ARTICLE IN PRESS

Social Science Research xxx (2016) 1–15

Contents lists available at ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

# Economic performance and public concerns about social class in twentieth-century books

Yunsong Chen [a, **, 1], Fei Yan [b, c, *, 1]

[a] Department of Sociology, Nanjing University, China
[b] Department of Sociology, Tsinghua University, China
[c] The Walter H. Shorenstein Asia-Pacific Research Center, Stanford University, USA

ARTICLE INFO

*Article history:*
Received 24 April 2015
Received in revised form 28 January 2016
Accepted 4 April 2016
Available online xxx

*Keywords:*
Big data
Social class
Economic misery
N-gram corpus
Co-integration

ABSTRACT

What is the association between macroeconomic conditions and public perceptions of social class? Applying a novel approach based on the Google Books N-gram corpus, this study addresses the relationship between public concerns about social class and economic conditions throughout the twentieth century. The usage of class-related words/phrases, or "literary references to class," in American English-language books is related to US economic performance and income inequality. The findings of this study demonstrate that economic conditions play a significant role in literary references to class throughout the century, whereas income inequality does not. Similar results are obtained from further analyses using alternative measures of class concerns as well as different corpora of English Fiction and the *New York Times*. We add to the social class literature by showing that the long-term temporal dynamics of an economy can be exhibited by aggregate class concerns. The application of massive culture-wide content analysis using data of unprecedented size also represents a contribution to the literature.



## 1. Introduction

One of the intriguing challenges of the "big data" agenda in social science is to investigate how socioeconomic cycles in the real world interact with the evolution of cultural behavior or political phenomena. Following the seminal study by Michel et al. (2011) with huge volumes of data analysis from the Google Books repository, recent interdisciplinary studies have related broad economic indicators to the word-usage dynamics. For example, the Dow Jones Industrial Average is reported to be predicted by the volume of Google queries regarding financial terminology, the viewing of Wikipedia financial pages, or Twitter mood words (Bollen et al., 2011; Moat et al., 2013; Preis et al., 2012). The gross domestic product (GDP) per capita of a nation is also found to be associated with the "forward looking" index extracted from Google Trends (Preis et al., 2012). The "literary mood," as measured by usage of various emotion-related words in twentieth-century English-language books, has likewise been revealed to be associated with the US "economic misery" index (Bentley et al., 2014).

In the current study, we examine the correlations between the frequency of word use and dynamics of economic performance based on massive content analysis from Google digitized books. We hypothesize significant links between

* Corresponding author. Department of Sociology, Tsinghua University, Beijing, 100084, China.
** Corresponding author. Department of Sociology, Nanjing University, Nanjing, 210093, China
*E-mail addresses:* yunsong.chen@nju.edu.cn (Y. Chen), feiyan@stanford.edu (F. Yan).
[1] The two authors contributed equally to this work

http://dx.doi.org/10.1016/j.ssresearch.2016.04.007

economic conditions and public concerns about "social class," a core concept in sociology referring to the hierarchy of socioeconomic status in which individuals are positioned. Working from the premise that the written language serves as the cultural component of the long-term accumulation of human knowledge, experiences, and attitudes (Hassanpour, 2013; Mullins et al., 2013), we use the appearance of class-related words in books as a proxy for aggregate class concerns. Specifically, we expect that prior economic conditions have affected the occurrence of class-related words in books, or "literary references to class," over the twentieth century. Our hypothesis builds on both classical and recent literature relating to social class that suggests that a person's objective or subjective social status at an individual level is fundamentally influenced by macroeconomic structures, performance, and cycles (Curtis, 2013; Evans and Kelley, 2004; Yaisha and Andersen, 2012).

Using the Google Book N-gram corpus, we construct an index of literary references to class (hereafter ***LC***) by tracking frequencies per year of a set of class keywords in American books throughout the last century. Then, we examine how well ***LC*** corresponds to: a) the US economic "misery index" (***EM***), namely, the inflation rate plus the unemployment rate, and b) the Gini coefficient (***GINI***), the widely used index measuring income inequality. We scrutinize the validity of using ***LC*** extracted from books to represent aggregate class concerns. Since books represent only one major cultural channel through which people express ideas, we also include other types of literary materials to test the consistency of our finding across publication types. Although we are cautious about making spurious correlations, our findings using large-scale data analysis suggest that the historical trajectory of the concept of "social class" is closely linked to the dynamics of economic practices. As such, the aim of this study is to show the potential of new methods such as massive content analysis for social science research, especially macro-level analyses with huge temporal and spatial scales.

