in periods of uncertainty only if they do not face short-term refinancing needs. Also, the reduction of guidance in periods of uncertainty takes place only when investors have relatively precise expectations about the pricing of governments' fiscal risk. Conversely, when investors have highly uncertain expectations governments do not reduce guidance in response to uncertainty shocks. At the same time, the value-added of guidance issued by transparent governments in periods of uncertainty is questionable: governments that do not reduce the amount of guidance in periods of uncertainty issue forecasts that turn out to be inaccurate and optimistic ex-post. In addition, these governments publish their guidance with substantial delay, and they present evidence of opportunistic reporting. Overall, the evidence suggests that periods of uncertainty give the government the incentive to use guidance somewhat opportunistically. Conversely, periods of fiscal uncertainty give governments the opportunity to reduce disclosure for relatively "benign" reasons (Leuz et al., 2008).

Through the evidence reported in the first chapter, the analysis contributes to the growing but limited literature on governments' disclosure choices (Kim et al., 2018). So far, studies have emphasized the importance of electoral and fiscal incentives in portraying a rosy but ultimately untruthful picture of fiscal *outcomes*. Others have shown that fiscal distress, economic conditions and political dynamics determine governments' willingness to be transparent about the budget *process*. By focusing on the understudied phenomenon of government guidance, this chapter examines a setting where the government *may* hold imprecise view about the future fiscal outlook. The strongest contribution is then that in the case

of guidance, less than full disclosure does not always depend on relatively sinister rationales. The application of a purely disclosure lens to the analysis of governments' guidance choices expands our understanding of the government's disclosure problem in general, and of the costs and benefits of fiscal transparency in particular. The chapter makes three additional contribution to extant literature in accounting. First, it examines the disclosure of fiscal information reported under the "statistical mode" of government accounting (Heald and Hodges, 2018) – a relatively understudied area. Second, the chapter adopts a treatment effect approach based on inverse-probability-weighting techniques that is relatively novel in the field. The method may be of interest to scholars working on the role of fiscal constraints on governments' disclosure choices (Chaney et al., 2002; Costello et al., 2017) because that these constraints effectively generate a rule for non-random assignment of units to treatment (Angrist and Kuersteiner, 2011). Finally, the empirical results of the first chapter provide a strong indication that various facets of transparency are driven by different rationales, confirming that important insights can be gained by studying the micro-aspects of disclosures indices (e.g., Copelovitch et al., 2018; de Renzio and Wehner, 2017).

The second chapter analyzes the equity markets response to guidance issued by the government on the future trajectory of the budget balance – the difference between government revenues and expenditures. The analysis suggests that government guidance does not reduce market uncertainty unless the government is credible. Hence, governments can take actions to anchor investors' expectations around their reported forecast. Indeed, while the results indicate that guidance by the government is an information-rich event in equity markets, the analysis shows the important role played by the credibility of guidance as a moderator of investors' response. In particular, the findings indicate that as governments announce that the budget balance will improve in the short term, investors remain generally skeptical. Indeed, investors interpret announcements of short-term improvements in the budget balance heterogeneously. The degree of heterogeneity in investors' response appears to be a function of the credibility of governments' forecasts. The evidence shows that investors' opinions diverge more when the government has been optimistic in its last forecast and when it

holds a weak political mandate. The analysis thus suggests that credibility plays an important role in moderating investors' response to governments' guidance. Surprisingly, the results indicate that credibility is negatively affected by the extent of disaggregation of guidance, suggesting that disaggregation may not increase the credibility of forecasts. The primary contribution of the second chapter is thus to test the claim that fiscal guidance reduces information asymmetries in markets. It is well-known that the desired reduction in information asymmetries has prompted international institutions to become vocal promoters of transparency in the fiscal and budgetary domain (IMF, 1998; Mehrpouya and Salles-Djelic, 2019; Mosley, 2003a; OECD, 2002; The World Bank, 1992). In this context, this chapter expands our understanding of the actual consequences of transparency in government guidance. The evidence complements extant studies that focus on the signaling role of transparency in capital markets. It also complements research on the value of fiscal forecasts in bond markets. With the exception of Bafundi, Columbano and Carrera (2018), who analyze the case study of Spanish SGP disclosures, no study tests the role of fiscal forecasts in equity markets. In addition, prior work does not examine trading volume and it does not attempt to estimate causal effects. The broader contribution of the second chapter is to extend the presently limited research on the consequences of central governments' disclosure choices (Kim et al., 2018). This subfield of accounting research is attracting growing interest, as mounting levels of public debt and deficit put the financial reporting "lever" of fiscal policy at the forefront (Naughton and Spamann, 2015). This chapter complements other efforts in the study of governments' use of accounting and disclosure tools, e.g. the opportunistic management of GDP figures, the electoral cycle in (mis)reporting the fiscal position, and the effect of fiscal stress on accounting discretion. The documented evidence about the conditions that make fiscal disclosures an effective channel to anchor expectations points to a novel and important responsibility for the financial disclosures of governments. Finally, this chapter contributes to the literature on disclosure regulation that has mostly focused on the effects of *corporate* disclosure mandates. In particular, the findings may assist in assessing the success of the SGP reporting mandate. They complement the work of others

who have examined the track record of governments in adhering to SGP requirements related to, e.g. the accuracy and bias of forecasts.

The third chapter examines the construction of the credit ratings of sovereign governments. The analysis shows that, as sovereign rating analysts attempt to infer whether a certain piece of information indicates that the risk of sovereign default has changed, they confront the challenge that sovereigns hardly default. It is thus challenging for analysts to determine whether publicly available information points to a higher or lower default risk: the small sample of sovereign default precludes drawing this inference statistically – and therefore, “objectively”. This implies that each government - and each event that it goes through - must be analyzed on a case-by-case basis. Then, statistical methods of inference cannot be applied because statistics only applies to classes of cases and never to individual instances (Guseva and Rona-Tas, 2001; Hoppe, 2007). Accordingly, exact numerical probabilities of default cannot be calculated. Instead, sovereign ratings represent “estimates” in the sense proposed by Frank Knight: “an estimate or intuitive judgment is somewhat like a probability judgment”. However, it is different from numerical probabilities because it rests on “*no valid basis of any kind for classifying instances.*” (Knight, 1964: 223–224, original emphasis). The possibility of assigning probabilities or making estimates is what distinguishes risk and uncertainty: “risk” and “uncertainty” denote not so much the possibility of attributing probabilities to future events but rather the possibility of classifying cases into categories. In the case of sovereign ratings, analysts mostly confront uncertainty and ratings themselves express estimates rather than probabilities. Thus, the primary contribution of the third chapter is to show that in the case of sovereign rating analysts, the statistical, model-based approach to valuation has not replaced analysts’ subjective judgment. Indeed, analysts’ judgement shapes both the construction and the use of valuation models. Overall, the findings highlight that judgment is a pervasive feature of sovereign credit ratings and that – somewhat contrary to what was claimed by Besedovski (2018) - sovereign rating analysts remain generally distrustful of automated models and statistical modes of inference. In addition, unlike Carruthers (2013), the findings suggest that sovereign rating analysts do not “transform” uncertainty into risk.

Rather, they seem to be aware that ratings express, at best, judgmental estimates. The sovereign rating industry thus represents a “paradigmatic case” (Patton, 2002) of a valuation challenge that – it has been proposed - confronts all valuation practices in accounting and finance. More broadly, the chapter contributes to the literature on valuation studies in accounting and finance by proposing that valuation practices that are forward-looking – such as those underpinning credit ratings – always require finding a solution to a classification problem. It is a classification problem that has an epistemic trait insofar as it unavoidably involves forming a view of what each case – a firm, a stock, a government - is most similar to. If the case to be valued is similar to other cases, prior knowledge of the outcome of those cases can be used to formulate a statistical, probabilistic prediction for the case at hand. Else, statistical approaches to make predictions cannot be made and one has to turn to qualitative estimates that have no statistical basis and no probabilistic character. In practice, cases will be never completely unique and never completely homogenous, so that a classification problem amounts to establishing or constructing such *relative* similarity and difference. The key point, however, is that such construction is the essence of forward-looking valuation such as the construction of ratings. The evidence presented qualifies the conclusion reached by prior work - that ratings *always* result from the assemblage of large samples of borrowers and the adoption of quantitative, model-based valuation techniques. The evidence also qualifies prior studies that suggest that the making of rankings neglects information that is available for some but not all entities, thus excluding dimensions that are either unquantifiable or idiosyncratic. The analysis qualifies this notion by positing that information that makes units comparable reduces uniqueness and allows to more easily attribute a unit to a common class, whereas idiosyncratic traits make this process relatively more difficult. Information that is unquantifiable, idiosyncratic and incomparable continues to be part of the valuation process but now has to be interpreted.

Collectively, the chapters have documented several phenomena that are interesting in light of prior literature. First, it has been shown that governments do not always become opaque in their disclosures for opportunistic reasons. Indeed,

there are times when governments may optimally reduce the amount of disclosures due to the limited benefit that they derive and the limited demand for information. Second, the analysis shows that in equity markets, the use of governmental disclosures is conditional on its credibility. In particular, the evidence indicates that investors often disagree on the probability that governments will translate policy plans into outcomes. Finally, it has been shown that in the sovereign rating setting, the analysis of the financial information of governments is permeated by the subjective judgment of analysts because of the idiosyncratic characteristics of governments and their limited history of default. Overall, this dissertation contributes to enriching our understanding of the financial reporting problem of governments – an understudied yet important manifestation of accounting.

CONCLUSIONES

Esta tesis doctoral ha proporcionado perspectivas novedosas sobre el problema de la información financiera de los gobiernos. Mediante el uso de estudios tanto cuantitativos como cualitativos y la creación de un conjunto de datos originales, se han explorado cuestiones en la intersección de la contabilidad, la economía política, la sociología y las finanzas.

En el primer capítulo se muestra que las decisiones de divulgación de los gobiernos en períodos de incertidumbre varían en función de las necesidades de información de los inversores en el mercado de bonos. Los gobiernos reducen la orientación sobre la evolución futura de la deuda y los activos en períodos de incertidumbre sólo si no se enfrentan a necesidades de refinanciación a corto plazo. Además, la reducción de la orientación en períodos de incertidumbre sólo tiene lugar cuando los inversores tienen expectativas relativamente precisas sobre la fijación de los precios del riesgo fiscal de los gobiernos. Por el contrario, cuando los inversores tienen expectativas muy inciertas, los gobiernos no reducen la orientación en respuesta a las crisis de incertidumbre. Al mismo tiempo, el valor añadido de la orientación emitida por gobiernos transparentes en períodos de incertidumbre es cuestionable: los gobiernos que no reducen la cantidad de orientación en períodos de incertidumbre emiten previsiones que resultan ser inexactas y optimistas a posteriori. Además, esos gobiernos publican sus orientaciones con un retraso considerable y presentan pruebas de informes oportunistas. En general, las pruebas sugieren que los períodos de incertidumbre dan al gobierno el incentivo de utilizar la orientación de manera un tanto oportunista. A la inversa, los períodos de incertidumbre fiscal dan a los gobiernos la oportunidad de reducir la divulgación por razones relativamente "benevolentes" (Leuz, Triantis y Yue 2008).

A través de las pruebas que se presentan en el primer capítulo, el análisis contribuye a la creciente pero limitada literatura sobre las opciones de divulgación de los gobiernos (Kim, Plumlee y Stubben 2018). Hasta ahora, los estudios han destacado la importancia de los incentivos electorales y fiscales para presentar una imagen optimista, pero, en última instancia, incorrecta de los

resultados fiscales. Otros han demostrado que las dificultades fiscales, las condiciones económicas y la dinámica política determinan la voluntad de los gobiernos de ser transparentes en el proceso presupuestario. Al centrarse en el fenómeno poco estudiado de la orientación gubernamental, en este capítulo se examina un entorno en el que el gobierno puede tener una visión imprecisa sobre las perspectivas fiscales futuras. La contribución más importante es entonces que, en el caso de la orientación, una divulgación parcial no siempre depende de justificaciones relativamente negativas. La aplicación de una lente puramente de divulgación al análisis de las opciones de orientación de los gobiernos amplía nuestra comprensión del problema de la divulgación del gobierno en general, y de los costos y beneficios de la transparencia fiscal en particular. El capítulo hace tres contribuciones adicionales a la literatura existente en materia de contabilidad. En primer lugar, examina la divulgación de la información fiscal presentada en el "modo estadístico" de la contabilidad gubernamental (Heald y Hodges 2018), un área relativamente poco estudiada. En segundo lugar, el capítulo adopta un enfoque del efecto de tratamiento basado en técnicas de ponderación de la probabilidad inversa que es relativamente novedoso en este campo. El método puede ser de interés para los académicos que trabajan sobre el papel de las restricciones fiscales en las opciones de divulgación de los gobiernos (Costello, Petacchi y Weber 2017; Chaney, Copley y Stone 2002) porque estas restricciones generan efectivamente una regla para la asignación no aleatoria de unidades al tratamiento (Angrist y Kuersteiner 2011). Por último, los resultados empíricos del primer capítulo proporcionan una fuerte indicación de que varias facetas de la transparencia están impulsadas por diferentes justificaciones, lo que confirma que se pueden obtener importantes conocimientos estudiando los micro-aspectos de los índices de revelación (por ejemplo, Copelovitch, Gandrud y Hallerberg 2018; de Renzio y Wehner 2017).

En el segundo capítulo se analiza la respuesta de los mercados de valores a las directrices publicadas por el gobierno sobre la trayectoria futura del equilibrio presupuestario, es decir, la diferencia entre los ingresos y los gastos del gobierno. El análisis sugiere que las orientaciones del gobierno no reducen la incertidumbre del mercado a menos que el gobierno sea creíble. Por

consiguiente, los gobiernos pueden adoptar medidas para anclar las expectativas de los inversores en torno a sus previsiones comunicadas. De hecho, si bien los resultados indican que la orientación del gobierno es un acontecimiento rico en información en los mercados de valores, el análisis muestra el importante papel que desempeña la credibilidad de la orientación como moderador de la respuesta de los inversores. En particular, los resultados indican que, a medida que los gobiernos anuncian que el equilibrio presupuestario mejorará a corto plazo, los inversores siguen siendo en general escépticos. De hecho, los inversores interpretan de manera heterogénea los anuncios de mejoras a corto plazo en el equilibrio presupuestario. El grado de heterogeneidad en la respuesta de los inversores parece depender de la credibilidad de las previsiones de los gobiernos. Las pruebas demuestran que las opiniones de los inversores divergen más cuando el gobierno ha sido optimista en su última previsión y cuando tiene un mandato político débil. Así pues, el análisis sugiere que la credibilidad desempeña un papel importante en la moderación de la respuesta de los inversores a las orientaciones de los gobiernos. Sorprendentemente, los resultados indican que la credibilidad se ve afectada negativamente por el grado de desagregación de la orientación, lo que sugiere que la desagregación puede no aumentar la credibilidad de las previsiones. La principal contribución del segundo capítulo es, pues, poner a prueba la afirmación de que la orientación fiscal reduce las asimetrías de información en los mercados. Es bien sabido que la reducción deseada de las asimetrías de información ha impulsado a las instituciones internacionales a convertirse en promotoras vocales de la transparencia en el ámbito fiscal y presupuestario (FMI 1998; OCDE 2002; Banco Mundial 1992; Mehrpouya y Salles-Djelic 2019; Mosley 2003a). En este contexto, el presente capítulo amplía nuestra comprensión de las consecuencias reales de la transparencia en la orientación de los gobiernos. Las pruebas complementan los estudios existentes que se centran en la función de señalización de la transparencia en los mercados de capital. También complementa la investigación sobre el valor de las previsiones fiscales en los mercados de bonos. Con la excepción de Bafundi, Columbano y Carrera (2018), que analizan el estudio de caso de las

divulgaciones del PEC español, ningún estudio pone a prueba el papel de las previsiones fiscales en los mercados de valores. Además, los trabajos anteriores no examinan el volumen de operaciones y no intentan estimar los efectos causales. La contribución más amplia del segundo capítulo consiste en ampliar la investigación, actualmente limitada, sobre las consecuencias de las opciones de divulgación de los gobiernos centrales (Kim, Plumlee y Stubben 2018). Este campo de la investigación de la disciplina contable está atrayendo cada vez más interés, ya que los crecientes niveles de deuda y déficit públicos sitúan en primer plano la "palanca" de información financiera de la política fiscal (Naughton y Spamann 2015). Este capítulo complementa otros esfuerzos en el estudio del uso de los instrumentos de contabilidad y divulgación por parte de los gobiernos, por ejemplo, el manejo oportunista de las cifras del PIB, el ciclo electoral en la incorrecta información de la posición fiscal y el efecto de la presión fiscal en la discreción contable. Las pruebas documentadas sobre las condiciones que hacen de las declaraciones fiscales un canal eficaz para anclar las expectativas apuntan a una novedosa e importante responsabilidad de las declaraciones financieras de los gobiernos. Por último, este capítulo contribuye a la bibliografía sobre la reglamentación de la divulgación de información que se ha centrado principalmente en los efectos de los mandatos de divulgación de información de las empresas. En particular, las conclusiones pueden ayudar a evaluar el éxito del mandato de presentación de informes del PEC. Complementan la labor de otros estudios que han examinado el historial de los gobiernos en cuanto a la adhesión a los requisitos del PEC relacionados, por ejemplo, con la exactitud y el sesgo de las previsiones.

