



UNIVERSITY OF SALERNO
DEPARTMENT OF ECONOMICS AND STATISTICS

Ph.D. Programme in
ECONOMICS AND POLICIES OF MARKETS AND FIRMS
Curriculum: *Statistical Methods*
Cycle XXXIV

Doctoral Thesis in

**Statistical Procedures to Predict
Economic Indicators through
Political Discourse Text Analysis**

Author

Dr. Alessandro GRIMALDI

Supervisor

Prof. Alessandra AMENDOLA

Coordinator

Prof. Alessandra AMENDOLA

A.Y. 2020/2021

Abstract

The aim of the thesis is to contribute to the field of textual analysis in economics by proposing text-based indices to quantify the impact of political debate on the economy and improve the forecast accuracy of traditional quantitative measures used to assess the state of the economy.

The thesis analyses parliamentary debates transcripts in the hypothesis that such type of texts may convey some qualitative, timelier, information about new regulations, benefits and taxes, public expenditure or debt, incentives towards certain technologies, and job market rules which are able to condition the expectations of firms, consumers, employees, their preferences and choices, hence the entire economy.

Chapter 2 and Chapter 3 are the main two parts of the thesis and are preceded by an Introduction explaining aim and motivation of the work. The Conclusions, instead, are dedicated to the discussion of the main contributions on possible future developments.

More in detail, Chapter 2 is concerned with Topic Modelling. After reviewing the main literature and applications, this chapter describes the step to analyse texts through a specific topic model algorithm and how its output is utilised to construct the proposed Textual Political Debate Indices. The chapter also describes the evaluation of the indices through a Model Confidence Set approach.

Chapter 3 is concerned with Sentiment Analysis and, after a review of the main literature and applications, describes a new method to extract words polarity in a data driven manner. The chapter also describes the procedure to construct and evaluate the proposed Textual Political Polarity Indices which reveal to convey a non-trivial information able to improve forecast accuracy compared to the benchmark autoregressive model.

Abstract

Lo scopo della tesi è contribuire al filone dell'analisi testuale in economia mediante la proposta di indicatori testuali in grado di quantificare l'impatto del dibattito politico sull'economia e migliorare l'accuratezza previsiva delle tradizionali misure quantitative impiegate per valutare lo stato dell'economia.

La tesi analizza i resoconti stenografici di dibattiti parlamentari ipotizzando che tale tipologia di testi possa racchiudere informazioni qualitative e tempistiche circa nuove regole, tasse e benefici, spesa pubblica e debito, incentivi verso certe tecnologie, e regole del mercato del lavoro le quali siano in grado di condizionare le attese di imprese, consumatori, lavoratori, le loro preferenze e scelte e, quindi, l'intera economia.

I Capitoli 2 e 3 costituiscono le parti principali della tesi e sono preceduti da una Introduzione la quale spiega motivazione e scopo del lavoro. Le Conclusioni, invece, sono dedicate alla discussione dei principali risultati e possibili futuri sviluppi.

Più in dettaglio, il Capitolo 2 è incentrato sul Topic Modelling. Dopo una rassegna della principale letteratura e applicazioni, il capitolo descrive i passi per analizzare i testi con uno specifico algoritmo e come il prodotto di tale algoritmo è impegnato per costruire gli Indici Testuali di Dibattito Politico proposti. Il capitolo si sofferma anche sulla valutazione degli indici attraverso un approccio basato sul Model Confidence Set.

Il Capitolo 3 è incentrato sulla Sentiment Analysis e, dopo una rassegna della principale letteratura e applicazioni, descrive un nuovo metodo per estrarre la polarità dei termini con approccio basato sui dati. Il capitolo descrive anche le procedure per costruire e valutare gli Indici Testuali di Polarità Politica proposti i quali rivelano incorporare informazione utile e in grado di migliorare l'accuratezza previsiva rispetto al modello autoregressivo di riferimento.