Journal of Behavioral and Experimental Finance 34 (2022) 100666

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Finance

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

Full length article

# Immigration narrative sentiment from TV news and the stock market

Stefano Mazzotta [*,1]

*Department of Economics, Finance, and Quantitative Analysis, and School of Data Science and Analytics, Kennesaw State University, United States of America*

ARTICLE INFO

*Article history:*
Received 28 October 2021
Received in revised form 18 April 2022
Accepted 24 April 2022
Available online 6 May 2022

*JEL classification:*
G40
B55

*Keywords:*
Sentiment
Immigration
Narrative economics

ABSTRACT

Often debated in the media, immigration is a contentious topic in the U.S. Shiller (2017), and Shiller (2019) posit that narratives can drive economics events. This paper investigates the relationship between the *immigration narrative sentiment* and the stock market using the sentiment extracted from of 1.3 million TV news transcripts. Results from Panel Vector Auto Regression (PVAR) estimations show that the immigration narrative sentiment is related to stock market indicators. A positive shock to the immigration narrative sentiment Granger-causes a statistically significant and economically meaningful increase in the stock prices, a decrease in implied volatility, and a statistically significant but economically small increase in trade volume. However, stock market variation does not affect the immigration narrative sentiment. The effect of the immigration narrative sentiment shock to market indicators is long lasting suggesting that the immigration narrative likely contains fundamental information about equities that has not been priced.



## 1. Introduction and motivation

This paper investigates the relationship between the *immigration narrative sentiment* and the stock market. Immigration has driven U.S. population growth and economic prosperity since the nineteen century. Some early immigrants arrived in search of religious freedom, but many came to America seeking greater economic opportunity. Today, about 13.7 per cent of people living in the U.S. were born in another country.

Labor economists have examined the effect of immigration and whether immigration and associated inflows of foreign labor reduce jobs available to natives. Borjas (1995) reports that natives' benefits from immigration are small, but would increase if the United States immigration policy favored more skilled immigrants. Peri (2012) finds no evidence that immigrants crowd out native employment. Immigrants instead promote efficient task specialization.

Despite the magnitude of the immigration phenomenon, studies on the relationship between immigration and financial markets remain sparse. Baker and Blau (2019) examine how immigration regulations granting legal status to large numbers of immigrants benefit stock prices of firms in the manufacturing and construction sectors. Their results suggest that the market perceive immigration as beneficial for those sectors. Nevertheless, immigration policy in US remains contentious and controversial rhetoric is not uncommon in the media and in the public discourse.

According to the Merriam-Webster dictionary, a narrative is "a way of presenting or understanding a situation or series of events that reflects and promotes a particular point of view or set of values". Narrative economics is a field of research that investigates how the spread of stories affect the economy. In a seminal contribution, Shiller (2017) notes "Though these narratives are deeply human phenomena that are difficult to study in a scientific manner, quantitative analysis may help us gain a better understanding".

Responding to Shiller's call, this paper investigates the relationship between sentiment of TV News related to the immigration narrative and the stock market during the period from October 2009 to December 2020. I develop a conceptual framework that connects the narrative economics approach of Shiller (2017, 2019) with sentiment extracted from language using Natural Language Processing (NLP). I preliminarily identify a cluster of n-grams related to the immigration narrative.[2] I hypothesize and test that directly connecting NLP sentiment and narrative

* Correspondence to: Department of Economics, Finance, and Q.A. Michael J. Coles College of Business, Kennesaw State University, 560 Parliament Garden Way, MD 0403, Kennesaw, GA 30144, United States of America.
*E-mail address:* smazzott@kennesaw.edu.

1 I thank Lucy Ackert, for insightful discussion, and Shashank Hebbar for research assistance. I also thank the editor and an anonymous referee for the valuable suggestions. This research was funded by Kennesaw State University, The Graduate College and did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. All the errors are mine.

2 An n-gram is a contiguous sequence of n items from a given sample of text or speech.

https://doi.org/10.1016/j.jbef.2022.100666
2214-6350/

allows to uncover the existence and measure the strength of the relationships between immigration narrative and stock market stock prices, volatility, and trade volume, after accounting for standard economics factors.

This study contributes to the literature in two ways. First, despite the magnitude of immigration flows, the relationship between financial markets and immigration is under-studied. This paper contributes to filling the gap.

Second, it uses a novel approach to quantify the relationship between the immigration narrative and the stock market. To do so, I define a *narrative n-gram cluster*. The immigration narrative n-gram cluster is a set of n-grams selected based on prior knowledge that plausibly signals that a text relates to the immigration narrative. I then search the Internet Archive TV news section for all designated n-grams and extract 1.3 million news transcripts. I use FLAIR, a state-of-the-art NLP algorithm, to classify the passages from the news transcripts and to assigns a sentiment score. The scores are then aggregated to form monthly time series that capture the sentiment dynamics and used to model the fluctuation of stock market variables.

Results from Panel Vector Auto Regression (PVAR) estimations show that the immigration narrative is related to the stock market. In particular, a positive shock to the immigration narrative sentiment Granger-causes a statistically significant and economically meaningful increase in the stock prices, a decrease in implied volatility, and a statistically significant but economically small increase in trade volume. In addition, the dynamics of sentiment and market indicators are such that the effect of a shock to the narrative on market indicators does not revert over time suggesting that the immigration narrative likely contains fundamental information about equities that has not been priced. However, shocks to returns, implied volatility, and trade volume do not have a discernible effect on the immigration narrative sentiment.

The remainder of this paper is organized as follows. Section 2 reviews the literature. Section 3 describes the quantitative methods used to measure the relationship between narrative and stock market indicators. Section 4 describes the results. Section 5 presents main results and robustness tests. Section 6 discusses the findings and Section 7 offers concluding remarks.

## 2. Literature review

The following sub-sections review two streams of research. The first is related to sentiment. The second develops around the immigration theme.

### 2.1. Sentiment in the financial literature

Among the behavioral effects determined by investors' departures from rationality, the importance of sentiment is firmly established in the finance literature. In recent years, sentiment derived from text analysis developed by computational linguistics, and in particular from Natural Language Processing (NLP) has augmented the approach based on proxies common in finance research.[3]

In an early contribution, De Long et al. (1990) show that sentiment affects investment decisions and creates risk that arbitrage cannot eliminate. Baker and Wurgler (2007) contend that ''investors' sentiment is a belief about future cash flows and investment risks that is not justified by the facts at hand''. Brown and Cliff (2004) suggest that sentiment is related to excessive optimism or pessimism. In this strand of work, the limit to arbitrage from the rational agents combined to a change in sentiment on the part of irrational ones is what determines mispricing. Shleifer and Vishny (1997) suggest that betting against sentimental investors is costly and risky. Consequently, rational market participants are unable or unwilling to force prices to fundamentals.

Several attempts have been made to predict financial performance using sentiment. Brown and Cliff (2004) find that market returns predict subsequent levels and changes of both individual and institutional sentiment, but only weak evidence exists that sentiment measures are useful in predicting future stock returns over short horizons. Chan (2003) studies the effect of news on stock and finds that investors under react to new information, and especially bad news. Tetlock (2007) studies the interaction between a daily column from the Wall Street Journal and the stock market. He finds that negative sentiment has significant temporary impact on future stock returns that is reversed in one week and unusually high or low pessimism predict high market trading volume.

Related to this paper, Uhl (2014) use sentiment from the news as measured with a proprietary Thomson Reuters neural network. He finds that daily sentiment predicts changes in stock returns better than macroeconomic factors. Similarly, Heston and Sinha (2017) use Thomson Reuters data and find that daily sentiment is predictive in the short term whereas weekly news sentiment predicts stock returns for up to one quarter.

Soo (2018) analyzes text from newspaper. The study classifies text as positive or negative based on a word list. It finds that local housing media sentiment predicts future home prices. Related to this paper, Chen et al. (2022) use a computational linguistic approach to identify and assess the financial risk from the COVID-19 narrative. They find that around severe market declines the narrative fragments into competing topics and its virality increases. Closest to this work in spirit and methods, Ackert and Mazzotta (2021) find that sentiment related to the American home ownership narrative significantly explains movements in home prices after accounting for known economic factors.

### 2.2. Immigration literature

Economists have focused their attention mostly on themes related to the effects of immigrant labor and to a significantly lesser extent on financial services targeted to immigrants.[4] Barbu and Iustina (2018) offer a qualitative investigation of the immigrants' impact on European financial market. Singular in their undertaking, Baker and Blau (2019) conduct an event study to examine the stock prices of firms most likely to be affected by immigration. The study shows that, relative to the entire market, the stock prices of agricultural, construction, and manufacturing firms increase significantly during the period surrounding these two events. Manufacturing and construction drive the positive stock price response and the market perception is that firms in these industries benefit from immigration.

Overall, the literature on the relationship between immigration and the stock market remains scant. The present study contributes to filling this gap.

## 3. Empirical methods

Narrative and sentiment are intimately related. Narratives exist in the realm of language; sentiments in the realm of emotions. This paper contends that a narrative drives economics events

[3] This article does not attempt to provide a complete overview of the vast finance literature related to sentiment. Recent reviews are provided in Gupta et al. (2020) and Algaba et al. (2020).

[4] For surveys of the theory and empirical economics work on immigration see Borjas (2014), Card and Peri (2016), and Kerr and Kerr (2011).

when it elicits sentiments that motivate individuals to take actions with economic consequences. In this study, the immigration narrative is the organizing principle that connects sentiment measures from TV news transcripts. The empirical methods described in the following section make this contention operational.

### 3.1. Measurement of the narrative

Shiller (2017) notes that narratives are difficult to study in a scientific manner. Natural Language Processing (NLP) algorithms are developed in computational linguistics, artificial intelligence, in order to perform high level tasks, such as image captioning, fake news and hate speech detection, or, relevant to this study, sentiment scoring. In computational linguistics, sentiment is typically defined as the *emotional tone* behind a body of text. This paper relies on the same notion of sentiment.[5]

Blei et al. (2003) propose the Latent Dirichlet Allocation (LDA) approach to provide an explicit representation of the topics in linguistic corpora. For instance, (Li et al., 2017) use LDA to quantify financial report data, Bastani et al. (2019) use LDA to identify narratives in Consumer Financial Protection Bureau consumers complaints, Chen et al. (2022) use LDA to identify important economic narratives overtime. The LDA produces a distribution of topics, or clusters, with associated probabilities. These clusters are typically subjected to further analysis. The tone of a textual passage i.e. its sentiment, can be estimated using the bag-of-word tone analysis, and a dictionary.[6] The above family of methodologies provides an effective and efficient way to help identify ex-ante undetermined topics in large corpora, which may be a prohibitively laborious task for humans.