El tercer capítulo examina la construcción de las calificaciones crediticias de los gobiernos soberanos. El análisis muestra que, a medida que los analistas de calificación soberana tratan de inferir si un determinado dato indica que el riesgo de impago soberano ha cambiado, se enfrentan al reto de que los soberanos difícilmente incumplen sus obligaciones. Así pues, para los analistas es difícil determinar si la información disponible públicamente apunta a un riesgo de incumplimiento mayor o menor: la pequeña muestra de incumplimiento soberano impide extraer esta inferencia estadísticamente -y, por tanto, "objetivamente".

Esto implica que cada gobierno -y cada evento que atraviesa- debe ser analizado caso por caso. Entonces, los métodos estadísticos de inferencia no pueden aplicarse porque las estadísticas sólo se aplican a clases de casos y nunca a casos individuales (Hoppe 2007; Guseva y Rona-Tas 2001). Por consiguiente, no se pueden calcular probabilidades numéricas exactas de incumplimiento. En cambio, las calificaciones soberanas representan "estimaciones" en el sentido propuesto por Frank Knight: "una estimación o juicio intuitivo es algo así como un juicio de probabilidad". Sin embargo, es diferente de las probabilidades numéricas porque se basa en "ninguna base válida de ningún tipo para clasificar las instancias". (Knight 1964, 223-24, énfasis original). La posibilidad de asignar probabilidades o hacer estimaciones es lo que distingue el riesgo y la incertidumbre: "riesgo" e "incertidumbre" denotan no tanto la posibilidad de atribuir probabilidades a eventos futuros sino más bien la posibilidad de clasificar los casos en categorías. En el caso de las calificaciones soberanas, los analistas se enfrentan sobre todo a la incertidumbre, y las propias calificaciones expresan estimaciones más que probabilidades. Así pues, la principal contribución del tercer capítulo es mostrar que, en el caso de los analistas de calificación soberana, el enfoque estadístico de la valoración basado en modelos no ha sustituido el juicio subjetivo de los analistas. De hecho, el juicio de los analistas determina tanto la construcción como el uso de los modelos de valoración. En general, las conclusiones ponen evidencian que el juicio es una característica omnipresente de las calificaciones crediticias de los soberanos y que -al contrario de lo que afirma Besedovski (2018)- los analistas de calificación soberana siguen desconfiando en general de los modelos automatizados y los modos estadísticos de inferencia. Además, a diferencia de Carruthers (2013), las conclusiones sugieren que los analistas de calificación soberana no "transforman" la incertidumbre en riesgo. Más bien parecen ser conscientes de que las calificaciones expresan, en el mejor de los casos, estimaciones de juicio. El sector de la calificación soberana representa, pues, un "caso paradigmático" (Patton 2002) de un desafío de valoración que -se ha propuesto- se enfrenta a todas las prácticas de valoración en contabilidad y finanzas. En términos más generales, el capítulo contribuye a la bibliografía sobre estudios de valoración en

contabilidad y finanzas proponiendo que las prácticas de valoración con visión de futuro -como las que sirven de base a las calificaciones crediticias- siempre requieren encontrar una solución a un problema de clasificación. Se trata de un problema de clasificación que tiene un rasgo epistémico en la medida en que implica inevitablemente formarse una opinión de lo que más se asemeja a cada caso - una empresa, una acción, un gobierno. Si el caso que se va a valorar es similar a otros casos, el conocimiento previo del resultado de esos casos puede utilizarse para formular una predicción estadística y probabilística del caso en cuestión. De lo contrario, no se pueden hacer enfoques estadísticos para hacer predicciones y hay que recurrir a estimaciones cualitativas que no tienen base estadística ni carácter probabilístico. En la práctica, los casos nunca serán completamente únicos y nunca completamente homogéneos, de modo que un problema de clasificación equivale a establecer o construir esa similitud y diferencia relativas. El punto clave, sin embargo, es que esa construcción es la esencia de la valoración prospectiva, como la construcción de clasificaciones. Las pruebas presentadas matizan la conclusión a la que se ha llegado en trabajos anteriores: que las calificaciones son siempre el resultado del ensamblaje de grandes muestras de prestatarios y de la adopción de técnicas de valoración cuantitativas basadas en modelos. Las pruebas también matizan los estudios anteriores que sugieren que la elaboración de las clasificaciones descuida la información disponible sobre algunas entidades, pero no todas, excluyendo así las dimensiones no cuantificables o idiosincrásicas. La información no cuantificable, idiosincrática e incomparable sigue formando parte del proceso de valoración, pero ahora debe ser interpretada. En el caso de las valoraciones soberanas, el predominio de las cualidades idiosincráticas hace más probable que los analistas recurran a estimaciones en lugar de a modos de inferencia basados en estadísticas.

En conjunto, los capítulos han documentado varios fenómenos que son interesantes a la luz de la literatura anterior. En primer lugar, se ha demostrado que los gobiernos no siempre se vuelven opacos en sus revelaciones por razones oportunistas. De hecho, hay ocasiones en que los gobiernos pueden reducir óptimamente la cantidad de revelaciones debido al limitado beneficio que

obtienen y a la escasa demanda de información. En segundo lugar, el análisis muestra que, en los mercados de valores, el uso de la información gubernamental está condicionado a su credibilidad. En particular, las pruebas indican que los inversores suelen discrepar sobre la probabilidad de que los gobiernos traduzcan los planes de política en resultados. Por último, se ha demostrado que, en el marco de la calificación soberana, el análisis de la información financiera de los gobiernos está impregnado del juicio subjetivo de los analistas debido a las características idiosincrásicas de los gobiernos y a su limitado historial de incumplimiento. En general, esta tesis contribuye a enriquecer nuestra comprensión del problema de la información financiera de los gobiernos, una manifestación poco estudiada pero importante de la contabilidad.

APPENDIX A. THE CONSTRUCTION OF THE SCP DATASET

The process to create the dataset containing the forecasts used in [Chapter 1](#) and [Chapter 2](#) involved three steps. In what follows, I describe each step in turn.

1. Retrieving the universe of Stability/Convergence Programmes

Articles 4 and 7 of the first of the two Stability and Growth pact (SGP) regulations (EU, 1997a) impose that European Union (EU) governments publish one Stability/Convergence Programme (SCP) every year, around autumn for the period 1998-2009 and around spring for the period 2011-2018. The mandate thus allows me to observe all the SCP published by EU governments. I restrict my attention to those published over the period 2001-2018 because it was only with the 2001 vintage that the EU published guidelines on the format and content of SCPs.²⁰ I retrieve all SCPs from the EC online archives dedicated to the European Semester.²¹ In the dataset, I identify each unique SCP through the counter *scpnr*.

Table A.1 reports the number of unique SCPs per country that are present in the dataset. In Column 2, I list the number of unique SCPs issued by each member state. There is heterogeneity in this number for three reasons. First, while in the period 2001-2003 there were 15 member states of the EU, the number of member states increased over the years. Specifically, a first wave of EU enlargement took place in 2004, when 10 countries joined the EU. A second wave of enlargement took place in 2007 when Bulgaria joined the EU together with Romania. A final addition to the EU membership was Croatia in July 2013. As the obligation to submit SCPs only applies to EU member states, the number of country-specific observations partly depends on the date of EU accession,

²⁰ I return to this issue below.

²¹ As of the date of writing, all SCPs published between 1998 and 2015 are available at this link: http://ec.europa.eu/economy_finance/economic_governance/sgp/convergence/programmes/index_en.htm. The remaining SCPs and the CoC are retrievable at: https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/stability-and-growth-pact_en. (Last accessed: 30 April 2019). After triangulating this search with governments-provided information as described below, I estimate that there were two additional SCPs that were submitted in February 2000 by France and Germany. I could not retrieve these SCPs. In addition, there were six addenda to SCPs published by governments following the European Recover Programme of 2008, which contained no new forecasts, however. I exclude these from the dataset.

reported in column 5 and coded in the dataset as *eu_accessiondate*.²² Second, governments have published additional SCPs beyond those required on an ad hoc basis. Between 2001 and 2018, 434 SCPs were made public, of which 418 were original SCPs and 16 were ad hoc publications as indicated in columns 3 and 4, respectively. The ad hoc publications increase the number of SCP observations for the related government. I identify these publications in the dataset through the variable *scp_original*, which takes value 0 if the SCP represents an ad hoc publication, and 1 otherwise.

Table A.2 lists the ad hoc publications and their rationales. I have analyzed each of them in depth to understand their rationale. My analysis indicates that three took place following a request by the Council that the government resubmits the SCP because inadequate or insufficient information was disclosed in the prior issuance. Whether the Council makes such request can be observed by retrieving the Council Opinion that closes the SGP cycle, which is public information. Four additional ad hoc publications represent updates of the SCP published between December 2008 and February 2009 to incorporate the EU-wide strategy of fiscal expansion that followed the 2008 financial crisis. Six ad hoc publications represent resubmissions of earlier SCPs issued by governments voluntarily, typically because of a change in the parliamentary majority or in the cabinet after an election. Finally, one occasional publication represents Portugal's voluntary submissions of a fiscal plan in the years 2012-2014. In those years, Portugal was exempted from mandatory reporting of a SCP because it was under a financial assistance program. These SCPs are included in the dataset because they follow the same structure of all other SCPs and because the European Commission (EC) assessed them. This was an exception, however. Indeed, the third and final reason why there is heterogeneity in the number of SCPs per government is that governments halt the publication of SCPs when they are under external financial assistance. This is the reason why the number

²² The accession dates were retrieved from: https://ec.europa.eu/neighbourhood-enlargement/policy/from-6-to-28-members_en (last access: 21 January 2020 14:48 CET).

of observations for Greece and Cyprus is lower than that of other governments that began mandatory guidance in the same year.

The first step thus resulted in the retrieval of a total of 434 unique SCPs published between 2001 and 2018 as a result of the SGP mandatory guidance regime.

Table A.1. Stability/Convergence Programmes (SCPs) by country

Country	Number of SCPs	Number of original SCPs	Number of ad hoc SCPs	EU Accession date
Austria	17	17	0	01/01/1995
Belgium	18	17	1	01/01/1957
Bulgaria	12	12	0	01/01/2007
Croatia	5	5	0	01/07/2013
Cyprus	12	11	1	01/05/2004
Czech Republic	15	15	0	01/05/2004
Denmark	17	17	0	01/01/1973
Estonia	15	15	0	01/05/2004
Finland	17	17	0	01/01/1995
France	17	17	0	01/01/1957
Germany	19	17	2	01/01/1957
Greece	11	10	1	01/01/1981
Hungary	16	15	1	01/05/2004
Ireland	18	17	1	01/01/1973
Italy	17	17	0	01/01/1957
Latvia	15	14	1	01/05/2004
Lithuania	15	15	0	01/05/2004
Luxembourg	18	17	1	01/01/1957
Malta	15	15	0	01/05/2004
Netherlands	19	16	3	01/01/1957
Poland	16	15	1	01/05/2004
Portugal	17	16	3	01/01/1986
Romania	12	12	0	01/01/2007
Slovakia	15	15	0	01/05/2004
Slovenia	15	15	0	01/05/2004
Spain	17	17	0	01/01/1986
Sweden	17	17	0	01/01/1995
UK	17	17	0	01/01/1973
Total	434	418	16	

This table reports the number of Stability/Convergence Programmes (SCP) published by the EU member states between 2001 and 2018. Columns 3 and 4 report the number of original and ad hoc SCPs, respectively. Column 5 lists the date of accession into the EU.

Table A.2. Description of ad hoc Stability/Convergence Programmes

Country	Year	Description	Rationale
Belgium	2009	Restatement	Request from the Council of the EU
		Addendum with updated	Addendum following the European Economic
Cyprus	2009	projections	Recovery Plan
Germany	2004	Revision	Revision following political negotiations
		Addendum with updated	Addendum following the European Economic
Germany	2009	projections	Recovery Plan
			Revision following concerns related to budgetary
Greece	2005	Revision	statistics
Hungary	2006	Restatement	Request from the Council of the EU
		Addendum with updated	Addendum following the European Economic
Ireland	2009	projections	Recovery Plan
Latvia	2009	Revision	Revision following economic downturn
		Addendum with updated	Addendum following the European Economic
Luxembourg	2009	projections	Recovery Plan
Netherlands	2003	Revision	Revision following political negotiations
		Addendum with updated	Addendum following the European Economic
Netherlands	2008	projections	Recovery Plan
Netherlands	2012	Restatement	Request from the Council of the EU
Poland	2008	Revision	Revision of the budgetary strategy
Portugal	2012	Voluntary Submission	Voluntary Submission of fiscal strategy
Portugal	2013	Voluntary Submission	Voluntary Submission of fiscal strategy
Portugal	2014	Voluntary Submission	Voluntary Submission of fiscal strategy

This table reports the rationale for the publication of the 16 ad hoc SCPs that contain guidance. In column 2 the year of publication of ad hoc SCPs is provided. Column 3 describes the nature of the ad hoc SCP. Column 4 provides a brief indication of the rationale for the publication of the ad hoc SCP.

2. The publication dates of Stability/Convergence Programmes

The second step of the data collection process involved estimating the dates at which each of these SCPs was published. No study has previously assembled a dataset of these dates at the day/month/year level. The publication date is particularly relevant to scholars interested in knowing when the content of the SCP was first made available to the public.

I began by contacting the relevant administration of each EU government by e-mail requesting the day/month/year each of the SCPs was first released to the public. While the relevant administration was typically the Ministry of Finance of the government, the final source of information varied somewhat.²³ A total of 10 out of 28 governments responded, for a 35.7 percent response rate. Of these responses, six contained the information and one pointed to a webpage containing the information. The remaining three provided an incomplete answer. In the absence of a (complete) response, the competent administration was contacted by phone. Following this second round of requests, the publication dates were most often obtained by email. Alternatively, I was directed to a web source that indicated how the information could be retrieved. At the end of the process, which lasted approximately two months between March and May 2016, 21 out of 28 governments responded to the request for information, for a 75 percent positive response rate. Overall, the interaction with governments returned 176 SCPs associated with unique publication dates, which I denote in the dataset with *pubdate*.

The second step involved a news search conducted using Thomson Reuters EIKON with the app “Newsroom”. Newsroom was interrogated for any piece of news containing the words “Stability Programme, Stability Program, Stability Report, Stability Plan” in a period starting on the first day of the month preceding the month of submission of the document to the EC and ending on the day of the Council Opinion.²⁴ As explained below, I retrieved the latter two sets of dates from

²³ When collecting these data, I included the SCPs published in the period 1998-2000. The numbers reported throughout refer only to the SCPs included in the final dataset.

²⁴ The word “Stability” was replaced with “Convergence” when the country was not a euro adopter.

the EC online archives. The news search returned 209 articles that allow me to date the publication of the SCP. In the dataset, I denote the press-retrieved publication date as *press_pubdate*. There are 73 cases in which both government sources and the press search resulted in a publication date. In 31 out of these 73 cases (42.4 percent), the publication date is exactly the same according to both the government and the press. On average, the government-provided publication date follows the press-estimated publication date by 1.47 days. The estimates from an intercept-only regression reported in column 1 of Table A.3 do not allow rejecting the hypothesis of this difference being zero. Thus, when the publication date was available either from the press or from both sources, I use the press-provided publication date as *pubdate*. I use the government-provided date to replace missing cases only when the press publication date was not available.

