Social Dynamics in Lending Opinion *

Hengyu Kuang†

University of California, Berkeley

April 28, 2017

Advisor: Professor Raymond Hawkins

Abstract

This paper provides model based empirical support for bank lending behaviors and credit dynamics in US, Euro Area, and Japan. Using the Opinion Formation Model and machine learning techniques, we provide stable model estimations based on central banks' lending surveys and a few selected Macro variables. We reconstruct banks' lending opinions by model simulation, and the result is stable for both forward prediction and backward propagation, following closely to the actual data. Our analysis indicates banks have asymmetric response to good and bad economic information, and we find banks adapt to their peers' opinions when changing lending policies.

*The codes and dataset used in this paper are maintained in github.com/yu45020.

†Email: hkuang@berkeley.edu. I would like to thank Professor Raymond Hawkins for his two years mentoring throughout my time in Berkeley and the weekly discussions for this research

In Post-Keynesian economic thought, money is endogenously determined by economic activities in a modern economy. The central theme is "loans create deposite" and " deposits generate reserve". When a bank underwrites loans, by double entry bookkeeping, it records the loans in the asset side of its balance sheet and automatically record the same amount as deposit for the borrowers in the liability side. Firms rely on credit prior to production and pay back part of revenues to banks and wage labors. Household adjust their deposit ratios as portfolio choice. Thus, new money is created by lending, and it is destroyed when loans are paid back. Banks are not intermediaries that lend deposits and bounded by reserves, but they are central actors of creating money in the forms of loans. Understanding their lending behavior will cast light on understanding economic fluctuation.

Central bankers have long recognized that banks create money. The Central Bank of Canada governor, Towers (1939), observed that "[e]ach and every time a bank makes a loan, new bank credit is created - new deposits - brand new money". The New York Federal Reserve Vice President, Holmes (1969), recognized the Fed's operational constrains on money supply: "In the real world, banks extend credit, creating deposits in the process, and look for the reserve later". McLeay, Radia, and Thomas (2014) from the Bank of England clarifies misconceptions on money creation in modern economy:" banks do not act simply as intermediaries, lending out deposits that savers place with them, and nor do they 'multiply up' central bank money to create new loans and deposits." A recent empirical demonstration, probably the first empirical experiment at its time, by Werner (2014) details the entire process from loan creation to money transfer in a Germany bank, and he found that the bank's loan officers confirmed that they didn't check their deposit balance nor existing reserve before and after the experiment.

Much research has been done analyzing loans and bank lending. In his "The Debt Deflation Theory of Great Depression", Fisher (1933) detailed the causal consequence from over-indebtedness to deflation. Since loans are underwritten in nominal terms, at a state of over-indebtedness, a distress selling caused by debt liquidation leads to contraction on currency and fall in price level, causing a real burden to borrowers. As profit falls investment shrinks and bankruptcies rise on the horizon, confidence collapses, leading to greater desire on cash and slower money circulation. Minsky (1976) developed the financial instability hypothesis and gives a more detail on verbal description on the rise and fall of an economy by adding a capitalist financial system. A boom in the economy moves borrowing from hedge purpose to Ponzi financing, and then instability rises within the system. A recent mathematical model of Minsky's approach is done by Keen (2013) whose model features on nonlinear relation so that the Great Recession can be generated after the great modernization without exogenous shock.

Other approaches focusing on endogenous money include Moore (1979), Moore (1988), Lavoie (1996) who advocate the Horizontalist's view that banks have a horizontal, but truncated, loan supply curve constrained by borrowers' credit worthiness. Later development to refine Horizontalists , or known as Structuralist, by Pollin (1991), and Wray (1995), and Dow (1996) to emphasize liquidity preference and expectation in the system, to name a few for both side. Fontana (2004) and Fontana (2009) in recent years theoretically bridge the two sides by treating Horizontalists as single period analysis and Structuralists as continuous analysis.

In this paper, we aim to provide model based behavior support for the lending dynamics in a capitalist economy. The goal is to discover the evolution of lending dynamics and find potential drivers for the changes. The analysis is based on the Opinion Formation Model from Quantitative Sociology by Weidlich and Haag (1983). The model imposes a social network effect on banks on top of rational risk calculations. Banks can't stand alone in the market because they need to borrow from counter-parties in order to clear payments and balance. Banks make their lending decisions independently but also adapt to their peers opinions. Hawkins (2011) calibrates this model to illustrate the rise and fall on non-agency mortgage backed securities in the U.S. Ghonghadze and Lux (2016) applies it on Fed's Senior Loan Officer Opinion Survey on Bank Lending Practices (SLOOS).

We follow their approach and extend the model from Fed to European Central Bank (ECB) and Bank of Japan (BOJ)'s lending surveys. These surveys ask banks whether they changed credit standard and credit spread in the past quarter. We found the opinion formation model is easily over-fitted because of its non linearity. To manage this issue and improve the model predictabilities, we use random forest, a state of art in machine learning developed by Professor Leo Breiman in UC Berkeley, to select variables that have strong predicting power. In addition, from the survey data we find a very clear phase transition for banks when they change lending policies due to different economic factors. Uncertainty is the main driving force in tightening credit while competition, especially from other banks, is the main force for easing credit. Other factors such as liquidity and capital position have little effect.

We start with the US data and get good results based on selected variables. We are also able to extrapolate the model for both forward prediction and backward propagation. The model solution is stable over time. From the model simulation, we find banks have asymmetric response to economic factors that change their lending decisions.

We then extend the model to ECB and BOJ's data. ECB's survey data produces very similar result to the U.S. data and show that banks' lending decisions are impacted by their peers. The model has worse performance for Japan. One explanation is Japan's two decades stagnation and non-performing loan problem.

The paper is organized as follows: Section 1 develops the Opinion Formation Model. The derivation follows closely from Weidlich and Haag (1983) and Ghonghadze and Lux (2016). In section 2, we apply the model to US data and then extend it to ECB and BOJ's data. We provide forward prediction and backward propagation in order to examine the model stability. Finally, we close in section 3 with a discussion and summary.

1 The Opinion Formation Model

Our model derivation is adapted from Weidlich and Haag (1983), Hawkins (2011), and Ghonghadze and Lux (2016). In the context of lending decision, assume there are $2N$ bankers fixed in a continuous time horizon T. All bankers have equal weights and face two opinions: to lend or not to lend. Let n_t^+ be the amount of bankers who choose to lend at time t, and n_t^- be the opposite, s.t. $n_t^+ + n_t^- = 2N$. Then the state of opinion can be represented by an integer:

$$2n_t = n_t^+ - n_t^-, \qquad \text{where n} \in [-N, N].$$

When $n_t = N$, all bankers choose to lend, and when $n_t = -N$ is the opposite. An opinion lending index, or the average lending sentiment, at time t can also be defined in the same way:

$$x_t := \frac{n_t}{N} = \frac{n_t^+ - n_t^-}{2N}.$$

Let $p(n;t)$ be the probability of the state of opinion at time t. Then by the normalization,

$$\sum_{n=-N}^{n=N} \frac{\partial p(n;t)}{\partial t} = 1 \qquad \text{for } \forall\ t \in T. \tag{1}$$

Because change of opinion can happen in any time, the state of opinion n can be interpreted as the difference of all influx to the state n and the outflux from n. Let $w(j \to i)$ be the transition rate of changing from state j to i for all $i, j \in [-N, N]$.

Hence, $p(n;t)$ can be rewritten as a special case of the Master Equation:

$$\frac{\partial p(n;t)}{\partial t} = \sum_{j=-N}^{j=N} [w(i \leftarrow j)p(j;t) - w(i \to j)p(i;t)]. \tag{2}$$

To simplify the discussion, assume $\forall\ \Delta t < \epsilon, \epsilon \in \Re^+$, only one banker changes opinion:

$n \to (n+1)$ or $n \to (n-1)$.[1] Then

$$w(n' \to n) = 0 \quad \text{for } \forall n' \neq n \pm 1 \ . \tag{3}$$

If we define the transition probability for the transition rate:

$$\begin{aligned} w_{\uparrow}(n) &:= w(n \to n+1), \\ w_{\downarrow}(n) &:= w(n \to n-1), \end{aligned}$$

Then equation (2) becomes:

$$\frac{\partial p(n;t)}{\partial t} = [w_{\uparrow}(n-1)p(n-1;t) + w_{\downarrow}(n+1)p(n+1;t)] - [(w_{\uparrow}(n)p(n;t) + w_{\downarrow}(n)p(n;t)]. \tag{4}$$

The first sum is the influx to the state n and the second sum is the outflux from n.

Recall the average lending sentiment $x_t := \frac{n_t}{N}$. Then $\Delta x_t = \frac{\Delta n_t}{N} = \frac{1}{N} = \epsilon$. Divide the equation (4) by N and introduce the new probability function $P(x;t)$ for x_t s.t. $P(x;t) = Np(n;t)$. The following can be shown to be equivalent:

$$\begin{aligned} \frac{\partial p(n;t)}{\partial t} = \frac{\partial P(x;t)}{\partial t} &= [w_{\uparrow}(x - \frac{1}{N})P(x - \frac{1}{N};t) + w_{\downarrow}(x + \frac{1}{N})P(x + \frac{1}{N};t)] \\ &\quad - [(w_{\uparrow}(x)P(x;t) + w_{\downarrow}(x)P(x;t)]. \end{aligned} \tag{5}$$

To approximate the equation above for large n, apply Taylor Expansion on $n$up to $O(n^2)$, then

$$\begin{aligned} \frac{\partial P(x;t)}{\partial t} &\approx \frac{\partial}{\partial x}[w_{\downarrow}P(x;t) - w_{\uparrow}P(x;t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[w_{\downarrow}P(x;t) + w_{\uparrow}P(x;t)] \\ &= -\frac{\partial}{\partial x}[K(x)P(x;t)] + \frac{\epsilon}{2}\frac{\partial^2}{\partial x^2}[Q(x)P(x;t)], \end{aligned} \tag{6}$$

which is also a standard form of the Fokker Planck equation in one dimension.

By taking the first and second moment of $\frac{\partial P(x;t)}{\partial t}$ on x and take the limit of $\Delta t \to 0$, $K(x)$ and $\epsilon \ Q(x)$ can be shown as the mean and variance of average sentiment. In terms of stochastic differential equation, $K(x)$ is the drift term, and $\epsilon \ Q(x)$ is the diffusion coefficient.

The choice of the transition probability and the exact solution of the equation have various forms and depend on further assumption on behavior (Weidlich and Haag (1983, P31-44)). To limit the discussion and make the model reflect actual behavior, three more

1. A more general discussion without this assumption can be found in Weidlich and Haag (1983, CP 3,4)

assumptions on behavior are introduced (Weidlich and Haag (1983, 41)):

a) bankers make their decisions independently, relying on inherent personal observations;

b) bankers are willing to adapt to the peers opinion once it becomes the majority.

c) preference and the willingness to adapt may vary over time.

The transition probability is assumed to be :

$$\begin{aligned} w_{\uparrow} &= \frac{n^-}{2N} v \exp(U(x_t)) = (1-x) v \exp(U(x_t)); \\ w_{\downarrow} &= \frac{n^+}{2N} v \exp(-U(x_t)) = (1+x) v \exp(-U(x_t)), \\ U(x_t) &= \alpha_0 + \alpha_1 x_t. \end{aligned} \tag{7}$$

where $\frac{n^+}{2N}$ and $\frac{n^-}{2N}$ measure the attitude towards lending; v $\exp(U(x_t))$ is the speed of changing opinions; α_0 measures the independent preference on lending decision; α_1 measures the degree of herding. They are assumed to be constant within a short period with contrast to the whole process of opinion evolution.

By applying hyperbolic transformation on (7), the drift and diffusion coefficients are re-written as:

$$K(x_t, Z_t; \theta) = 2v \cosh(U(x_t, Z_t; \theta))(\tanh(U(x_t, Z_t; \theta)) - x_t); \tag{8}$$

$$Q(x_t, Z_t; \theta) = 2v \cosh(U(x_t, Z_t; \theta))(1 - x_t \tanh(U(x_t, Z_t; \theta)))/N; \tag{9}$$

$$U(x_t, Z_t; \theta) = \alpha_0 + \alpha_1 x_t + \sum_{i=1}^{m} \beta_i Z_i, \tag{10}$$

where x_t is the average lending sentiment at time t; N is the amount of bankers; Z_t includes a series of exogenous variables such as GDP and unemployment rate that affect lending decisions; θ is a collection of unknown coefficients of variables to be measured.

The closed form solution exists when “only one highly peaked maximum$[P(x;t)]$” exists,a linear K(x), and a constant Q(x) (P27), which fail to capture the non-linear change of risk preference and self-variation for peer pressure and shocks.

A general solution is to numerically approximate P, which is discussed later. The corresponding solution of the dynamic opinion, x_t, can be extended by applying standard Ito calculus and approximate x_t as a stochastic differential equation (SDE) (Ghonghadze and Lux 2016 and Ghonghadze and Lux 2012)

$$dx_t = K(x_t, Z_t; \theta)dt + \sqrt{Q(x_t, Z_t; \theta)}dW_t, \tag{11}$$

where W_t is the standard Wiener process.

This model measures the evolution of average lending sentiment in a group. The assumptions on banker's behavior on lending are simplistic but valid from a sociological perspective. People interact with a social structure that has norms for group members' common behaviors and that has network on controlling the flow of information; they rely on other group members to justify their actions, which is also known as "social embeddedness" (Granovetter 2005). Keynes (1937) in his summary of The General Theory also emphasizes that in a world of uncertainty, people rely on the crowd:

> "Knowing that our own individual judgment is worthless, we endeavor to fall back on the judgment of the rest of the world which is perhaps better informed. That is, we endeavor to conform with the behavior of the majority or the average. The psychology of a society of individuals each of whom is endeavoring to copy the others leads to what we may strictly term a conventional judgment."

