ESSAYS ON HIGH FREQUENCY MACROECONOMIC MONITORING

Xiang LI

May 29, 2020

- Real time tracking of the present state of macroeconomic activity, particularly for tracking recessions in real time, is of great interest to firms, workers, financial market participants and policy makers
- Despite significant research focus on forecasting and nowcasting macroeconomic activity, there are still substantial delays in identifying key macroeconomic fluctuations
- For example, the December 2007 peak of the Great Recession was not identified until mid-to-late 2008 by statistical tracking models in real time
- In this dissertation, the dominant theme is evaluating techniques and developing novel datasets for improved high frequency monitoring of the macroeconomy
- The first paper stands apart from the other chapters in its focus; however, they have some connection in methods, particularly in the use of dynamic factor models

Are the Effects of Monetary Policy Asymmetric in China?

Introduction

- China has become the second largest economic engine in the world measured by nominal GDP; however, there has been relatively little work done on understanding the effects of Chinese monetary policy
- In other countries, substantial attention has been paid to state-dependence, or asymmetry, in the effects of policy over the business cycle
  - Thoma (1994), Garcia & Schaller (2002), Lo & Piger (2005), Weise (1999), Kaufmann (2002), Peersman & Smets (2001)
  - Tenreyro & Thwaites (2016)
- I find evidence that monetary policy shocks have larger impacts on output growth in low-growth states, and that monetary policy shocks have larger effects on inflation in high-growth states
- My paper is the first to study asymmetric effects of monetary policy on the Chinese economy over the business cycle

Methodology

Step #1: Extracting the Latent Dynamic Factor

- Motivation:
  - Output in China cannot be measured directly with industrial production or GDP, due to the poor quality of the officially published Chinese economic data (Rawski (2001), Maier (2011), Wallace (2014), Holz (2014))
  - One of the method to evaluate Chinese economy when data is suspected to be inaccurate is to extract latent factors from a large panel of underlying time series (Mehrotra & Paakkonen (2011), He, Leung, & Chong (2013), Fernald, Spiegel, & Swason (2014))
- Advantages:
  - Dynamic factors extracted from a large number of underlying variables convey more information regarding Chinese economic activity than the reported GDP data alone
  - Need no prior knowledge to determine which variables to include in the model

Methodology

Step #1: Extracting the Latent Dynamic Factor

- Data: downloaded from the CEIC China Premium Database
  - 30 fundamental series that correlate with output
  - 4 price indexes
- Pre-processing:
  - removing effects of the Lunar New Year
  - adjusting for seasonality
  - taking monthly growth rates of each series
  - handling missing values and outliers by an iterative expectation-maximization algorithm
  - standardizing
  - removing a local mean from each series using a biweight kernel

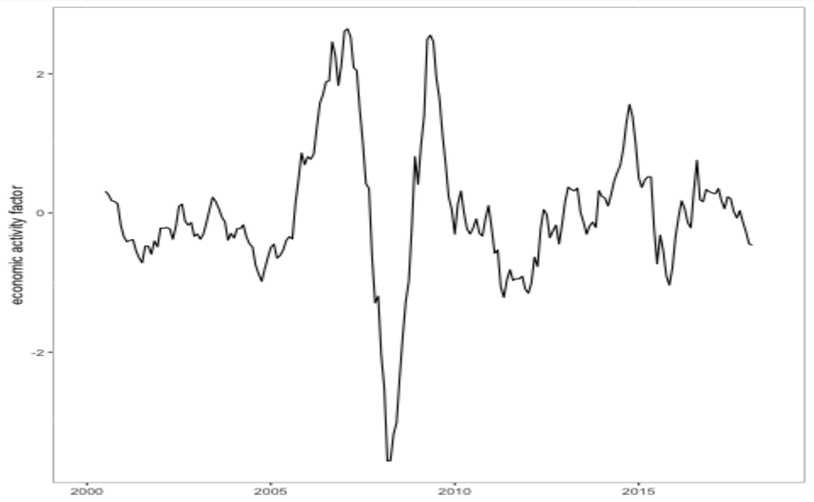
Methodology

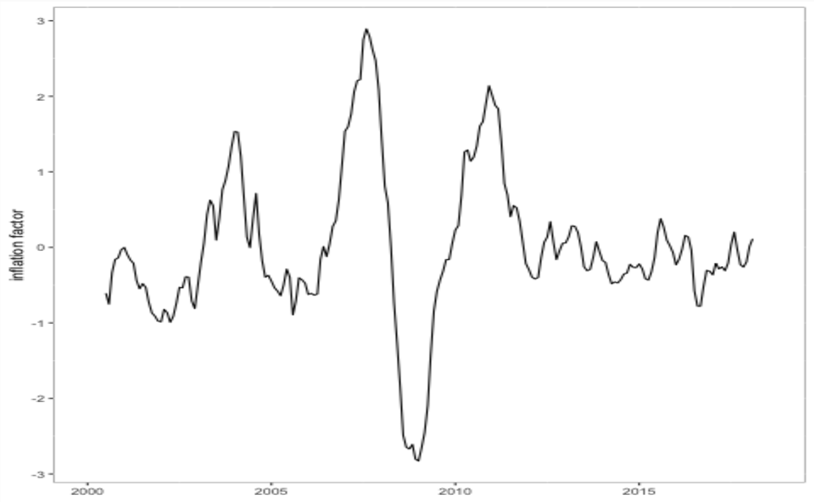
Step #1: Extracting the Latent Dynamic Factor