When neither the press nor the government provided me with a publication date, I replace the missing dates with the day on which the government approved the SCP, denoted with *cabdate*. Information on cabinet-approval dates is contained either in the SCP itself or it was provided by governments. Once more, the regression results reported in column 2 of Table A.3 reveal that the difference between the cabinet approval date and the press publication date has a mean value of 0.66, not significantly different from zero. Therefore, the cabinet approval date was used to fill in *pubdate* in the absence of a publication date. The data collection procedure was finalized by consulting EC online archives.²⁵ I retrieve the date at which the government transmitted each SCP to the EC. This date – denoted with *subdate* – may differ from the cabinet approval or the publication date. Informal conversation with staff involved in the process indicated that governments typically submit the SCP to the EC about three days after the approval of the document by the cabinet. Column 3 of Table A.3 shows that the difference between the submission date and the press-provided publication date

²⁵ As of the date of writing, the SCPs published between 1998 and 2015 are available at this link: http://ec.europa.eu/economy_finance/economic_governance/sgp/convergence/programmes/index_en.htm. The remaining SCPs and the CoC (see below) are retrievable at: https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/stability-and-growth-pact_en. (Last access: 30 April 2019).

indeed returns a mean value of 3.09 days, which I estimate to be statistically significant. I replace the publication date with the submission date when neither the cabinet approval date nor the publication date is available. The resulting *date* variable represents my best estimate of the publication date for the entire sample of SCPs. Researchers can use the other dates reported in the dataset for specific purposes. At the end of this second step, I am able to attribute a publication date to each of the 434 SCPs included in the dataset between 2001 and 2018. In the 8 cases in which the publication date falls on a weekend, I attribute it to the following Monday.

The dataset also includes the date at which the European Commission made its recommendation to the Council of the EU - denoted with *recdate* - as well as the date at which the latter released its Opinion – denoted with *opdate*. The first refers to the date contained in the EC recommendations. The second refers to the day the Council Opinion was first published in the Official Gazette of the European Union. These dates are both public information. I estimate an average difference of 32 days between the EC recommendation and the date the Council releases its own Opinion. In turn, the difference between the EC recommendation date and the estimated publication date is on average 44 days. Hence, the whole SGP cycle runs for 76 days, on average.

Table A.3. Differences in event dates

	(1)	(2)	(3)
<i>Constant</i>	1.47 [1.14]	0.66 [0.75]	3.09*** [0.77]
R-squared	0.030	0.011	0.124
Observations	73	123	203
Cluster	Country	Country	Country

This table reports the results of an intercept-only OLS regression that tests whether there are significant differences in the dates at which governments approve, submit and publish their Stability/Convergence Programmes (SCP). Column 1 reports the difference between the publication date reported by the government and that reported by the press. Column 2 reports the difference between the date at which the government approves the SCP and the publication date. Column 3 reports the difference between the date at which the government submits the SCP to the European Commission and the publication date. Standard errors clustered by country in brackets. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The availability of the submission date allows constructing *delay*, a measure of the number of days between the date of submission of the SCP to the EC and the deadline for the submission itself. This measure can be of interest to researchers because it may reflect incumbents' willingness to disclose or withhold information or alternatively, weak governance mechanisms (Andersen et al., 2014; Edmonds et al., 2017). To construct *delay*, I attribute a deadline to each SCP submission. These deadlines are imposed by the EU and are public information. From 2011 onwards, SCPs have to be submitted to the EC by April 30. Between 2001 and 2010, they could be submitted any time between October 15 and December 1. In addition, specific government-years had specific deadlines. For instance, exceptional deadlines were set for the 2008 and 2009 submissions, which could be transmitted to the EC by the end of December and the end of January, respectively (EC, 2008; EU, 2009a). In addition, the Eastern European governments that entered the EU in 2004 had their first submission deadline set on 15 May 2004. Until 2010, the governments of the UK and Ireland could submit their SCP by the time of the approval of the budget and pre-budget statements – usually discussed in Parliament by the first week of December. For the UK, I retrieve these dates from parliamentary sources (House of Commons Information Office, 2010; House of Commons Library, 2018).²⁶ For Ireland, information is available at the government-dedicated webpage.²⁷ Other ad hoc deadlines were established in case the Council requested the resubmissions of a non-compliant SCP. The variable *deadline* reflects each of these idiosyncrasies and it captures the SCP-specific deadline. The difference *delay* is calculated as the submission date less the deadline, i.e. *subdate-deadline*, and it is available for 412 observations. The difference *pubdelay* uses the publication date in place of the submission date, and it is available for 423 observations. The two measures are highly correlated (over 95 percent).

²⁶ The 2018 date of the spring budget was retrieved from: <https://www.gov.uk/government/topical-events/spring-statement-2018>. (Last access: 12 April 2020 at 14:09 GMT)

²⁷ See <http://budget.gov.ie/Budgets/2011/2011.aspx>. (Last access: 12 April 2020 at 14:14 GMT)

3. Organizing the content of Stability/Convergence Programmes

As noted earlier, SCPs are published approximately once every year, although late filings and ad hoc submissions may translate into an imbalance in the number of annual SCPs publications. Since the first year of SCP submission – 1998 – the content of SCPs has changed, raising the question of how to organize their content into a dataset that is comparable over the years. I chose to organize the content of SCP reports by anchoring each SCP to the format established in the Guidelines on the format and content of Stability and Convergence Programmes, also known as Codes of Conduct (hereafter referred to as “CoC”).

The CoC are documents issued by the ECOFIN section of the Council.²⁸ It is important to clarify the place of CoC within the Stability and Growth Pact. The CoC are a “complement” to the SGP (EU, 2005). Specifically, the CoC complement the SGP provisions related to the disclosure requirements set out originally with Regulation 1466/97 to monitor governments’ fiscal plans. This Regulation – which has legal force – dictates the minimum amount of information that governments must release. Specifically, the Regulation (EU, 1997a, Article 3.2a) imposes on government the reporting of the outlook for the government budget balance - the difference between revenues and expenditures – and the outlook for the ratio of government debt over gross domestic product.²⁹ It also imposes the publication of the “main assumptions about expected economic developments and important economic variables which are relevant to the realization of the stability programme *such as* government investment expenditure, real gross domestic product (GDP) growth, employment and inflation” (ibid., Article 3.2b, emphasis added).³⁰ Note that, while the Regulation is clear as to the budgetary variables to be included in the SCP reports, the list

²⁸ The ECOFIN is the forum where the heads of the Ministries of Finance of the EU member states are represented.

²⁹ The budget balance is known in the Government Finance Statistics Manual as the “net lending/net borrowing” of the general government. The debt ratio is an indicator created by dividing the total amount of loans and bonds by gross domestic product at market prices. See IMF (2014 Chapter 4). In the EU, these variables include some specific items as determined by Eurostat. See Savage (2005) for a detailed discussion of the accounting treatment of budgetary figures in the SGP process.

³⁰ Article 7 applies these requirements to Convergence Programmes issued by governments of countries not adopting the euro as their currency.

of “main assumptions” and “important economic variables” to be reported is indicative. The CoC approved shortly before the introduction of the SGP mandate thus served the purpose to clarify the set of additional economic and financial information that governments should include in their annual SCP reports. In this sense the CoC represent “guidelines” on the content of SCPs. In turn, the CoC also clarify the “format” for presenting the information. In what follows, I describe the format and content and how they assist me in giving structure to my dataset.

All CoC were retrieved from the online archives of the EC. There have been eight CoC issued since 1998. Each iteration of the CoC slightly modified the required information set. All CoC, except for the 1998 one, specify the tabular format to be used for the submission of SCPs. They further contain guidelines over the applicable measurement basis of economic and budgetary items, and the “required” vs. “optional” nature of each item. The first CoC, applicable to the submission between 1998 and 2000, did not contain disclosure requirements that are sufficiently unambiguous to generate a coherent list of items that were expected to be disclosed. Beginning with the second stage of the CoC, the number of items to be disclosed was harmonized and a list of required items was made explicit. At the date of writing, the applicable CoC has reached its eighth vintage, approved in May 2017.

Because EU sources include the date at which each CoC has been approved, I am able to match the disclosure requirements of each SCP with the date of publication estimated above. I am therefore able to use the format introduced in each CoC as a template that sets out the expected format and content of each SCP in each particular year. I use this template to give structure to my data collection process as follows.

I build a spreadsheet that contains as many worksheets as the number of years of SCP submissions, ranging from 2001 to 2018. Because information has to be disclosed in each year t at the horizons $t-1$, t , $t+1$, $t+2$, and $t+3$, plus information on $t-1$ levels (for a limited set of items), each of the worksheets contains up to six observations per government and as many governments as there were Member States of the EU in year t . If a government issues multiple

SCP in the same calendar year, that government appears more than once in the worksheet, unless a change in the CoC applicable took place in the meanwhile. Each worksheet is anchored, therefore, to one applicable CoC and its columns are organized according to the tabular format set out therein. Hence, in each worksheet, each column contains one item that the applicable CoC recommends publishing and the number of columns across worksheets changes in line with the number of items required by the CoC over the years.

I only exclude three categories of items from the data collection process. These are “Government Expenditure by Function”, which only contain data points related to years $t-2$ and $t+3$, and “Long-Term Projections” which contain forecasts for decades ahead. The horizon of the forecasts for the items that compose these categories is not aligned with the others and thus is not comparable. I also do not collect the forecasts contained in the table named “Comparison with Prior Update” because it contains the forecasts for real GDP growth, the debt and the budget balance already found in other tables and compares them to those disclosed in the prior SCP vintage. I also exclude one item – the budget balance of state governments, which only applies to countries with a federalist structure and may thus over-or under-estimate the level of transparency of these countries. For the same reason, I exclude items that were required only for a limited set of governments to create a comparable index.

In the dataset, I attribute a name to each of the items in each of the column that resemble the item name itself, while using variable labels to provide a more accurate description. A number precedes variable names, denoting the number of the table that should contain the item based on the CoC guidelines. For example, real GDP growth is typically found in table 1, while the governments’ budget balance is found in table 2, etc.³¹

Finally, within each worksheet, each cell contains one numerical value that reflects the forecast for that item as disclosed in the related SCP. If the forecast

³¹ The only exception is that for the years 2001-2005 I renumber tables 3, 4, and 5 to 4, 5, and 6, respectively to allow comparison with later submissions of the SCP. This is because beginning in Stage 3, the third table of the CoC includes “Government expenditure by function”, which I exclude from the data collection.

is missing, I set the cell to missing. With this procedure, if an item is disclosed, say, for forecast horizon t , $t+1$ and $t+2$, but not $t+3$, the cells that correspond to that item in t , $t+1$ and $t+2$ receive a numerical value, while its cell in $t+3$ is set to missing. In this way, the dataset preserves within-government-SCP variation in the number of forecasts issued at different forecast horizons.

The inclusion of data points in the dataset required some instances of judgment. First, when a government has a fiscal year that differs from the calendar year and reports forecasts as referring to, say the $t/t+1$ period, that forecast is attributed to year t in line with both corporate and national accounting practices if the fiscal year ends before June 30. Second, when forecasts are presented as ranges, approximate values, or as average of multiple horizons, the corresponding values were noted in the dataset and coded to missing afterwards. The rationale for the choice is that the CoC does not allow publishing range or approximate forecasts, nor it allows for the collapsing of information that pertain to multiple forecast horizons. When items are reported as the sum of multiple items, they are coded as missing because the CoC does not allow collapsing items. When a forecast is presented as referring to an item indicate in the related CoC, but the government specifies that its value refers to an item not included in the CoC guidelines, the cell is set to missing. Also, when the CoC allows the same item (say, government revenue) to be presented multiple times in the SCP, the data points are coded as not missing if the item is present at least once to avoid double-counting. This procedure applies only when the CoC explicitly states that the item code refers to the same measure in all the tables where it appears. The “item code” is the official European System of Accounts (ESA) code. When items are reported outside of the table where they should be presented, their values count as not missing. In addition, values that are not reported although they would have been equal to zero, had they been reported, are set to missing. Finally, to guarantee uniformity within the dataset I eliminate all items reported for years $t-1$, which do not represent forecasts. The resulting complete dataset contains 434 unique SCPs and 1,674 observations, for a final number of 195,858 data-points disclosed by up to 28 countries and that relate to

117 unique items forecasted up to 4 years ahead. The list of items is shown in Table A.4

These data allow me to compute *Transparency* at each country-date-horizon triple by summing across all items that are present in a SCP and divide the sum by the number of items that the guidelines recommended to disclose on the date at which the SCP submission took place. Denoting with $i \in (0,1)$ each item on which a forecast should be disclosed within the set N , $k = 0, \dots, 3$ each forecast horizon, d the date of publication of SCPs, and c each country, I calculate:

$$Transparency_{c,k,d} = 100 * \left(\frac{1}{N} \sum_{i=1}^N i_{c,k,d} \right)$$

This measure varies at the country, forecast horizon, and publication date level.

4. Distribution of SCP publication dates

Table A.5 reports the distribution of the full sample of SCPs by year, quarter, month and day of the week. The table reports substantial heterogeneity in the disclosure dates. This heterogeneity partly reflects some institutional idiosyncrasies of the SGP process that are important to consider when using the data in empirical work.

The increase in the number of SCPs registered in 2004 and 2007 is due to the enlargement of EU membership to several Eastern European nations, which was accompanied by two SCP issuances, one in spring and the other in autumn.³² In 2010, SCP submissions were suspended as part of a reform to the SGP monitoring cycle known as the “European Semester”. Yet, the table reports 25 SCPs published in 2010. This is due to the fact that most governments published in early 2010 those SCPs that belonged to the 2009 cycle, for which the EU authorized the exceptional deadline of January 31, 2010 (EU, 2009a).³³ Between 2012 and 2018, Greece did not publish its SCP because it went under a financial

³² See the factsheet at: https://ec.europa.eu/neighbourhood-enlargement/sites/near/files/pdf/publication/factsheet_en.pdf. (Last access: 26 November 2019)

³³ In 2008, the deadline for submission of SCPs was moved to December 31. See European Commission (2008).

assistance program that had its own disclosure requirements. Cyprus was subject to the same exemption in the years 2013-2015, hence the number of observations declines in those years. In terms of distribution of SCP by quarters, I document a clustering of submission dates in the second and fourth quarter. This is to be expected because unless explicit exceptions were granted to governments, SCPs had to be submitted between October 15 and December 1 in the 2001-2009 period. From 2011 onwards, the deadline was moved to April 30 to align SCP submissions with the European Semester. Accordingly, the large number of submissions in the first quarter of the year partly reflects late filers publishing in January. The monthly breakdown of observations reflects these idiosyncrasies. Note that the somewhat large number of observations in the month of May is primarily due to the year 2004, when the first SCP of the new EU member states was to be submitted. In terms of weekdays, the publication of SCPs appears to be relatively uniform, with a somewhat larger number of SCPs published on Wednesday and Thursdays

Table A.4. List of items contained in Stability/Convergence Programmes (by category)

#	Transparency - Macro	#	Transparency - Fiscal Performance	#	Transparency - Fiscal Position	#	Transparency - Assumptions
1	Real GDP growth	36	Net lending/borrowing of the general government sector (budget balance)	83	Gross debt	100	Short term interest rate
2	Nominal GDP growth	37	Net lending/borrowing the central government	84	Change in the debt ratio	101	Long term interest rate
3	GDP deflator	38	Net lending/borrowing of local governments	85	Primary balance	102	US: Short term interest rate
4	Harmonized Index of Consumer Prices	39	Net lending/borrowing of social security funds	86	Nominal GDP growth	103	US: Long term interest rate
5	Employment growth	40	Total revenues	87	Stock flow adjustment	104	USD/EUR exchange rate
6	Labor productivity growth	41	Total expenditures	88	of which: privatization receipts pro memoria: implicit interest rate on debt	105	Effective exchange rate of the euro area
7	Private consumption expenditure	42	Net interest payments	89		106	Effective exchange rate of the EU
8	Government consumption expenditure	43	Primary balance	90	of which: cash-accruals adjustment	107	National Currency/EUR exchange rate
9	Gross fixed capital formation	44	Taxes	91	of which: acquisition of financial assets	108	World GDP growth
10	Changes in inventory and acquisition of valuables	45	Social contributions	92	of which: valuation effects	109	US GDP growth
11	Exports of goods and services	46	Interest income	93	Liquid financial assets	110	Japan GDP growth
12	Imports of goods and services	47	Other revenues	94	Net financial debt	111	EU15 GDP growth
13	Final domestic demand	48	Collective consumption	95	Debt amortization	112	GDP growth of relevant foreign markets
14	External balance	49	Social transfers in kind	96	Percentage of debt in foreign currency	113	Global imports - volume
15	Private consumption deflator	50	Social transfers not in kind	97	Average maturity of debt	114	Global imports - prices
16	Public consumption deflator	51	Interest expenditure	98	Public guarantees	115	Commodity prices
17	Gross fixed capital formation deflator	52	Subsidies	99	of which: to the financial sector	116	Oil prices

Table B.1 (continued)

18	Exports deflator	53	Gross fixed capital formation	117	Effective exchange rate
19	Imports deflator	54	Other expenditures		
20	Employment - hours worked	55	Cyclical budgetary component		
21	Unemployment rate	56	Cyclically adjusted budget balance		
22	Labor productivity per hour	57	Cyclically adjusted primary balance		
23	Compensation of employees	58	pro memoria: FISIM		
24	Net lending/borrowing - Rest of the World	59	Taxes on production and imports		
25	Balance of Goods and Services	60	Taxes on income and wealth		
26	Balance of income	61	Capital taxes		
27	Capital account	62	Property income		
28	Net lending/borrowing of the private sector	63	pro memoria: Tax burden		
29	Statistical discrepancy	64	Social payments		
30	Potential GDP growth	65	pro memoria: compensation of employees		
31	Output gap	66	of which: unemployment benefits		
32	Potential GDP growth - contribution from labor	67	Capital transfers		
33	Potential GDP growth - contribution from capital	68	Total revenues at unchanged policies		
34	Potential GDP growth - contribution from productivity	69	Total expenditures at unchanged policies		
35	Compensation per employee	70	Expenditures on EU programmes		
		71	of which: investments on EU programmes		
		72	Cyclical unemployment benefits		
		73	Effect of discretionary revenues		

Table B.1 (continued)

74	Revenue mandated by law
75	One-off
76	of which: one-off expenditures
77	of which: one-off revenues
	Compensation of employees plus
78	intermediate consumption
79	Compensation of employees
80	Intermediate consumption
	pro memoria: Government
81	consumption
82	Structural budget balance

This table lists the items contained in the dataset used in [Chapter 1](#) and [Chapter 2](#). The column names refer to the sub-score of *Transparency* to which each item contributes.