## 2. Economic performance, inequality, and social class

Following the Marxist theory of class, which recognizes the economic basis of class systems, there is an extensive body of literature that relates class structure to income inequality (Andrews and Leigh, 2008; Blanden, 2011; Bossia and Gumus, 2013; Erikson and Goldthorpe, 1992; Milios, 2000; Tyree et al., 1979; Weeden and Grusky, 2012; Wolff and Zacharias, 2013). According to this line of argument, income inequality is indicative of the class structure of a society. If inequality between classes is high, people are more likely to see differences in class positions and thus are more likely to distinguish themselves along class lines. For example, Wright and Perrone (1977) show that class position was powerfully correlated with inequality in American society. Tyree, Semyonov, and Hodge (1979) also find that class consciousness is positively associated with income inequality. If inequality is high, those at the top of the stratification system have a strong incentive to maintain the status quo, because the consequences of downward mobility for elites are more severe than when distances between social classes are small. Conversely, when inequality is low, elites are less likely to mobilize resources and networks to secure their vested privilege. At the other end of the class spectrum, those who are at the bottom of the income distribution are more concerned about their relative disadvantage when inequality is high and thus are more likely to identify with lower-class positions (Andersen and Curtis, 2012). In the United States, the so-called American dream has long inspired working-class citizens to believe in unlimited opportunities and upward mobility channels for all, including those at the bottom of society (Alesina and Glaeser, 2004). Using 1999 International Social Survey Program (ISSP) data on occupational earnings, Osberg and Smeeding (2006) find American attitudes toward economic inequality to be much more polarized than those of other nations—people at the top end of the American income distribution are less tolerant of inequality, while people at the bottom end have become more accepting of inequality.

Others, however, suggest that economic inequality is not the only macro-level factor affecting people's sense of social class (Breen and Luijkx, 2004; Curtis, 2013; Evans and Kelley, 2004; Yaisha and Andersen, 2012); they note, instead, that class identities and social mobility rates are also influenced by economic conditions and performance. Some scholars point out the positive relationship between economic cycles and class consciousness. For example, Goldthorpe et al.'s (1969) classic study of affluent workers finds that improved economic conditions lifted workers' perceptions of their class position in mid-twentieth-century British society. Evans and Kelley's (2004) examination of class identification in twenty-one nations between 1987 and 1998 similarly reveals that economic prosperity significantly strengthens people's sense of class identity. Comparative studies of acceptance of economic inequality also show that white collar workers and retired people in the United States, Australia, Germany, and Norway all prefer larger income differences than blue collar workers and the younger generations (Svallfors, 1997). On average, people in richer countries tend to be more likely to report higher class identities than those in poorer countries.

Another area of research suggests that economic prosperity actually has a negative influence on class identity. For example, Kikkawa (2000) reports a decline in class consciousness in Japan that corresponds with a period of vast economic growth since 1975. There is also evidence of a curvilinear relationship between the two variables, where the influence of economic conditions on public attitudes toward class are effective only until a certain level of development is reached (Raftery, 1983; Wilkinson and Pickett, 2009). In short, these studies tend to agree that macroeconomic performance and context are associated with class concerns, generating varied expectations for consumption and social mobility.

Yet, despite the wealth of different arguments regarding the role of economic conditions, few empirical studies have attempted to address the relationship between economic conditions and class consciousness/concern directly over a long time scale, especially given yawning wealth gaps in the United States and abroad over the past thirty years (Piketty, 2014; Piketty and Zucman, 2014). Moreover, class consciousness in a society at the macro-level is difficult to quantify and seldom measured by time-series data. As Janmaat (2013: 385) points out, "Studies investigating macro-level conditions

Please cite this article in press as: Chen, Y., Yan, F., Economic performance and public concerns about social class in twentieth-century books, Social Science Research (2016), http://dx.doi.org/10.1016/j.ssresearch.2016.04.007