Institutionally, the requirement on prudent lending from FDIC regulation guarantees various lending practice must converges. Prudent lending means that lending practices are generally accepted [2], and banks' risk taking are evaluated against the average of industrial practice. Significant deviation, either for good or bad purposes, may subject to punishment. For example, 20% mortgage down payment was a common practice in the 20^{th} century, but a bank that imposed such a policy before the great recession would be considered outside of the norm . For bankers, "it is better to fail conventionally than to succeed unconventionality". Thus, it is reasonably to consider bankers value peers' opinion on lending.

An important features of this model is the existence of unstable equilibrium. Multiple opinions can co-exist at the same time, and the average sentiment can switch suddenly and without warming. Real life examples include the sudden collapse in the financial market and violate jump of US LIBOR rate right after Lehman Brothers' bankruptcy.

Empirical application with this model requires estimating the parameters θ. Ghonghadze and Lux (2012) numerically approximate the solution of the Fokker-Planck equation (6) by Crank-Nicolson finite difference, and then use log Maximum Likelihood on the solution $P(x;t)$ to estimate θ. This method is rigorous but computationally expensive. Ghonghadze and Lux (2016) later introduced the use a less rigorous but more efficient method of Quasi Maximum Likelihood Estimate (QLME)[3]. P is approximated as a conditional normal distri-

2. Federal Regulation Code, Title 13, CFR 307.8
3. QMLE gives consistent parameters estimation even when the Gaussian assumption is violated in dynamic models (Bollerslev and Wooldridge 1992).

bution with further assumptions on x_t:

$$E(x_{t+1}|x_t) = x_t + K(x_t, Z_t; \theta)\Delta t \tag{12}$$

$$\mathrm{Var}(x_{t+1}|x_t) = Q(x_t, Z_t; \theta)\Delta t. \tag{13}$$

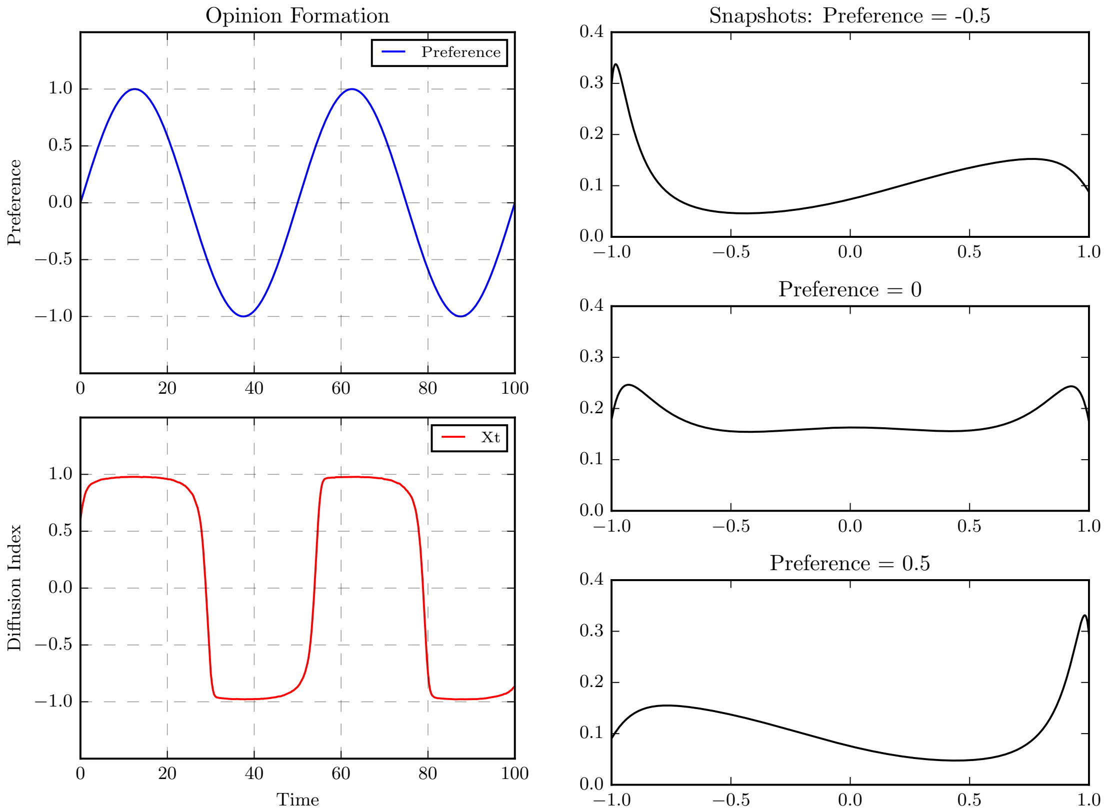
Suppose there are T observations $\{x_1, \cdots, x_T\}$, then θ is estimated by maximizing the log-likelihood function

$$L = \sum_{t=1}^{T-1} \ln(N(x_{t+1}|x_t, Z_t; \theta)), \tag{14}$$

where $N(x_{t+1}|x_t, Z_t; \theta)$ is normal density with mean $E(x_{t+1}|x_t)$ and variance $\mathrm{Var}(x_{t+1}|x_{t=0})$. We numerically solve L by Newton Conjugate Gradient method using Scipy's Generic likelihood model and set the stopping criteria to be 10^{-12}. Ghonghadze and Lux (2016) solve L by specifying its analytical Gradient and Hessian matrices. We replicate their model estimations and test our code on their dataset. We obtain comparable results: the parameters agree at least 4 decimal points, and the variances have at most a 7% difference. A numerical simulation is shown below.

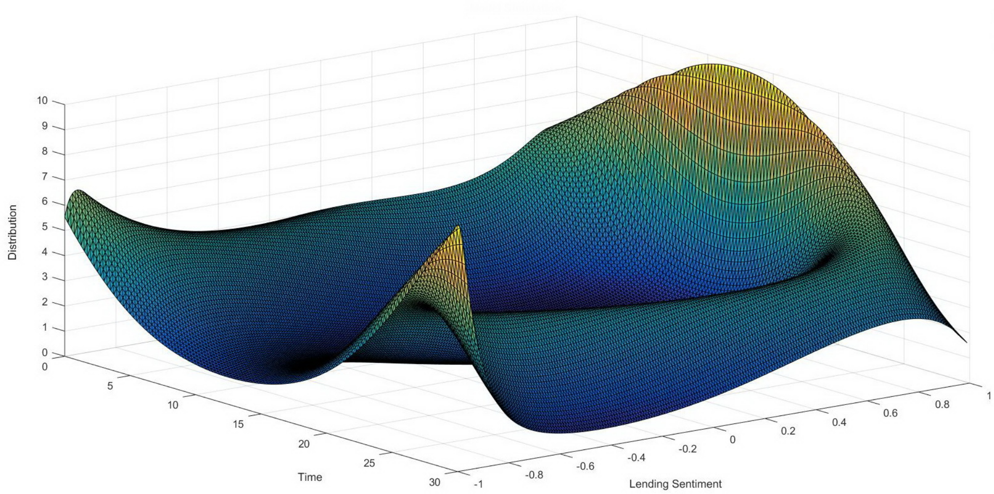
Suppose there are no exogenous effect and consider 25 banker's preference over time. Assume the preference on lending changes over time ($\alpha_0 = \sin(2t\pi/50)$). At t=0, the average sentiment favors lending is high, and bankers value peers' opinions ($x_{t0} = 0.6$, $k = 1.5, v = 1, dt = 0.25$ (This example is adapted from Weidlich and Haag (1983, P50)). The SDE (11) is simulated by Shoji-Ozaki's method, and the solution is bounded to [-1,1]. Euler's method is found to be unstable for the real data because Δt is fixed at 0.25 (real data are measured quarterly). A comparison of these two methods with algorithms can be found in Iacus (2009).

Figure 1: Numerical Simulation



In figure 1 the left panels show the independent preference and corresponding average lending sentiment. The upper plot is the preference function specified above, and the lower plot is the mean of a thousand simulations on lending sentiment. Positive sign means favor lending. When the sign switches (when time near 25, 50, 75), the majority lending sentiment quickly switch from to lend to not to lend or vice verse. Accurate estimation for the future may be impossible because opinion changes so rapidly. The right hand side of the figure 1 is the snap shots for the probability density of bankers with different opinions. Independent preference shifts from -0.5 to 0.5. When bankers have no preference ($\alpha_0 = 0$), there are two groups with opposing opinions. Being neutral is rare. And once bankers favor lending($\alpha_0 = 0.05$), the hight goes to the lending side, meaning the majority shifts towards lending. An overall distribution on lending up to time 30 is shown in figure 2. It is generated by approximating the probability P. At time equals 0, although the majority choose to lend, there is a small group of people choose not to lend. And when time is near 25, the lending sentiment is shifted to be negative, the distribution also moves quickly towards not to lend. At time equals 30, lending becomes a minority opinion.

Figure 2: Model Simulation on Lending Distribution



2 Cross Countries Analysis on Credit Dynamics

Our datasets are central bank lending surveys. The Fed, Bank of Japan, European Central Bank, and Bank of England publish quarterly survey of bank's lending practice. The US data is the longest, dating back to 1990. Japan followed suit a decade later. The European survey launched in 2003 with details on each countries within the Union. England started their survey after the Great Recession. All surveys ask banks whether they change credit standards and spreads for firms and households. In addition, the questionnaires also ask banks why they change their lending policies.

Given the length of each survey data availability and prior analysis from Ghonghadze and Lux (2016), the primary focus is on the U.S. data. We find that the opinion formation model has a strong predicting power based solely on the past quarter' survey result. The estimated coefficients are stable over time. In addition, banks have asymmetric responses during business cycles. When searching possible exogenous variables to explain these dynamics, we use random forests combined with the Boruta method to select a subset of relevant variables that are stable and have strong predictive power. The model simulation, based solely on the initial value of the credit spread and the estimated coefficients, follows the actual data closely. The analysis is applied to European data and Japan data. BOE's survey data is weighted by the level of importance, so it is incompatible with the model assumption. Given its raw data is not available, it will be considered in the future.

2.1 US Credit Dynamics

The Fed's Senior Loan Officer Opinion Survey on Bank Lending Practices (SLOOS) is conducted at the beginning of each calendar quarter, covering up to 60 large domestic commercial banks and up to 24 large foreign banks. All of them have asset greater than $3 billions, or 5% or less, of commercial and industrial loans over their total assets. We focus on question (2.d) about bank's credit spread on commercial and industrial loans for middle and large firms:

> *2. For applications for C&I loans or credit lines-other than those to be used to finance mergers and acquisitions- from large and middle-market firms and from small firms that your bank currently is willing to approve, how have the terms of those loans* ***changed over the past three months*** *? (emphasis added)*
>
> *(d) Spread of loan rates over your bank's cost of funds (wider spreads=tightened, narrower spread = eased)*

Five options are available: tightened considerably, tightened somewhat, remained basically unchanged, eased somewhat, and eased considerably. On average, there are around 70 banks answering this question. The diffusion index (DI) is the net percentage of eased minus tightened [4]. Figure 3 shows banks collectively tightened credit spread well before the last two recessions, and they raise spread sharply during the boom period.

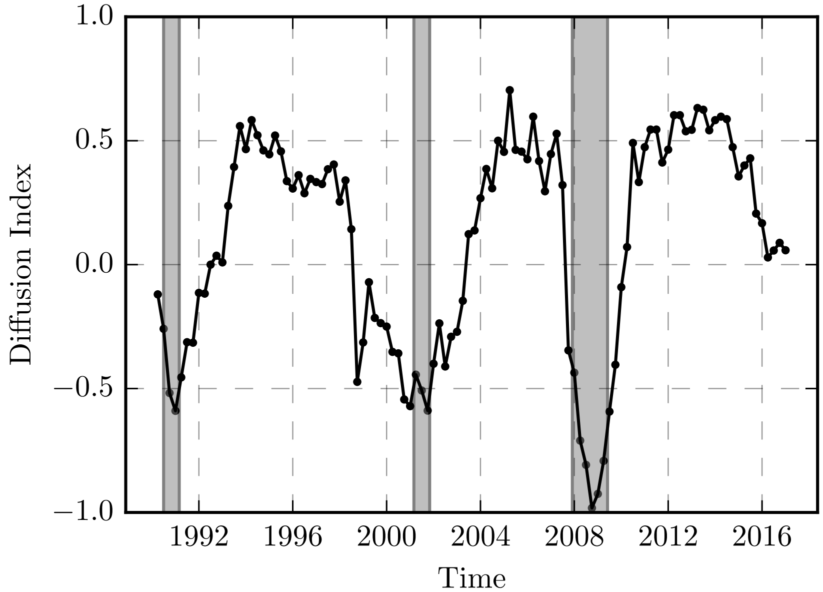
The survey lists a series of factors for banks to choose if they change lending policies in the past quarter. 6 out of 8 factors are consistently reported since 1999 :

> *If your bank has tightened or eased its credit standards or its terms for C&I loans or credit lines over the past three months, how important have been the following possible reasons for the change? (Please respond to either A, B, or both as appropriate and rate each possible reason using the following scale: 1 = not important, 2= somewhat important, 3 = very important)*