I measure factors f_t with the first principal component from a dynamic factor model:

$$X_t = \lambda(L) f_t + e_t$$

$$f_t = \Psi(L) f_{t-1} + \eta_t$$





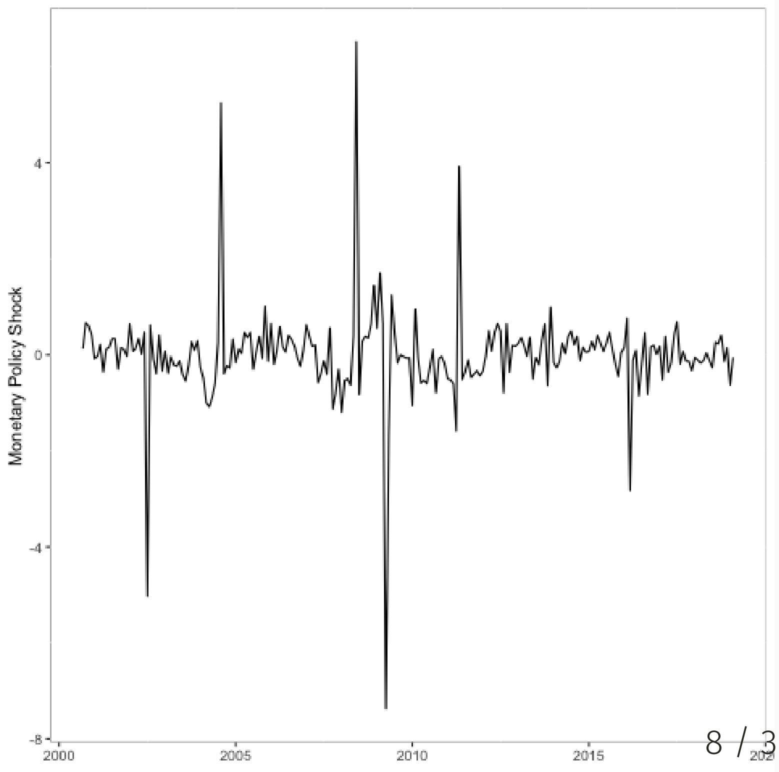
Methodology

Step #2: Measuring Monetary Policy Shocks

I Measure structural shocks of monetary policy from Choleski decomposition of residuals Σ from the following Factor Augmented Vector Autoregression (FAVAR) model:

$$\begin{bmatrix} f_t^e \\ f_t^p \\ y_t \end{bmatrix} = A(L) \begin{bmatrix} f_{t-1}^e \\ f_{t-1}^p \\ y_{t-1} \end{bmatrix} + u_t$$

$$u_t \overset{iid}{\sim} N(0, \Sigma)$$

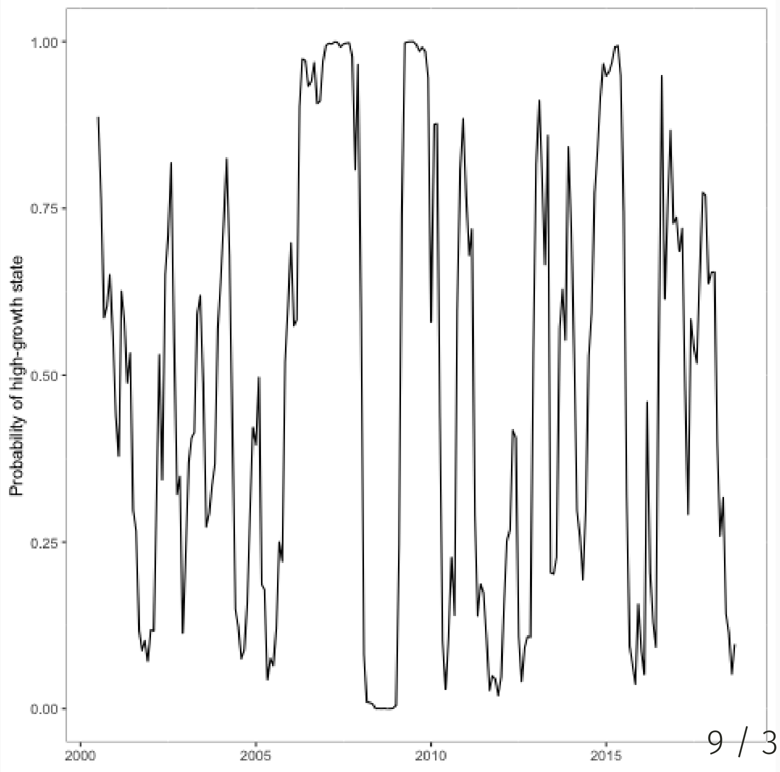


Methodology

Step #3: Identifying High and Low Growth Phases

As in Tenreyro & Thwaites (2016), I measure probabilities of the unobservable states of the economy z_t being in high-growth states using a smooth transition logistic function:

$$F(z_t) = \frac{exp(\theta\frac{z_t-c}{\sigma_z})}{1+exp(\theta\frac{z_t-c}{\sigma_z})}$$



Methodology

Step #4: Estimating Impulse Response Functions using Local Projections

- Following Tenreyro & Thwaites (2016), the baseline model specifies the impulse response of the standardized economic activity factor at horizon g in state $j \in \{high, low\}$ to a shock u_t as the coefficient β_g^j:

$$f^e_{t+g} = \tau t + F(z_t)(\alpha^h_g + \beta^h_g u_t + \gamma^h_{1,g} f^e_{t-1} + \gamma^h_{2,g} b_{t-1})$$

$$+(1 - F(z_t))(\alpha^l_g + \beta^l_g u_t + \gamma^l_{3,g} f^e_{t-1} + \gamma^l_{4,g} b_{t-1}) + \nu_{t+g}$$

- Same specification for the standardized inflation factor:

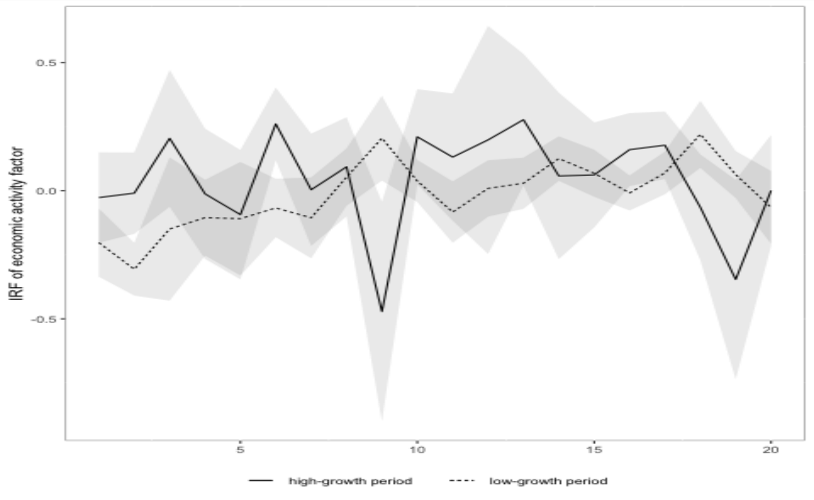
$$f^p_{t+g} = \tau t + F(z_t)(\alpha^h_g + \beta^h_g u_t + \gamma^h_{1,g} f^p_{t-1} + \gamma^h_{2,g} b_{t-1})$$

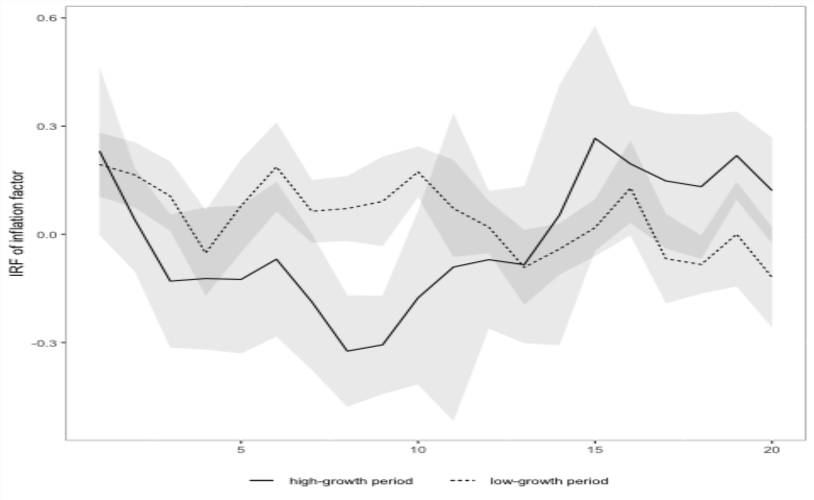
$$+(1 - F(z_t))(\alpha^l_g + \beta^l_g u_t + \gamma^l_{3,g} f^p_{t-1} + \gamma^l_{4,g} b_{t-1}) + \nu_{t+g}$$

Methodology

Step #4: Estimating Impulse Response Functions using Local Projections

- Following Tenreyro & Thwaites (2016), I measure the response at period g by regressing dependent variables over period $g+1, g+2, \ldots, T$ on independent variables over period $1, 2, \ldots, T-g$
- Monetary policy shocks having larger impacts on output growth during low-growth states and have larger impacts on inflation during high-growth states





Future Directions

Price factors:

- This chapter has only included CPI series into the inflation factor
- I plan to incorporate more prices indicators, such as PPI and house price

Policy variable:

- This chapter has used M2 growth rate as the monetary policy instrument. After the financial crisis in 2008, People's Bank of China started to change its policy framework from the benchmark interest rate to the inflation-targeting interest rate
- I plan to measure the monetary policy shock following Chen, Higgins, Waggoner & Zha (2016)

Nowcasting Business Cycle Phases with High and Mixed Frequency Data

Introduction

- The business cycle consists of alternating periods of expansion and recession, which are not explicitly observed
- Identification of business cycle turning points generally occurs long after the fact
    - For example, the December 2007 peak of the Great Recession was not announced by the National Bureau of Economic Research (NBER) until December 1, 2008
- An active literature has worked to develop statistical techniques to “nowcast” business cycle turning points toward the end of the observable sample period
    - Chauvet & Piger (2008), Chauvet & Hamilton (2006), Fossati (2016), Giusto & Piger (2017)
    - Uses relatively low frequency data (monthly) to nowcast turning points
- My paper is the first to systematically evaluate the ability of high frequency data (daily, weekly) to improve upon the timeliness with which new expansions and recessions can be identified

Methodology

Step #1: Pseudo Real-Time Dataset

- Data:
  - daily yield curve term premium, defined as the difference between daily 10-year and daily 3-month U.S. Treasury yields
  - weekly initial claims for unemployment insurance
  - monthly nonfarm payroll employment
  - quarterly real GDP
- Pseudo Real-Time Dataset:
  - From Jan 1, 1979 to April 30, 2020, on every Sunday (called "analysis date", denoted by T), I updated the above variables
  - I stripped each series of a linear and a quadratic trend, and standardized the residuals

Methodology

Step #2: Dynamic Factor Model at Daily Frequency

- Following Aruoba, Diebold, & Scotti (2009), I propose a dynamic factor model at daily frequency
- x_t: the underlying real economic activity factor on day t, assumed to evolve daily with AR(1) dynamics:

$$x_t = \rho x_{t-1} + e_t$$

- y_t^1: term premium (daily), a stock variable

$$\begin{aligned} y_t^1 &= \begin{cases} \beta_1 x_t + u_t^1 \\ NA \end{cases} \\ &= \begin{cases} \beta_1 x_t + \gamma_1 u_{t-1}^1 + \zeta_t & y_t^1 \text{is observed} \\ NA & y_t^1 \text{is not observed} \end{cases} \end{aligned}$$