Table A.5. Distribution of SCP publication dates

Year	N	Publication quarter	N	Publication month	N	Publication day	N
2001	14	1	61	1	36	Monday	60
2002	12	2	225	2	13	Tuesday	71
2003	17	3	2	3	12	Wednesday	100
2004	33	4	146	4	189	Thursday	109
2005	27	<i>Total</i>	<i>434</i>	5	31	Friday	94
2006	27			6	5	<i>Total</i>	<i>434</i>
2007	31			7	0		
2008	18			8	1		
2009	18			9	1		
2010	25			10	6		
2011	27			11	58		
2012	27			12	82		
2013	25			<i>Total</i>	<i>434</i>		
2014	26						
2015	26						
2016	27						
2017	27						
2018	27						
<i>Total</i>	<i>434</i>						

This table shows the distribution of all Stability/Convergence Programmes (SCP) by year, quarter, month and day of publication.

APPENDIX B. VALIDATION OF THE TREATMENT ASSIGNMENT MODEL OF CHAPTER 1

The propensity score approach is valid under the assumption that selection takes place only on observable conditions. This assumption is likely to be met in the context of the deficit rule applicable to EU governments, which gives strong priors about the process of treatment assignment. At the same time, the assumption is not exactly testable because other unobserved factors can affect treatment assignment and remain excluded from the model (Caselli and Wingender, 2018). For this reason, the diagnostics of propensity score models concern themselves with testing the extent to which the propensity score stage succeeds in designing a sample where treated and untreated units are similar in terms of observables. There are typically three categories of model diagnostics: graphical, numerical and statistical (Austin and Stuart, 2015). All three types of diagnostics aim at examining the extent of covariate balancing in the subpopulation of treated and untreated units. I show evidence that the treatment assignment mechanism can be modeled based on observables in Section 1. Section 2 reports the diagnostics of the propensity score stage.

1. Validation of the propensity score assumptions

It is important to validate the claim that the budget balance forecasts reported by governments are predictable based on observable conditions – a prerequisite for the adoption of IPW method.

The EU fiscal rules dictate that governments facing a deficit should *report* that they will return the budget towards balance over the medium-term, and towards less than three percent of GDP annually. This rule leads governments to forecast a rapid improvement in their budget balance, particularly when governments run a deficit. The rule also creates a mechanical relation between the lagged fiscal performance of the government and the reported, *expected* performance throughout the forecast horizon. Table B.1 shows that the EU fiscal rule leads governments to predictably report that the budget balance will come close to the target by the end of the forecast horizon. Column 1 reports that the average

forecasted budget deficit declines monotonically from 2.3 percent of GDP to 0.6 as one moves from the forecast for year t to $t+3$. Hence, governments on average report that they will make considerable progress towards the balanced budget target. Column 2 further shows that over a third of governments report that the budget balance will be equal to or above zero by the end of the forecast horizon, compared to twenty percent at the end of year t . This is consistent with claims that medium-term fiscal plans give incentives to governments to postpone costly fiscal adjustments (Alesina and Perotti, 1996). Finally, a striking 97 percent of governments' forecasts of the budget deficit at the longer-term forecast horizon fall below the 3 percent rule. As shown in columns 4 and 5, this percentage is driven both by governments that in the year prior to the publication of the fiscal plan comply with the deficit rule and by those that do not. Between 83 and 99 percent of governments that comply with the deficit rules in year t report that they will remain compliant throughout the forecast horizon. A striking 92 percent of *non-compliant* governments report that they will comply by the end of year $t+3$. Therefore, in the sample the forecasts for the budget balance follow the pattern that is predictable given EU rules. Governments forecast that the budget balance will improve over the medium-term even if their budgetary performance falls significantly short of EU requirements.

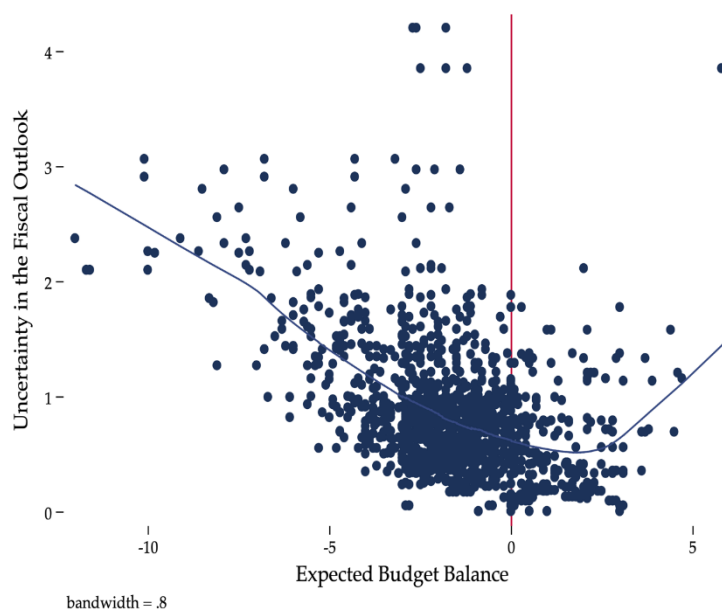
This dynamic leads to a mechanical relation between past fiscal performance and the standard deviation of the forecast budget balance as shown in Figure B.1. The figure shows that the standard deviation of budget balance forecasts increases exponentially with the amount of the expected deficit. Budget balance forecasts are more volatile, the larger the forecasted deficit.

Table B.1. Governments' budget balance forecasts follow a predictable pattern

Variable	N	Forecasted budget balance (% of GDP)	Frequency of expected budget surplus	Frequency of expected budget balance above the excessive deficit	of which: issued by governments that have excessive deficit	of which: issued by governments that do not have excessive deficit
Horizon= t	309	-2.28	0.20	0.67	0.83	0.37
Horizon= $t+1$	309	-1.88	0.20	0.79	0.97	0.91
Horizon= $t+2$	309	-1.33	0.23	0.90	0.94	0.63
Horizon= $t+3$	264	-0.59	0.36	0.97	0.99	0.92

This table shows that SCP budget balance forecasts as reported by governments follow a predictable pattern. Column 1 reports the budget balance forecast at horizon $t+k$ for a government releasing the forecast in year t and $k=0, \dots, 3$. A negative sign denotes a budget deficit. Column 2 reports the frequency of forecasted budget surpluses. Column 3 reports the frequency of forecasted budget balances that exceed the EU rule that the budget deficit should be less than 3 percent of Gross Domestic Product. Column 4 (5) replicates Column 3 for the subsample of governments that in year t have a lagged budget deficit (not) exceeding 3 percent of GDP. The budget balance is the difference between the revenues and expenditures of the government. The number of observations in columns 4 is 136, 173, 136 and 173 as the forecast horizon changes from t to $t+3$. The number of observations in column 5 is 136, 173, 107 and 157 as the forecast horizon changes from t to $t+3$.

Figure B.1. Uncertainty increases in the size of the forecasted budget deficit



This figure plots the *Expected Budget Balance* reported by governments in Stability/Convergence Programmes (SCP) for continuous levels of uncertainty, calculated as the standard deviation of the *Expected Budget Balance* over a four-year forecast horizon.

Note that there is no reason per se why *large* forecasted *deficit* should translate into a large standard deviation of the forecasts. Table B.2 reports an example of the effect that different forecasted levels of the budget balance have on *Uncertainty* at varying levels of lagged fiscal performance. Recall that the standard deviation is calculated over a four-year forecast horizon. Now compare government A, which expects the deficit to be 5% of GDP in each year between t and $t+3$, to government B, which expects the deficit to be 1% in each year between t and $t+3$. One sees immediately that the magnitude of the expected deficit does not imply that the standard deviation of the budget forecasts be high. Absent the deficit rule, the government continues to forecast a stable deficit, with no effect on the standard deviation of the forecast. Similarly, government C is in surplus and needs not adjust its budget balance to comply with the deficit rule. The EU deficit rule is asymmetric; there is no penalty for governments running “excessive” surpluses. Hence, governments with large surpluses can forecast that the surplus will continue indefinitely.³⁴ Governments in deficit cannot: the presence of the deficit rule ensures that large forecasted deficits cannot be *reported* for the entire four-year forecast horizon. Consider the examples of governments D, E, and F, which all have deficits in excess of the deficit rule. Given the rule that deficits should return to less than three percent of GDP by year $t+3$, assume that all governments plan to distribute the adjustment towards the target in equal amounts over the forecast horizon. It is clear from the table that the standard deviation of reported budget balance forecast increases in the amount of the *Fiscal Gap*, i.e. the difference between the lagged budget balance and the deficit limit of three percent of GDP.

³⁴ In fact, the figure shows that governments that expect surpluses, too, display a larger standard deviation of the forecasts. This may be due to the fact that governments in surplus choose to reduce the amount of surplus over time while remaining compliant with the deficit rule, perhaps owing to political pressure to redistribute the surplus (Caselli and Wingender, 2018).

Table B.2. The larger the lagged budget deficit, the greater the standard deviation of budget balance forecasts in presence of a fiscal rule

Government	Lagged budget balance	Budget balance forecast, t	Budget balance forecast, t+1	Budget balance forecast, t+2	Budget balance forecast, t+3	Standard deviation of budget balance forecasts
A	-6,0%	-5,0%	-5,0%	-5,0%	-5,0%	0,0%
B	-6,0%	-1,0%	-1,0%	-1,0%	-1,0%	0,0%
C	2,5%	2,5%	2,5%	2,5%	2,5%	0,0%
D	-4,5%	-3,5%	-2,9%	-1,5%	-0,5%	1,4%
E	-0,8%	-7,0%	-4,0%	-3,0%	-1,5%	2,3%
F	-10,0%	-6,0%	-3,5%	-2,0%	1,0%	2,9%

This table shows an example of how the presence of a fiscal rule according to which the budget deficit should be less than 3 percent of GDP mechanically affects the standard deviation of budget balance forecasts that underlie the *Uncertainty* measure.

Therefore, governments in the EU can be said to be subject to the following rule: when having a fiscal performance that is below the EU deficit limit, the government will report that the budget deficit will diminish to a level close to three percent of GDP by the end of the forecast horizon. Following the rule means that certain governments – namely governments with large deficits, are more likely to report a large standard deviation of the budget balance forecasts. Because other things equal, high values of the standard deviation of the budget balance forecasts are coded here as episodes of uncertainty – denoting the condition of a government being “treated” in a quarter-date – the deficit rule means that governments are not randomly assigned to treatment. The IPW approach serves to reweight observations by the inverse of the probability of treatment following the estimation of the propensity score model.

2. Diagnostics of the propensity score stage

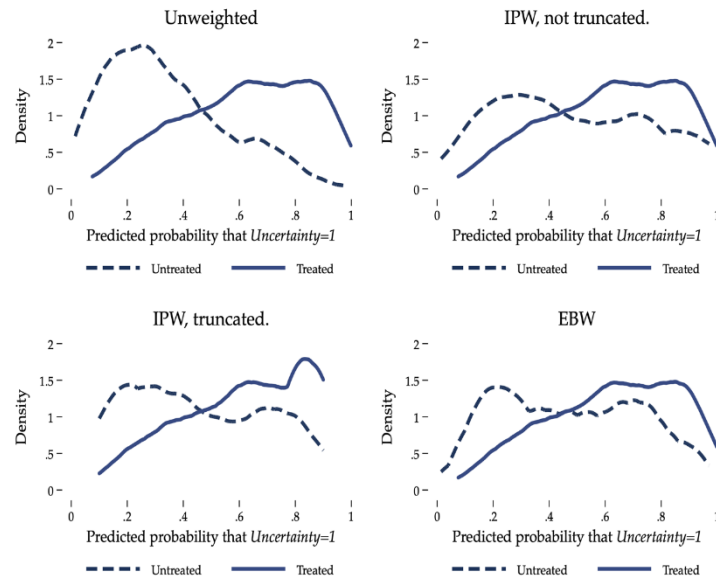
Figure B.2 plots the density of treated and untreated observations across the predicted probability of treatment. The two densities should have most of their masses distributed over a common support with limited mass at the extremes (Busso et al., 2014). When plotting the density of predicted probabilities of treatment prior to weighting (top-left panel), we see that the mass of treatment and control groups is located at different parts of the probability distribution. Namely, observations for which $Uncertainty = 1$ are observations for which the conditional probability of treatment is somewhat high, while the opposite is true for observations for which $Uncertainty = 0$. Reassuringly, none of the distributions exhibit a peak at either the values of 0 or 1, however, which provides comfort as to the existence of common support to estimate the ATE of $Uncertainty$. The top-right panel shows that the overlap improves considerably as observations are weighted on the basis of their estimated IPWs. In particular, untreated observations with a relatively low probability of being treated no longer exhibit extreme mass around the 0.2-0.3 segment, and much more mass is now present at higher probability levels. Because reweighting can generate implausibly large weights if observations have a predicted probability of treatment close to 1 or 0, the bottom-left panel shows the same densities, but now IPWs

are truncated at the estimated conditional probabilities below or above 0.1 and 0.9, respectively (Jordà and Taylor, 2016). There are no visible improvements in the overlap, as should be expected given that only few observations have extremely high or low estimated probabilities of treatment.

The bottom-right panel shows the distribution of conditional probabilities based on entropy-balancing weights (Hainmueller, 2012).³⁵ In accounting, entropy balancing weights (EBW) are becoming popular (McMullin and Schonberger, 2020; Shroff et al., 2017). The weights derived from entropy balancing approaches are based on an optimization criterion that aims for achieving the smallest standardized difference in the mean, standard deviation and skewness of the covariates in treated and untreated subpopulations. Graphically, the EBW approach produces results very similar to the IPW approach, although it does increase the mass of untreated units that have high probabilities of being treated at the expense of also increasing the mass of untreated units that have low probability of being treated. Table B.3 presents the results of EBW.

³⁵ To compute entropy-balancing weights, I implement the ebalance routine in Stata ©.

Figure B.2. Evidence of sufficient overlap between treated and untreated units (N=1191)



This figure displays the distribution of predicted probabilities that $Uncertainty=1$ for treated and untreated observations. Treated (untreated) observations are those for which $Uncertainty=1$. Predicted probabilities are calculated based on the estimated coefficients from the propensity score model. IPW, not truncated indicates non-truncated inverse-probability weights, as described in the text. IPW, truncated indicates IPW truncated at the extreme probabilities below (above) 0.1 (0.9). EBW denotes entropy-balancing weights as in Hainmueller (2012).