(a) *Deterioration(improvement) in current or expected capital position*
(b) *Less(more) favorable or more(less) uncertain economic outlook*
(c) *Worsening(improvement) of industry-specific problems*
(d) *Less (more) aggressive competition from others* [5]
(e) *Reduced (increased) risk tolerance for risk*
(f) *Decreased (increased) liquidity in the secondary market for these loans*

4. In the Fed's report, this diffusion index is constructed by the net percentage of tighten minus eased.

Figure 3: US Diffusion Index of Credit Spread



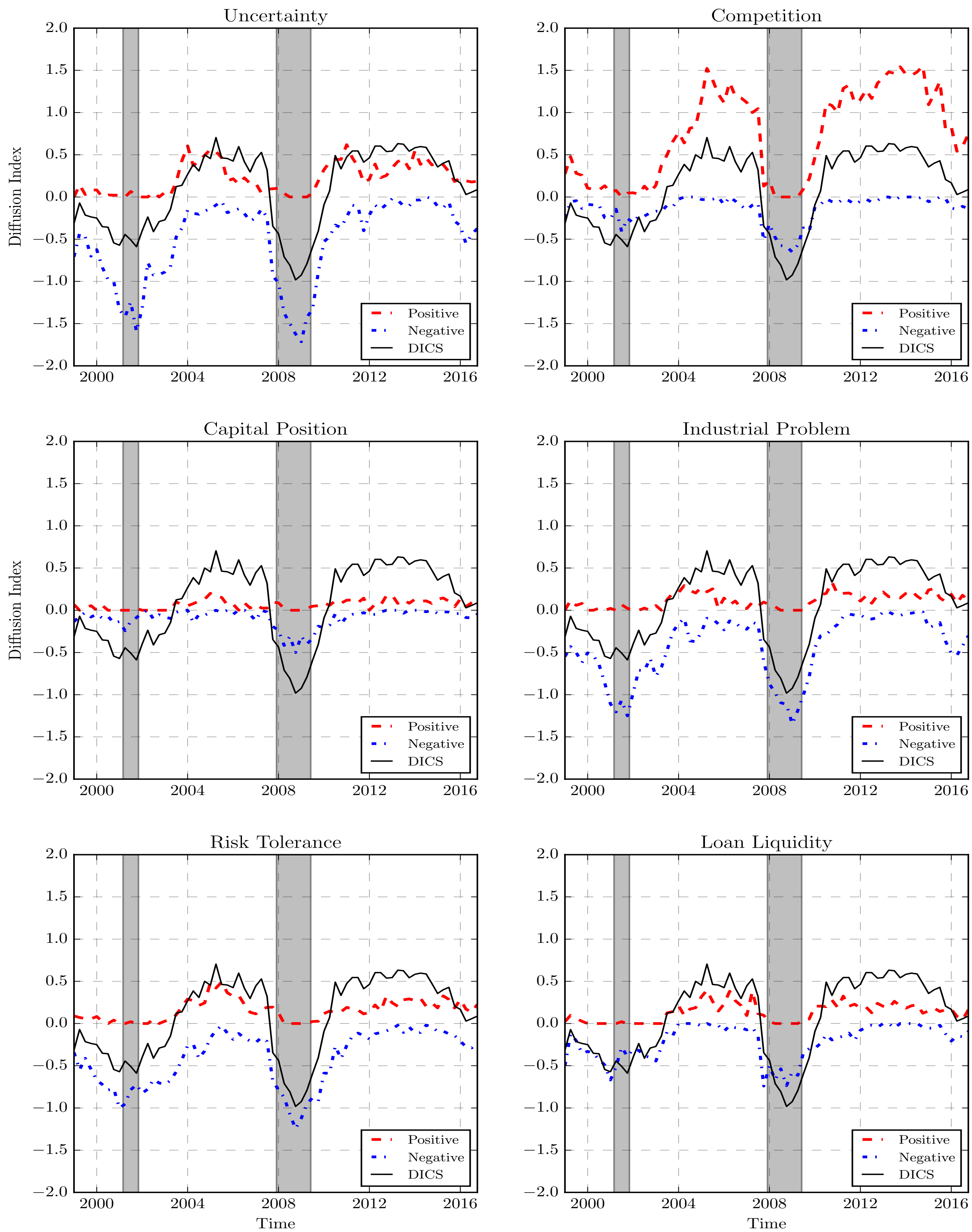
These data is list in Fed's pdf reports only, so we extra the data from 1999 to 2017 reports and recalculate the mean for each reason. They are weighted by the scale of importance (0,1,2) as well as their fraction of the total number of responding. The reasons contributing to tighten lending policies are set to be negative. If a reason has value 0, it means either (i) no bank ease or tighten lending or (ii) all banks think the reason is not important. If it is 2 (-2), all banks consider it is very important. 1 (-1) means banks think it is somewhat important. Values in between mean some banks eased/tightened lending and have mixed opinions.

Figure 4 shows a 'phase transition' when banks change lending policies over the last two decades. The dark line in the middle is the diffusion index of credit spread, the dashed line represents the weight on a reason for easing credit, and the dashdot line is the weight for tightening credit. When banks tighten lending, uncertainty is the main driving force, but it has almost no effect in easing credit. By contrast, competition is the main driving force when banks ease lending. Its effect is the most obvious before and after the Great Recession. Risk tolerance and industrial problems (borrowers' creditworthiness) have a much weaker effect in tightening credit, and neither of them have strong effect in easing credit. Capital position and loan liquidity are the least important factors in changing lending policies.

Such repeated asymmetric responses during business cycles provide an empirical support

5. The surveys before 2001 Q3 split this reasons into 2 parts: competition from nonbank lenders and other banks. We take the average of their value.

Figure 4: US Reasons for Changing Lending Policies



for the assumptions of the opinion model. Banks have their analysis on economic outlook, but they are also impacted by their peers' decisions. When they are not sure their counterparties' performance, they tighten lending. But once their peers are controlling the Street, the state of economy become less important in driving credit expansion, and their lending decisions are much less dependent on capital positions and loan liquidity.

Next we turn to the opinion formation model. Suppose there are no exogenous effect and banks only look at their peers, then

$$\text{Model 1:} \qquad U(x_t, Z_t; \theta) = \beta_0 + \beta_1 x_t, \qquad \theta = \{\beta_0, \beta_1\}, \tag{15}$$

$$\text{Model 2:} \qquad U(x_t, Z_t; \theta) = \beta_0 + \beta_1 x_t^+ + \beta_2 x_t^-, \qquad \theta = \{\beta_0, \beta_1, \beta_2\}, \tag{16}$$

$$\text{where} \qquad x_t^+ = \begin{cases} x_t & x_t > 0 \\ 0 & x_t \leq 0 \end{cases} \qquad x_t^- = \begin{cases} x_t & x_t \leq 0 \\ 0 & x_t > 0 \end{cases}$$

In our algorithm, parameter coefficients are estimated first, and the variances are then calculated by the Hessian matrix in the last step. Model 1 indicates the current lending opinion has significant impact. The current diffusion index of credit spread (DICSxt0) is significantly larger than 0, meaning peers' opinions have large weight. The opinion turn-over rate (v) suggests banks change opinions fast. Model 2 is used to estimate the asymptomatic effect from peers. The coefficients on positive and negative lending opinion are slightly different, but they overlap within two-standard- deviation region. Model 2's Akaike Information Criterion (AIC) and Bayesian Information Criterion(BIC) are slightly better than Model 1.

Figure 5 is a simulation of M2. The predicted values follows closely the actual data. During the simulation the credit spread (x_{t+}, x_{t-}) in equation (16) are updated via the survey data rather than the simulated values. The result is surprisingly well. It means the current opinion index of credit spread has strong predictive power over the credit dynamics for the next quarter.

Table 1: US Model 1

Dep. Variable:	DICSxt1	**Log-Likelihood:**	44.737
Model:	QMLE	**AIC:**	-85.47
Method:	Maximum Likelihood	**BIC:**	-80.13
No. Observations:	107	**Number of Banks:**	70

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
v	1.6657	0.204	8.160	0.000	1.266 2.066
constant	0.0009	0.020	0.047	0.963	-0.038 0.040
DICSxt0	1.0440	0.049	21.126	0.000	0.947 1.141

Table 2: US Model 2

Dep. Variable:	DICSxt1	**Log-Likelihood:**	46.101
Model:	QMLE	**AIC:**	-88.20
Method:	Maximum Likelihood	**BIC:**	-82.86
No. Observations:	107	**Number of Banks:**	70

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
v	1.6767	0.207	8.101	0.000	1.271 2.082
constant	0.0651	0.044	1.485	0.138	-0.021 0.151
DICSxt0+	0.8793	0.113	7.810	0.000	0.659 1.100
DICSxt0-	1.2276	0.120	10.191	0.000	0.991 1.464

Figure 6 examines the stability of Model 2's coefficients over data points. The lines labeled 'Ordered' use the original data in time order; the lines labeled 'Random' shuffle the data and sample points to estimate the model. The lines after 100 sample points use sampling with replacement and equal probability. It is a bootstrap technique to explore future sample data. The result is twofold. First, the coefficients are stable over time, meaning banks are constantly looking at each other. Splitting the diffusion index may not be able to capture banks' asymmetric reaction to their peers. Second, the model estimation becomes stable after 60 data points. It is also time independent once it is conditioned on the current quarter. Given these results and the conjuncture that only the current opinions have much stronger impact on lending decisions, most of existing machine learning algorithms are worth considering in the analysis of this diffusion index.

Figure 5: Model 2 Simulation

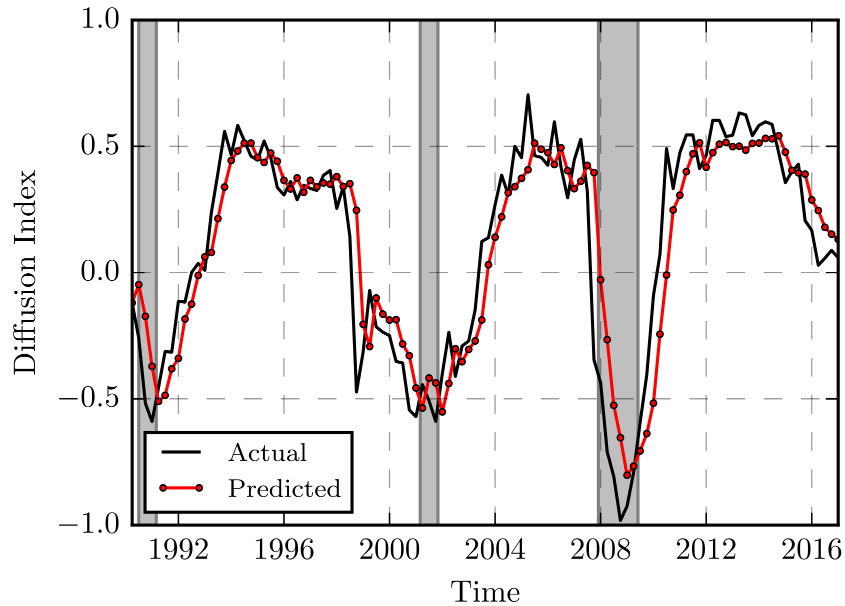
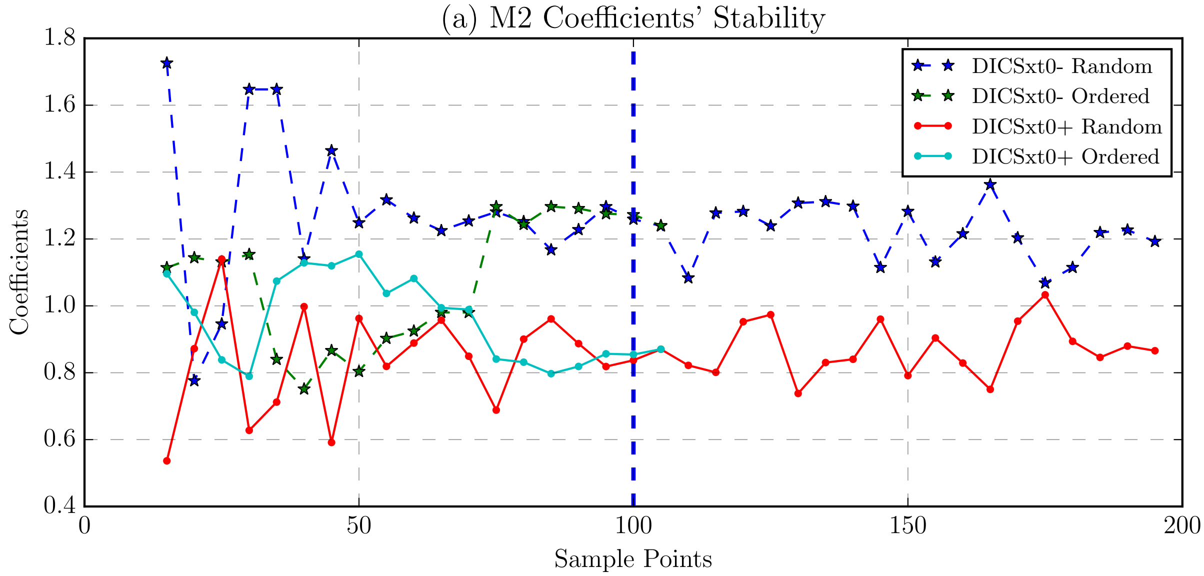


Figure 6: Model 2 Stability Examination

(a) M2 Coefficients' Stability



As a comparison, Model 3 includes the past quarter's credit spread. The equation (16) is extended as the below:

$$\text{Model 3:} \qquad U(x_t, Z_t; \theta) = \beta_0 + \beta_1 x_t^+ + \beta_2 x_t^- + \beta_3 x_{t-1}^+ + \beta_4 x_{t-1}^-, \quad \theta = \{\beta_0, \beta_1, \beta_2\}; \tag{17}$$

$$\text{where} \qquad x_{ti}^+ = \begin{cases} x_{ti} & x_{ti} > 0 \\ 0 & x_{ti} \leq 0 \end{cases} \qquad x_{ti}^- = \begin{cases} x_{ti} & x_{ti} \leq 0 \\ 0 & x_{ti} > 0 \end{cases} \tag{18}$$

The sum of coefficients for both positive and negative credit spread are close to Model 2's result. This suggests banks' decision strongly rely on the one period lending opinion. However, Model 3 becomes worse in terms of BIC. Its coefficients of current and past DI credit spread have conflicting signs, which is a signal of over-fitting.