This study uses a different approach. It hypothesizes that the contour of a narrative can be defined by a cluster of n-grams selected based on prior knowledge. Unlike in the LDA approach where the topics are an output of the analysis, here the narrative and the associated topics are the primary input and the starting point of the empirical study design. The two methodologies may be considered complementary but there are some finer differences worth noticing. The latent clusters of contents identified by the LDA need to be examined ex-post by a human to be interpreted as meaningful topics, and this may not always be possible. The n-grams in this paper are instead selected ex-ante because of their meaning. Potential shortcomings of the n-gram approach is that some important topic may be inadvertently left out from the search. Alternatively, the search for some interesting candidate n-grams may not turn out a sufficient number of textual passages to subject to the analysis. However, these limitations work in favor of the null hypothesis that sentiment from the immigration news does not matter for the stock market, and therefore statistical evidence to the contrary that may emerge from the analysis can be regarded as conservative.

Once the tentative n-grams list is defined, a search script is used to pull the text transcripts that contain the n-grams from the internet archive. The text passages are then subjected to sentiment analysis using the FLAIR classifier, a state-of-the-art NLP algorithm proposed by Akbik et al. (2019) that models language as distributions over characters. The algorithm internalizes linguistic concepts such as words, sentences, sub-clauses, and even sentiment.[7] FLAIR's "contextual string embeddings" contextualize words by surrounding text. Thus, the meaning of a word depends on its "embedding" in the context of the text passage.

The FLAIR algorithm is pre-trained to identify positive or negative tone, and tone's intensity within the text passage. When a text passage is submitted for classification, it is scored as positive or negative and assigned a score that ranges between zero and one. The polarity captures the positive or negative emotional tone of the text passage. The numerical score captures the intensity of the tone. For instance, a piece of news with a positive tone receives a positive sentiment score. A positive piece of news with an enthusiastic tone, as perhaps suggested by intensifiers such as "great" or "very", will receive a higher score with a value closer to one. Therefore, the sentiment score in this paper measures the emotional tone and intensity of a news transcript passage related to the immigration narrative.

It should be clear that the tone of a textual passage depends to a varying extent on the anchors' subjective analysis and style. One and the same fact presented with different tones scores differently.[8] It is also worth noticing that the news passages are pulled from the internet archive because they contain one of the target n-grams, regardless of any economic content. Thus, the sentiment score strictly refers to the text passage fed to the NLP classifier and only incidentally can be influenced by economic commentary in the news passage.

Naturally, the accuracy of a computerized sentiment classifier needs to be assessed. Note first that even human sentiment polarity annotators do not agree all the time. Human agreement baseline is in fact 80–85 per cent. For the purpose of sentiment classification, the FLAIR model is pre-trained over a mix of polarized positive and negative text samples. The algorithm is considered state-of-the-art because its sentiment scoring accuracy reaches or exceeds the human baseline.[9]

To summarize, the steps to obtain the immigration narrative sentiment data set are as follows:

1. Select the n-grams potentially signaling a TV news is related to the immigration narrative based on prior knowledge.
2. Search the Internet Archive TV news database for each of the immigration n-grams for every day in the sample period.
3. Extract the news transcripts from the archive website that contain the n-gram.
4. Clean the transcripts from extraneous coding and web control characters.
5. Feed each text passage containing the n-gram with up to one hundred tokens before and after the n-gram for sentiment classification to FLAIR.[10]
6. Collect the sentiment score for each passage computed by the FLAIR algorithm.
7. Aggregate the sentiment scores for each n-gram by month and compute monthly count, mean, and other descriptive statistics.
8. Construct the monthly time series of sentiment average for each n-gram and news count for further analysis.

As noted, the algorithm classifies the passage's sentiment based on context and returns a score with a negative or positive polarity and a value ranging from zero to one. The sentiment measure object of study in the next sections are the series of immigration narrative sentiment monthly average for each of the 15 n-grams.

[5] Financial economists define sentiment as a relative measure of deviation between noise traders' beliefs and rational arbitrageurs' beliefs. See e.g. De Long et al. (1990), Baker and Wurgler (2007), Brown and Cliff (2004), Shleifer and Vishny (1997) and Tetlock (2007). A survey based indicator is the Consumer Sentiment index released by the University of Michigan. See University of Michigan (2018).

[6] See e.g. Loughran and McDonald (2011)

[7] Akbik et al. (2018).

[8] For opinion programs in television news the divergence can be extreme. See e.g. Bursztyn et al. (2022)

[9] See e.g. Wilson et al. (2005), Otter et al. (2021), and Li et al. (2018).

[10] A token is an individual occurrence of a linguistic unit in speech or writing.

### 3.2. Quantifying the relationship

The goal of this study is to examine the relationship between stock market indicators and the immigration narrative sentiment. The methods to quantify the immigration sentiment variable are described above. The following subsections describe the methods to investigate the relationship between sentiment and stock market indicators.

#### 3.2.1. Market activity and the immigration sentiment narrative

As a first step, it is interesting to ascertain whether the immigration narrative sentiment predicts stocks returns and other market indicators. A vector autoregression (VAR) model augmented with a measure of the average immigration narrative sentiment is therefore estimated first. The VAR model can be represented as follows:

$$\mathbf{Y_t} = \mathbf{c} + \sum\nolimits_{\mathbf{l=1}}^{\mathbf{m}} \mathbf{A_l Y_{t-l}} + \boldsymbol{\epsilon}_\mathbf{t}, \tag{1}$$

where $\mathbf{Y_t} = [RET_t\ DY_t\ RB_t\ DEF_t\ VIX_t\ VL_t\ ASENT_t]'$ is the vector of market variables and average sentiment at time $t$. $\mathbf{c}$ is a vector of constants, $\mathbf{A_l}$ is a matrix of coefficients for lag $l$, and $\boldsymbol{\epsilon}_\mathbf{t}$ is a white noise error term. The specification of the model is similar to Campbell and Shiller (1988) and Campbell (1991). The variables $RET$, $DY$, $RB$, and $DEF$ are the market returns, the dividend yield, the interest rate, and the default spread.[11] Following Tetlock (2007) and Chen et al. (2022) the implied volatility ($VIX$), and trade volume ($VL$) variables are also included. Standard specification tests indicate that $m = 2$ is the optimal lag structure. Finally, the monthly average narrative sentiment score across the 15 immigration n-grams ($ASENT$) is included as well.

The model represented in Eq. (1) treats all the variables as endogenous. However, theoretical models make different predictions about the relationship between sentiment and market indicators such as volatility and trade volume.[12] Therefore, to remove the impact of known sources of predictability from one indicator to another, it is desirable to estimate a preliminary VAR regression before considering the relationship between market indicators and immigration sentiment. The variables vector reduces to $\mathbf{Y_t} = [RET_t\ DY_t\ RB_t\ DEF_t\ VIX_t\ VL_t]'$. The vector of innovations $\epsilon_t = [e\_RET_t\ e\_DY_t\ e\_RB_t\ e\_DEF_t\ e\_VIX_t\ e\_VL_t]'$ can then be used to investigate the effect of the immigration narrative sentiment on each variable of interest net of its the predictable component. The preliminary VAR's residuals $e\_RET_t$, $e\_VIX_t$, $e\_VL_t$ respectively from returns, implied volatility, and trade volume equations represent the variation in each indicator that is not predicted by the other fundamental economic variables. They are the object of study in the following subsections.

#### 3.2.2. Bivariate vector autoregressions

To quantify the relationship between market indicators and immigration narrative sentiment it is possible to estimate bivariate VAR models between the average narrative sentiment variable ($ASENT_t$) and the VAR innovations from market return ($e\_RET_t$), the implied volatility ($e\_VIX_t$), and the trade volume ($e\_VL_t$) equations in turn. These results are uninformative and are not reported for brevity, but available from the author.

The Panel Vector Auto Regression (PVAR) model developed in Holtz-Eakin et al. (1988) and implemented in Abrigo and Love (2016) is estimated instead. The panel structure preserves the contribution of each individual n-gram time series and thereby avoids the loss of information caused by the cross sectional averaging of the immigration narrative sentiment variable $ASENT_t$ in the non-panel VAR.

The PVAR model allows for non-stationarity, optimal lag length selection, parameters estimation and computation of test statistics using the generalized method of moments. The PVAR is estimated with the forward orthogonal deviation (FOD) transformation of Arellano and Bover (1995) which minimizes information loss.

The three PVAR models considered therefore are:

$$\begin{aligned} e\_RET_{i,t} &= \gamma_{1i} + \sum_{j=1}^{m} \beta_{1j} e\_RET_{i,t-j} + \sum_{j=1}^{m} \delta_{1j} S_{i,t-j} + \epsilon_{1i,t} \\ S_{i,t} &= \gamma_{2i} + \sum_{j=1}^{m} \beta_{2j} e\_RET_{i,t-j} + \sum_{j=1}^{m} \delta_{2j} S_{i,t-j} + \epsilon_{2i,t} \end{aligned} \tag{2}$$

$$\begin{aligned} e\_VIX_{i,t} &= \gamma_{1i} + \sum_{j=1}^{m} \beta_{1j} e\_VIX i,t-j + \sum_{j=1}^{m} \delta_{1j} S_{i,t-j} + \epsilon_{1i,t} \\ S_{i,t} &= \gamma_{2i} + \sum_{j=1}^{m} \beta_{2j} e\_VIX_{i,t-j} + \sum_{j=1}^{m} \delta_{2j} S_{i,t-j} + \epsilon_{2i,t} \end{aligned} \tag{3}$$

$$\begin{aligned} e\_VL_{i,t} &= \gamma_{1i} + \sum_{j=1}^{m} \beta_{1j} e\_VL_{i,t-j} + \sum_{j=1}^{m} \delta_{1j} S_{i,t-j} + \epsilon_{1i,t} \\ S_{i,t} &= \gamma_{2i} + \sum_{j=1}^{m} \beta_{2j} e\_VL_{i,t-j} + \sum_{j=1}^{m} \delta_{2j} S_{i,t-j} + \epsilon_{2i,t} \end{aligned} \tag{4}$$

where $e\_RET_{it} = e\_RET_t$, $e\_VIX_{it} = e\_VIX_t$, and $e\_VL_{it} = e\_VL_t$ for all $i$'s are the stock market return, the implied volatility and the trade volume variables residuals described in the previous section respectively. $S_{it}$ is the monthly average sentiment for n-gram $i$, with $i = 1 \dots 15$, at time $t$. The lag length, $m$, is selected such that the error term is white noise. Specification tests based on the overall coefficient of determination indicate that $m = 1$ is the optimal lag structure for the three PVAR models considered. PVAR Eqs. (2)–(4), allow to test the hypothesis that the n-grams related to the immigration narrative explain the variation of stock price, volatility, and trade volume that is not accounted for by the other non-simultaneous economic variables' and own lags in turn.

## 4. Data

### 4.1. TV news textual data

The source of the immigration TV News transcripts is the Internet Archive, a non-profit library of books, video, audio, and images in digital form. The Internet Archive is an ideal source for immigration TV news.[13] Transcripts are available for the October 2009 through December 2020 sample period. The resulting corpus for this study includes 1.38 million news text passages, with an average count of 682 news and median of 187 news per month.