Table B.3. Covariate balancing after entropy weighting (N=318)

Treated units: 148 Total of weights 148
 Untreated units: 170 Total of weights 148

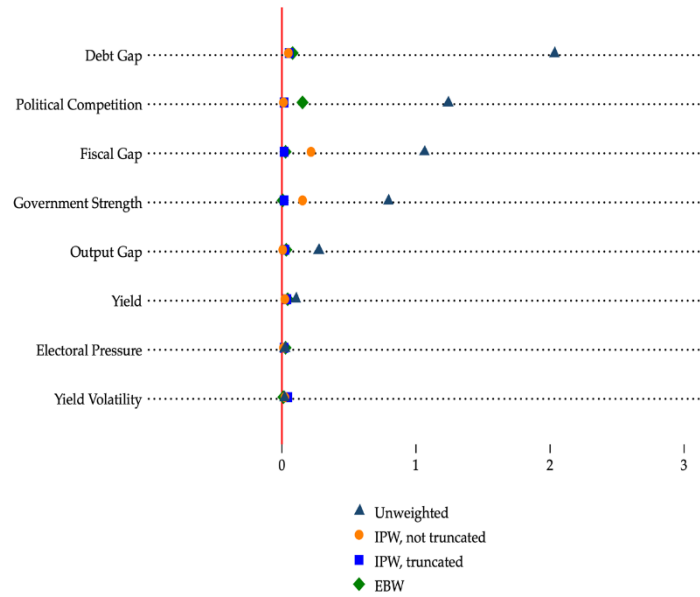
	<i>Before: without weighting</i>								
	Treated				Untreated				
	Mean	Variance	Skewness	Mean	Variance	Skewness	Mean	Variance	Skewness
<i>Yields</i>	3,41	5,42	1,03	3,14	4,05	0,43			
<i>Yields Volatility</i>	0,17	0,02	1,34	0,16	0,02	1,38			
<i>Government Strength</i>	54,97	50,33	-0,05	57,01	79,75	0,22			
<i>Political Competition</i>	69,22	101,90	-0,01	72,07	128,60	-0,57			
<i>Electoral Pressure</i>	1,87	1,70	0,04	1,92	1,44	-0,13			
<i>Output Gap</i>	-0,74	9,82	-0,50	-0,27	5,07	0,35			
<i>Fiscal Gap</i>	0,97	18,07	1,66	-1,09	6,20	0,58			
<i>Debt Gap</i>	7,46	983,00	0,44	-2,30	565,00	0,47			

	<i>After: EBW as the weighting variable</i>								
	Treated				Untreated				
	Mean	Variance	Skewness	Mean	Variance	Skewness	Mean	Variance	Skewness
<i>Yields</i>	3,41	5,42	1,03	3,41	5,42	0,41			
<i>Yields Volatility</i>	0,17	0,02	1,34	0,17	0,02	1,47			
<i>Government Strength</i>	54,97	50,33	-0,05	54,97	50,34	0,32			
<i>Political Competition</i>	69,22	101,90	-0,01	69,22	101,90	-0,33			
<i>Electoral Pressure</i>	1,87	1,70	0,04	1,87	1,70	-0,18			
<i>Output Gap</i>	-0,74	9,82	-0,50	-0,74	9,82	0,33			
<i>Fiscal Gap</i>	0,97	18,07	1,66	0,97	18,07	-0,02			
<i>Debt Gap</i>	7,46	983,00	0,44	7,46	983,00	0,13			

This table reports the results of the entropy-balancing weighting approach (Hainmueller, 2012). Observations in the treated and untreated subpopulations are those for which *Uncertainty* is equal to 1 and 0, respectively. The entropy-balancing algorithm reweights observations so that the reweighted observable characteristics of the subpopulations are similar in terms of mean, variance and skewness.

The next test evaluates directly the extent of covariate balancing. To evaluate covariate balancing, I first present the analysis of unstandardized differences in the weighted and unweighted means of the covariates. Figure B.3 plots absolute standardized differences of the covariates included in the propensity score. The blue triangles report unweighted differences, i.e. the absolute value of the difference in the means of the covariates between treated and untreated observations. Five of the eight covariates exhibit a pronounced difference in the two group, among which the differences in fiscal and political condition appear substantial. Next, the orange dots report differences in the covariates weighted by the non-truncated IPW, i.e. weights that include extreme probabilities, while blue squares report differences weighted by the truncated IPW. The entropy-balancing weighted differences are denoted with green diamonds. Under all weighting approaches the covariates are almost perfectly balanced in the sample of treated and untreated units. None of the approaches appear superior at visual inspection. For instance, EBW are unable to fully balance on *Political Competition*, while non-truncated IPWs are unable to fully balance on *Fiscal Gap*, and truncated IPWs are unable to fully balance the *Debt Gap* and *Yields*. Nevertheless, the lack of balancing is minimal after reweighting. Table B.4 reports the numerical differences. A perfectly balanced sample would show average differences of exactly zero. The non-truncated IPW achieve superior balancing the majority of times, with minimal average differences in five out the eight covariates, while the truncated IPW achieves superior balancing only once. In turn, the EBW approach achieves superior balancing twice. Overall, all reweighting approaches deliver a substantial improvement over the unweighted approach as evidenced by the extreme unweighted differences in *Debt Gap*, *Political Competition*, *Fiscal Gap*, *Government Strength* and *Output Gap*. In the rest of the analysis, the superior, non-truncated IPW this adopted.

Figure B.3. Absolute standardized differences – treated vs. untreated (N=304)



This figure plots the absolute value of the standardized difference in the means of the covariates listed in the figure between treated and untreated units. Treated (untreated) observations are those for which *Uncertainty*=1. The covariates listed are those included in the propensity score model. IPW, not truncated indicates non-truncated inverse-probability weights, as described in the text. IPW, truncated indicates IPW truncated at the extreme probabilities below (above) 0.1 (0.9). EBW denotes entropy-balancing weights as in Hainmueller (2012).

Table B.4. Absolute standardized differences – treated vs. untreated (N=304)

Variable	Unweighted	IPW, Not truncated	IPW, truncated	EBW
<i>Yield</i>	0,108	0,025	0,044	0,050
<i>Yield Volatility</i>	0,021	0,026	0,050	0,018
<i>Government Strength</i>	0,797	0,158	0,022	0,010
<i>Political Competition</i>	1,242	0,018	0,021	0,157
<i>Electoral Pressure</i>	0,025	0,014	0,019	0,031
<i>Output Gap</i>	0,277	0,013	0,025	0,039
<i>Fiscal Gap</i>	1,066	0,221	0,019	0,033
<i>Debt Gap</i>	2,036	0,052	0,060	0,082

This table reports the absolute value of the standardized difference in the means of the covariates listed in the table between treated and untreated units. Treated (untreated) observations are those for which *Uncertainty*=1. The covariates listed are those included in the propensity score model. IPW, ntr. indicates non-truncated inverse-probability weights, as described in the text. IPW, truncated indicates IPW truncated at the extreme probabilities below (above) 0.1 (0.9). EBW denotes entropy-balancing weights as in Hainmueller (2012).

As a final test, a balancing test of covariates is run that expressly tests for the statistical difference in means between treated and untreated units under the null of no difference. This test represents a formal alternative to the visual inspection of standardized differences as above because it relies on significance testing. Note, however, that the aim of this test may be somewhat misleading because the IPW approach does not aim to test if treated and untreated units are similar at the population level, as a statistical test does. Rather, the IPW approach seeks the minimal differences in the sample. Nevertheless, significance tests of differences in covariates' (first) moments remain popular (Imai and Ratkovic, 2014). The results are presented in Table B.5, which reports the coefficients of an intercept-only regression of each of the covariates of the propensity score stage on the binary *Uncertainty* variable. The test of no difference in means through a regression is preferable over a simple t-test because it allows to cluster the standard errors at the country level. Panel A (B) shows the coefficients when observations are unweighted (weighted). Panel A results indicate that *Political Competition* is weaker in treated compared to untreated units, while both the amount of the deficit and the amount of debt are significantly larger. These differences are statistically significant. On the contrary, Panel B confirms that after IPW reweighting, none of the covariate differences between the treated and the untreated units is statistically significant at conventional levels.

Collectively, the evidence provides sufficient comfort that the propensity score model achieves satisfactory balancing in the covariates.

Table B.5. Test of covariate balancing

Panel A: Unweighted covariates

	(1) Yields	(2) Yields Volatility	(3) Government Strength	(4) Political Competition	(5) Electoral Pressure	(6) Output Gap	(7) Fiscal Gap	(8) Debt Gap
<i>Uncertainty=1</i>	0.19	0.01	-2.18	-3.91*	-0.04	-0.46	1.86***	9.99+
	[0.32]	[0.02]	[1.48]	[1.78]	[0.10]	[0.52]	[0.46]	[5.03]
Constant	3.20***	0.16***	57.17***	73.15***	1.91***	-0.26	-1.09**	-2.66
	[0.30]	[0.01]	[1.60]	[2.36]	[0.10]	[0.26]	[0.35]	[4.95]
Observations	309	309	309	309	309	309	309	309

Panel B: IPW-weighted covariates

	(1) Yields	(2) Yields Volatility	(3) Government Strength	(4) Political Competition	(5) Electoral Pressure	(6) Output Gap	(7) Fiscal Gap	(8) Debt Gap
<i>Uncertainty=1</i>	-0.03	-0.00	1.42	0.22	0.01	-0.04	-0.84	-1.98
	[0.33]	[0.02]	[1.93]	[0.71]	[0.11]	[0.54]	[1.15]	[3.26]
Constant	3.45***	0.16***	54.43***	71.47***	1.92***	-0.11	0.19	3.94
	[0.31]	[0.01]	[1.95]	[1.95]	[0.12]	[0.26]	[1.02]	[5.75]
Observations	309	309	309	309	309	309	309	309

This table reports the OLS estimates of an intercept-only regression of the covariate listed on top of each model of the binary measure of *Uncertainty*. The covariates listed are those included in the propensity score model. Standard errors in brackets are clustered at the country level. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

APPENDIX C. DETAILS ON THE FILTERING PROCESS FOR THE COMPUSTAT SAMPLE USED IN CHAPTER 2

This section describes the filtering process used to construct the firm-level measures of abnormal return volatility *ABVAR* and trading turnover *ABT* as in [Chapter 2, section 4](#).

The dependent variables are countrywide cumulative absolute abnormal stock returns and trading volume calculated from daily firm-level data. Firm-level data are obtained through Compustat Global-Security Daily files through WRDS. Prior research using market data on non-U.S. firms documents that the quality of raw data provided by the vendor can be significantly improved by applying appropriate filtering and screening procedures (Gao, Parsons, and Shen 2018). Following earlier studies, several filters are applied at the issue, issuer, stock market and country level in order to obtain a final sample of stocks that exhibit similar characteristics, display sufficient data quality, and is representative of the national stock market (Amihud et al., 2015; Gao et al., 2018).

Following Gao et al. (2018), only common shares are retained and amongst these, stocks are dropped that do not correspond to the primary issue of the firm. Following Amihud et al. (2015), only stocks traded in EU stock exchanges and denominated in EU currencies including national currencies pre-Euro are retained. To obtain comparable estimates of the effect of country-specific fiscal guidance, only firms incorporated, located, and traded in the same jurisdiction are kept.

To obtain a usable returns and volume series, it is necessary to deal with the known problem that Compustat Global suffers from sparse coverage: some securities disappear from the sample only to reappear later (Gao et al., 2018). For example, simple net stock returns calculated as the percentage change in closing prices between two consecutive observations are daily returns if the consecutive observations are sampled over two consecutive trading days. Conversely, they are not daily returns if the firm disappears from the sample only to reappear several days later (i.e. if there are firm-specific gaps in the Compustat closing price series). Thus, market variables created as transformation of data

spanning multiple days may be sampled at different frequencies. Based on the analysis of the data, there are several cases where the gap exists, and in a non-trivial number of cases, this gap is large. To address this issue, for each firm a variable is generated that contains the difference between the calendar day number recorded by Compustat (0=Sunday; 1=Monday, 2=Tuesday, etc.) and the calendar day number of the previous observation. For properly spaced observations, this difference is equal to 1 for all days of the week with the exception of Monday, as its prior observation is Friday (recorded as day=5 by Compustat).

When the newly created variable is not equal to 1 or -4, therefore, either one of two scenarios occur. In the first scenario, the firm has a gap in the series. In the second scenario, the day preceding the focal day was a non-trading day. The two scenarios are fundamentally different, and it is important to retain stock returns that follow a non-trading day because while these returns are not daily, they are comparable across firms that trade in the same stock exchange. On the other hand, non-daily returns caused by a firm-specific gap in the Compustat series should not be retained. To distinguish between the two scenarios, an indicator variable is created that takes value 1 if the difference between two consecutive calendar days is either 1 or -4, and zero otherwise. Summing across this variable for each stock market-day pair and each day gives the number of observations that display a normal gap. This number is then compared to the number of firms that trade in that stock exchange on that specific day. When no firms trading in that stock exchange has a normal gap in a specific day, that day a non-trading day. Instead, observations are set to missing that do not have a normal gap and do not occur in a non-trading day. Returns and volume data are further set to missing in country-days with less than 10 valid observations. Next, a filter is imposed on the data before computing abnormal returns and trading volume. Firm-event pairs are set to missing if they have missing volume or return data for less than 80 percent of 21 days in the event window and 90 days in the estimation window. Finally, return and volume data are winsorized at the 1st and 99th percentile cutoff.

APPENDIX D. DETAILS ON THE CREDIT RATING INDUSTRY ANALYZED IN CHAPTER 3

The precursor to the modern credit rating industry was Louis Tappan who in 1841 began to rate the ability of merchants to honor their financial obligations. These ratings were based on the allocation of merchants to credit categories. Until the early 20th century, mercantile credit agencies sold their services to clients who used the information as a marker of value in making credit decisions. Thereafter, ratings spread from short-term trade credit to the much larger industry of corporate debt. Moody's began covering the U.S. railroad bonds market in 1909, the utilities and industrial bonds industry in 1910, and government debt in 1918, a practice that was soon followed by other leading agencies. By the 1930s, CRAs had achieved complete coverage of the debt market of advanced economies. After World War II, the market for sovereign rating came to a halt as governments replaced private-sector sources of funds with official credit lines granted by the Bretton Woods institutions. By the mid-70s, both S&P and Fitch had suspended their sovereign rating activity. This recovered only in the late 80s driven by renewed borrowing by Latin American nations, followed by bond issuances in the African region and the opening of the bond market in Eastern Europe after the collapse of the USSR. By 2011, the three largest CRAs each rated over a hundred governments, representing about half of all sovereign nations (Gaillard, 2012).

Based on official data, two agencies known as Standard and Poor's (S&P) and Moody's Investors Service (Moody's) currently dominate the credit rating industry. At end-2015, S&P and Moody's held a combined share of 76.3% of the European market for rating activities and ancillary services, followed by Fitch (16.6%) and DBRS (1.9%) (ESMA, 2016: 6). Across the various classes of financial instruments—debt issued by financial firms and nonfinancial corporations, structured finance products, sovereign governments, etc.—the market shares vary and in the US other CRAs command important positions in the market. In the EU, Fitch and DBRS are significant competitors of S&P and Moody's. At end-2015, these four agencies were the only ones (out of twenty-six registered CRAs) that provided rating services in all asset classes (ESMA, 2016).

Still, it is customary to describe the industry as a regulated oligopoly centered on the ratings of S&P and Moody's (EC, 2016). As Bruner and Abdelal (2005: 192) argued, "[n]o sovereign government would dare to issue debt without being rated by one or both of the agencies." While that statement is less true now than it used to be, the European Commission states that to-date "the EU-wide credit market does not show any trend toward decreasing concentration" (EC, 2016: 10).

Today, the credit rating industry is largely institutionalized (Ruef and Patterson, 2009). CRAs have been regulated in the US since 1975 and in the European Union (EU) since 2009, and the G20 meeting held in Sao Paulo in November 2008 included CRAs in the list of systemically important institutions. Since 2009, the European Securities and Markets Authority (ESMA) has been regulating both the ECB-approved and the other registered CRAs providing services in the EU (EU, 2010). Heightened regulatory scrutiny followed the alleged untimely and aggressive response by CRAs to the financial crisis and the subsequent European sovereign debt crisis, when ratings were significantly downgraded through a series of "bulk rating actions" (ESMA, 2013). For example, the downgrading of Greece by S&P following the financial assistance agreement between the IMF and European governments led to severe criticism of CRAs (Staikouras, 2012). These critiques led to calls for the establishment of a supervisory system of CRAs' operations in Europe and to the issuance of Regulation 1060/2009, known as CRA I (EU, 2009b). The Regulation acknowledged both the widespread view that CRAs "have failed, first, to reflect early enough in their credit ratings the worsening market conditions, and second, to adjust their credit ratings in time following the deepening market crisis." (Recital 10). The regulation also enshrined into law the systemic relevance of credit ratings, noting their "significant impact on the operation of the markets and on the trust and confidence of investors and consumers" (Recital 1).