Table 3: US Model 3

Dep. Variable:	DICSxt1	**Log-Likelihood:**	48.216
Model:	QMLE	**AIC:**	-86.43
Method:	Maximum Likelihood	**BIC:**	-73.11
No. Observations:	106	**Number of Banks:**	70

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
v	1.6088	0.201	8.015	0.000	1.215 2.002
constant	0.0597	0.048	1.243	0.214	-0.034 0.154
DICSxt0+	1.0599	0.206	5.157	0.000	0.657 1.463
DICSxt0-	1.5344	0.206	7.463	0.000	1.131 1.937
DICSxt-1+	-0.1808	0.200	-0.906	0.365	-0.572 0.211
DICSxt-1-	-0.3545	0.207	-1.715	0.086	-0.760 0.051

Next we include exogenous variables. A list of variables is selected based on SLOOS' reasons for changing lending policies,related work by Ghonghadze and Lux (2016), and empirical research from ECB's lending surveys ((Köhler-Ulbrich, Hempell, and Scopel 2016) and (Altavilla, Darracq Paries, and Nicoletti 2015)). Since the opinion formation model is nonlinear and easily over-fitted, we select an optimal and relevant subset of variables via Random forest combined with Boruta method. Random forest is the current state of art in machine learning algorithms, and Boruta is a wrapper method to select all relevant variables.

The survey asks whether banks change credit spread in the past quarter, so data are collected with attention on their release data. We assume banks make decisions at the mid of each quarter, so they are able to access public data released near that time. But they don't have access to the data released near the end of quarters. Because banks have asymmetric responses to economy fluctuation as shown in figure 4, exogenous variables are de-meaned by their one year exponential moving average (EMA), and their positive and negative values are considered separately. The EMA emphasizes the momentum effects and weights more heavily on the most recent events [6]. Moving average, rather than Hodrick-Prescott filter, are widely used in business organization (Osborn 1995). It is also a momentum based trading strategy (Lemperiere et al. 2014), (Menkhoff et al. 2012). Also, as figure 4 shows, both the diffusion index of credit spread and reasons (uncertainty and competition) for changing lending have momentum. Full period de-trending, which uses 'future' data to demean any quarter data, violates our assumption that banks don't have full knowledge about the future.

Table 4: US List of Variables for US

Variable	Description	Release Date	Season. Adj
DICS	Diffusion index of credit spread. Large & middle firms	Mid of next quarter	N
DILD	Diffusion index of loan demand. Large & middle firms	Mid of next quarter	N
EBP	Excessive bond premium. Quarter averaged, change	Monthly	Y
NASDAQ	NASDAQ index. Quarter average, percent change	Daily	N
VIX	CBOE volatility index. Quarter average, change	Daily	N
NPL	Non performing commercial loans rate. De-mean by its 1Y EMA	Mid of next quarter	N
RGDP	RGDP growth rate. De-mean by its 1Y EMA	End of next quarter	Y
SPF RGDP	SPF RGDP growth rate. De-demean by its 1Y EMA; col 3	Mid of current quarter	Y
Co. Profit	Corporate profit growth after tax. De-mean by its 1Y EMA	End of next quarter	Y
SPF Co.Profit	SPF Co. profit growth[7]. De-demean by its 1Y EMA; col 3	Mid of current quarter	Y
Unemp	Unemployment rate U3, percent change. De-mean by its 1Y EMA	Monthly	Y
SPF Unemp	SPF unemp rate, percent change. De-demean by its 1Y EMA; col 3	Mid of current quarter	Y
CPI	All item. Percent change	Monthly	Y
SPF CPI	SPF CPI, mean. Percent change; col 3	Mid of current quarter	Y

The diffusion index of loan demand (DILD) is from question 4A in the SLOOS. It asks how the C&I loan demand from large and middle market size firms changed over the past quarter. Excessive Bond Premium (EBP) measures the investors' risk appetite in the corporate bond market ((Favara et al. 2016, Fred Note), (Gilchrist and Zakraj 2012)). This data is available in FRED Notes page. It assumes the spread contains information on the expected default risk and independent risk preference. It is constructed by a simple unweighted average of credit spreads in corporate bonds, and then subtracted each bonds' expected default risk implied in the spread through linear regression. It is found to be a statistically significant leading indicator for recessions (Favara et al. 2016). Altavilla, Darracq Paries, and Nicoletti (2015) construct a EU version from bank-level data in its Bank Lending Survey and find it statistically significant for credit supply. In this paper, EBP is used to approximate the aggregate risk preference.

NASDAQ is a major stock index. VIX is the implied volatility for stock index option prices, which is also known as "fear index". Non-performing loan (NPL) rate measures the performance of bank loans. Loan defaults take times to be observed, so 3 year and 5 year

6. We also examine various years as window size for EMA but fail to find significant change on results
7. After 2006 Q1, the level forecast includes IVA and CCAdj

average cumulative default rate are key indicators for S&P credit rating (Hawkins 2011). In addition, Bouwman and Malmendier (2015) use call reports to find that the banks' past history of under-capitalization change their risk preference, and 1-6 years are statistically significant.

Real GDP (RGDP) is lagged two quarters because of its release data. The first quarter's GDP is only available by the end of second quarter. Advance estimates are available at the first month of second quarter, which usually provides fair estimation (Fixler, Greenaway-McGrevy, and Grimm 2014). The Survey of Professional Forecasters (SPF) data (mean response) are used to approximate banks' estimation for the future quarter[8]. If the past quarter RGDP is approximated by SPF's survey result [9], the model results don't have significant change.

To select relevant variables, we use random forest combined with Boruta method. The random forest approach is an ensemble method that aggregates multiple decision trees. For a classification problem, a decision tree tries to split the data into groups based on input variables (or 'features'). For a regression problem, it takes the average of all points within a group.

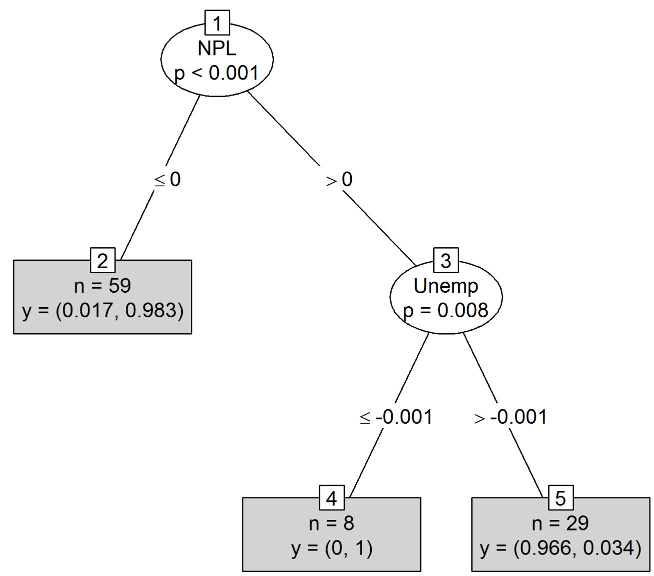
An example of decision tree is in figure 7. It is built by part of the dataset from table 4. The tree has one node and depth of two. Positive values in diffusion index of credit spread are labeled as 1 (to lend), and negative values are labeled as 0 (not to lend). The decision tree starts with change of non-performing loan rate as an initial split, and unemployment rate is used to further partition the result.

Random forest combines various of independent decision trees and randomly chooses the initial variable, decision boundaries, and variables used in each node when building trees. It makes decisions by the majority vote from the end nodes of each tree. Breiman (2001), one of major contributor to this method, proves by averaging all decision trees, the result is unbiased and has low variance. In addition, the result always converges and its accuracy is usually immune from irrelevant variables. Caruana and Niculescu-Mizil (2006) compare multiple surprised learning algorithms, such as SVM, Neural Nets, Random Forest, Logistic Regression, etc. Each method has a edge on one or two large datasets, but random forest and neural net have the best overall performance on all 11 large datasets. Kane et al. (2014) use Random Forest to predict outbreaks of H5N1 and finds it out-performs ARIMA. Khaidem, Saha, and Dey (2016) use Random Forest to predict trends in stock market and obtain 85% to 95% accuracy for one to three months prediction.

Random forest also access variable importance. It randomly permutes variables among

8. This data is labeled under column 3 in SPF's data spread sheet.
9. This data is labeled under column 2 in SPF's data spread sheet.

Figure 7: An Example of Decision Tree



nodes and calculates their loss of accuracy, or Z score. In classification, the Z score is mean decrees accuracy or Gini impurity. In regression, it is mean squared error. Then the variables importance is rank by the Z score. However, if two variables are highly correlated and equally important, one of them will be ranked high while the other one will be low (Breiman 2001). It also biases towards variables with more categories (Strobl et al. 2007). To circumvent these problems, we combine it with Boruta method, an 'all-relevant wrapper feature selection method' (Nilsson et al. 2007). It is used in many biomedical research such as genes (e.g. (Kursa 2014)) for robust variable selection. This method creates 'shadow copies' for each variable by randomly shuffling its order to remove correlation, and then it fits the original data and all 'shadow copies' to Random Forest and compare their Z scores. A variable is important (relevant) if its Z score is better than all shadow copies, and if not, it is removed. This method aggressively compare every variable and return a subset of relevant variables . As a comparison, forward and backward stepwise feature selection tend to give a subset of optimal variables that best fit the model, leaving some potential important variables.

We relax the criteria for Boruta' selection because we have too few data points compared to thousands of sample data used in biomedical research where this method is applied. A variable is labeled as important if its Z score out-performances 90% shadow copies. If it is lower than the threshold but higher than 5%, it is labeled as weak In addition, we randomly sample 70% of the date to fit the model. In the last two decays, U.S. has two major recessions but their durations are relatively short, thus some 'bad time' variables may have much lower frequencies in the whole data and have fewer chances to be selected in Random Forest. We then count the number of times each variable is labeled as important or weak.

The result of variable selection is in table 5. The variables with positive or negative are values above or below their means and have different signals. Both credit spread (DICS), loan demand (DILD), non-performing loans below mean (NPL negative), and excessive bond premium (EBP) are constantly selected as important. Unemployment rate below mean (Unemp negative) also has high ranks. It may not be a significant factor for loan officers, but it has a high correlation with the state of economy. Surprisingly, Real GDP growth is not important. Possible reasons may be its 2 lags and exponential moving average. Given the relaxed selection criteria and repeated sampling, We use the variables in the left column to fit the opinion formation model because they have much higher score of importance.

Table 5: US Variable Importance

Var	Important	Weak	Var	Important	Weak
DICS	20	0	SPF unemp positive	3	1
Unemp positive	20	0	SPF Corp negative	2	0
NPL negative	20	0	corp negative	1	2
EBP	19	0	SPF unemp negative	1	0
NPL positive	16	2	SPF Corp positive	0	0
DILD	16	1	Unemp negative	0	0
VIX	13	5	Corp positive	0	0
CPI	12	4	SPF RGDP positive	0	0
SPF CPI	11	4	RGDP negative	0	1
NASDAQ	8	1	RGDP positive	0	0
			SPF RGDP negative	0	1

The model coefficients are in Table 6. The diffusion index of credit spread (DICS) has a very significant effect. Non-performing loan rate (NPL) are also significantly larger than 0. Their large coefficients are a data scaling issue. The excessive bond premium (EBP) and unemployment that's above its one year EMA also reveal significant impact on lending decision.