### 4.2. Stock market data

Stock market data are from Robert Shiller's web page and are the same data used in Shiller (2015).[14] The data includes the monthly stock returns, dividend yield, interest rate, and default premium. In addition, following Tetlock (2007) and Chen et al. (2022) the implied volatility index from stock options (VIX) and trade volume are also included. The VIX monthly variable is obtained by aggregating the daily VIX volatility, the trade volume variable (VL) is the natural logarithm of the total number of shares traded in all U.S. markets each month. They are both from the Cboe.

11 The variables specification is as in Hodrick (1992), page 365.

12 A concise discussion of theories and implications is in Tetlock (2007).

13 https://archive.org/.

14 http://www.econ.yale.edu/~shiller/data.htm - Stock Market Data file. Please see details therein.

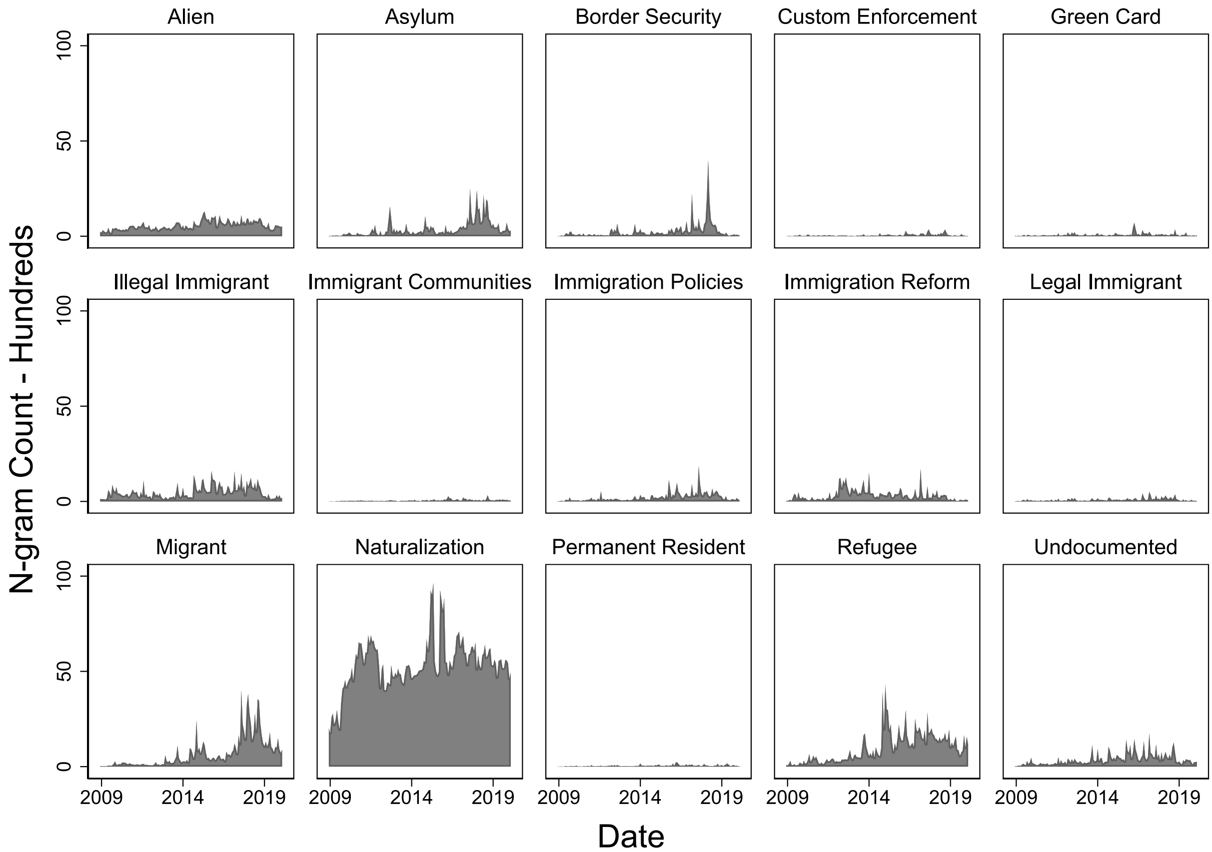


**Fig. 1.** Frequency of n-grams. Appearance by Month, October 2009–December 2020.

### *4.3. The N-grams. Selection and descriptives*

From a preliminary selection of thirty candidate n-grams only fifteen receive sufficiently consistent news coverage in the Internet News Archive over the sample period to generate viable time series. Some ex-ante relatively common n-grams such as *working visa*, *border enforcement*, or *citizenship requirements*, have months of no news and thus produce time series with missing observations. The final list of n-grams includes only mono-grams or bi-grams. They are *Alien*, *Asylum*, *Border Security*, *Custom Enforcement*, *Green Card*, *Illegal Immigrant*, *Immigrant Communities*, *Immigration Policies*, *Immigration Reform*, *Legal Immigrant*, *Migrant*, *Naturalization*, *Permanent Resident*, *Refugee*, and *Undocumented*.

Fig. 1 shows the monthly count for each of the 15 n-gram in the TV news. TV news display clustering of intense coverage and periods of lower interest. Fig. 2 shows the histograms of the sentiment measure for all the n-grams. It appears that the sentiment of TV news for some n-grams, such as Naturalization or Illegal Immigrant is less susceptible to swings. Other n-grams such as Immigrant Community show a more spread out and mostly positive sentiment. Fig. 3 shows plots of the measured n-gram sentiment for each month in the sample. Sentiment varies over time and, as noted, the average monthly scores are mostly negative.

Table 1 shows the summary statistics for the count of news by n-gram. Table 2 shows descriptive statistics for the sentiment score from the immigration narrative news. All n-grams but Immigrant Community display a negative monthly average sentiment. Moreover, almost half n-grams never foray in the positive sentiment territory. Commentators frequently note the negative slant in news content. One potential explanation is that negative news are the result of a newsmaking process that favors new and exciting information. In addition, Trussler and Soroka (2014) suggest that the demand side may be an important factor and that news may be negative because people are more interested in that type of news.

**Table 1**
Summary statistics for the news count. The table reports summary statistics for the TV news transcripts containing each n-gram for the October 2009 through December 2020 sample period. The sample includes 138 monthly observations for each of the 15 n-grams. The table reports the total count of news over the entire period, the monthly mean number of news, the median, the standard deviation, the minimum, and maximum across sample months.

| Ngram | Total news | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|---|
| All | 1,382,658 | 682.79 | 187 | 1394.9 | 2 | 9646 |
| Alien | 78,742 | 583.27 | 550 | 230.32 | 151 | 1297 |
| Asylum | 54,926 | 406.86 | 247 | 473.51 | 15 | 2521 |
| Border Security | 37,627 | 278.72 | 142 | 467.32 | 7 | 4004 |
| Custom Enforcement | 8,254 | 61.14 | 39 | 61.8 | 3 | 363 |
| Green Card | 11,186 | 82.86 | 58 | 89.15 | 7 | 719 |
| Illegal Immigrant | 68,538 | 507.69 | 422 | 341.23 | 92 | 1610 |
| Immigrant Communities | 7,854 | 58.18 | 46 | 47.85 | 2 | 333 |
| Immigration Policies | 30,965 | 229.37 | 136 | 248.35 | 12 | 1859 |
| Immigration Reform | 45,980 | 340.59 | 257 | 305.02 | 20 | 1723 |
| Legal Immigrant | 12,727 | 94.27 | 74 | 82.79 | 9 | 510 |
| Migrant | 97,384 | 721.36 | 407 | 839.36 | 20 | 4025 |
| Naturalization | 731,399 | 5417.77 | 5347 | 1362.53 | 1766 | 9646 |
| Permanent Resident | 6,902 | 51.13 | 39 | 41.38 | 2 | 239 |
| Refugee | 136,371 | 1010.16 | 864 | 807.59 | 65 | 4317 |
| Undocumented | 53,803 | 398.54 | 300 | 324.92 | 20 | 1758 |

## 5. Results

### *5.1. Vector autoregressions*

Table 3 presents results from the predictability VAR model in Eq. (1). The model includes returns ($RET_t$), dividend yield ($DY_t$),