CRA I implemented a European system of registration and mandatory licenses for CRAs to offer services within the EU borders. In particular, the regulation imposed that any EU entity that wished to adopt CRA ratings for regulatory purposes could only do so if the ratings used were either issued or endorsed by

EU-registered rating agencies (Articles 2 and 3). In practice, this provision forced US-based agencies to open branches or otherwise separate legal entities in EU jurisdictions to sell their services to EU-based investors.³⁶ CRA I further established requirements regarding rating methodologies, demanding that CRAs show the extent to which these are “rigorous, systematic, continuous and subject to validation based on historical experience, including backtesting” (Article 8), while also imposing the need to review methodologies at least once a year. The amendments set out in Regulation 513/2011 – known as CRA II - and especially Article 21 - transferred the supervision of CRAs to the European Securities and Markets Authority (ESMA) (EU, 2011). CRA II also clarified ESMA’s authority regarding the on-site inspection, regular monitoring of rating methodologies and imposition of fees and other penalties to CRAs that are found to infringe at least one of the provisions included in CRA I. The subsequent and final amendment to CRA I took place in 2013 with Regulation No 462/2013, known as CRA III (EU, 2013). CRA III included specific provisions regarding sovereign ratings because of the perceived market damage that CRAs can provoke when changing their assessment of governments’ creditworthiness.

Because of regulatory interventions, sovereign ratings are now subject to separate and additional provisions compared to ratings of other financial instruments. According to CRA III, CRAs issuing sovereign ratings must take into account in their rating assessments “the individual specificity of a particular Member State” of the EU (Article 8a). In addition, the dates on which a revision of sovereign ratings should take place have to be preannounced in a publicly available calendar released by CRAs by the end of December each year. Only on these dates can CRAs announce a rating action, except when material and unpredictable circumstances affect sovereign creditworthiness. These dates must occur on a Friday after market hours, and deviations are allowed only in exceptional circumstances.

³⁶ Following the Brexit vote, similar provisions are being put in place to preserve EU-based monitoring of CRAs selling their services in the EU. See the relevant ESMA statement at: https://www.esma.europa.eu/sites/default/files/library/esma33-5-735_public_statement.pdf. (Last access: 18 March 2019)

REFERENCES

- Abadie A, Athey S, Imbens G, et al. (2017) *When Should You Adjust Standard Errors for Clustering?* NBER Working Paper Series 24003.
- Abhayawansa S, Aleksanyan M and Bahtsevanoglou J (2015) The use of intellectual capital information by sell-side analysts in company valuation. *Accounting and Business Research* 45(3): 279–306.
- Afonso A, Gomes P and Rother P (2011) Short-and Long-Run Determinants of Sovereign Debt Credit Ratings. *International Journal of Finance and Economics* 16: 1–15.
- Agnello L and Sousa RM (2013) Fiscal policy and asset prices. *Bulletin of Economic Research* 65(2): 154–177.
- Ajinkya BB and Gift MJ (1984) Corporate Managers' Earnings Forecasts and Symmetrical Adjustments of Market Expectations. *Journal of Accounting Research* 22(2): 425–444.
- Alesina A and Perotti R (1996) Fiscal Discipline and the Budget Process. *American Economic Review* 86(2): 401–407.
- Alesina A, Perotti R and Tavares J (1998) The political economy of fiscal adjustments. *Brookings Papers on Economic Activity* 1: 197–266.
- Alesina A, Favero C and Giavazzi F (2015) The output effect of fiscal consolidation plans. *Journal of International Economics* 96: S19–S42.
- Alt JE and Lassen DD (2006) Fiscal transparency, political parties, and debt in OECD countries. *European Economic Review* 50: 1403–1439.
- Alt JE and Lowry RC (1994) Divided Government, Fiscal Institutions, and Budget Deficits: Evidence from the States. *American Political Science Review* 88(4): 811–828.
- Alt JE, Lassen DD and Rose S (2006) The Causes of Fiscal Transparency: Evidence from the U.S. States. *IMF Staff Papers* 53(Special Issue): 30–57.
- Alt JE, Lassen DD and Wehner J (2014) It isn't just about Greece: Domestic politics, transparency and fiscal gimmickry in Europe. *British Journal of Political Science* 44(4): 707–716.
- Amihud Y, Hameed A, Kang W, et al. (2015) The illiquidity premium: International evidence. *Journal of Financial Economics* 117(2): 350–368.
- Andersen AL, Lassen DD and Nielsen LHW (2014) The impact of late budgets on state government borrowing costs. *Journal of Public Economics* 109: 27–35.
- Anessi-Pessina E, Barbera C, Sicilia M, et al. (2016) Public sector budgeting : a European review of accounting and public management journals. *Accounting, Auditing & Accountability Journal* 29(3): 491–519.
- Angrist JD (2010) Treatment Effect. In: Durlauf SN and Blume LE (eds) *Microeconometrics. The New Palgrave Economics Collection*. London: Palgrave Macmillan, pp. 329–338.
- Angrist JD and Hahn J (2004) When to control for covariates? Panel asymptotics for estimates of treatment effects. *Review of Economics and Statistics* 86(1): 58–72.
- Angrist JD and Kuersteiner GM (2011) Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. *Review of Economics and Statistics* 93(3): 725–747.
- Angrist JD and Pischke J-S (2009) *Mostly Harmless Econometrics : An*

- Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Angrist JD, Jordà Ò and Kuersteiner GM (2018) Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited. *Journal of Business and Economic Statistics* 36(3): 371–387.
- Anilowski C, Feng M and Skinner DJ (2007) Does earnings guidance affect market returns? The nature and information content of aggregate earnings guidance. *Journal of Accounting and Economics* 44(1–2): 36–63.
- Arapis T and Reitano V (2018) Examining the Evolution of Cross-National Fiscal Transparency. *American Review of Public Administration* 48(6): 550–564.
- Ardagna S (2009) Financial markets' behavior around episodes of large changes in the fiscal stance. *European Economic Review* 53(1): 37–55.
- Armeliu H, Hull I and Stenbacka Köhler H (2017) The timing of uncertainty shocks in a small open economy. *Economics Letters* 155: 31–34.
- Austin PC and Stuart EA (2015) Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine* 34: 3661–3679.
- Badertscher B, Shroff N and White HD (2013) Externalities of public firm presence: Evidence from private firms' investment decisions. *Journal of Financial Economics* 109(3): 682–706.
- Bafundi A, Columbano C and Carrera N (2018) Divulgación de información por parte del gobierno y reacción del mercado de valores : análisis del caso español. *Cuadernos de Información Económica* 266: 59–72.
- Baker SR, Bloom N and Davis SJ (2016) Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics* 131(4): 1593–1636.
- Balakrishnan K, Billings MB, Kelly B, et al. (2014) Shaping Liquidity : On the Causal Effects of Voluntary Disclosure. *The Journal of Finance* 49(5): 2237–2278.
- Ball R and Shivakumar L (2008) How Much New Information Is There in Earnings? *Journal of Accounting Research* 46(5): 975–1016.
- Ball R, Sadka G and Sadka R (2009) Aggregate earnings and asset prices. *Journal of Accounting Research* 47(5): 1097–1133.
- Ballard-Rosa C, Mosley L and Wellhausen RL (2019) Contingent Advantage? Sovereign Borrowing, Democratic Institutions and Global Capital Cycles. *British Journal of Political Science*: 1–21.
- Bamber LS, Barron OE and Stober TL (1997) Trading volume and different aspects of disagreement coincident with earnings announcements. *Accounting Review* 72(4): 575–597.
- Bamber LS, Barron OE and Stober TL (1999) Differential Interpretations and Trading Volume. *Journal of Financial and Quantitative Analysis* 34(3): 369–386.
- Barker R, Hendry J, Roberts J, et al. (2012) Can company-fund manager meetings convey informational benefits? Exploring the rationalisation of equity investment decision making by UK fund managers. *Accounting, Organizations and Society* 37(4): 207–222.
- Barro RJ (1974) Are Government Bonds Net Wealth? *Journal of Political Economy* 82(6): 1095–1117.
- Barro RJ (1989) The Ricardian Approach to Budget Deficits. *Journal of*

- Economic Perspectives* 3(2): 37–54.
- Bastida F, Guillamón MD and Benito B (2017) Fiscal transparency and the cost of sovereign debt. *International Review of Administrative Sciences* 83(1): 106–128.
- Beaver WH (1968) The Information Content of Annual Earnings Announcements. *Journal of Accounting Research* 6(3): 67–92.
- Begg I (2015) Fiscal Policy Transparency. In: Forssbaeck J and Oxelheim L (eds) *The Oxford Handbook of Economic and Institutional Transparency*. Oxford University Press, pp. 98–115.
- Belo F and Yu J (2013) Government investment and the stock market. *Journal of Monetary Economics* 60(3): 325–339.
- Belo F, Gala VD and Li J (2013) Government spending, political cycles, and the cross section of stock returns. *Journal of Financial Economics* 107(2): 305–324.
- Benito B, Guillamón MD and Bastida F (2016) The Impact of Transparency on the Cost of Sovereign Debt in Times of Economic Crisis. *Financial Accountability & Management* 32(3): 309–334.
- Bernoth K and Wolff GB (2008) Fool the markets? creative accounting, fiscal transparency and sovereign risk premia. *Scottish Journal of Political Economy* 55(4): 465–487.
- Berthoin Antal A, Hutter M and Stark D (eds) (2015) *Moments of Valuation : Exploring Sites of Dissonance*. Oxford: Oxford University Press.
- Besedovsky N (2018) Financialization as calculative practice : the rise of structured finance and the cultural and calculative transformation of credit rating agencies. *Socio-Economic Review* 16(1): 61–84.
- Beunza D and Garud R (2007) Calculators, lemmings or frame-makers ? The intermediary role of securities analysts. *The Sociological Review* 55: 13–39.
- Beyer A, Cohen DA, Lys TZ, et al. (2010) The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics* 50(2–3): 296–343.
- Białkowski J, Gottschalk K and Wisniewski TP (2008) Stock market volatility around national elections. *Journal of Banking and Finance* 32(9): 1941–1953.
- Billings MB, Jennings R and Lev B (2015) On guidance and volatility. *Journal of Accounting and Economics* 60(2–3): 161–180.
- Blomberg J (2016) The multiple worlds of equity analysts: valuation, volume, and volatility. *Journal of Cultural Economy* 9(3): 277–295.
- Bolton P, Freixas X and Shapiro J (2012) The credit ratings game. *The Journal of Finance* 67(1): 85–112.
- Bonsall SB, Bozanic Z and Fischer PE (2013) What do management earnings forecasts convey about the macroeconomy? *Journal of Accounting Research* 51(2): 225–266.
- Borensztein E, Cowan K and Valenzuela P (2013) Sovereign ceilings ‘lite’? The impact of sovereign ratings on corporate ratings. *Journal of Banking and Finance* 37(11): 4014–4024.
- Borio CE V. and McCauley RN (1996) *The Economics of Recent Bond Yield Volatility*. BIS Economic Papers 45. Basel.

- Botosan CA and Plumlee MA (2002) A Re-examination of Disclosure Level and the Expected Cost of Equity Capital. *Journal of Accounting Research* 40(1): 21–40.
- Boudreaux DJ and Holcombe RG (1989) The Coasian and Knightian Theories of the Firm. *Managerial and Decision Economics* 10: 147–154.
- Brender A and Drazen A (2008) How Do Budget Deficits and Economic Growth Affect Reelection Prospects? Evidence from a Large Panel of Countries. *American Economic Review* 98(5): 2203–2220.
- Brogaard J and Detzel A (2015) The Asset-Pricing Implications of Government Economic Policy Uncertainty. *Management Science* 61(1): 3–18.
- Brown JL, Lin K, Moore JA, et al. (2017) *Tax Policy Uncertainty and Stock Return Volatility (January 20, 2017)*. Available at SSRN: <https://ssrn.com/abstract=2348414>.
- Bruner CM and Abdelal R (2005) To judge Leviathan: Sovereign credit ratings, national law, and the world economy. *Journal of Public Policy* 25(2): 191–217.
- Bushman RM and Landsman WR (2010) The pros and cons of regulating corporate reporting: A critical review of the arguments. *Accounting and Business Research* 40(3): 259–273.
- Bushman RM, Piotroski JD and Smith AJ (2004) What Determines Corporate Transparency? *Journal of Accounting Research* 42(2): 207–252.
- Busso M, DiNardo J and McCrary J (2014) New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators. *Review of Economics and Statistics* 96(5): 885–897.
- Callon M and Muniesa F (2005) Economic Markets as Calculative Collective Devices. *Organization Studies* 26(8): 1229–1250.
- Cantor R and Packer F (1994) The Credit Rating Industry. *Federal Reserve Bank of New York Quarterly Review* 19(2): 10–34.
- Cantor R and Packer F (1996) Determinants and Impact of Sovereign Credit Ratings. *Economic Policy Review - Federal Reserve Bank of New York* 2(2): 37–53.
- Carruthers BG (2013) From uncertainty toward risk: The case of credit ratings. *Socio-Economic Review* 11(3): 525–551.
- Caselli F and Wingender P (2018) *Bunching at 3 Percent: The Maastricht Fiscal Criterion and Government Deficits (August 2018)*. IMF Working Paper WP/18/182.
- Chae J (2005) Trading volume, information asymmetry, and timing information. *The Journal of Finance* 40(1): 413–442.
- Chaney BA, Copley PA and Stone MS (2002) The effect of fiscal stress and balanced budget requirements on the funding and measurement of state pension obligations. *Journal of Accounting and Public Policy* 21(4–5): 287–313.
- Chen SS, Chen HY, Chang CC, et al. (2016) The relation between sovereign credit rating revisions and economic growth. *Journal of Banking and Finance* 64: 90–100.
- Cheng RH (1992) An Empirical Analysis of Theories on Factors Influencing State Government Accounting Disclosure. *Journal of Accounting and Public Policy* 11: 1–42.

- Cicatiello L, De Simone E and Gaeta GL (2017) Political determinants of fiscal transparency: a panel data empirical investigation. *Economics of Governance* 18(4): 315–336.
- Citro F, Cuadrado-Ballesteros B and Bisogno M (2019) Explaining budget transparency through political factors. *International Review of Administrative Sciences* 0(0): 1–20. DOI: 10.1177/0020852319847511.
- Copelovitch M, Gandrud C and Hallerberg M (2018) Financial Data Transparency, International Institutions, and Sovereign Borrowing Costs. *International Studies Quarterly* 62: 23–41.
- Corsetti G, Kuester K, Meier A, et al. (2013) Sovereign Risk, Fiscal Policy, and Macroeconomic Stability. *Economic Journal* 123(566): F99–F132.
- Corwin SA and Schultz P (2012) A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance* 67(2): 719–760.
- Costello AM, Petacchi R and Weber JP (2017) The impact of balanced budget restrictions on states' fiscal actions. *Accounting Review* 92(1): 51–71.
- D'Agostino A and Lennkh A (2016) *Euro Area Sovereign Ratings: An Analysis of Fundamental Criteria and Subjective Judgement*. ESM Working Paper Series. 14.
- Darrat AF (1988) On Fiscal Policy and the Stock Market. *Journal of Money, Credit and Banking* 20(3): 353–363.
- Darrat AF (1990) Stock Returns, Money, and Fiscal Deficits. *The Journal of Financial and Quantitative Analysis* 25(3): 387–398.
- Daston L and Galison P (2010) *Objectivity*. Zone Books - MIT.
- DBRS (2016) *Rating Sovereign Governments*.
- de Renzio P and Wehner J (2017) The Impacts of Fiscal Openness. *The World Bank Research Observer* 32(2): 185–210.
- Desrosieres A (1998) *The Politics of Large Numbers : A History of Statistical Reasoning*. Cambridge, Massachusetts, and London, England: Harvard University Press.
- Desrosieres A (2001) How Real Are Statistics? Four Possible Attitudes. *Social Research* 68(2): 339–355.
- Dewey J (1943) Theory of Valuation. In: Neurath O, Carnap R, and Morris CW (eds) *International Encyclopedia of Unified Science - Foundations of the Unity of Science (Volumes I-II of the Encyclopedia)*. Chicago: The University of Chicago Press, pp. vii–65.
- Dye RA (1985) Disclosure of Nonproprietary Information. *Journal of Accounting Research* 23(1): 123–145.
- Dye RA (2001) An evaluation of 'essays on disclosure' and the disclosure literature in accounting. *Journal of Accounting and Economics* 32(1–3): 181–235.
- Dye RA (2017) Some recent advances in the theory of financial reporting and disclosures. *Accounting Horizons* 31(3): 39–54. DOI: 10.2308/acch-51717.
- EC (1996) *Ensuring Budgetary Discipline in Stage Three of EMU - Note for the Monetary Committee*. 19 July 1996. II/409/96-EN. Brussels.
- EC (2008) *A European Economic Recovery Plan*. Communication from the Commission to the European Council. 26.11.2008. COM(2008) 800 final. Brussels.
- EC (2016) *Study on the State of the Credit Rating Market Final Report*.