Methodology

Step #2: Dynamic Factor Model at Daily Frequency

- Following Aruoba, Diebold, & Scotti (2009), I propose a dynamic factor model at daily frequency
- y_t^2: initial claims for unemployment insurance (weekly), a flow variable

$$y_t^2 = \begin{cases} \beta_2 C_t^W + \gamma_2 y_{2,t-7} + u_t^2 & y_t^2 \text{is observed} \\ NA & y_t^2 \text{is not observed} \end{cases}$$

$$C_t^W = \xi_t^W C_{t-1}^W + x_t = \xi_t^W C_{t-1}^W + \rho x_{t-1} + e_t$$

$$\xi_t^W = \begin{cases} 0 & \text{if t is the first day of a week} \\ 1 & \text{otherwise} \end{cases}$$

Methodology

Step #2: Dynamic Factor Model at Daily Frequency

- Following Aruoba, Diebold, & Scotti (2009), I propose a dynamic factor model at daily frequency
- y_t^3: nonfarm payroll employment (monthly), a stock variable

$$y_t^3 = \begin{cases} \beta_3 x_t + \gamma_3 y_{3,t-30} + u_t^3 & y_t^3 \text{is observed} \\ NA & y_t^3 \text{is not observed} \end{cases}$$

Methodology

Step #2: Dynamic Factor Model at Daily Frequency

- Following Aruoba, Diebold, & Scotti (2009), I propose a dynamic factor model at daily frequency
- y_t^4: real GDP (quarterly), a flow variable

$$y_t^4 = \begin{cases} \beta_4 C_t^Q + \gamma_4 y_{4,t-90} + u_t^4 & y_t^4 \text{is observed} \\ NA & y_t^4 \text{is not observed} \end{cases}$$

$$C_t^Q = \xi_t^Q C_{t-1}^Q + x_t = \xi_t^Q C_{t-1}^Q + \rho x_{t-1} + e_t$$

$$\xi_t^Q = \begin{cases} 0 & \text{if t is the first day of a quarter} \\ 1 & \text{otherwise} \end{cases}$$

Methodology

Step #2: Dynamic Factor Model at Daily Frequency

$$\underbrace{\begin{bmatrix} y_t^1 \\ y_t^2 \\ y_t^3 \\ y_t^4 \end{bmatrix}}_{\boldsymbol{\Upsilon_t}} = \underbrace{\begin{bmatrix} \gamma_1 & \beta_1 & 0 & 0 & 0 \\ 0 & 0 & \beta_2 & 0 & \gamma_2 \times y_{2,t-7} \\ 0 & \beta_3 & 0 & 0 & \gamma_3 \times y_{3,t-30} \\ 0 & 0 & 0 & \beta_4 & \gamma_4 \times y_{4,t-90} \end{bmatrix}}_{\boldsymbol{FF_t}} \times \underbrace{\begin{bmatrix} u_{t-1}^1 \\ x_t \\ C_t^W \\ C_t^Q \\ 1 \end{bmatrix}}_{\boldsymbol{\theta_t}} + \underbrace{\begin{bmatrix} \zeta_t \\ u_t^2 \\ u_t^3 \\ u_t^4 \end{bmatrix}}_{\boldsymbol{\nu_t}}$$

$$\underbrace{\begin{bmatrix} u_{t-1}^1 \\ x_t \\ C_t^W \\ C_t^Q \\ 1 \end{bmatrix}}_{\boldsymbol{\theta_t}} = \underbrace{\begin{bmatrix} \gamma_1 & 0 & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 & 0 \\ 0 & \rho & \xi_t^W & 0 & 0 \\ 0 & \rho & 0 & \xi_t^Q & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\boldsymbol{GG_t}} \times \underbrace{\begin{bmatrix} u_{t-2}^1 \\ x_{t-1} \\ C_{t-1}^W \\ C_{t-1}^Q \\ 1 \end{bmatrix}}_{\boldsymbol{\theta_{t-1}}} + \underbrace{\begin{bmatrix} \zeta_{t-1} \\ e_t \\ e_t \\ e_t \\ 0 \end{bmatrix}}_{\boldsymbol{\omega_t}}$$

Methodology

Step #2: Dynamic Factor Model at Daily Frequency

- The model can be represented in time-varying state-space form as:

$$\boldsymbol{\Upsilon_t} = \boldsymbol{FF_t} \times \boldsymbol{\theta_t} + \boldsymbol{\nu_t}$$
$$\boldsymbol{\theta_t} = \boldsymbol{GG_t} \times \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t}$$

- I use the Kalman filter to form the likelihood function. Then I estimate paramteres using Maximum Likelihood Estimation with all available data. Then I plug in estimates of parameters into equations and use the Kalman smoother to extract a daily index of real economic activity $\hat{x}_t$

Methodology

Step #2: Dynamic Factor Model at Daily Frequency

Methodology

Step #3: Supervised (Markov-switching) Regime Classifications

- $S_t = 0$: day t is a expansion regime; $S_t = 1$: day t is a recession regime
- $p_{ji} = Pr(S_t = j | S_{t-1} = i)$
- Fitting the first difference of $\hat{x}_t$ to a univariate Markov-switching AR(0) process with a switching mean:

$$\begin{aligned} \Delta\hat{x}_t &= \beta_{S_t} + \epsilon_t \\ \epsilon_t &\sim N(0, \sigma^2) \end{aligned}$$

- The parameters of the model: $\Omega = (\beta_0, \beta_1, p_{00}, p_{11}, \sigma)'$

Methodology

Step #3: Supervised (Markov-switching) Regime Classifications

- On each analysis date T, using non-parametric techniques, I estimate parameters of the model using data up to one year from T:
    - Estimating β_0 and β_1 as the mean of $\Delta\hat{x}_t$ in each NBER regime
    - Estimating transition probabilities as the mean of transitions using the NBER regimes
    - Estimating variance of the disturbance terms as the residuals of the regression
- Given these estimates, I run the Hamilton smoother through data to the end of T in order to obtain the recession probabilities, denoted $\hat{P}(S_t = 1|\Psi_T)$