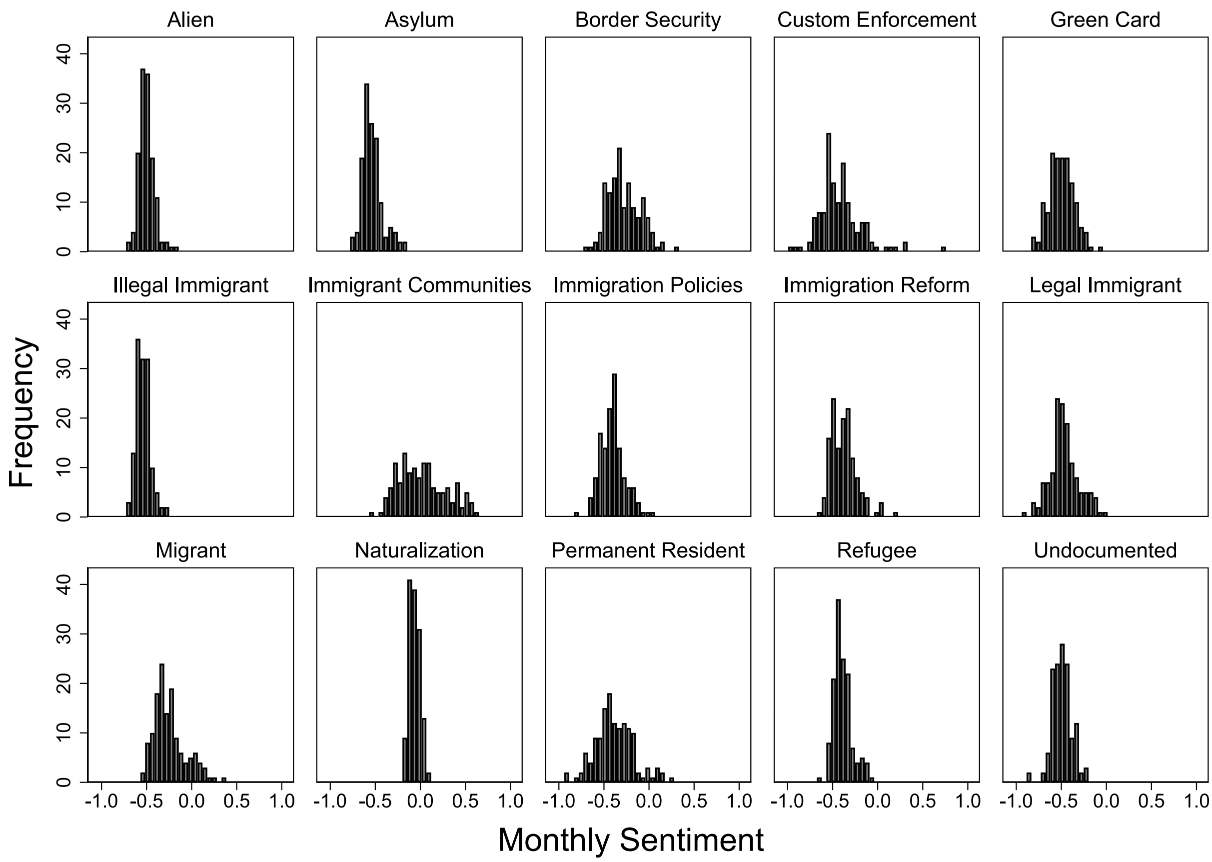


**Fig. 2.** N-grams sentiment measure histograms. October 2009–December 2020.

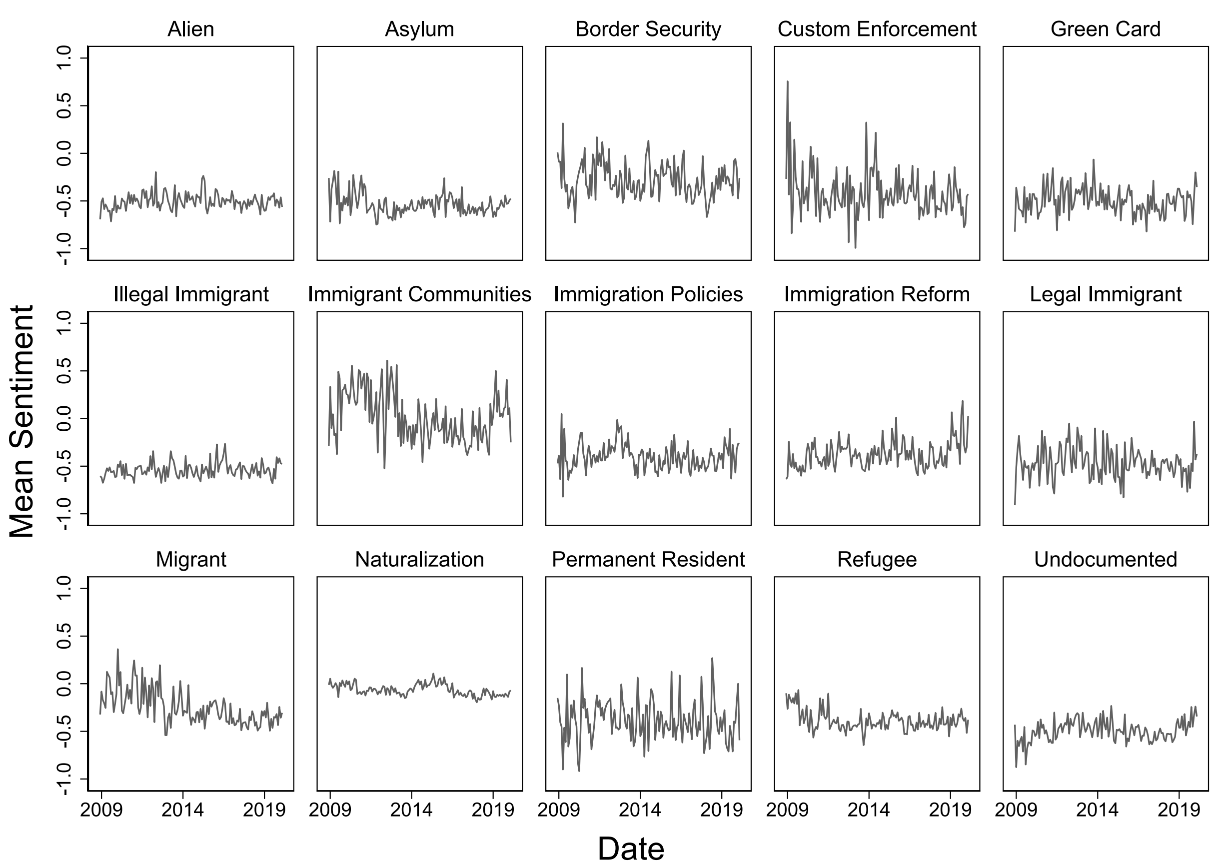


**Fig. 3.** n-gram Sentiment by Month, October 2009–December 2020.

**Table 2**
Sentiment descriptive statistics for the 15 n-grams. The table reports summary information for each n-gram sentiment measure for the October 2009 through December 2020 sample period. The sample includes 138 monthly observations for each of the 15 n-grams. It reports the monthly average sentiment extracted from the TV news, median, standard deviation, minimum, and maximum across sample period. The sentiment measure is bounded between −1 and +1.

| Ngram | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|
| All | −0.37 | −0.42 | 0.23 | −0.99 | −0.76 |
| Alien | −0.50 | −0.51 | 0.08 | −0.71 | −0.20 |
| Asylum | −0.53 | −0.55 | 0.11 | −0.75 | −0.18 |
| Border Security | −0.28 | −0.31 | 0.18 | −0.73 | −0.31 |
| Custom Enforcement | −0.42 | −0.46 | 0.24 | −0.99 | −0.76 |
| Green Card | −0.50 | −0.50 | 0.14 | −0.82 | −0.06 |
| Illegal Immigrant | −0.54 | −0.55 | 0.08 | −0.68 | −0.26 |
| Immigrant Communities | −0.02 | −0.00 | 0.26 | −0.52 | −0.61 |
| Immigration Policies | −0.40 | −0.40 | 0.14 | −0.82 | −0.05 |
| Immigration Reform | −0.38 | −0.38 | 0.15 | −0.64 | −0.18 |
| Legal Immigrant | −0.47 | −0.48 | 0.16 | −0.91 | −0.03 |
| Migrant | −0.25 | −0.29 | 0.18 | −0.54 | −0.36 |
| Naturalization | −0.06 | −0.07 | 0.06 | −0.20 | −0.11 |
| Permanent Resident | −0.38 | −0.40 | 0.22 | −0.92 | −0.27 |
| Refugee | −0.39 | −0.41 | 0.10 | −0.64 | −0.07 |
| Undocumented | −0.49 | −0.49 | 0.11 | −0.88 | −0.23 |

**Table 3**
The table reports summary results from a market VAR model with average sentiment. The model includes the market return (RET), the dividend yield (DY), the interest rate (RB), the default spread (DEF), the implied volatility (VIX), the trade volume (VL), and the average sentiment across the 15 n-grams (ASENT). All the variables are demeaned.

| Equation | Parms | RMSE | R-sq | $\chi^2$ | p-val |
|---|---|---|---|---|---|
| RET | 15 | 0.032 | 0.180 | 29.70 | 0.008 |
| DY | 15 | 0.001 | 0.769 | 450.12 | 0.000 |
| RB | 15 | 0.076 | 0.195 | 32.65 | 0.003 |
| DEF | 15 | 0.002 | 0.214 | 36.65 | 0.001 |
| VIX | 15 | 0.046 | 0.617 | 217.78 | 0.000 |
| VL | 15 | 0.001 | 0.585 | 190.18 | 0.000 |
| ASENT | 15 | 0.057 | 0.210 | 35.86 | 0.001 |

interest rate ($RB_t$), default spread ($DEF_t$), implied volatility ($VIX_t$), trade volume ($VL_t$), and average sentiment across the 15 n-gram ($ASENT_t$). All the equations in the VAR model are significant at any conventional level. The $R^2$ for all the equations suggests that a sizable portion of all indicators' variation is predictable.

Table 4 shows Granger-causality tests statistics. The causality tests for all market indicators are jointly significant, with the exception the interest rate. The tests are however unable to pin down any significant Granger-causation from the average sentiment to any other variable. The tests confirm known predictability properties of market variables. In addition, none of the market variables, individually as well as jointly, Granger-cause average sentiment. In other words, the average immigration narrative sentiment does not significantly Granger-cause any market indicator nor is caused by any of them. This first set of results suggests that the relationship between immigration narrative sentiment and market activity can be investigated separately from market predictability.

Table 5 presents results from the standard predictability VAR model in Eq. (1), with no average sentiment variable ($ASENT_t$). All the equations in the VAR model are significant at any conventional level. The exclusion of the sentiment variable results in a small decrease of the $R^2$ for market, implied volatility, and trade volume equations from 18 to 16.7, 61.7 to 60.5, and 58.5 to 58.3 per cent respectively. Granger-causality tests in Table 6 closely mirror results in Table 4.

All of the above tests suggest that a sizable portion of market indicators' variation is predictable, whether the $ASENT_t$ variable is included or not. Consequently, given the lack of Granger-causality to and from cross sectional average sentiment $ASENT$, the pair-wise relationship between the sentiment and returns, volatility, and trade volume is studied separately next.

### 5.2. Panel vector autoregressions results

This subsection presents results from Panel Vector Autoregression pair-wise estimations represented in Eqs. (2)–(4). The models include the immigration narrative sentiment panels and the residuals from stock return ($e\_RET_t$), implied volatility ($e\_VIX_t$), the trade volume ($e\_VL_t$) equations in turn. The residuals are generated from predictability VAR model (1) without the sentiment variable ($ASENT_t$). All the variables are stationary, and the systems are stable.

#### 5.2.1. Market returns results

Table 7 reports estimation results for the market PVAR in Eq. (2). The market variable is the innovation $e\_RET_t$. The sentiment variable $S_{i,t}$ is the monthly average of for n-gram $i$ with $i = 1 \ldots 15$ for the 15 immigration n-grams. The market coefficient $\beta_{1,1}$ is small and not statistically significant. This is expected and due to the preliminary VAR treatment of the market return variable. The coefficient $\delta_{1,1}$ is positive and significant at any conventional level. Thus, the lagged immigration narrative sentiment variable significantly explains the market variation that is not captured by non-contemporaneous known economic factors. The sign, magnitude, and significance of $\delta_{1,1}$ indicate that a positive change in the immigration TV news sentiment predicts a statistically significant and economically meaningful increase in the stock market return. This is a central finding in this paper.

In the Sentiment Equation, the market lagged return coefficient $\beta_{2,1}$ is not significant. This indicates that immigration sentiment is largely unaffected by changes in the stock market. The coefficient $\delta_{2,1}$ is positive and significant indicating that sentiment is positively correlated with own first order lag.

Table 8 reports the PVAR Granger-causality test results. The test rejects the exclusion of the Sentiment variable from the Market Equation at any conventional level and indicates that Sentiment Granger-cause market movement. To the contrary, the market does not Granger-cause change in the immigration narrative sentiment. This result indicates that the immigration narrative sentiment significantly drives the stock returns component not explained by non-contemporaneous economic variables.

Table 9 shows the Forecast Error Variance Decomposition from the PVAR model (2). The variance decomposition indicates the amount of information that the market variable and sentiment variable contribute to the other variable respectively. The dynamics are such that the sentiment variable drives up to 13 per cent of the market variation unexplained by non-contemporaneous economic variables by the third month of the forecast horizon. The contribution of the market to the sentiment variation appears negligible.