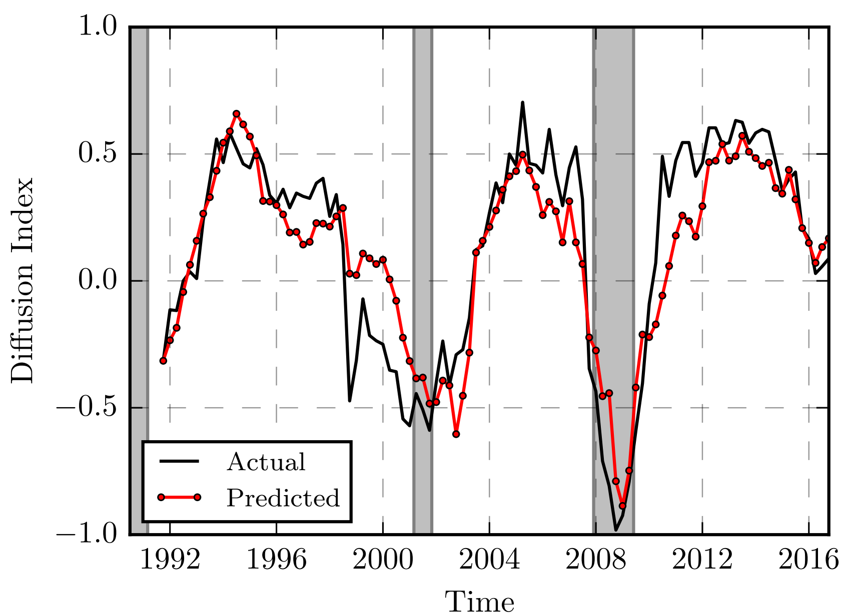
The model simulation for Model 4 is in figure 8. The first 5 observations are omitted because the DI loan demand is unavailable during that time. The simulation shares the same initial value with the actual data, but subsequent values in the U function (equation 16) are updated with calculated value. It captures the general trend of lending over the last two decades. But it fails at two significant movement near 1998 and 2007. The reason for this

Table 6: US Model 4

Dep. Variable:	DICS			**Log-Likelihood:**	76.409
Model:	QMLE			**AIC:**	-130.8
Method:	Maximum Likelihood			**BIC:**	-102.2
No. Observations:	100			**Number of Banks:**	70

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
v	0.8859	0.135	6.561	0.000	0.621 1.151
constant	0.2298	0.077	2.989	0.003	0.079 0.380
DICS	0.6213	0.137	4.534	0.000	0.353 0.890
NPL positive	-45.9799	27.395	-1.678	0.093	-99.674 7.714
NPL negative	-71.6126	24.658	-2.904	0.004	-119.942 -23.283
CPI	-26.0254	8.433	-3.086	0.002	-42.554 -9.497
EBP	-0.4983	0.141	-3.523	0.000	-0.776 -0.221
DILD	0.1102	0.168	0.656	0.512	-0.219 0.439
VIX	-1.6297	0.844	-1.932	0.053	-3.283 0.024
Unemp positive	-9.7948	2.593	-3.777	0.000	-14.877 -4.712
SPF CPI	-0.0035	0.152	-0.023	0.982	-0.302 0.295
NASDAQ	-0.2662	0.355	-0.749	0.454	-0.962 0.430

Figure 8: Model 4 Simulation



failure might lie in a spike in uncertainty given our observation in figure 4.

The exact date for the first drop of DI credit spread was in Q3 1998, when the credit spread dropped from 0.143 to -0.473. This is when Russia had its financial crisis, or known as Russian Flu (Aug 17th, 1998) , (Kindleberger and O'Keefe 2011, P 95-96). At the same time, Long-Term Capital Management had a long position on Russian market bonds and had business with most of U.S. firms during the summer of 1998. It went under as the Russian market crashed, and the US banks were waiting the Fed to clear market stress. Two quarters later, the credit spread was back to-0.071. During this period, loan demand increased, and unemployment rate continued to be lower its mean for 1 year. But the non-performing loan rates increased above its average.

The second significant drop of credit spread was on July 1st, 2007 when the spread dropped from 0.321 to -0.346. This was a quarter after the bank run on Countrywide, Northern Rock also had bank run on Sep 14th 2007 (86-87). Banks might feel the storms well before they were in headlines.

To appreciate the effect of variable reduction provided by the random forest procedure, we compare the model fitted by all variables. The model estimation is on table 7. Coefficients seem to have expected signs except the corporate profit. However, Model 5 is stretching parameters to over-fit the model and become sensitive noise. This is made clear when we perform cross validation for both model 4 and model 5 simulations as shown in figure 9.

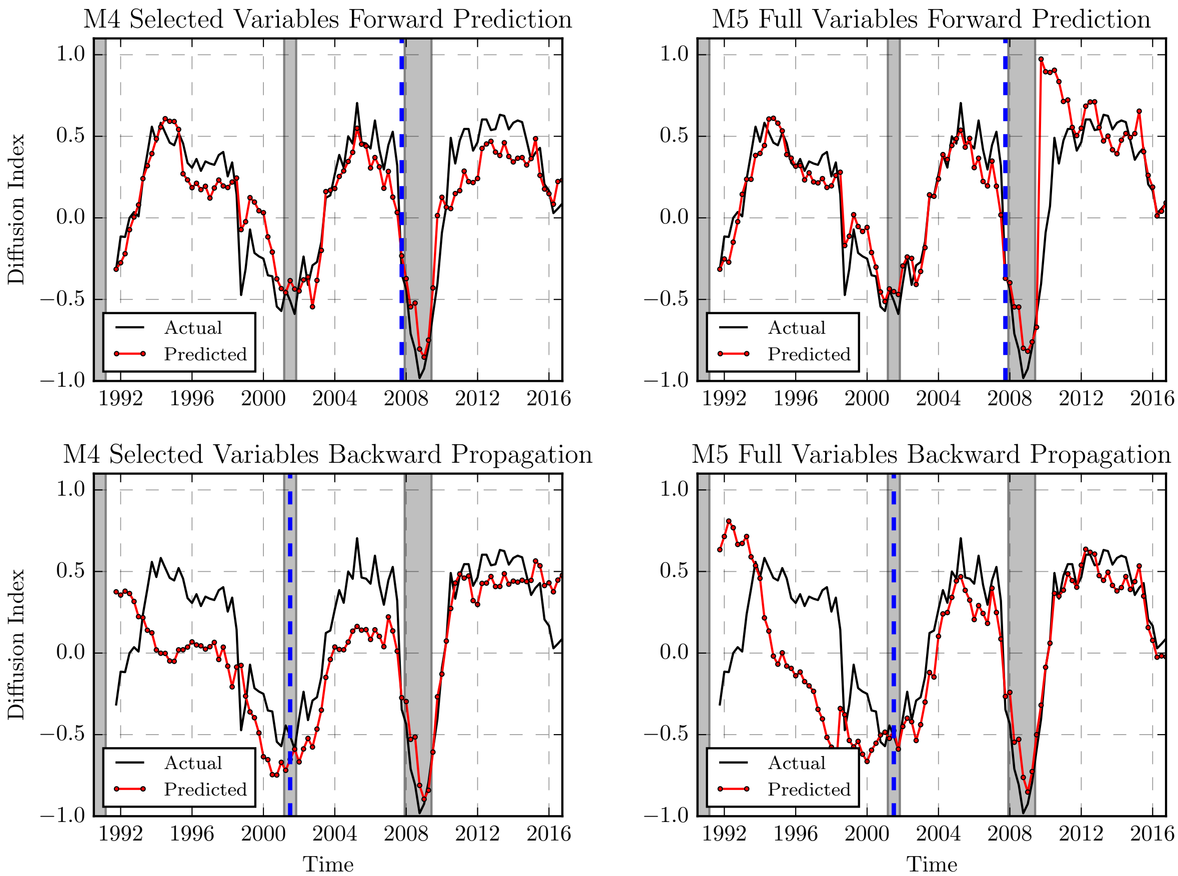
For the forward 'prediction', we use the first 70 points to estimate the model and 'predict' the actual diffusion index as we are updating the actual data (except the DI credit spread) for the exogenous variables. For the backward propagation, we use the last 70 points to build the model and update the data (except the DI credit spread) 'backward'. The exact date is chosen near the financial crisis. Model 4 has relatively stable simulation for forward 'prediction' as the economy is moving out of recession. Even thought it fails the backward propagation, the simulation still loosely follow the actual path. The unexpected "Russian Flue' may have large impact. Model 5's simulation is very unstable. Its forward prediction fits well for the data before 2008, but it explodes near 2009. Its backward propagation also has a very good fit for data after 2004, but the back 'prediction' losses almost all predictive power.

Table 7: US Modle 5

Dep. Variable:	DICS	**Log-Likelihood:**	93.970
Model:	QMLE	**AIC:**	-143.9
Method:	Maximum Likelihood	**BIC:**	-86.63
No. Observations:	100	**Number of Banks**	70

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
v	0.5944	0.099	6.031	0.000	0.401 0.788
constant	0.0291	0.101	0.288	0.773	-0.169 0.227
DICS	0.6263	0.168	3.718	0.000	0.296 0.956
DILD	0.3454	0.232	1.490	0.136	-0.109 0.800
EBP	-0.7833	0.195	-4.008	0.000	-1.166 -0.400
VIX	-2.0566	1.075	-1.912	0.056	-4.164 0.051
NASDAQ	-0.1577	0.453	-0.348	0.728	-1.045 0.730
CPI	-26.2813	10.057	-2.613	0.009	-45.992 -6.571
SPF CPI	0.1437	0.195	0.735	0.462	-0.239 0.527
RGDP positive	48.7490	21.948	2.221	0.026	5.732 91.766
RGDP negative	-34.3133	21.856	-1.570	0.116	-77.151 8.524
SPF RGDP positive	-2.3051	24.012	-0.096	0.924	-49.367 44.757
SPF RGDP negative	-31.1695	28.640	-1.088	0.276	-87.303 24.964
Unemp positive	-10.3816	3.741	-2.775	0.006	-17.714 -3.049
Unemp negative	-7.3362	4.259	-1.722	0.085	-15.684 1.012
SPF unemp positive	3.1802	4.008	0.793	0.427	-4.675 11.036
SPF unemp negative	-3.5309	3.277	-1.077	0.281	-9.954 2.892
Corp positive	-1.2457	1.061	-1.174	0.240	-3.325 0.834
Corp negative	4.0447	1.484	2.725	0.006	1.136 6.954
SPF Corp positive	3.7323	2.456	1.519	0.129	-1.082 8.547
SPF Corp negative	1.4935	3.165	0.472	0.637	-4.709 7.696
NPL positive	-48.8259	42.971	-1.136	0.256	-133.048 35.396
NPL negative	-73.9522	31.240	-2.367	0.018	-135.182 -12.723

Figure 9: M4 & M5 Comparison



2.2 Euro Area Credit Dynamics

The European Central Bank launched its quarterly bank lending survey in 2003, covering around 90 banks in all European areas. The sample expands to around 140 in 2016. Most of them are large banks, but specialized small banks are also included. The sample size for each country depends on its share of loan to private non-financial sector. The survey results are aggregated at the country level and complied to Euro area with weights on each country's loan share (Köhler-Ulbrich, Hempell, and Scopel 2016). The survey questions are similar to the U.S. SLOOS, covering lending to firms and households and factors of changing lending policies. They are available in the ECB's Statistical Data Warehouse. We focus on the credit spread for average loans to firms:

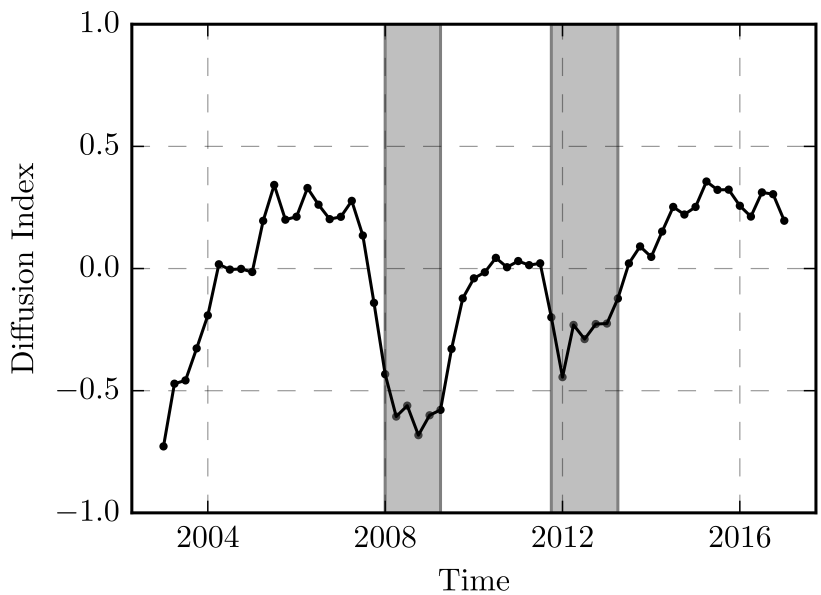
> *3. Over the past three months, how have your bank's terms and conditions for new loans or credit lines to enterprises changed? Please rate the overall terms and conditions for this loan category and each factor using the following scale: 1.tight-*

ened considerably, 2. tightened somewhat, 3. remained basically unchanged, 4. eased somewhat, 5. eased considerably

(c) Your banks' loan margin (i.e. the spread over a relevant market reference rate) on average loans (wider spread=tightened, narrower spread = eased)

We choose the data series on average loans to all firms because it is the longest. Its diffusion index (DI), constructed by the net percentage of banks that eased minus banks that tightened credit spread, weighted by each country's share of outstanding loans is in figure 10; the grey bars are Euro Area recessions indicated by OECD Composite Leading Indicators. The Euro Area diffusion index behaves similarly to the US data: banks tighten lending immediately before recession and ease credit before the economy moves out of recession. An example of this is seen in the period from 2003 to 2007, when most Euro banks switched from tightening to easing, but the DI spread dropped significantly in 2008.