Methodology

Step #4: Business Cycle Phases Dating Procedures

- Identifying a new recession
  - Suppose that the last NBER turning point date that was announced is a business cycle trough
  - I search for the first analysis date T for which the average value of $\hat{P}(S_t = 1|\Psi_T)$ over the 12 weeks prior to T exceeds 0.8
- Identifying a new expansion
  - Suppose that the last NBER turning point date that was announced is a business cycle peak
  - I search for the first analysis date T for which the average value of $\hat{P}(S_t = 1|\Psi_T)$ over the 12 weeks prior to T is below 0.2
- I also produce results for an alternative threshold of 0.9 as a robustness check

Preliminary Results of the Dynamic Factor Markov Switching Model at Daily Frequency (DFMSDF)

NBER Peak Date	First Day of Recession	Date Peak Call Available - DFMSDF	NBER's Business Cycle Dating Committee Lag	Giusto & Piger (2017) Lag	DFMSDF Lag
Jan-1980	2/1/1980	5/11/1980	123	92	100
Jul-1981	8/1/1981	11/1/1981	158	126	92
Jul-1990	8/1/1990	7/29/1990	267	78	-3
Mar-2001	4/1/2001	7/2/2000	239	216	-273
Dec-2007	1/1/2008	12/9/2007	335	158	-23
Average			**224**	**134**	**-21**
		3/15/2020			

NBER Trough Date	First Day of Expansion	Date Trough Call Available - DFMSDF	NBER's Business Cycle Dating Committee Lag	Giusto & Piger (2017) Lag	DFMSDF Lag
Jul-1980	8/1/1980	8/10/1980	341	127	9
Nov-1982	12/1/1982	12/12/1982	219	136	11
Mar-1991	4/1/1991	6/2/1991	631	443	62
Nov-2001	12/1/2001	8/19/2001	593	308	-104
Jun-2009	7/1/2009	5/31/2009	446	157	-31
Average			**446**	**234**	**-11**

Preliminary Baseline Results

- False Recessions and False Expansions Identified in Real-time (Threshold=0.8):

False Recessions	**False Expansions**
2/16/1979-12/30/1979	2/28/1982-3/14/1982
10/7/1984-11/11/1984	
1/24/1988-2/7/1988	
7/9/1989-8/6/1989	
4/23/1995-6/18/1995	
4/13/2003-5/4/2003	

- These are all short in that they only send a false signal for a few weeks
- In future work I plan to incorporate additional series into the analysis to reduce the number of false recessions and expansions

Future Directions

A broader vintage real-time dataset:

- Modifying the method to incorporate additional series, such as the unemployment rate, industrial production, and real manufacturing and trade sales
- The current data set incorporates data revisions, since it was obtained in April 2020
  - I plan to compile the vintage data into a real-time dataset that would have been available on each day
  - By using the vintage real-time dataset, I search for new business phases turning points as if I were an analyst each day beginning on Jan 1, 1979

A broader variety of supervised regime classifications:

- I plan to train a variety of supervised machine learning classification techniques
  - the Naive Bayes classifier, the k-nearest neighbor classifier, learning vector quantization, random forest, and boosting

A New High Frequency, News Based, Indicator of Macroeconomic Activity

Introduction

- The information encoded in text has been recently used in empirical economics research as a complement to the more structured macroeconomic and financial data traditionally used
  - Gentzkow, Kelly, & Taddy (2019), Kelly, Manela, & Moreira (2019), Shapiro, Sudhof, & Wilson (2020), Bybee, Kelly, Manela, & Xiu (2020)
- Text selected from news articles arrives daily and is not revised, making textual data an ideal candidate to build more accurate nowcasting models about aggregate economic activity in real time
- In this paper I propose a text-based approach to create a high-frequency News-Based Sentiment Indicator (NBSI) regarding aggregate economic conditions, and evaluate the predictive ability of the NBSI
  - use NBSI to track a wide range of economic activity measures
  - estimate the contribution of NBSI to identify U.S. recessions in real time by incorporating NBSI in models developed in the previous chapter

Methodology

Step #1: Data

From Factiva Database, I collect a large sample of leading paragraphs of 410,601 articles published at daily frequency in the Wall Street Journal from April 2, 1991 to April 30, 2020 that have the following subjects:

- News about commodity, debt, bond, equity, money and currency markets
- Analysts' comments or recommendations about corporates and industries
- Economic performance or indicators, government finance, monetary policy, trade or external payments

Methodology

Step #1: Data

The leading paragraph is represented at each point in time. An example of this sample:

Date	Leading Paragraph
03-28-2006	Jon Corzine, New Jersey's new Governor, isn't the first politician not to follow through on a campaign promise. But rarely is such dishonesty later presented as a virtue. The question for voters to contemplate is whether this is also an indication of what to expect if Democrats gain control of Congress in November.Mr. Corzine won the Trenton statehouse last year by running as a tax cutter who'd raise property tax rebates by 40% over four years. "I'm not considering raising taxes. It's not on my agenda. We have a very high-rate tax structure. I'm not considering it," the then-U.S. Senator had vowed in October.
03-29-2006	A growing number of homeowners, riding the crest of the real-estate boom, are getting hit by an unpleasant surprise when they sell: a hefty tax bill.This development stems from a 1997 law that Treasury Department officials said at the time would eliminate capital-gains taxes for nearly everyone selling their primary residence. Under that law, most married couples who file jointly can exclude as much as $500,000 of their gain. For most singles, the limit is $250,000.
03-29-2006	The airline most often viewed as the strongest, healthiest and best-run of the pack may be one of the weaker bets for investors hoping to profit from a budding turnaround in the industry.A strengthening market has sent some beleaguered airline stocks soaring in recent months, but shares in industry stalwart Southwest Airlines have been more sluggish. In the past year, the stock of American Airlines parent AMR Corp. has risen 168% on the New York Stock Exchange. Rival hub-and-spoke carrier Continental Airlines saw its stock price jump 142% on the Big Board during the 12-month period. But the same tide of good news has lifted Southwest's stock just 25% on the NYSE, though that gain still handily outpaced the Dow Jones Industrial Average, which was up 7% over the same period.