The impulse response function in Fig. 4 shows how sentiment and stock prices dynamically affect each other. The impulse response function shows the cumulative expected response to the impact of a one standard deviation orthogonalized shock to the first variable on the second over a six-month time horizon with 95 per cent confidence intervals bands.[15] The top right figure shows that the effect of a shock to sentiment causes a significant lasting positive impact on stock returns. This is direct quantitative evidence that the immigration narrative as measured by TV news

[15] In general, innovations in the PVAR are correlated contemporaneously and a shock on one variable is accompanied by a shock on the other. The orthogonalized shocks are uncorrelated by construction and facilitate the interpretation of the impulse response function.

**Table 4**
Panel VAR-Granger causality Wald test
Ho: Excluded variable does not Granger-cause Equation variable
Ha: Excluded variable Granger-causes Equation variable.

| Equation | Excluded | $\chi^2$ | df | p-val | Equation | Excluded | $\chi^2$ | df | p-val |
|---|---|---|---|---|---|---|---|---|---|
| RET | DY | 4.65 | 2 | 0.098 | DEF | RET | 0.50 | 2 | 0.780 |
| RET | RB | 1.07 | 2 | 0.585 | DEF | DY | 5.59 | 2 | 0.061 |
| RET | DEF | 4.23 | 2 | 0.121 | DEF | RB | 4.94 | 2 | 0.084 |
| RET | VIX | 11.11 | 2 | 0.004 | DEF | VIX | 4.13 | 2 | 0.127 |
| RET | VL | 2.82 | 2 | 0.245 | DEF | VL | 1.56 | 2 | 0.458 |
| RET | ASENT | 2.24 | 2 | 0.326 | DEF | ASENT | 1.69 | 2 | 0.429 |
| RET | ALL | 26.58 | 12 | 0.009 | DEF | ALL | 23.18 | 12 | 0.026 |
| DY | RET | 4.31 | 2 | 0.116 | VIX | RET | 3.69 | 2 | 0.158 |
| DY | RB | 1.39 | 2 | 0.499 | VIX | DY | 7.51 | 2 | 0.023 |
| DY | DEF | 5.52 | 2 | 0.063 | VIX | RB | 2.05 | 2 | 0.358 |
| DY | VIX | 10.61 | 2 | 0.005 | VIX | DEF | 0.58 | 2 | 0.747 |
| DY | VL | 2.24 | 2 | 0.327 | VIX | VL | 3.24 | 2 | 0.198 |
| DY | ASENT | 2.73 | 2 | 0.255 | VIX | ASENT | 4.26 | 2 | 0.119 |
| DY | ALL | 24.09 | 12 | 0.020 | VIX | ALL | 29.60 | 12 | 0.003 |
| RB | RET | 0.18 | 2 | 0.914 | VL | RET | 0.25 | 2 | 0.881 |
| RB | DY | 1.02 | 2 | 0.602 | VL | DY | 10.18 | 2 | 0.006 |
| RB | DEF | 3.86 | 2 | 0.145 | VL | RB | 1.69 | 2 | 0.430 |
| RB | VIX | 1.17 | 2 | 0.557 | VL | DEF | 0.00 | 2 | 0.998 |
| RB | VL | 0.04 | 2 | 0.978 | VL | VIX | 8.60 | 2 | 0.014 |
| RB | ASENT | 0.30 | 2 | 0.860 | VL | ASENT | 0.62 | 2 | 0.734 |
| RB | ALL | 8.52 | 12 | 0.743 | VL | ALL | 27.56 | 12 | 0.006 |
| | | | | | ASENT | RET | 0.59 | 2 | 0.744 |
| | | | | | ASENT | DY | 1.13 | 2 | 0.569 |
| | | | | | ASENT | RB | 3.14 | 2 | 0.209 |
| | | | | | ASENT | DEF | 2.59 | 2 | 0.274 |
| | | | | | ASENT | VIX | 2.76 | 2 | 0.252 |
| | | | | | ASENT | VL | 0.47 | 2 | 0.791 |
| | | | | | ASENT | ALL | 12.41 | 12 | 0.413 |

**Table 5**
The table reports summary results from a market VAR model. The model includes the market return (RET), the dividend yield (DY), the interest rate (RB), the default spread (DEF), the implied volatility (VIX), and the trade volume (VL).

| Equation | Parms | RMSE | R-sq | $\chi^2$ | p-val |
|---|---|---|---|---|---|
| RET | 13 | 0.032 | 0.167 | 27.01 | 0.008 |
| DY | 13 | 0.001 | 0.765 | 438.51 | 0.000 |
| RB | 13 | 0.076 | 0.193 | 32.27 | 0.001 |
| DEF | 13 | 0.002 | 0.204 | 34.52 | 0.001 |
| VIX | 13 | 0.046 | 0.605 | 206.98 | 0.000 |
| VL | 13 | 0.001 | 0.583 | 188.70 | 0.000 |

sentiment drives the stock price. The estimates indicate that one standard deviation increase in the immigration news sentiment Granger-cause a stock market long lasting return increase close to 1.5 per cent of one standard deviation of the market return variable. This result suggests that the immigration narrative sentiment likely contains fundamental information about equities that has not been priced.

The bottom left panel shows that the market may not significantly affect sentiment as zero is contained in the 95 per cent confidence bands.

#### *5.2.2. Implied volatility results*

Table 10 reports estimation results for the implied volatility PVAR model in Eq. (3). The implied volatility own lag coefficient $\beta_{1,1}$ is not statistically significant due to the preliminary VAR treatment of the variable, as expected. The coefficient $\delta_{1,1}$ is negative and significant at any conventional level. Thus, the lagged immigration narrative sentiment variable significantly explains the implied volatility that is not captured by non-contemporaneous stock returns and other economic indicators. The negative sign, and significance of $\delta_{1,1}$ indicate that a positive change in the immigration TV news sentiment predicts a statistically significant decrease in the stock market implied volatility.

Table 11 reports the PVAR Granger-causality test results. The test rejects the exclusion of the Sentiment variable from the implied volatility equation at any conventional level and indicates that Sentiment Granger-cause implied volatility movement. To the contrary, the implied volatility does not Granger-cause change in the immigration narrative sentiment.

Table 12 shows the Forecast Error Variance Decomposition from the PVAR model (3). The dynamics are such that the sentiment variable drives up to 5 per cent of the implied volatility variation unexplained by economic variables by the third month of the forecast horizon. The contribution of the implied volatility to the sentiment variation appears negligible.

The impulse response function in Fig. 5 shows how sentiment and implied volatility dynamically affect each other. The top right figure shows that the effect of an orthogonalized shock to sentiment causes a significant lasting negative impact on implied volatility. The effect is economically meaningful at about 1.2 per cent of one standard deviation of the implied volatility innovation. Hence, the immigration narrative sentiment significantly drives in the opposite direction the implied volatility not explained by non-contemporaneous economic variables. These results complement and corroborate those in the previous subsection.

#### *5.2.3. Trade volume results*

Table 13 reports estimation results for the trade volume PVAR model in Eq. (4). The trade volume own lag coefficient $\beta_{1,1}$ is not statistically significant due to the preliminary VAR treatment of the variable, as expected. The coefficient $\delta_{1,1}$ is positive and significant at any conventional level. Thus, the lagged immigration narrative sentiment variable significantly explains the trade volume that is not captured by non-contemporaneous stock returns and other economic factors. The positive sign and significance of $\delta_{1,1}$ indicate that a positive change in the immigration TV news sentiment predicts a statistically significant increase in the trade volume.

**Table 6**
Panel VAR-Granger causality Wald test
Ho: Excluded variable does not Granger-cause Equation variable
Ha: Excluded variable Granger-causes Equation variable.

| Equation | Excluded | $\chi^2$ | df | p-val | Equation | Excluded | $\chi^2$ | df | p-val |
|---|---|---|---|---|---|---|---|---|---|
| RET | DY | 3.97 | 2 | 0.137 | DEF | RET | 0.45 | 2 | 0.800 |
| RET | RB | 1.18 | 2 | 0.555 | DEF | DY | 4.88 | 2 | 0.087 |
| RET | DEF | 3.49 | 2 | 0.175 | DEF | RB | 4.16 | 2 | 0.125 |
| RET | VIX | 10.22 | 2 | 0.006 | DEF | VIX | 3.76 | 2 | 0.153 |
| RET | VL | 2.24 | 2 | 0.326 | DEF | VL | 1.23 | 2 | 0.541 |
| RET | ALL | 23.94 | 10 | 0.008 | DEF | ALL | 21.22 | 10 | 0.020 |
| DY | RET | 4.29 | 2 | 0.117 | VIX | RET | 3.51 | 2 | 0.173 |
| DY | RB | 1.57 | 2 | 0.457 | VIX | DY | 6.07 | 2 | 0.048 |
| DY | DEF | 4.57 | 2 | 0.102 | VIX | RB | 1.12 | 2 | 0.570 |
| DY | VIX | 9.69 | 2 | 0.008 | VIX | DEF | 0.26 | 2 | 0.876 |
| DY | VL | 1.71 | 2 | 0.425 | VIX | VL | 2.12 | 2 | 0.347 |
| DY | ALL | 20.93 | 10 | 0.022 | VIX | ALL | 24.57 | 10 | 0.006 |
| RB | RET | 0.17 | 2 | 0.920 | VL | RET | 0.36 | 2 | 0.834 |
| RB | DY | 1.06 | 2 | 0.589 | VL | DY | 9.80 | 2 | 0.007 |
| RB | DEF | 3.76 | 2 | 0.153 | VL | RB | 1.49 | 2 | 0.475 |
| RB | VIX | 1.18 | 2 | 0.555 | VL | DEF | 0.02 | 2 | 0.988 |
| RB | VL | 0.03 | 2 | 0.987 | VL | VIX | 9.51 | 2 | 0.009 |
| RB | ALL | 8.20 | 10 | 0.609 | VL | ALL | 26.82 | 10 | 0.003 |

**Table 7**
Panel VAR estimation results. The table reports coefficient estimates for the PVAR estimation including the stock market return residual variable $e\_RET_t$ and the immigration narrative sentiment panel variable $S_{i,t}$. Each equation includes one lag of each variable and is estimated with 15 panels of 1,995 observations after adjustment for lags. Robust standard errors and two-sided p-values appear to the right of each estimated coefficient.

| | Lag Variables | Coefficient | SE | p-val |
|---|---|---|---|---|
| Market Residual Equation - $e\_RET_t$ | $\beta_{1,1}$ | 0.003 | 0.021 | 0.877 |
| | $\delta_{1,1}$ | 0.081 | 0.014 | 0.000 |
| Sentiment Equation - $S_{i,t}$ | $\beta_{2,1}$ | −0.074 | 0.107 | 0.488 |
| | $\delta_{2,1}$ | 0.187 | 0.064 | 0.003 |
| | Panels | 15 | Obs. | 1995 |