Figure 10: Euro Area Diffusion Index of Credit Spread



The ECB survey also asks banks to rank factors that affect lending polices:

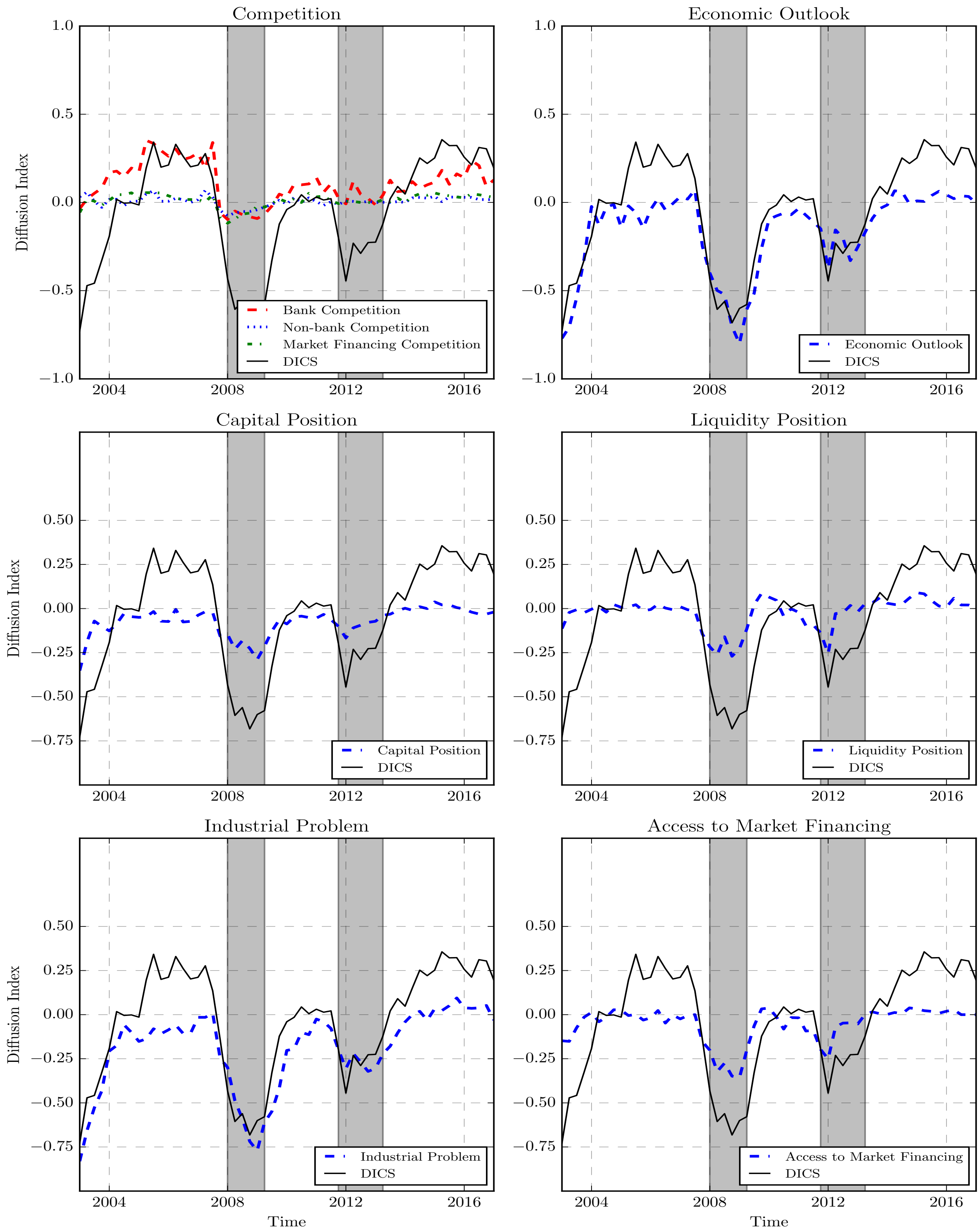
4. Over the past three months, how have the following factors affected your bank's credit terms and conditions as applied to new loans or credit lines to enterprises (as defined in the notes to question 3)? Please rate the contribution of the following factors to the tightening or easing of credit terms and conditions using the following scale: 1.contributed to tighten considerably, 2. contributed to tighten *somewhat, 3. contributed basically unchanged, 4. contributed to eased somewhat, 5. contributed to eased*

(a) *Cost related to your bank's capital position*
(b) *Your bank's ability to access market financing*
(c) *Your banks' liquidity position*
(d) *Competition from other banks*
(e) *Competition from non-banks*
(f) *Competition from market financing*
(g) *General economic situation and outlook*
(h) *Industry or firm-specific outlook/borrowers' creditworthiness*
(j) *Risk related to the collateral demanded*

This data series start in the second quarter of 2015. The survey also asks the same questions for credit standards since the beginning of the survey. Since both credit spread and credit standard have similar shapes, we choose the data for credit standard to approximate the impact on credit spread.

The figure 11 shows diffusion indexes on factors that impact lending policies. The index is the net percentage of eased minus tightened, weighted by the shares of each country's outstanding loans. They also show the same asymmetric effect on lending decisions as seen in US. When banks ease credit, competition, especially from other banks, is the main force. It explains the sharp increase in DI credit spread near 2005. Other reported factors have almost no impact. However, when banks are tightening lending, the general economic outlook is dragging lending. Industrial problems and borrowers' creditworthiness are also important in tightening credit but have almost no impact in easing credit. Surprisingly, liquidity position and capital position, which are generally regarded as important in lending decision, have far less important.

Figure 11: Euro Area Reasons for Changing Lending Policies



We apply the opinion formation on Euro Area's diffusion index of credit spread. Suppose we only consider peer effect only. Table 8 and 9 compare the asymmetric effect from peers. Quite similar to the US result, there is no big difference from the positive and negative on lending opinions. Model 2's AIC and BIC are even worse than model 1. This means separating this diffusion index by signs may not be an optimal choice. The simulation on M2 is in figure 12. The DI credit spread in U function is updated by the actual data. The simulation misses the great recession in 2007-8, but it has the similar trend with the actual data.

Table 8: Euro Area Model 1

Dep. Variable:	ECB DICS	**Log-Likelihood:**	40.611
Model:	QMLE	**AIC:**	-77.22
Method:	Maximum Likelihood	**BIC:**	-73.17
No. Observations:	56	**Number of banks:**	140

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
v	2.0156	0.371	5.439	0.000	1.289	2.742
constant	0.0053	0.017	0.317	0.751	-0.027	0.038
ECB DICSx0	0.9512	0.062	15.235	0.000	0.829	1.074

Table 9: Eura Area Model 2

Dep. Variable:	ECB DICS	**Log-Likelihood:**	40.745
Model:	QMLE	**AIC:**	-75.49
Method:	Maximum Likelihood	**BIC:**	-69.41
No. Observations:	56	**Number of Banks:**	140

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
v	2.0352	0.378	5.382	0.000	1.294	2.776
constant	0.0176	0.029	0.607	0.544	-0.039	0.074
ECB DICSx0+	0.8795	0.152	5.769	0.000	0.581	1.178
ECB DICSx0-	0.9912	0.098	10.125	0.000	0.799	1.183

The list of variables for the Euro Area is selected based on the U.S. data result. In the Euro Area, the VIX, 'fear index', is replaced by EURO STOXX 50, obtained from Factset database. The forecasted data for GDP, unemployment rate, and inflation rare are available

Figure 12: Euro Area Model 2 Simulation

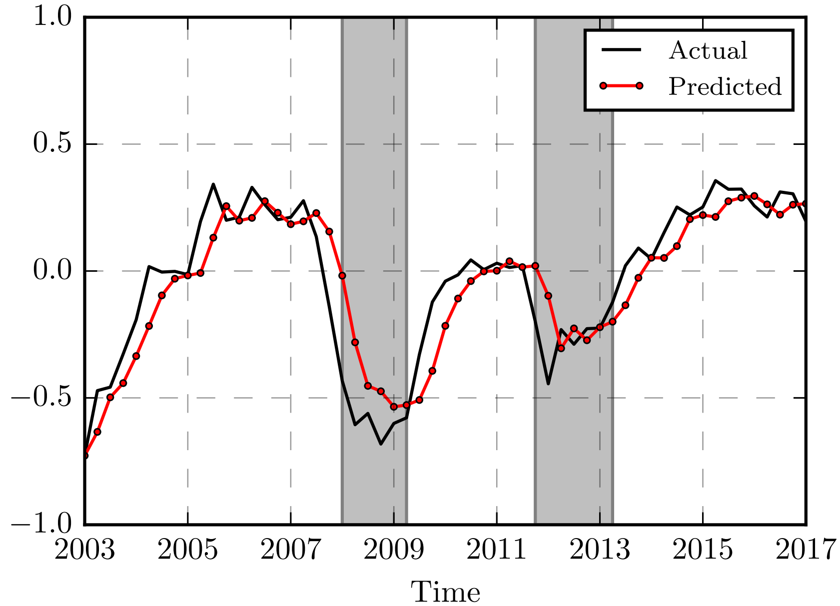


Table 10: Euro Area List of Variable

Variable	Description	Release Date	Season. Adj
US	Diffusion index if credit spread, average loans	Mid of next quarter	N
DILD	Diffusion index of loan Demand	Mid of next quarter	N
STOXX	Euro STOXX 50 volatility index. Quarter average, change	Daily	N
RGDP	RGDP growth rate. De-mean by its EMA	End of next quarter	Y
Unemp	Unemployment rate Euro Area, percent change. De-mean by its EMA	monthly	Y
SPF RGDP	SPF RGDP growth rate, next year. De-demean by its EMA	Mid of current quarter	Y
SPF Unemp	SPF unemp rate, percent change, mean. De-demean by its EMA	Mid of current quarter	Y
Inflation	Overall index (HICP). Percent change	Monthly	N
SPF Inflation	SPF CPI, next year. Percent change	Mid of current quarter	N

from ECB [10]. Corporate profit in the U.S. data has little impact on lending decisions, so it is excluded from the Euro data. The non-performing loan rate for the Euro Area is available, but it starts from Q2 2015 and therefore not considered here. Excessive bond premium for the Euro Area is absent because bank level information is not public.

The variable importance obtained from the same setting of the Boruta method is shown in table 11. In the absence of excessive bond premium and non performing loan rates, unemployment is an important variable. Real GDP, with two period lags (because of data release date), and future the forecast are less important. By the BOruta method, We choose the variables that are larger than 0 in the left column because of their relevance.

10. These data are scraped from ECB's SPF website instead of its data warehouse.

Table 11: Euro Area Variable Importance

Var	Important	Weak	Var	Important	Weak
DICS	20	0	Unemp negative	3	3
SPF Unemp positive	17	2	SPF Unemp negative	1	1
DILD	16	2	RGDP lag2 positive	0	0
Unemp positive	12	3	RGDP lag2 negative	0	0
STOXX	13	4	SPF RGDP positive	0	0
Inflation	6	2	SPF RGDP negative	0	0
SPF Infla	3	8			

Table 12: Euro Area Selected Variable Model 3

Dep. Variable:	DICS	**Log-Likelihood:**	55.909
Model:	QMLE	**AIC:**	-91.82
Method:	Maximum Likelihood	**BIC:**	-71.74
No. Observations:	55	**Number of Banks**	140

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
v	1.1854	0.225	5.259	0.000	0.744 1.627
constant	0.1879	0.194	0.970	0.332	-0.192 0.567
DICS	0.5242	0.188	2.788	0.005	0.156 0.893
DILD	0.0797	0.202	0.395	0.693	-0.316 0.475
STOXX	0.3259	0.294	1.110	0.267	-0.250 0.902
Inflation	-9.5982	5.539	-1.733	0.083	-20.454 1.258
Unemp positive	-3.6388	4.773	-0.762	0.446	-12.994 5.716
Unemp negative	-0.2153	5.103	-0.042	0.966	-10.218 9.787
SPF Unemp positive	-5.0701	2.889	-1.755	0.079	-10.733 0.592
SPF Unemp negative	-0.6924	2.122	-0.326	0.744	-4.851 3.466
SPF Infla	1.2303	15.288	0.080	0.936	-28.735 31.195

The model estimation for selected variables in table 12 gives consistent results, and variables have the expected signs. However, the unemployment rate may be noisy because of the diverse economy situation across European countries. For example, Greek has unemployment rate well above 20% in the last 5 years, but Germany maintains around 4% - 5% of unemployment rate. The Euro Area reports a around 10% of unemployment during the same time period. A country-level model analysis may have a better insight. The simulation is in figure 13. The value of DI credit spread in the U function is updated by calculated value. The predicted line miss the sudden increase of DI credit spread in 2005. Some exogenous factors may have impact on banks' decisions.

As we did for the US data, we also compare the full variable estimation as a stability check. The full data estimation is in table 13, and we also compare the forward prediction starting and backward propagation. For the forward prediction, we use the data before Q3 2013 to estimate the parameters and then update real data (except DI credit spread) to compute the prediction. For the backward propagation, we use the data after Q3 2006 to build the model and use that date as the initial position. Then we backward update the data to compute the prediction. The forward prediction is relatively stable, but the back propagation is very unstable. On potential explanation is the data length. The U.S. data requires around 60 points to obtain stable parameters, but these two simulations use 40 points for parameter estimation. The EU data may be too short to give a stable estimation.

Table 13: Euro Area All Variables Model 4

Dep. Variable:	DICS		**Log-Likelihood:**		56.516	
Model:	QMLE		**AIC:**		-89.03	
Method:	Maximum Likelihood		**BIC:**		-64.94	
No. Observations:	55		**Number of Banks:**		140	

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
v	2.2894	0.425	5.382	0.000	1.456	3.123
constant	0.0666	0.104	0.641	0.522	-0.137	0.270
DICS	0.8334	0.099	8.442	0.000	0.640	1.027
DILD	-0.0159	0.107	-0.148	0.882	-0.226	0.194
STOXX	0.1780	0.155	1.148	0.251	-0.126	0.482
Inflation	-3.7667	2.940	-1.281	0.200	-9.529	1.996
SPF_Infla	-0.0845	8.002	-0.011	0.992	-15.768	15.599
RGDP_lag2 positive	0.4055	525.526	0.001	0.999	-1029.606	1030.417
RGDP_lag2 negative	-0.9045	2193.465	-0.000	1.000	-4300.017 [11]	4298.208
Unemp positive	-2.3672	2.493	-0.949	0.342	-7.254	2.520
Unemp negative	0.3466	2.676	0.130	0.897	-4.898	5.591
SPF_RGDP positive	0.4055	525.526	0.001	0.999	-1029.606	1030.417
SPF_RGDP negative	-0.9045	2193.464	-0.000	1.000	-4300.015	4298.206
SPF_Unemp positive	-2.4533	1.500	-1.635	0.102	-5.394	0.487
SPF_Unemp negative	-0.8982	1.126	-0.798	0.425	-3.106	1.309

11. The variance for Real GDP is large. The algorithm fails to invert the Hessian matrix when the stopping criteria is set at 1e-12. We relax it and modify the algorithm to make the Hessian invertible. The Hessian matrix is obtained after estimating the parameters, so those values change very little when we test various stopping criteria. In addition, when we do model simulation, only the variable coefficients are used.