Methodology

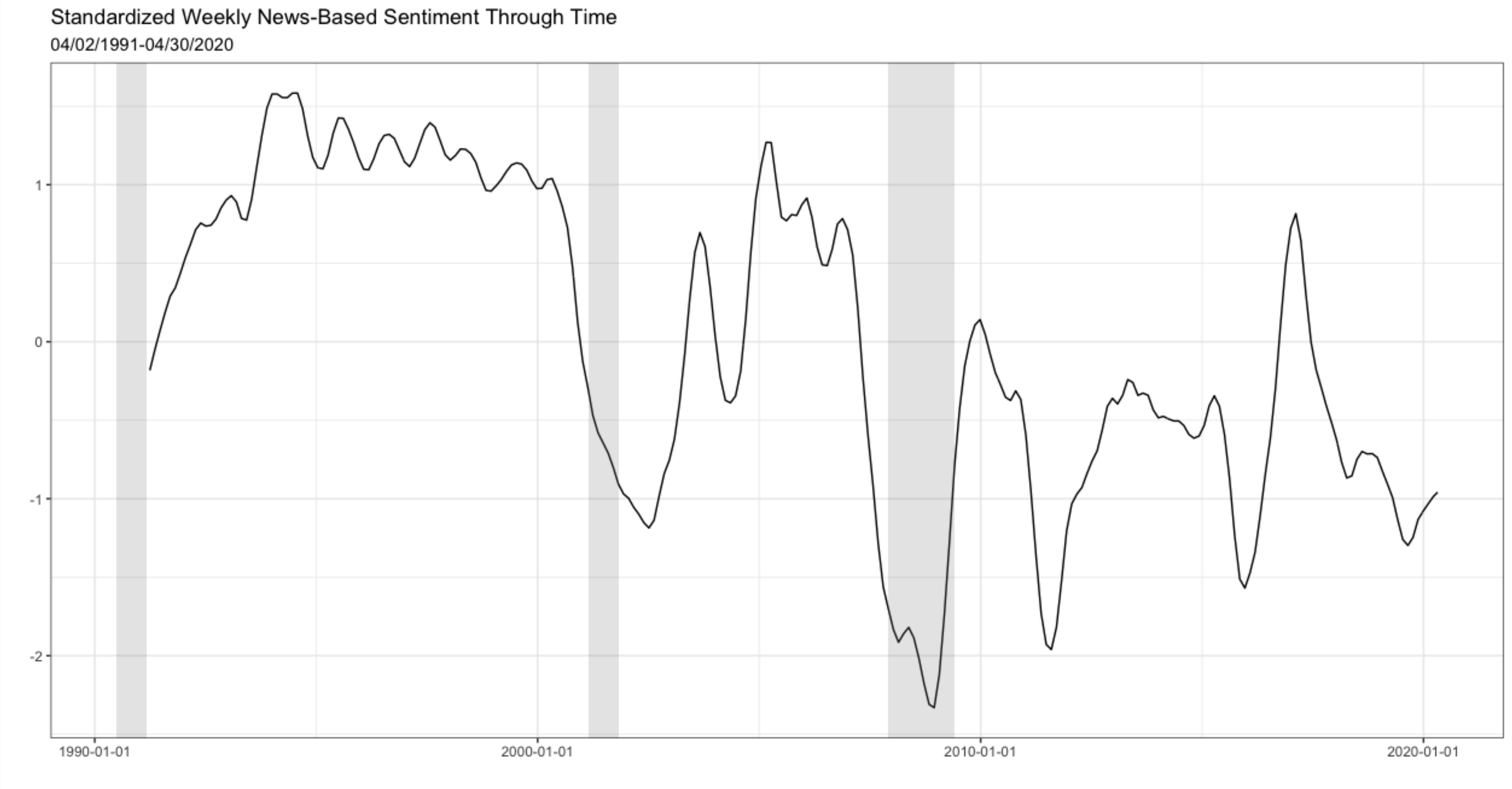
Step #2: Dictionary methods

The raw text is extracted as a manageable high-dimensional numerical array using process criteria in the following order:

- Exclude leading paragraphs with less than 50 characters
- Tokenize the text into individual words and transform it to a tidy data structure
- Strip stopwords such as ‘a’, ‘the’, ‘to’, ‘for’ out of token lists
- Replace words with their root such that "economic", "economics", and "economically" are all replaced by the stem "economic"
- Assign "positive" or "negative" sentiment to words based on a general-purpose "Bing" lexicon which categorizes words in a binary fashion
- Reverse sentiment of words preceded by negation words “no”, “not”, “never”, “without”
- Measure sentiment scores for each leading paragraph with $\frac{n_{pos}-n_{neg}}{n_{words}}$
- Average sentiment scores of the same day to obtain a daily sentiment index
- Average daily sentiment to obtain a weekly and monthly sentiment index