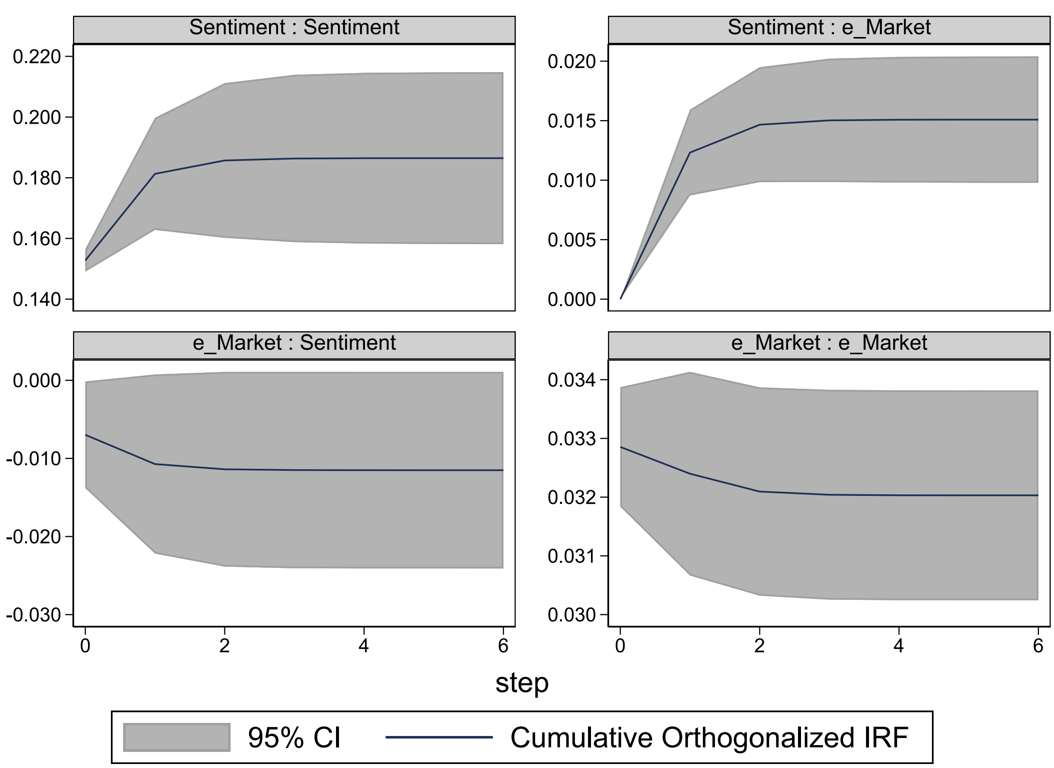


**Fig. 4.** Impulse response functions of Market and Sentiment with 95 per cent confidence bands.

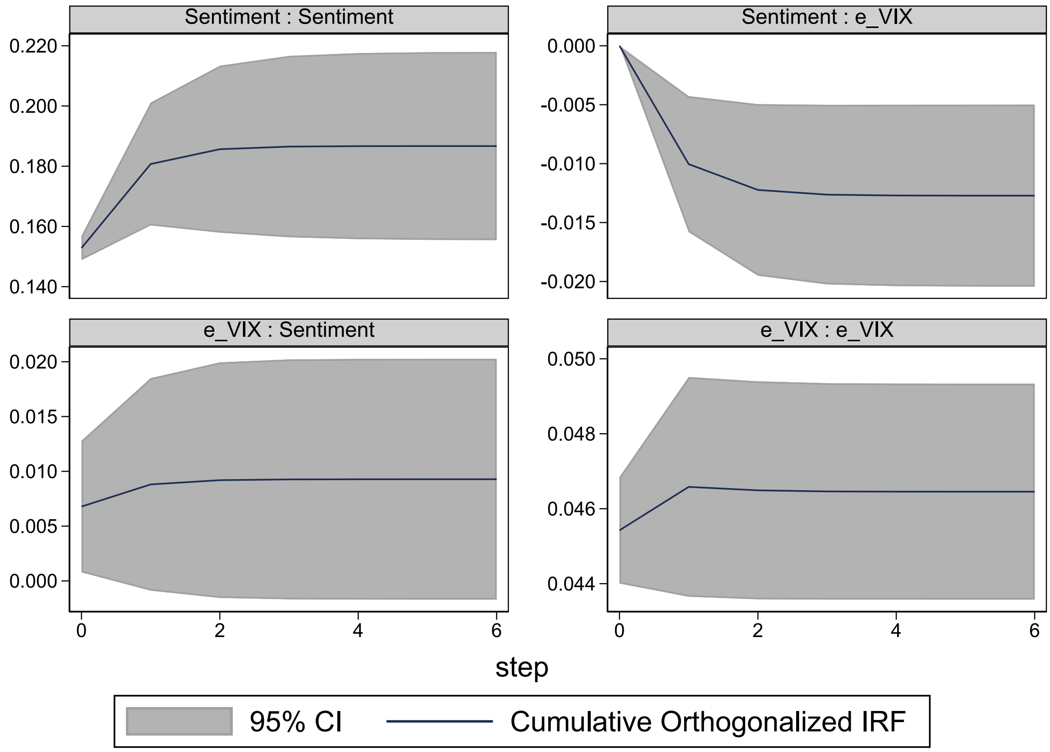


**Fig. 5.** Impulse response functions of Implied Volatility Innovation and Sentiment with 95 per cent confidence bands.

**Table 8**
The table includes results from a PVAR Granger-causality test, chi-square, and $p$-value.
Panel VAR-Granger-causality Wald test
Ho: Excluded variable does not Granger-cause Equation variable
Ha: Excluded variable Granger-causes Equation variable
Degrees of Freedom 1.

| Equation | Excluded | $\chi^2$ | p-val |
|---|---|---|---|
| Market | Sentiment | 32.072 | 0.000 |
| Sentiment | Market | 0.480 | 0.488 |

**Table 9**
Forecast Error Variance Decomposition for Stock Market and Immigration Narrative Sentiment. The table reports the response of Market and Sentiment to an orthogonal shock to Market and Sentiment, respectively.

| | Response Variable Market | | Response Variable Sentiment | |
|---|---|---|---|---|
| | Impulse Variable | | Impulse Variable | |
| Forecast Horizon | Market | Sentiment | Market | Sentiment |
| 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 1 | 1.0000 | 0.0000 | 0.0021 | 0.9979 |
| 2 | 0.8767 | 0.1233 | 0.0026 | 0.9974 |
| 3 | 0.8728 | 0.1272 | 0.0026 | 0.9974 |
| 4 | 0.8727 | 0.1273 | 0.0026 | 0.9974 |
| 5 | 0.8727 | 0.1273 | 0.0026 | 0.9974 |
| 6 | 0.8727 | 0.1273 | 0.0026 | 0.9974 |

**Table 10**
Panel VAR estimation results. The table reports coefficient estimates from a PVAR estimation including the implied volatility residual variable $e\_VIX_t$ and the immigration narrative sentiment panel variable $S_{i,t}$. Each equation includes one lag of each variable and is estimated with 15 panels of 1,995 observations after adjustment for lags. Robust standard errors and two-sided p-values appear to the right of each estimated coefficient.

| | Lag Variables | Coefficient | SE | p-val |
|---|---|---|---|---|
| VIX Residual Equation - $e\_VIX_{i,t}$ | $\beta_{1,1}$ | 0.035 | 0.030 | 0.239 |
| | $\delta_{1,1}$ | −0.066 | 0.019 | 0.001 |
| Sentiment Equation - $S_{i,t}$ | $\beta_{2,1}$ | 0.017 | 0.071 | 0.810 |
| | $\delta_{2,1}$ | 0.183 | 0.064 | 0.004 |
| | Panels | 15 | Obs. | 1995 |

**Table 11**
The table includes results from a PVAR Granger-causality test, chi-square, and $p$-value.
Panel VAR- Granger-causality Wald test
Ho: Excluded variable does not Granger-cause Equation variable
Ha: Excluded variable Granger-causes Equation variable
Degrees of Freedom 1.

| Equation | Excluded | $\chi^2$ | p-val |
|---|---|---|---|
| VIX Residual | Sentiment | 11.787 | 0.001 |
| Sentiment | VIX Residual | 0.058 | 0.810 |

Table 14 reports the Granger-causality test results for the trade volume PVAR. The test rejects the exclusion of the Sentiment variable from the trade volume equation at any conventional level and indicates that Sentiment Granger-cause trade volume movement. To the contrary, the trade volume does not Granger-cause change in the immigration narrative sentiment.

Table 15 shows the Forecast Error Variance Decomposition from the PVAR model (4). The dynamics are such that the sentiment variable drives up to 11 per cent of the trade volume variation unexplained by economic variables by the third month of the forecast horizon. The contribution of trade volume variation unexplained by economic indicators to the sentiment variation appears negligible.

The impulse response function in Fig. 6 shows how sentiment and trade volume dynamically affect each other. The top right figure shows that the effect of a shock to sentiment causes a significant lasting positive impact on trade volume. The effect is however one order of magnitude smaller as compared the market returns and volatility.

Overall, the impulse response functions of the market, implied volatility, and volume show that the responses to a shocks to the immigration narrative sentiment accumulate and are long lasting,

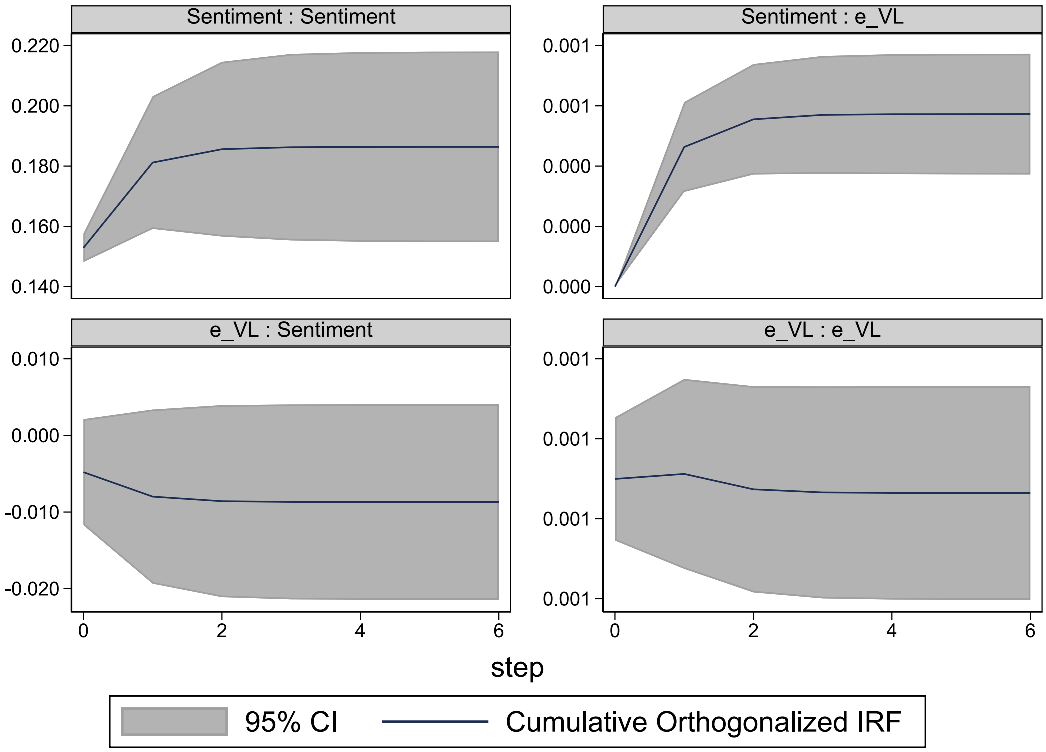


**Fig. 6.** Impulse response functions of Trade Volume Innovation and Sentiment with 95 per cent confidence bands.

**Table 12**
Forecast Error Variance Decomposition for Implied Volatility and Immigration Narrative Sentiment. The table reports the response of Implied Volatility Innovation and Sentiment to an orthogonal shock to Implied Volatility Innovation and Sentiment, respectively.