Figure 13: Euro Model 3 Simulation

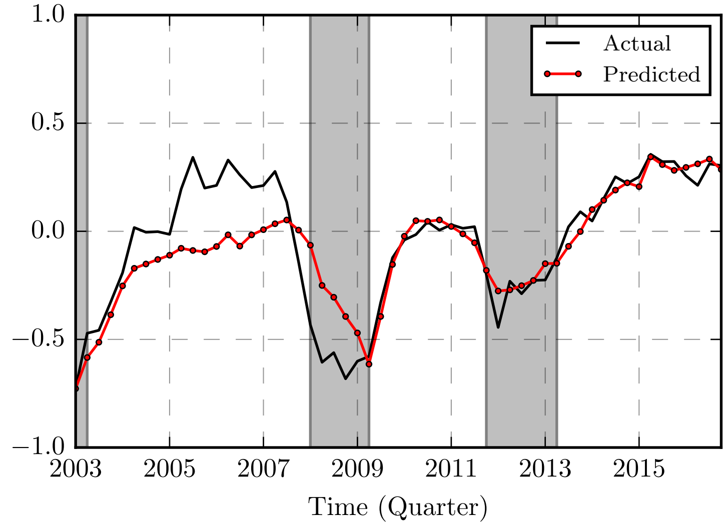
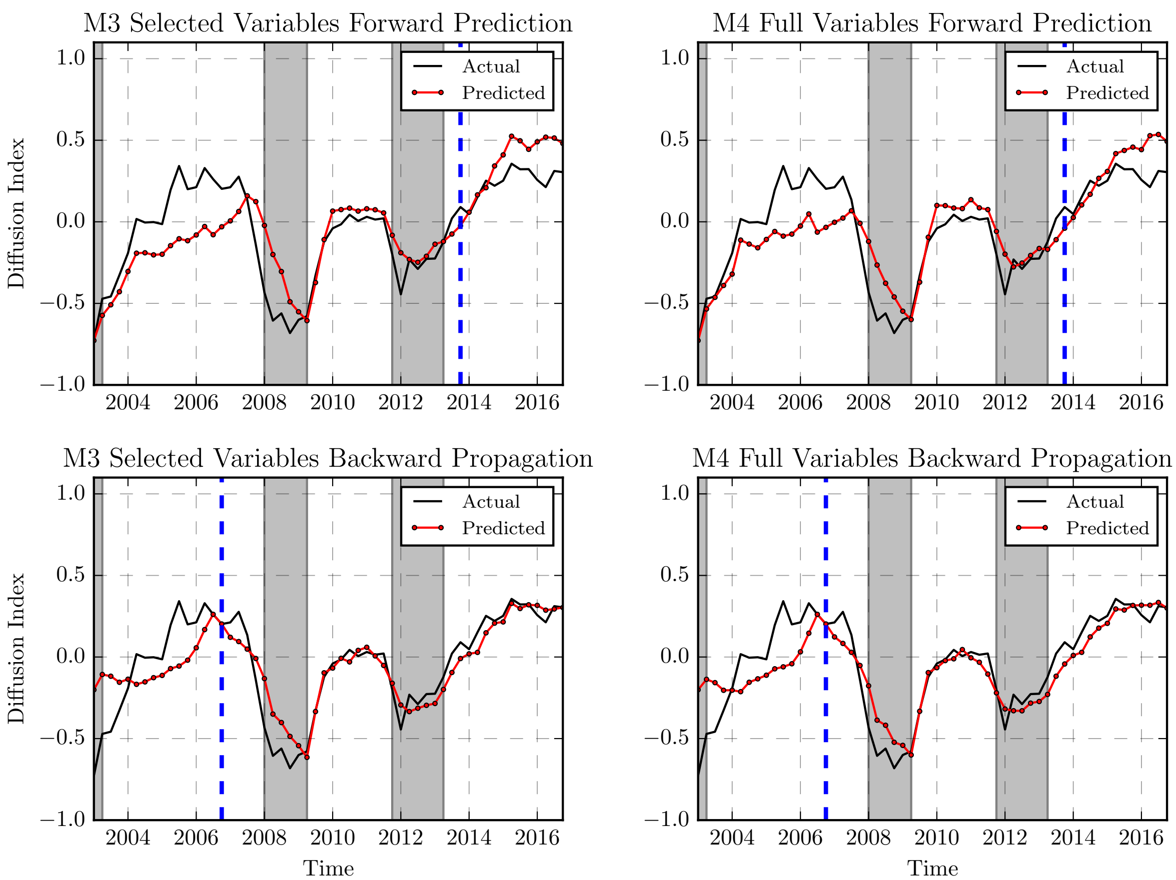


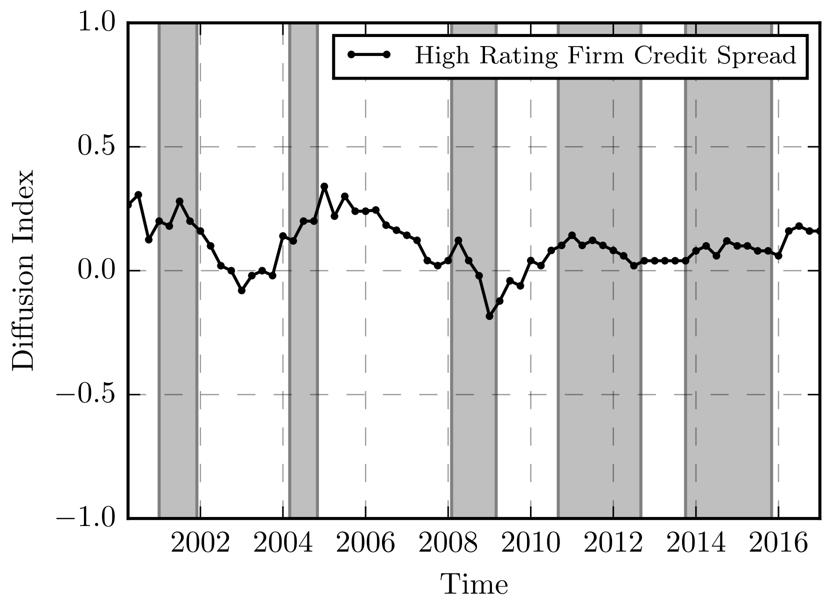
Figure 14: Euro Area Model 3 & Model 4 Stability Check



2.3 Japan Credit Dynamics

Japan's bank lending survey's structure and questions are very similar to the US. However, The credit spread data is not available in the central bank's database. We extra the data from its reports and calculate their values by taking the net difference of eased minus tightened. The credit spread for high rating firms is in figure 15 . The gray bars are OECD based recession period for Japan.

Figure 15: Japan Diffusion Index of Credit Spread



The list of potential relevant variables are in table 14. They are selected based on US and ECB's model result.

Table 14: Japan List of Variables

Variable	Description	Release Date	Season. Adj
DICS	Diffusion index of credit spread for high rating firms	Mid of next quarter	N
DILD	Diffusion index of loan demand from large firms	Mid of next quarter	N
Business Forecast	TANKAN Business Forecast from Banks (DI)	Mid of next quarter	Y
Employment Forecast	TANKAN Employment Forecast from Banks (DI)	Mid of next quarter	N
VXJ	OSE Volatility index of Japan. Quarter average, change	Daily	N
Nikkei 225	Nikkei Stock Average. Percent change	Daily	N
Inflation	All item. Percent change	Monthly	N
Unemp	Unemployment rate. Percent change. De-mean by its 1Y EMA	Monthly	Y
RGDP	Real GDP, 2 lags. Percent change;	Mid of next quarter	Y

The TANKAN, a short-term economic survey of enterprises starting from Q2 2004, is used to approximate banks' future expectation. Part of the survey asks banks opinions on

Table 15: Japan Model 1

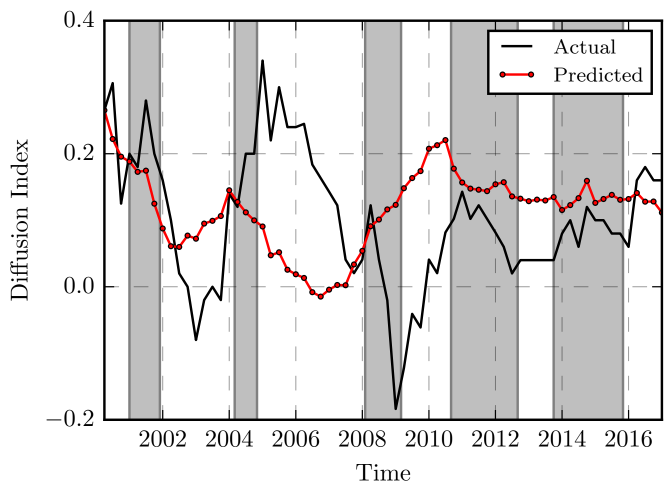
Dep. Variable:	High rating credit spread	**Log-Likelihood:**	94.712
Model:	QMLE	**AIC:**	-185.4
Method:	Maximum Likelihood	**BIC:**	-181.0
No. Observations:	67	**Number of Banks**	50

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
v	0.1646	0.032	5.156	0.000	0.102 0.227
constant	0.2424	0.128	1.888	0.059	-0.009 0.494
DICS	-1.6313	0.989	-1.649	0.099	-3.570 0.307

the future business and employment conditions. Their diffusion index (favorable minus unfavorable) are listed in the BOJ's database. The VIX equivalent for Japan is VXJ, obtained from Osaka University, Japan. The index follows CBOE's method for calculating VIX. Nonperforming loans for Japan is a serious problem. But the actual quarterly data, especially the non performing commercial loans, are hard to find [12].

Model 1 considers the peer effect only, and its coefficient for DI credit spread is quite different from the U.S. and ECB's results. It is in negative signs. In Model 1's simulation, DI credit spread is updated via the actual data, and the predicted result has barely resembles the estimated data.

Figure 16: Japan Model 1 Simulation



12. Technically, we could use the annual data from Japan government's Financial Services Agency and apply K-Nearest Neighbor (K-NN), a machine learning algorithm, to impute quarterly data, or we could also use random forest to compute the probability distribution for the quarterly data and calculate their weighted mean. But this topic will be explored in the future.

The relevant variables obtained from the Boruta method are in table 16. Real GDP below mean becomes important while the unemployment rate seems to be less relevant. This result seems to reconcile parts of the Japanese economy: it labor market is tight and the economy suffers from the Lost Two Decades. We use variables on the left column of table 16 because they have high score of importance, which means they are highly relevant.

Table 16: Japan Variable Importance

Var	Important	Weak	Var	Important	Weak
DICS	20	0	Employment Forecast	0	0
VJX	14	3	Inflation	0	0
Nikkei	14	3	RGDP lag2 positive	0	0
Business Forecast	13	3	Unemp positive	0	0
RGDP lag2 negative	4	2	Unemp negative	0	0
DIHRLD	3	5			

In Model 2, the coefficient for Real GDP is large and has the expected sign. The VJX, 'fear index', also has significant effect. The state of economy does has strong impact on banks' lending decisions. Model 3 uses all variables. Variables have expected signs, and the DI credit spread still behaves different from the US and Euro Area because it is significantly negative.

Finally, we compare Model 2 and Model 3's predicting power in figure 17 and 18. The simulation starts two year later because the business and employment forecast starts at 2004. For forward prediction, we use data before Q2 2014 to estimate model and update the real data (except DI credit spread). For the backward propagation, we use Q4 2006 as the initial value and data afterward to estimate the model, then we backward update the real data (except DI credit spread). DI credit spread is updated by the calculated value. Including more variables make the in sample prediction better as shown in figure 17. But including all variable makes the forward prediction worse. The predicted value moves in the opposite direction compared to the actual data. Given the poor performance for the Japan simulations, we conjecture that the non-performing loans might be a very significant factors for banks' lending policies.