- MARKT/2014/257/F4/ST/OP. Brussels.
- Edmonds CT, Edmonds JE, Vermeer BY, et al. (2017) Does timeliness of financial information matter in the governmental sector? *Journal of Accounting and Public Policy* 36(2): 163–176.
- Eichengreen B and Wyplosz C (1998) The Stability Pact: More than a minor nuisance? *Economic Policy* 13(26): 65–113.
- Ellahie A and Ricco G (2017) Government purchases reloaded: Informational insufficiency and heterogeneity in fiscal VARs. *Journal of Monetary Economics* 90: 13–27.
- ESMA (2013) *Credit Rating Agencies. Sovereign ratings investigation. ESMA's assessment of governance, conflicts of interest, resourcing adequacy and confidentiality controls. 2 December 2013. ESMA/2013/1175.*
- ESMA (2016) *Competition and choice in the credit rating industry. Market share calculation required by Article 8d of Regulation 1060/2009 on Credit Rating Agencies as amended. 16 December 2016. ESMA/2016/1662.*
- ESMA (2019) *Report on CRA Market Share Calculation. 29 November 2019. ESMA33-9-340.*
- Espeland WN and Sauder M (2007) Rankings and Reactivity : How Public Measures Recreate Social Worlds. *American Journal of Sociology* 113(1): 1–40.
- Espeland WN and Stevens ML (1998) Commensuration as a Social Process. *Annual Review of Sociology* 24(1): 313–343.
- Ettmeier S, Kim CH and Kriwoluzky A (2020) Economic policy and financial market expectations during COVID-19. Available at: <https://voxeu.org/article/covid-19-economic-policy-and-market-expectations> (accessed 11 May 2020).
- Etzioni A (2010) Is Transparency the Best Disinfectant? *Journal of Political Philosophy* 18(4): 389–404.
- Etzioni A (2014) The Limits of Transparency. *Public Administration Review* 74(6): 687–688.
- EU (1997a) *Council Regulation (EC) No 1466/97 of 7 July 1997 on the strengthening of the surveillance of budgetary positions and the surveillance and coordination of economic policies. Official Journal of the European Communities. 2.8.97 L 209/1.*
- EU (1997b) *Council Regulation (EC) No 1467/97 of 7 July 1997 on speeding up and clarifying the implementation of the excessive deficit procedure. Official Journal of the European Communities. 2.8.97 L 209/6.*
- EU (1997c) *Resolution of the European Council on the Stability and Growth Pact. Amsterdam, 17 June 1997. Official Journal of the European Communities. 2.8.97 C 236/1.*
- EU (2005) *Draft Council Conclusions on the Revised Code of Conduct Relating to the Stability and Growth Pact. 3 October 2005. 12822/05. Brussels.*
- EU (2009a) *Press Release 2967th Council meeting Economic and Financial Affairs. 20 October 2009. 14601/09 (Presse 294). Luxembourg.*
- EU (2009b) *Regulation (EC) No 1060/2009 of the European Parliament and of the Council of 16 September 2009 on credit rating agencies. Official Journal of the European Union 17.11.2009 L 302/1.*
- EU (2010) *Regulation (EU) No 1095/2010 of the European Parliament and of*

- the Council of 24 November 2010 establishing a European Supervisory Authority (European Securities and Markets Authority), amending Decision No 716/2009/EC and repealing Commission Decision 20.
- EU (2011) Regulation (EU) No 513/2011 of the European Parliament and of the Council of 11 May 2011 amending Regulation (EC) No 1060/2009 on credit rating agencies. *Official Journal of the European Union* 31.5.2011 L 145/30.
- EU (2013) Regulation (EU) No 462/2013 of the European Parliament and of the Council of 21 May 2013 amending Regulation (EC) No 1060/2009 on credit rating agencies. *Official Journal of the European Union* 31.5.2013 L 146/1.
- European Fiscal Board (2019) *Assessment of EU fiscal rules with a focus on the six and two-pack legislation*.
- Farlow A (2015) Financial indicators and the global financial crash. In: Rottenburg R, Merry SE, Park S-J, et al. (eds) *The World of Indicators : The Making of Governmental Knowledge through Quantification - Cambridge Studies in Law and Society*. Cambridge: Cambridge University Press, pp. 220–253.
- Fernández-Villaverde J, Guerrón-Quintana P, Kuester K, et al. (2015) Fiscal volatility shocks and economic activity. *American Economic Review* 105(11): 3352–3384.
- Ferreira MA and Gama PM (2007) Does sovereign debt ratings news spill over to international stock markets? *Journal of Banking and Finance* 31(10): 3162–3182.
- Ferri G, Liu L-G and Stiglitz JE (1999) The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis. *Economic Notes by Banca Monte dei Paschi di Siena SpA* 28(3): 335–355.
- Fitch (2016a) *Reply form for the Consultation Paper on Guidelines on validation and review of Credit Rating Agencies ' methodologies*.
- Fitch (2016b) *Sovereign Rating Criteria*.
- Flick U (2009) *An Introduction to Qualitative Research*. 4th ed. London: Sage.
- Flyvbjerg B (2006) Five Misunderstandings about case-study research. *Qualitative Inquiry* 12(2): 219–245.
- Foss NJ (1993) More on Knight and the Theory of the Firm. *Managerial and Decision Economics* 14: 269–276.
- Fourcade M and Healy K (2013) Classification situations: Life-chances in the Neoliberal Era. *Accounting, Organizations and Society* 38: 559–572.
- Frankel J (2011) Over-optimism in forecasts by official budget agencies and its implications. *Oxford Review of Economic Policy* 27(4): 536–562.
- Frankel J and Schreger J (2013) Over-optimistic official forecasts and fiscal rules in the eurozone. *Review of World Economics* 149(2): 247–272.
- Frieden JA (1988) Capital Politics : Creditors and the International Political Economy. *Journal of Public Policy* 8(3/4): 265–286.
- Frieden JA (1991) Invested Interests: The Politics of National Economic Policies in a World of Global Finance. *International Organization* 45(4): 425–451.
- Frieden JA (2016) The Governance of International Finance. *Annual Review of Political Science* 19(1): 33–48.
- Froud J (2003) The Private Finance Initiative : risk, uncertainty and the state. *Accounting, Organizations and Society* 28: 567–589.

- Gabriele C, Erce A, Athanasopoulou M, et al. (2017) *Debt Stocks Meet Gross Financing Needs: A Flow Perspective into Sustainability*. ESM Working Paper Series. 24.
- Gaillard N (2012) *A Century of Sovereign Ratings*. New York: Springer.
- Gale WG and Orszag PR (2003) Economic Effects of Sustained Budget Deficits. *National Tax Journal* 46(3): 463–485.
- Gan Y and Qiu B (2019) Escape from the USA: Government debt-to-GDP ratio, country tax competitiveness, and US-OECD cross-border M&As. *Journal of International Business Studies* 50: 1156–1183.
- Gande A and Parsley DC (2005) News spillovers in the sovereign debt market. *Journal of Financial Economics* 75(3): 691–734.
- Gao P, Parsons CA and Shen J (2018) Global relation between financial distress and equity returns. *Review of Financial Studies* 31(1): 239–277.
- Garfinkel JA (2009) Measuring investors' opinion divergence. *Journal of Accounting Research* 47(5): 1317–1348.
- Garfinkel JA and Sokobin J (2006) Volume, opinion divergence, and returns: A study of post-earnings announcement drift. *Journal of Accounting Research* 44(1): 85–112.
- Gärtner M, Griesbach B and Jung F (2011) PIGS or Lambs? The European Sovereign Debt Crisis and the Role of Rating Agencies. *International Advances in Economic Research* 17(3): 288–299.
- Gelos RG and Wei S (2005) Transparency and International Portfolio Holdings. *The Journal of Finance* 40(6): 2987–3020.
- Giavazzi F, Jappelli T and Pagano M (2000) Searching for non-linear effects of fiscal policy: Evidence from industrial and developing countries. *European Economic Review* 44(7): 1259–1289.
- Gietzmann MB and Ostaszewski AJ (2014) Why managers with low forecast precision select high disclosure intensity: An equilibrium analysis. *Review of Quantitative Finance and Accounting* 43(1): 121–153.
- Giorgi S and Weber K (2015) Marks of Distinction: Framing and Audience Appreciation in the Context of Investment Advice. *Administrative Science Quarterly* 60(2): 333–367.
- Glaser B and Strauss AL (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine.
- Glennerster R and Shin Y (2008) Does Transparency Pay? *IMF Staff Papers* 55(1): 183–209.
- Gore AK (2004) The effects of GAAP regulation and bond market interaction on local government disclosure. *Journal of Accounting and Public Policy* 23(1): 23–52.
- Graham JR, Harvey CR and Rajgopal S (2005) The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40: 3–73.
- Gulati M and Scott RE (2013) *The Three and a Half Minute Transaction: Boilerplate and the Limits of Contract Design*. Chicago: University of Chicago Press.
- Guseva A and Rona-Tas A (2001) Uncertainty, Risk, and Trust: Russian and American Credit Card Markets Compared. *American Sociological Review* 66(5): 623–646.

- Hainmueller J (2012) Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20(1): 25–46.
- Hallerberg M, Strauch R and von Hagen J (2009) *Fiscal Governance in Europe*. New York: Cambridge University Press.
- Hardouvelis GA, Karalas GI, Karanastasis DI, et al. (2018) Economic Policy Uncertainty, Political Uncertainty and the Greek Economic Crisis (second draft). Available at: https://www.policyuncertainty.com/media/HKKS_Greece_EPU.pdf.
- Hayn C (1995) The information content of losses. *Journal of Accounting and Economics* 20: 125–153.
- Heald D (2006) Transparency as an Instrumental Value. *Transparency: The Key to Better Governance?* 135: 59–73.
- Heald D and Hodges R (2018) Accounting for government guarantees: perspectives on fiscal transparency from four modes of accounting. *Accounting and Business Research* 48(7): 782–804.
- Hemming R (2013) The Macroeconomic Framework for Managing Public Finances. In: Allen R, Hemming R, and Potter BH (eds) *The International Handbook of Public Financial Management*. Palgrave Macmillan, pp. 17–37.
- Hernán MA and Robins JM (2020) *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.
- Hill P, Brooks R and Faff R (2010) Variations in sovereign credit quality assessments across rating agencies. *Journal of Banking and Finance* 34: 1327–1343.
- Hirano K, Imbens GW and Ridder G (2003) Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71(4): 1161–1189.
- Hirst DE, Koonce L and Venkataraman S (2007) How disaggregation enhances the credibility of management earnings forecasts. *Journal of Accounting Research* 45(4): 811–837.
- Hirst DE, Koonce L and Venkataraman S (2008) Management Earnings Forecasts : A Review and Framework. *Accounting Horizons* 22(3): 315–338.
- Holthausen RW and Verrecchia RE (1990) The Effect of Informedness and Consensus on Price and Volume Behavior. *Accounting Review* 65(1): 191–208.
- Hood C (2006) Beyond Exchanging First Principles? Some Closing Comments. *Transparency: The Key to Better Governance?* 135: 211–225.
- Hood C (2007) What happens when transparency meets blame-avoidance? *Public Management Review* 9(2): 191–210.
- Hood C and Heald D (eds) (2006) *Transparency: The Key to Better Governance?* Series: Proceedings of the British Academy (135). Oxford, U.K.: Oxford University Press for The British Academy.
- Hoppe H-H (2007) The Limits of Numerical Probability: Frank H. Knight and Ludwig von Mises and the Frequency Interpretation. *The Quarterly Journal of Austrian Economics* 10(1): 1–20.
- House of Commons Information Office (2010) Budgets and Financial

- Documents - Factsheet P5 Procedure Series Revised August 2010.
- House of Commons Library (2018) Budget debates and Finance Bills since 1968. *Brief Paper Number 02271, 29 October 2018*.
- Imai L and Ratkovic M (2014) Covariate Balancing Propensity Score. *Journal of the Royal Statistical Society* 76(1): 243–263.
- Imam S and Spence C (2016) Context, not predictions: a field study of financial analysts. *Accounting, Auditing & Accountability Journal* 29(2): 226–247.
- IMF (1998) Code of Good Practices on Fiscal Transparency - Declaration on Principles. 16 April.
- IMF (2014) *Government Finance Statistics Manual*. International Monetary Fund.
- Ingram RW (1984) Economic Incentives and the Choice of State Government Accounting Practices. *Journal of Accounting Research* 22(1): 126–144.
- Jeacle I and Carter C (2011) In TripAdvisor we trust : Rankings , calculative regimes and abstract systems. *Accounting, Organizations and Society* 36: 293–309.
- Johansson J (2007) Sell-side analysts' creation of value – key roles and relational capital. *Journal of Human Resource Costing & Accounting* 11(1): 30–52.
- Jollineau SJ, Tanlu LJ and Winn A (2014) Evaluating proposed remedies for credit rating agency failures. *Accounting Review* 89(4): 1399–1420.
- Jordà Ò and Taylor AM (2016) The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy. *Economic Journal* 126(590): 219–255.
- Jung W-O and Kwon YK (1988) Disclosure when the Market is Unsure of Information Endowment of Managers. *Journal of Accounting Research* 26(1): 146–153.
- Kandel E and Pearson ND (1995) Differential Interpretation of Public Signals and Trade in Speculative Markets. *Journal of Political Economy* 103(4): 831–872.
- Karpik L (2010) *Valuing the Unique : The Economics of Singularities*. Princeton: Princeton University Press.
- Kato K, Skinner DJ and Kunimura M (2009) Management forecasts in Japan: An empirical study of forecasts that are effectively mandated. *Accounting Review* 84(5): 1575–1606.
- Khagram S, de Renzio P and Fung A (2013) Overview and Synthesis: The Political Economy of Fiscal Transparency, Participation, and Accountability around the World. In: Khagram S, Fung A, and de Renzio P (eds) *Open Budgets: The Political Economy of Transparency, Participation, and Accountability*. Brookings Institution Press, p. 275.
- Khalil EL (1997) Chaos theory versus Heisenberg' s uncertainty: Risk, uncertainty and economic theory. *American Economist* 41(2): 27–40.
- Kido N, Petacchi R and Weber JP (2012) The Influence of Elections on the Accounting Choices of Governmental Entities. *Journal of Accounting Research* 50(2): 443–476.
- Kim O and Verrecchia RE (1991a) Market reaction to anticipated announcements. *Journal of Financial Economics* 30: 273–309.
- Kim O and Verrecchia RE (1991b) Trading Volume and Price Reactions to Public Announcements. *Journal of Accounting Research* 29(2): 302–321.