| | Response Variable *e_VIX* | | Response Variable Sentiment | |
|---|---|---|---|---|
| | Impulse Variable | | Impulse Variable | |
| Forecast Horizon | *e_VIX* | Sentiment | *e_VIX* | Sentiment |
| 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | 1.0000 | 0.0000 | 0.0020 | 0.9980 |
| 3 | 0.9534 | 0.0466 | 0.0021 | 0.9979 |
| 4 | 0.9513 | 0.0487 | 0.0021 | 0.9979 |
| 5 | 0.9513 | 0.0487 | 0.0021 | 0.9979 |
| 6 | 0.9513 | 0.0487 | 0.0021 | 0.9979 |

**Table 13**
Panel VAR estimation results. The table reports coefficient estimates for the Panel VAR estimation including the trade volume residual variable $e\_VL_t$ and the immigration narrative sentiment panel variable $S_{i,t}$. Each equation includes one lag of each variable and is estimated with 15 panels of 1,995 observations after adjustment for lags. Robust standard errors and two-sided p-values appear to the right of each estimated coefficient.

| | Lag Variables | Coefficient | SE | p-val |
|---|---|---|---|---|
| VL Residual Equation - $e\_VL_{i,t}$ | $\beta_{1,1}$ | 0.013 | 0.021 | 0.533 |
| | $\delta_{1,1}$ | 0.003 | 0.001 | 0.000 |
| Sentiment Equation - $S_{i,t}$ | $\beta_{2,1}$ | −1.739 | 2.759 | 0.529 |
| | $\delta_{2,1}$ | 0.185 | 0.064 | 0.004 |
| | Panels | 15 | Obs. | 1995 |

**Table 14**
The table includes results from a PVAR Granger-causality test, chi-square, and *p*-value.
Panel VAR- Granger-causality Wald test
Ho: Excluded variable does not Granger-cause Equation variable
Ha: Excluded variable Granger-causes Equation variable
Degrees of Freedom 1.

| Equation | Excluded | $\chi^2$ | p-val |
|---|---|---|---|
| VL Residual | Sentiment | 25.936 | 0.000 |
| Sentiment | VL Residual | 0.397 | 0.529 |

**Table 15**
Forecast Error Variance Decomposition for Trade Volume and Immigration Narrative Sentiment. The table reports the response of Trade Volume Innovation and Sentiment to an orthogonal shock to Trade Volume Innovation and Sentiment, respectively.

| | Response Variable *e_VL* | | Response Variable Sentiment | |
|---|---|---|---|---|
| | Impulse Variable | | Impulse Variable | |
| Forecast Horizon | *e_VL* | Sentiment | *e_VL* | Sentiment |
| 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | 1.0000 | 0.0000 | 0.0010 | 0.9990 |
| 3 | 0.8910 | 0.1090 | 0.0014 | 0.9986 |
| 4 | 0.8872 | 0.1128 | 0.0014 | 0.9986 |
| 5 | 0.8871 | 0.1129 | 0.0014 | 0.9986 |
| 6 | 0.8871 | 0.1129 | 0.0014 | 0.9986 |

which in turn suggests that the narrative sentiment proxies for information about the fundamental value of equities currently not incorporated into prices.

### 5.3. Robustness

As noted, the linguistic contour of a narrative is blurry. The limited availability of TV news from the internet archive constrains the time period and selection of n-grams. Nevertheless, to insure that the results reported in the previous section are robust to the selection of n-gram characterizing the narrative, several alternative PVAR specifications are estimated using different sets of n-grams. First, 5 additional n-grams with a small number of missing observation were added to the 15 that have complete series. The added n-grams are *children of immigrant*, *customs and border protection*, *path to citizenship*, *immigration cases*, and *illegal entries*. Since it is plausible that in absence of

**Table 16**
Panel VAR estimation results. The table reports coefficient estimates from the PVAR estimation including the Stock Market Return variable and the Immigration N-gram count variable. $C_{i,t}$ is the monthly news count in thousands. Each equation includes one lag of each variable and is estimated with 15 panels of 1,995 observations after adjustment for lags. standard errors and two-sided p-values appear to the right of each estimated coefficient.
$R_{i,t} = \gamma_{1i} + \beta_{1,1}R_{i,t-1} + \delta_{1,1}C_{i,t-1} + \epsilon_{1i,t}$
$C_{i,t} = \gamma_{2i} + \beta_{2,1}R_{i,t-1} + \delta_{2,1}C_{i,t-1} + \epsilon_{2i,t}$.

| | Lag Variables | Coefficient | SE | p-val |
|---|---|---|---|---|
| Market Residual Equation - $e\_RET_t$ | $\beta_{1,1}$ | −0.006 | 0.023 | 0.785 |
| | $\delta_{1,1}$ | −0.004 | 0.002 | 0.084 |
| Count Equation - $C_{i,t}$ | $\beta_{2,1}$ | 2.319 | 2.545 | 0.362 |
| | $\delta_{2,1}$ | 0.510 | 0.337 | 0.129 |
| | Panels | 15 | Obs. | 1995 |

news the sentiment stays unchanged, the missing observations are filled carrying forward the non-missing observation.

Additional robustness exercises were performed excluding 3 n-grams at a time from the set of complete series and estimating the PVAR with a panel of 12 n-grams. The results remain qualitatively unchanged. The Granger-causality tests show that in all specifications sentiment Granger-cause market variation. Variance decomposition results are also qualitatively unchanged but the proportion of variance explained by the sentiment ranges from mid single digit to about 20 per cent.

A simple measure that captures the spread of a narrative is based on the n-gram counts from a text corpus. There is however evidence that social dynamics and investor sentiment moves asset markets. This paper argues that a simple n-gram count may not be able to capture the effect a narrative on the emotions that motivate weaker or stronger behavioral response. Nevertheless, as a robustness exercise, this section discusses a bi-variate panel VAR specification with n-gram counts instead of the sentiment variable and a tri-variate PVAR specifications with the count in addition to the immigration narrative sentiment variable.

The bi-variate system is estimated with one lag and is stable. Results are reported in Table 16. The coefficient of the lagged Count variable is negative with a 0.084 p-val. All the other parameters are insignificant and, as a consequence, a shock to the TV news Count has a short term negative impact on the market return but its effects do not significantly accumulate over time.

A tri-variate PVAR, including Market, Count and Sentiment, not reported for brevity, confirms qualitatively the results presented thus far. The Immigration Narrative Sentiment is the only variable that has a significant impact of the and which effect does not fade over time. Depending on the impulse response function variable ordering, Sentiment and Count explain between 26 to about 35 per cent of the market variable variation. However, given the large standard error of the parameters estimates this figures should be taken with caution.

Thus, the results presented in Section 5 are overall robust. The following section discusses the results and puts them in the context of related contributions.

## 6. Discussion

The result reported above have similarities with Tetlock (2007) but also important differences. He finds that negative sentiment has and impact on future stock returns. The impact is however reversed in one week which indicates that sentiment does not contain new information about equities. While the monthly frequency in this study does not allow for short term reversals, the long term effect of immigration narrative sentiment appears to be a genuinely different outcome. This may be due to the intrinsic difference in the nature of the topical content between the Wall Street Journal column corpus in Tetlock (2007) vis-a-vis the TV news related to immigration used here.

The approach in this study is similar to Soo (2018), but also different in two important respects. First, here the sentiment measure here does not depend on a dictionary.[16] Unlike a dictionary, the n-gram clusters used in this paper is a vehicle to screen the universe of TV news for reports that touch upon themes related to the immigration narrative. Thus, it does not directly affect the polarity of the sentiment extracted by the sentiment classifier. Second, this paper uses the n-grams cluster to identify immigration stories within the TV news archive. Unlike Soo who studies how the sentiment from local newspapers affect home prices, the goal here is to examine whether the immigration narrative can explain any of the stock price, volatility, and trade volume variation during the sample period.

This study confirm (Baker and Blau, 2019) results despite their approach is focused on a labor economics channel and few sectors. It also confirm findings in Ackert and Mazzotta (2021) who find that sentiment related to the American home ownership narrative significantly explains movements in home prices. Uhl (2014), and Heston and Sinha (2017) too find that sentiment predicts returns. Unlike here however, their sentiment is directly referred to the stock news and independent of any specific narrative.

The results diverge from Brown and Cliff (2004) who find that market returns predict subsequent levels and changes of sentiment, but sentiment measures do not seem to predict future stock returns with high significance. They also contrasts with Mudinas et al. (2019) who report that sentiment does not Granger-cause stock price changes and Renault (2020) who report high correlation between sentiment and stock returns but do not find that investor sentiment derived from messages sent on social media helps in predicting large capitalization stocks return at a daily frequency. The also differ from Chen et al. (2022) who find that the COVID-19 pandemic-relevant narrative affects the stock market but no evidence that the COVID-19 narratives Granger-cause stock market conditions. It should be emphasized that methods and data in these studies differ along several important dimensions, including the very notion and proxies for sentiment. Therefore, a direct comparison may be difficult.

## 7. Concluding remarks

Stories are difficult to quantify, yet they motivate people to take actions with economic consequences. Immigration stories are often controversial and news coverage takes at times contentious tones. This paper relies on narrative economics, an approach proposed by Robert Schiller that studies how stories drive economics events. It measures the mutual relationship between the immigration narrative sentiment and stock market fluctuation.

This study contributes to the field of narrative economics by quantifying the relationship between the immigration narrative and the stock market. Moreover, the relationship between financial markets and immigration is under studied and this paper contributes to filling the gap.

The paper hypothesizes that the contour of a narrative can be defined by a cluster of n-grams that signals the presence of the narrative and quantified by the sentiment extracted from TV news transcripts using FLAIR, a state-of-the-art NLP algorithm.