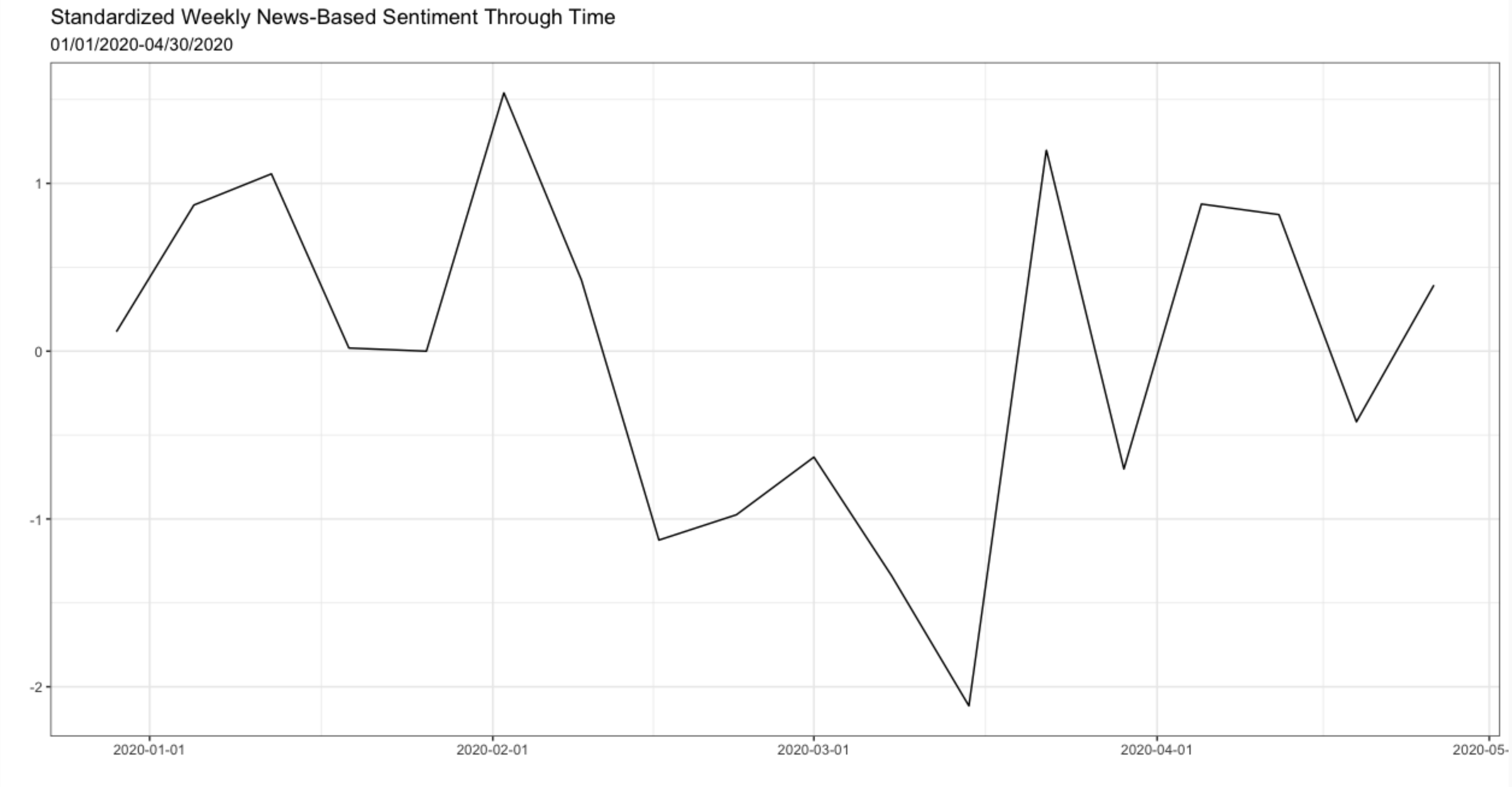
Preliminary Results - weekly NBSI

The index drops sharply before the start of the two recessions in the sample period, suggesting that the index might be a leading indicator with respect to recessions and might be used to nowcast or even forecast recessions