- Kim O and Verrecchia RE (1994) Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics* 17: 41–67.
- Kim WJ, Plumlee MA and Stubben SR (2018) Kim, Won Jung and Plumlee, Marlene A. and Stubben, Stephen, *Overview of U.S. State and Local Government Financial Reporting: A Reference for Academic Research* (February 29, 2020). Available at SSRN: <https://ssrn.com/abstract=3286448>.
- Klein PG (2009) Risk, Uncertainty, and Economic Organization. In: Hülsmann JG and Kinsella S (eds) *Property, Freedom, and Society: Essays in Honor of Hans-Hermann Hoppe*. Ludwig von Mises Institute, pp. 325–337.
- Knight FH (1946) Immutable Law in Economics: Its Reality and Limitations. *American Economic Review* 36(2): 93–111.
- Knight FH (1964) *Risk, Uncertainty and Profit*. New York: Augustus M. Kelly, Bookseller.
- Knights D and Vurdubakis T (1993) Calculations of Risk: Towards an Understanding of Insurance as a Moral and Political Technology. *Accounting, Organizations and Society* 18(7/8): 729–764.
- Knorr Cetina K (2011) Financial Analysis: Epistemic Profile of an Evaluative Science. In: Camic C, Gross N, and Lamont M (eds) *Social Knowledge in the Making*. Chicago and London: The University of Chicago Press.
- Kok S, Kroese L and Parlevliet J (2015) Beleidsonzekerheid in Nederland” [“Policy uncertainty in the Netherlands”]. *Economisch-Statistische Berichten (ESB)*.
- Konchitchki Y (2016) Accounting Valuation and Cost of Capital Dynamics: Theoretical and Empirical Macroeconomic Aspects. Discussion of Callen. *Abacus* 52(1): 26–34.
- Kopits G and Craig J (1998) *Transparency in Government Operations*. Occasional Paper. International Monetary Fund.
- Kornberger M and Carter C (2010) Manufacturing competition: How accounting practices shape strategy making in cities. *Accounting, Auditing & Accountability Journal* 23(3): 325–349.
- Kornberger M, Justesen L, Madsen AK, et al. (eds) (2015) *Making Things Valuable*. Oxford University Press.
- Kothari SP and Warner JB (2007) Econometrics of Event Studies. In: Eckbo BE (ed.) *Handbook of Corporate Finance, Empirical Corporate Finance: Volume 1*. Elsevier/North-Holland, pp. 1–36.
- Kreiner K and Mouritsen J (2005) The analytical interview. Relevance beyond reflexivity. In: Tengblad S, Solli R, and Czarniawska B (eds) *The Art of Science*. Liber & Copenhagen Business School Press, pp. 153–176.
- Landsman WR and Maydew EL (2002) Has the information content of quarterly earnings announcements declined in the past three decades? *Journal of Accounting Research* 40(3): 797–808.
- Lansford B, Lev B and Wu Tucker J (2013) Causes and Consequences of Disaggregating Earnings Guidance. *Journal of Business Finance and Accounting* 40(1–2): 26–54.
- Laubach T (2009) New Evidence on the Interest Rate Effects of Budget Deficits and Debt. *Journal of the European Economic Association* 7(4): 858–885.
- Leblond P (2006) The political Stability and Growth Pact is dead: Long live the

- economic Stability and Growth Pact. *Journal of Common Market Studies* 44(5): 969–990.
- Leeper EM (2009) *Anchoring Fiscal Expectations*. NBER Working Paper Series 15269.
- LeRoy SF and Singell LD (1987) Knight on Risk and Uncertainty. *Journal of Political Economy* 95(2): 394–406.
- Leuz C and Wysocki PD (2016) The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research. *Journal of Accounting Research* 54(2): 525–622.
- Leuz C, Triantis A and Yue T (2008) Why do firms go dark? Causes and economic consequences of voluntary SEC deregistrations. *Journal of Accounting and Economics* 45: 181–208.
- Li N, Richardson S and Tuna I (2014) Macro to micro: Country exposures, firm fundamentals and stock returns. *Journal of Accounting and Economics* 58(1): 1–20.
- Lukka K and Vinnari E (2014) Domain theory and method theory in management accounting research. *Accounting, Auditing & Accountability Journal* 27(8): 1308–1338.
- Lunceford JK and Davidian M (2004) Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine* 23(19): 2937–2960.
- Lyu C, Wang K, Zhang F, et al. (2018) GDP management to meet or beat growth targets. *Journal of Accounting and Economics* 66(1): 318–338.
- Mallard G and Pénét P (2013) Seeing Like a Credit Rating Agency: The Constitution of Financial Uncertainties during the Greek Sovereign Debt Crisis. In: Hawthorne C (ed.) *Financial Crises: Identification, Forecasting and Effects on Transition Economies*. New York: Nova Publishers, pp. 164–174.
- March JG, Sproull LS and Tamuz M (1991) Learning from Samples of One or Fewer. *Organization Science* 2(1): 1–13.
- Mathis J, McAndrews J and Rochet J-C (2009) Rating the raters: Are reputation concerns powerful enough to discipline rating agencies? *Journal of Monetary Economics* 56(5): 657–674.
- McKinney J (1977) Frank H. Knight on Uncertainty and Rational Action. *Southern Economic Journal* 43(4): 1438–1452.
- McMenamin I, Breen M and Muñoz-Portillo J (2015) Austerity and credibility in the Eurozone. *European Union Politics* 16(1): 45–66.
- McMullin JL and Schonberger B (2020) Entropy-balanced discretionary accruals. *Review of Accounting Studies* 25: 84–119.
- Mehrpouya A and Salles-Djelic ML (2019) Seeing like the market; exploring the mutual rise of transparency and accounting in transnational economic and market governance. *Accounting, Organizations and Society* 76: 12–31.
- Mehrpouya A and Samiolo R (2016) Performance measurement in global governance: Ranking and the politics of variability. *Accounting, Organizations and Society* 55: 12–31.
- Mercer M (2004) How Do Investors Assess the Credibility of Management Disclosures. *Accounting Horizons* 18(3): 185–196.
- Merkley KJ, Bamber LS and Christensen TE (2013) Detailed management

- earnings forecasts: Do analysts listen? *Review of Accounting Studies* 18: 479–521.
- Merola R and Pérez JJ (2013) Fiscal forecast errors: Governments versus independent agencies? *European Journal of Political Economy* 32: 285–299.
- Meuser M and Nagel U (2009) The Expert Interview and Changes in Knowledge Production. In: Bogner A, Littig B, and Menz W (eds) *Interviewing Experts*. Palgrave Macmillan, pp. 17–42.
- Mikes A (2011) From counting risk to making risk count: Boundary-work in risk management. *Accounting, Organizations and Society* 36: 226–245.
- Milesi-Ferretti GM (2004) Good, bad or ugly? On the effects of fiscal rules with creative accounting. *Journal of Public Economics* 88(1–2): 377–394.
- Milgrom P and Stokey N (1980) Information, trade, and common knowledge. *Journal of Economic Theory* 26: 17–27.
- Moody's (2016) *Sovereign Bond Ratings. Rating Methodology*.
- Moody's (2020) *Rating symbols and definitions*.
- Mora N (2006) Sovereign credit ratings: Guilty beyond reasonable doubt? *Journal of Banking and Finance* 30(7): 2041–2062.
- Mosley L (2003a) Attempting global standards: National governments, international finance, and the IMF's data regime. *Review of International Political Economy* 10(2): 331–362.
- Mosley L (2003b) *Global Capital and National Governments*. Cambridge University Press.
- Mosley L and Singer DA (2008) Taking Stock Seriously: Equity-Market Performance, Government Policy, and Financial Globalization. *International Studies Quarterly* 52: 405–425.
- Mousavi S and Gigerenzer G (2014) Risk, uncertainty, and heuristics. *Journal of Business Research* 67(8): 1671–1678.
- Muniesa F (2011) A flank movement in the understanding of valuation. *The Sociological Review* 59: 24–38.
- Nagar V, Schoenfeld J and Wellman L (2019) The effect of economic policy uncertainty on investor information asymmetry and management disclosures. *Journal of Accounting and Economics* 67(1): 36–57.
- Nagel T (1986) *The View from Nowhere*. Oxford: Oxford University Press.
- Naughton J and Spamann H (2015) Fixing Public Sector Finances: The Accounting and Reporting Lever. *UCLA Law Review* 62(3): 572–620.
- Naughton J, Petacchi R and Weber JP (2015) Public pension accounting rules and economic outcomes. *Journal of Accounting and Economics* 59(2–3): 221–241.
- OECD (2002) OECD Best Practices for Budget Transparency. *OECD Journal on Budgeting* 1(3): 7–14.
- Paredes J, Pérez JJ and Perez-Quiros G (2015) *Fiscal targets. A guide to forecasters?* European Central Bank Working Paper Series No 1834 (August 2015) 1834.
- Pástor L and Veronesi P (2013) Political uncertainty and risk premia. *Journal of Financial Economics* 110(3): 520–545.
- Patell JM (1976) Corporate Forecasts of Earnings Per Share and Stock Price Behavior: Empirical Test. *Journal of Accounting Research* 14(2): 246.

- Patton MQ (2002) *Qualitative Research & Evaluation Methods*. 3rd ed. Thousand Oaks: Sage Publications.
- Pénet P (2015) Rating Reports as Figuring Documents: How Credit Rating Agencies Build Scenarios of the Future. In: Kornberger M, Justesen L, Mouritsen J, et al. (eds) *Making Things Valuable*, pp. 62–88.
- Penno MC (1997) Information Quality and Voluntary Disclosure. *Accounting Review* 72(2): 275–284.
- Pina AM and Venes NM (2011) The political economy of EDP fiscal forecasts: An empirical assessment. *European Journal of Political Economy* 27(3): 534–546.
- Pollock N and D’Adderio L (2012) Give me a two-by-two matrix and I will create the market: Rankings, graphic visualisations and sociomateriality. *Accounting, Organizations and Society* 37(8): 565–586.
- Poon M (2009) From new deal institutions to capital markets: Commercial consumer risk scores and the making of subprime mortgage finance. *Accounting, Organizations and Society* 34(5): 654–674.
- Porter TM (1994) Making things quantitative. *Science in Context* 7(3): 389–407.
- Porter TM (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Porter TM (2015) The flight of the indicator. In: Rottenburg R, Merry SE, Park S-J, et al. (eds) *The World of Indicators: The Making of Governmental Knowledge through Quantification - Cambridge Studies in Law and Society*. Cambridge: Cambridge University Press, pp. 34–55.
- Poterba J (1994) State Responses to Fiscal Crises: The Effects of Budgetary Institutions and Politics. *The Journal of Political Economy*.
- Power M (1997) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Power M (2004) Counting, control and calculation: Reflections on measuring and management. *Human Relations* 57(6): 765–783.
- Prat A (2006) The More Closely We Are Watched, the Better we Behave? In: Hood C and Heald D (eds) *Transparency: The Key to Better Governance?* Series: Proceedings of the British Academy (135). Oxford, U.K.: Oxford University Press for The British Academy, pp. 91–103.
- Qu SQ and Dumay J (2011) The qualitative research interview. *Qualitative Research in Accounting & Management* 8(3): 238–264.
- Rajan RG (1992) Insiders and Outsiders: The Choice between Informed and Arm’s-Length Debt. *The Journal of Finance* 47(4): 1367–1400.
- Ramey VA (2011) Can Government Purchases Stimulate the Economy? *Journal of Economic Literature* 49(3): 673–685.
- Reck JL, Wilson ER, Gotlob D, et al. (2004) Governmental Capital Markets Research in Accounting: A Review, Extension, and Directions for Future Research. *Research in Governmental and Nonprofit Accounting* 11.
- Ricci O (2015) The impact of monetary policy announcements on the stock price of large European banks during the financial crisis. *Journal of Banking and Finance* 52: 245–255.
- Ríos AM, Bastida F and Benito B (2016) Budget Transparency and Legislative Budgetary Oversight: An International Approach. *American Review of Public Administration* 46(5): 546–568.

- Robins JM, Rotnitzky A and Zhao LP (1994) Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 89(427): 846–866.
- Rodríguez Bolívar MP, Alcaide Muñoz L and López Hernández AM (2013) Determinants of Financial Transparency in Government. *International Public Management Journal* 16(4): 557–602.
- Rogers JL and Stocken PC (2005) Credibility of Management Forecasts. *Accounting Review* 80(4): 1233–1260.
- Rogers JL and Van Buskirk A (2013) Bundled forecasts in empirical accounting research. *Journal of Accounting and Economics* 55(1): 43–65.
- Rosenbaum PR and Rubin DB (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70(1): 41–55.
- Rottenburg R, Merry SE, Park S-J, et al. (eds) (2015) *The World of Indicators : The Making of Governmental Knowledge through Quantification - Cambridge Studies in Law and Society*. Cambridge: Cambridge University Press.
- Rubin DB (1974) Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* 66(5): 688–701.
- Rubin DB (1977) Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics* 2(1): 1–26.
- Rubin DB (2007) The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine* 26: 20–36.
- Ruef M and Patterson K (2009) Credit and Classification: The Impact of Industry Boundaries in Nineteenth-Century America. *Administrative Science Quarterly* 54(3): 486–520.
- S&P (2011) *General criteria: Principles of credit ratings*.
- S&P (2014a) *Sovereign Ratings Methodology*.
- S&P (2014b) *What are credit ratings and how do they work?*
- S&P (2015) *RatingsDirect. How We Rate Sovereigns*.
- Samiolo R (2012) Commensuration and styles of reasoning: Venice, cost-benefit, and the defence of place. *Accounting, Organizations and Society* 37(6): 382–402.
- Savage JD (2005) *Making the EMU : The Politics of Budgetary Surveillance and the Enforcement of Maastricht*. Oxford, U.K.: Oxford University Press.
- Schultz M, Mouritsen J and Gabrielsen G (2001) Sticky Reputation: Analyzing a Ranking System. *Corporate Reputation Review* 4(1): 24–41.
- Scope Ratings (2019a) *Public Finance Sovereign Ratings. Rating Methodology*.
- Scope Ratings (2019b) *Rating Definitions - Credit Ratings and Ancillary Services*.
- SEC (2008a) *Proposed Rules for Nationally Recognized Statistical Rating Organizations*.
- SEC (2008b) *Summary Report of Issues Identified in the Commission Staff's Examinations of Select Credit Rating Agencies*.
- Shroff N, Verdi RS and Yu G (2014) Information environment and the investment decisions of multinational corporations. *Accounting Review* 89(2): 759–790.

- Shroff N, Verdi RS and Yost BP (2017) When does the peer information environment matter? *Journal of Accounting and Economics* 64(2–3): 183–214.
- Sorić P and Lolić I (2017) Economic uncertainty and its impact on the Croatian economy. *Public Sector Economics* 41(4): 443–477.
- Spence C, Aleksanyan M, Millo Y, et al. (2019) Earning the ‘Write to Speak’ : Sell-Side Analysts and Their Struggle to Be Heard. *Contemporary Accounting Research* 36(4): 2635–2662.
- Staikouras PK (2012) A Theoretical and Empirical Review of the EU Regulation on Credit Rating Agencies: In Search of Truth, Not Scapegoats. *Financial Markets, Institutions & Instruments* 21(2): 71–155.
- Stiglitz JE (1999) On Liberty, the Right to Know, and Public Discourse: The Role of Transparency in Public Life. Oxford Amnesty Lecture. January 27, 1999. Oxford, UK.
- Suijs J (2007) Voluntary disclosure of information when firms are uncertain of investor response. *Journal of Accounting and Economics* 43(2–3): 391–410.
- Tavares J and Valkanov R (2003) Fiscal Policy and Asset Returns. Available at: <https://rady.ucsd.edu/faculty/directory/valkanov/pub/docs/fiscal-policy.pdf>.
- The World Bank (1992) *Governance and Development*. Washington, D.C.
- Themsen TN and Skærbæk P (2018) The performativity of risk management frameworks and technologies : The translation of uncertainties into pure and impure risks. *Accounting, Organizations and Society* 67: 20–33.
- Thorbecke W (1993) Why deficit news affects interest rates. *Journal of Policy Modeling* 15(1): 1–11.
- Urquiza FB, Abad Navarro MC, Trombetta M, et al. (2010) Disclosure theories and disclosure measures. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad* 39(147): 393–420.
- Vermeer TE, Styles AK and Patton TK (2012) Do local governments present required disclosures for defined benefit pension plans? *Journal of Accounting and Public Policy* 31(1): 44–68.
- Vernazza DR and Nielsen EF (2015) The Damaging Bias of Sovereign Ratings. *Economic Notes by Banca Monte dei Paschi di Siena SpA* 44(2): 361–407.
- Verrecchia RE (1990) Information Quality and Discretionary Disclosure. *Journal of Accounting and Economics* 12: 365–380.
- Verrecchia RE (2001) Essays on disclosure. *Journal of Accounting and Economics* 32(1–3): 97–180.
- von Hagen J and Wolff GB (2006) What do deficits tell us about debt? Empirical evidence on creative accounting with fiscal rules in the EU. *Journal of Banking and Finance* 30(12): 3259–3279.
- Wachtel P and Young J (1987) Deficit Announcements and Interest Rates. *American Economic Review* 77(5): 1007–1012.
- Wehner J and de Renzio P (2013) Citizens, Legislators, and Executive Disclosure: The Political Determinants of Fiscal Transparency. *World Development* 41(1): 96–108.
- White LJ (2010) Markets: The Credit Rating Agencies. *Journal of Economic Perspectives* 24(2): 211–226.

- Williams A (2015) A global index of information transparency and accountability. *Journal of Comparative Economics* 43(3): 804–824.
- Wisniewski TP and Moro A (2014) When EU Leaders Speak, the Markets Listen. *European Accounting Review* 23(4): 519–551.
- Zalla R (2016) Economic Policy Uncertainty in Ireland. Available at: https://www.policyuncertainty.com/media/Zalla_EPU_Ireland.pdf.
- Zhao B and Wang W (2006) *Transparency in State Debt Disclosure*. Federal Reserve Bank of Boston Working Papers 17–10.
- Zuckerman EW (1999) The categorical imperative: Securities analysts and the illegitimacy discount. *American Journal of Sociology* 104(5): 1398–1438.