I use Panel Vector Autoregression (PVAR) to empirically tests whether and to what extent the market and the immigration narrative are related. The PVAR allows to test the joint hypotheses that the n-grams are connected to the immigration narrative and

[16] See Loughran and McDonald (2011).

that the narrative sentiment of the designated n-gram jointly explains the variation of stock price unaccounted for by other known economic factors.

The results indicate that immigration narrative sentiment has a statistically significant and economic meaningful effect on the stock market. A one orthogonalized standard deviation positive shock to the immigration narrative sentiment Granger-cause a lasting increase in the stock market of about 1.5 per cent of one standard deviation of the marker return innovation. Likewise, one standard deviation positive shock to the immigration narrative sentiment Granger-cause a lasting decrease in the implied volatility of about 1.2 per cent of one standard deviation of the implied volatility innovation. The response to a sentiment shock Granger-cause a statistically significant but economically small change in the trade volume. Market variables shocks do not cause changes in the immigration narrative sentiment. These results overall indicate that the immigration narrative sentiment likely contains fundamental information about equities that has not been priced.

A limitation of this study is that it does not attempt to identify the channels through which the immigration narrative might affect the stock market. Also, it does not study the relationship between immigration policies and how they may affect the stock market or specific sectors. It likewise does not address how other important narratives relate to specific market sectors. These interesting themes are left for future research.

## CRediT authorship contribution statement

**Stefano Mazzotta:** Conceptualization, Methodology, Investigation, Software, Data curation, Writing, Visualization.

## References


Abrigo, M.R.M., Love, I., 2016. Estimation of panel vector autoregression in stata. Stata J. 16 (3), 778–804. http://dx.doi.org/10.1177/1536867X1601600314.

Ackert, L.F., Mazzotta, S., 2021. Homeownership for all: An American narrative. J. Risk Financ. Manag. (ISSN: 1911-8074) 14 (6), http://dx.doi.org/10.3390/jrfm14060240, URL https://www.mdpi.com/1911-8074/14/6/240.

Akbik, A., Bergmann, T., Blythe, D.A.J., Rasul, K., Schweter, S., Vollgraf, R., 2019. FLAIR: An easy-to-use framework for state-of-the-art NLP. In: NAACL-HLT.

Akbik, A., Blythe, D., Vollgraf, R., 2018. Contextual string embeddings for sequence labeling. In: Proceedings of the 27th International Conference on Computational Linguistics. Association for Computational Linguistics, Santa Fe, New Mexico, USA, pp. 1638–1649, URL https://aclanthology.org/C18-1139.

Algaba, A., Ardia, D., Bluteau, K., Borms, S., Boudt, K., 2020. Econometrics meets sentiment: An overview of methodology and applications. J. Econ. Surv. 34 (3), 512–547.

Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. J. Econometrics (ISSN: 0304-4076) 68 (1), 29–51. http://dx.doi.org/10.1016/0304-4076(94)01642-D, URL https://www.sciencedirect.com/science/article/pii/030440769401642D.

Baker, J., Blau, B.M., 2019. Immigration and the Stock Market : How Immigration Policy Affects the Stock Prices of Firms that Employ Low-Skilled Workers.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. J. Econ. Perspect. 21 (2), 129–152. http://dx.doi.org/10.1257/jep.21.2.129, URL https://www.aeaweb.org/articles?id=10.1257/jep.21.2.129.

Barbu, T., Iustina, B., 2018. Immigrants' impact on financial market – European countries' evidence. In: Zbornik Radova Ekonomskog Fakulteta U Rijeci-Proceedings of Rijeka Faculty of Economics, University of Rijeka, Faculty of Economics and Business, vol. 36. pp. 183–212. http://dx.doi.org/10.18045/zbefri.2018.1.183.

Bastani, K., Namavari, H., Shaffer, J., 2019. Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. Expert Syst. Appl. (ISSN: 0957-4174) 127, 256–271. http://dx.doi.org/10.1016/j.eswa.2019.03.001, URL https://www.sciencedirect.com/science/article/pii/S095741741930154X.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet allocation. J. Mach. Learn. Res. (ISSN: 1532-4435) 3 (null), 993–1022.

Borjas, G.J., 1995. The economic benefits from immigration. J. Econ. Perspect. 9 (2), 3–22. http://dx.doi.org/10.1257/jep.9.2.3, URL https://www.aeaweb.org/articles?id=10.1257/jep.9.2.3.

Borjas, G.J., 2014. Immigration Economics. Harvard University Press.

Brown, G.W., Cliff, M.T., 2004. Investor sentiment and the near-term stock market. J. Empir. Financ. (ISSN: 0927-5398) 11 (1), 1–27. http://dx.doi.org/10.1016/j.jempfin.2002.12.001, URL https://www.sciencedirect.com/science/article/pii/S0927539803000422.

Bursztyn, L., Rao, A., Roth, C., Yanagizawa-Drott, D., 2022. Opinions as facts. URL https://home.uchicago.edu/bursztyn/OpinionsAsFacts_Revised.pdf.

Campbell, J.Y., 1991. A variance decomposition for stock returns. Econ. J. 101 (405), 157–179, (ISSN: 00130133, 14680297). URL http://www.jstor.org/stable/2233809.

Campbell, J.Y., Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Rev. Financ. Stud. 1 (3), 195–228, (ISSN: 08939454, 14657368). URL http://www.jstor.org/stable/2961997.

Card, D., Peri, G., 2016. Immigration economics, by George J. Borjas: A review essay. J. Econ. Lit. 54 (4), 1333–1349. http://dx.doi.org/10.1257/jel.20151248, URL https://www.aeaweb.org/articles?id=10.1257/jel.20151248.

Chan, W.S., 2003. Stock price reaction to news and no-news: drift and reversal after headlines. J. Financ. Econ. (ISSN: 0304-405X) 70 (2), 223–260. http://dx.doi.org/10.1016/S0304-405X(03)00146-6, URL https://www.sciencedirect.com/science/article/pii/S0304405X03001466.

Chen, Y., Bredin, D., Potì, V., Matkovskyy, R., 2022. COVID risk narratives: a computational linguistic approach to the econometric identification of narrative risk during a pandemic. Digit. Finance (ISSN: 2524-6186) 4 (1), 17–61. http://dx.doi.org/10.1007/s42521-021-00045-3.

De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Noise trader risk in financial markets. J. Polit. Econ. 98 (4), 703–738, (ISSN: 00223808, 1537534X). URL http://www.jstor.org/stable/2937765.

Gupta, A., Dengre, V., Kheruwala, H.A., Shah, M., 2020. Comprehensive review of text-mining applications in finance. Financ. Innov. (ISSN: 2199-4730) 6 (1), 39. http://dx.doi.org/10.1186/s40854-020-00205-1.

Heston, S.L., Sinha, N.R., 2017. News vs. Sentiment: Predicting stock returns from news stories. Financ. Anal. J. 73 (3), 67–83. http://dx.doi.org/10.2469/faj.v73.n3.3.

Hodrick, R.J., 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. Rev. Financ. Stud. 53 (3), 357–386. http://dx.doi.org/10.1093/rfs/5.3.351.

Holtz-Eakin, D., Newey, W., Rosen, H.S., 1988. Estimating vector autoregressions with panel data. Econometrica (ISSN: 00129682, 14680262) 56 (6), 1371–1395, URL http://www.jstor.org/stable/1913103.

Kerr, S.P., Kerr, W.R., 2011. Economic Impacts of Immigration: A Survey. Working Paper Series 16736, National Bureau of Economic Research, http://dx.doi.org/10.3386/w16736, URL http://www.nber.org/papers/w16736.

Li, J., Sun, A., Han, J., Li, C., 2018. A survey on deep learning for named entity recognition. CoRR arXiv:1812.09449.

Li, G., Zhu, X., Wang, J., Wu, D., Li, J., 2017. Using LDA model to quantify and visualize textual financial stability report. In: 5th International Conference on Information Technology and Quantitative Management, ITQM 2017. Procedia Comput. Sci. (ISSN: 1877-0509) 122, 370–376. http://dx.doi.org/10.1016/j.procs.2017.11.382, URL https://www.sciencedirect.com/science/article/pii/S187705091732625X.

Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. J. Finance 66 (1), 35–65. http://dx.doi.org/10.1111/j.1540-6261.2010.01625.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2010.01625.x.

Mudinas, A., Zhang, D., Levene, M., 2019. Market trend prediction using sentiment analysis: Lessons learned and paths forward. CoRR arXiv:1903.05440.

Otter, D.W., Medina, J.R., Kalita, J.K., 2021. A survey of the usages of deep learning for natural language processing. IEEE Trans. Neural Netw. Learn. Syst. (ISSN: 2162-2388) 32 (2), 604–624. http://dx.doi.org/10.1109/TNNLS.2020.2979670.

Peri, G., 2012. The effect of immigration on productivity: Evidence from U.S. states. Rev. Econ. Stat. (ISSN: 0034-6535) 94 (1), 348–358. http://dx.doi.org/10.1162/REST_a_00137.

Renault, T., 2020. Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages. Digit. Finance (ISSN: 2524-6186) 2 (1), 1–13. http://dx.doi.org/10.1007/s42521-019-00014-x.

Shiller, R.J., 2015. Irrational Exuberance. In: Economics Books, vol. 10421, Princeton University Press, URL https://ideas.repec.org/b/pup/pbooks/10421.html.

Shiller, R.J., 2017. Narrative economics. Am. Econ. Rev. 107 (4), 967–1004, (ISSN: 00028282, 19447981). URL http://www.jstor.org/stable/44251584.

Shiller, R.J., 2019. Narrative Economics: How Stories Go Viral and Drive Major Economic Events. Princeton University Press, Princeton, New Jersey.

Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. J. Finance 52 (1), 35–55. http://dx.doi.org/10.1111/j.1540-6261.1997.tb03807.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1997.tb03807.x.

Soo, C.K., 2018. Quantifying sentiment with news media across local housing markets. Rev. Financ. Stud. (ISSN: 0893-9454) 31 (10), 3689–3719. http://dx.doi.org/10.1093/rfs/hhy036.

Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. J. Finance 62 (3), 1139–1168. http://dx.doi.org/10.1111/j.1540-6261.2007.01232.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2007.01232.x.
Trussler, M., Soroka, S., 2014. Consumer demand for cynical and negative news frames. Int. J. Press/Politics 19 (3), 360–379. http://dx.doi.org/10.1177/1940161214524832.
Uhl, M.W., 2014. Reuters sentiment and stock returns. J. Behav. Finance 15 (4), 287–298. http://dx.doi.org/10.1080/15427560.2014.967852.
University of Michigan, 2018. University of Michigan: Consumer Sentiment [UMCSENT], retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/UMCSENT, May 29, 2018.
Wilson, T., Wiebe, J., Hoffmann, P., 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural LanguageProcessing. pp. 347–354.