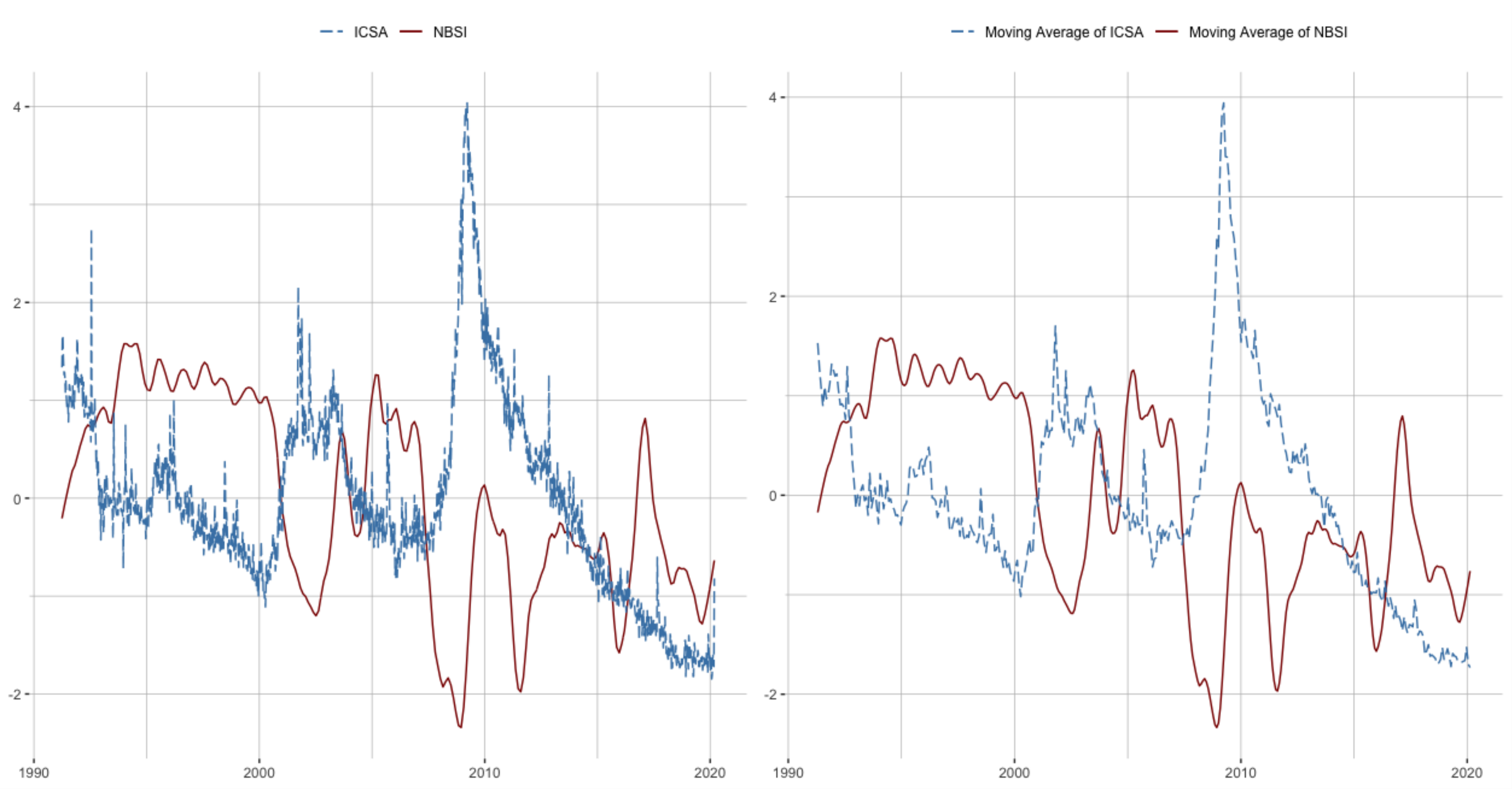


Preliminary Results - weekly NBSI

The standardized weekly NBSI for January through April in 2020 presents how NBSI picked up the bad economic outcomes in February 2020, prior to the real problems starting in the United States in March



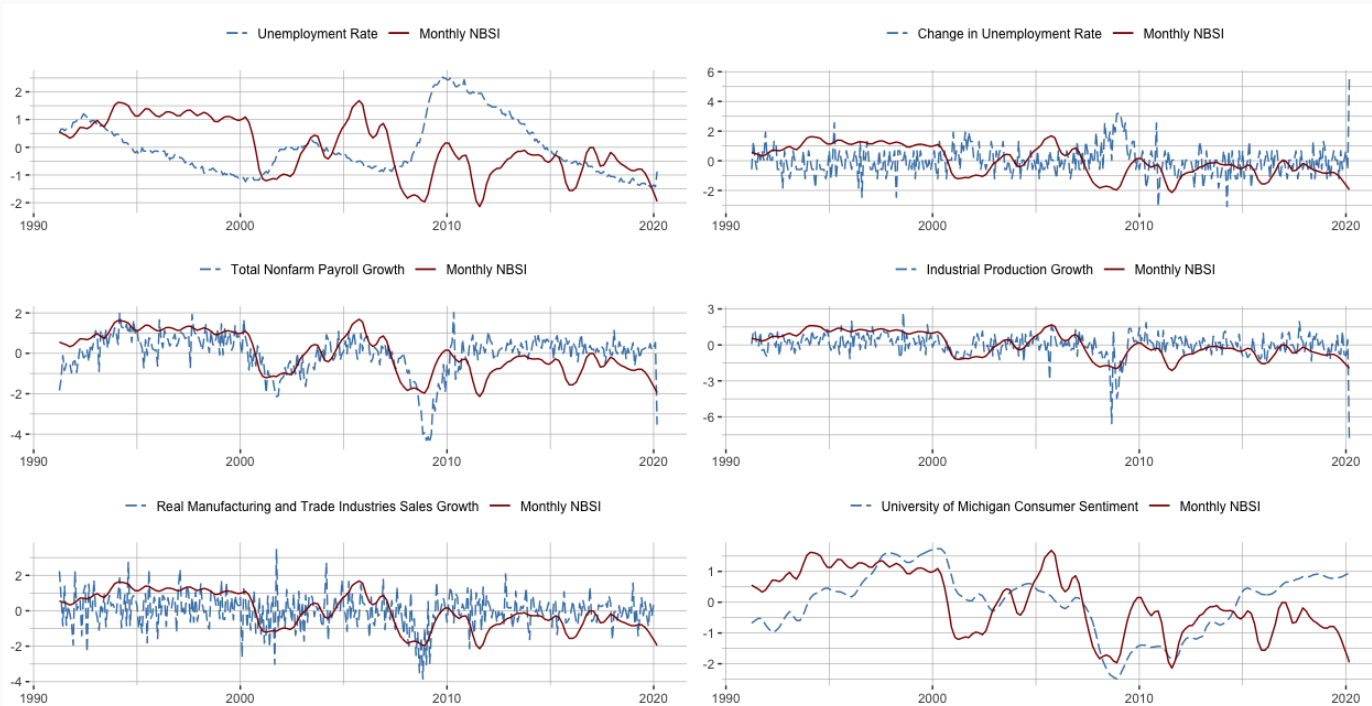
Preliminary Results - weekly NBSI with weekly Initial Claims (ICSA)



	NBSI	Moving Average of NBSI	Lagged NBSI	Moving Average of Lagged NBSI
ICSA	-0.099(<.001)		-0.101(<.001)	
Moving Average of ICSA		-0.097(.060)		-0.099(.055)

- All Pearson correlations have correct signs and fairly large magnitudes

Preliminary Results - monthly NBSI with monthly macroeconomic variables



	Pearson Correlation	P-value
Unemployment Rate	-0.1347044	0.0118927
Change in Unemployment Rate	-0.2390528	0.0000065
Payroll Employment Growth	0.4907129	0.0000065
Industrial Production Growth	0.3747335	0.0000065
Real Manufacturing and Trade Sales Growth	0.2464932	0.0000065

- All Pearson correlations have correct signs and fairly large magnitudes

Future Directions

Accounting for the context of text:

- I plan to manually rate the sentiment of a fraction of news articles, and then use Bidirectional Encoder Representations from Transformers (BERT), a more recently developed learning model developed at Google, to model context and sequential information in text

Applications using NBSI:

- Exploring nowcasting power of NBSI for monthly macroeconomic variables as the month evolves
- Exploring nowcasting power for recessions and expanions, within the framework proposed in previous paper