Table 17: Japan Selected Varialbes Model 2

Dep. Variable:	DICS	**Log-Likelihood:**	81.993
Model:	QMLE	**AIC:**	-150.0
Method:	Maximum Likelihood	**BIC:**	-136.5
No. Observations:	51	**Number of Banks:**	50

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
v	0.1087	0.025	4.332	0.000	0.060 0.158
constant	0.3936	0.246	1.602	0.109	-0.088 0.875
DICS	-2.5989	1.649	-1.576	0.115	-5.832 0.634
VJX	-3.6197	1.426	-2.538	0.011	-6.415 -0.825
Nikkei	-0.7526	1.500	-0.502	0.616	-3.692 2.187
Business Forecast	0.3591	0.953	0.377	0.706	-1.509 2.227
RGDP lag2 negative	53.6641	35.037	1.532	0.126	-15.008 122.336
DIHRLD	-0.0218	0.857	-0.025	0.980	-1.702 1.659

Figure 17: Japan Model 2 Simulation

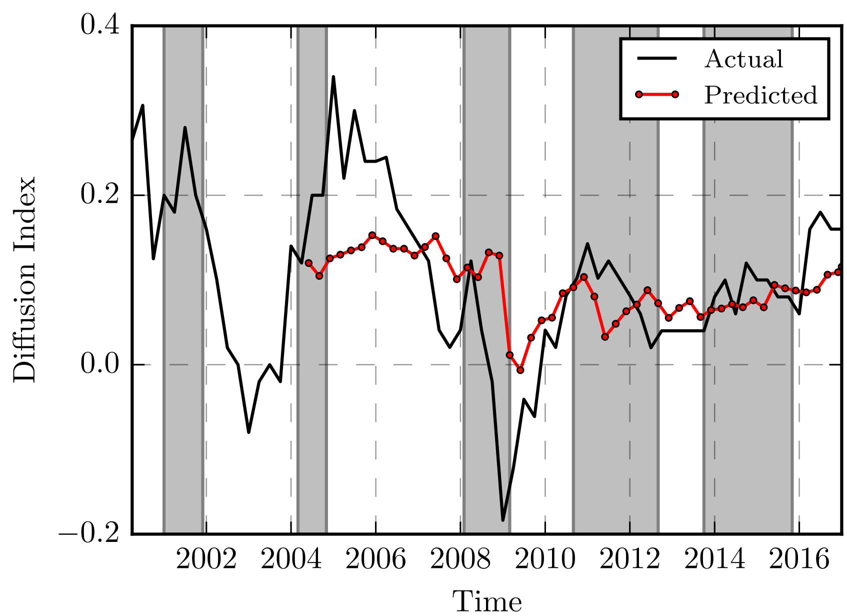
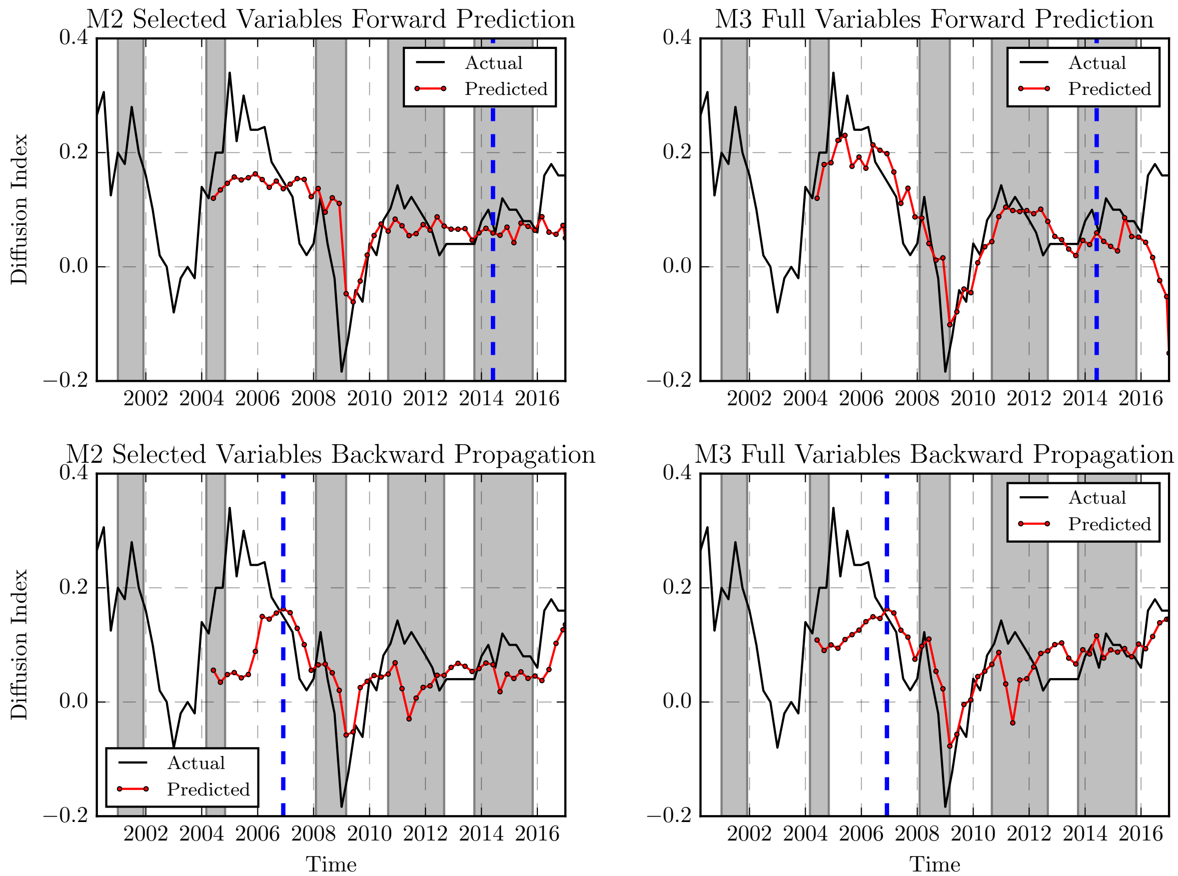


Table 18: Japan ALl Varialbes Model 3

Dep. Variable:	DICS	**Log-Likelihood:**	85.503
Model:	QMLE	**AIC:**	-147.0
Method:	Maximum Likelihood	**BIC:**	-123.8
No. Observations:	51	**Number of Banks:**	50

	coef	**std err**	**z**	**P>|z|**	**[95.0% Conf. Int.]**
v	0.0973	0.023	4.273	0.000	0.053 0.142
constant	0.0752	0.345	0.218	0.828	-0.602 0.752
DICS	-3.6806	2.110	-1.744	0.081	-7.816 0.455
DILD	1.1644	0.976	1.193	0.233	-0.748 3.077
Business Forecast	3.2089	1.889	1.699	0.089	-0.494 6.912
Employment Forecast	5.3428	2.732	1.956	0.051	-0.012 10.697
Inflation	-1.7535	5.096	-0.344	0.731	-11.742 8.235
VJX	-3.2922	1.446	-2.277	0.023	-6.126 -0.458
Nikkei	-1.1286	1.558	-0.724	0.469	-4.183 1.926
RGDP lag2 positive	-18.0797	27.857	-0.649	0.516	-72.678 36.519
RGDP lag2 negative	64.7337	41.294	1.568	0.117	-16.201 145.668
Unemp positive	3.1704	12.929	0.245	0.806	-22.170 28.510
Unemp negative	-13.7926	9.820	-1.405	0.160	-33.040 5.454

Figure 18: Japan Model 2 & Model 3 Simulation



3 Discussion and Summary

In this paper we extends the Opinion Formation Model to Euro Area and Japan. We use machine learning techniques for variable selection. The result for the US data is stable. We find banks react differently to economic factors for changing lending policies. Uncertainty and competition are two main factors that act differently: the former in bad times and the latter in good times. Capital position and loan liquidity for banks in the U.S. and Euro weight less. The data on diffusion index of credit spread reveal momentum effect. For both the US and the ECB's data, banks tighten the spread one or two quarters before economic downturn or unexpected event such as 'Russian Flue'. Their plots also has momentum: banks lend more if the past quarter lend more until uncertainty dominate the market, and then the liquidity freezes. The success of the simulations mean banks do consider their peers' lending opinions in addition to their own rational risk calculation.

Interestingly, unemployment rate has strong effect, but Real GDP has little impact on lending for these the US and Euro Area' data. Random forest ranks RGDP very low in terms of variables importance. According to OKun's law, both variables should have similar explanatory power. It might due to Real GDP's two quarter lags and their moving average. Banks might rely more on their economic forecasts. However, from Post-Keynesian's point of view, such finding is not unexpected. Loans are underwritten in nominal terms, so price level matters. Dropping money from helicopters or multiplying everything by 100 does have real effect unless loans are denominated in real term.

For the Euro Area, a more detailed analysis at the country level may reveal more insight on bank lending opinions. For example, Germany has the most banks represented in ECB's lending survey. The country also has data on non-performing loans. Its strong economic position inside the Euro might bring in side effects, but we are interested in examining how German banks's credit dynamics evolves compared to the US.

In the Japan's analysis, Real GDP, especially below its one year exponential moving average, has very significant impact on lending decisions. This is reasonable given Japan's long term stagnation. Japan also has a significant problem in non-performing loans, We conjecture that the nominal GDP might have a much larger weights than Real GDP and provides better insight. Japan might be a place to provide empirical support on lending behaviors for Fisher's debt deflation theory and Richard Koo's analysis on balance sheet recession ((Koo 2011), (Koo 2016)).

However, none of these model results have explanatory power comparable to a top banker's comment :

> I missed apiece of business, ..., I can live with that, but as soon as I hung up the phone someone else put up 10 times leverage. ***We cannot control ourselves. You have to step in and control the street***.
>
> —– John Mack, Morgan Stanley' Chief Executive, at the "Covering the Crisis" panel discussion with Bloomberg News in Nov,19,2009 (emphasis added).

References

Altavilla, Carlo, Matthieu Darracq Paries, Giulio Nicoletti, et al. 2015. *Loan supply, credit markets and the euro area financial crisis.* Technical report.

Bollerslev, Tim, and Jeffrey M Wooldridge. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews* 11 (2): 143–172.

Bouwman, Christa HS, and Ulrike Malmendier. 2015. Does a bank's history affect its risk-taking? *The American Economic Review* 105 (5): 321–325.

Breiman, Leo. 2001. Random forests. *Machine learning* 45 (1): 5–32.

Caruana, Rich, and Alexandru Niculescu-Mizil. 2006. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on machine learning,* 161–168. ACM.

Dow, Sheila C. 1996. Horizontalism: a critique. *Cambridge Journal of Economics* 20 (4): 497–508.

Favara, Giovanni, Simon Gilchrist, Kurt Lewis, and Egon Zakrajsek. 2016. Recession risk and the excess bond premium. `http://EconPapers.repec.org/RePEc:fip:fedgfn:2016-04-08`.

Fisher, Irving. 1933. The debt-deflation theory of great depressions. *Econometrica: Journal of the Econometric Society:* 337–357.

Fixler, Dennis J, Ryan Greenaway-McGrevy, and Bruce T Grimm. 2014. Revisions to gdp, gdi, and their major components.

Fontana, Giuseppe. 2004. Rethinking endogenous money: a constructive interpretation of the debate between horizontalists and structuralists. *Metroeconomica* 55 (4): 367–385.

———. 2009. *Money, uncertainty and time.* Routledge.

Ghonghadze, Jaba, and Thomas Lux. 2012. Modelling the dynamics of eu economic sentiment indicators: an interaction-based approach. *Applied Economics* 44 (24): 3065–3088.

———. 2016. Variation of lending standards and the financial instability hypothesis. *Under Review Process.*

Gilchrist, Simon, and Zakraj. 2012. Credit spreads and business cycle fluctuations. *The American Economic Review* 102 (4): 1692–1720.

Granovetter, Mark. 2005. The impact of social structure on economic outcomes. *The Journal of economic perspectives* 19 (1): 33–50.

Hawkins, Raymond J. 2011. Lending sociodynamics and economic instability. *Physica A: Statistical Mechanics and its Applications* 390 (23): 4355–4369.

Holmes, Alan. 1969. Operational constraints on the stabilization of money supply growth. *Controlling Monetary Aggregates,(Boston MA: Federal Reserve Bank of Boston):* 65–77.

Iacus, Stefano M. 2009. *Simulation and inference for stochastic differential equations: with r examples.* Springer Science & Business Media.

Kane, Michael J, Natalie Price, Matthew Scotch, and Peter Rabinowitz. 2014. Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks. *BMC bioinformatics* 15 (1): 276.

Keen, Steve. 2013. A monetary minsky model of the great moderation and the great recession. *Journal of Economic Behavior & Organization* 86:221–235.

Keynes, John Maynard. 1937. The general theory of employment. *The quarterly journal of economics:* 209–223.

Khaidem, Luckyson, Snehanshu Saha, and Sudeepa Roy Dey. 2016. Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003.*

Kindleberger, Charles Poor, and Robert O'Keefe. 2011. *Manias, panics, and crashes.* Springer.

Köhler-Ulbrich, Petra, Hannah S Hempell, and Silvia Scopel. 2016. The euro area bank lending survey: role, development and usein monetary policy preparation.

Koo, Richard. 2011. The world in balance sheet recession: causes, cure, and politics. *Real-world economics review* 58 (12): 19–37.

———. 2016. Tthe other half of macroeconomics and three stages of economic development. *Real-world economics review* 75 (1).

Kursa, Miron Bartosz. 2014. Robustness of random forest-based gene selection methods. *BMC bioinformatics* 15 (1): 8.

Lavoie, Marc. 1996. Horizontalism, structuralism, liquidity preference and the principle of increasing risk. *Scottish Journal of Political Economy* 43 (3): 275–300.

Lemperiere, Yves, Cyril Deremble, Philip Seager, Marc Potters, and Jean-Philippe Bouchaud. 2014. Two centuries of trend following. *arXiv preprint arXiv:1404.3274.*

McLeay, Michael, Amar Radia, and Ryland Thomas. 2014. Money in the modern economy: an introduction. *Bank of England Quarterly Bulletin:* Q1.

Menkhoff, Lukas, Lucio Sarno, Maik Schmeling, and Andreas Schrimpf. 2012. Currency momentum strategies. *Journal of Financial Economics* 106 (3): 660–684.

Minsky, Hyman P. 1976. *John Maynard Keynes.* Springer.

Moore, Basil J. 1979. The endogenous money stock. *Journal of Post Keynesian Economics* 2 (1): 49–70.

———. 1988. *Horizontalists and verticalists: the macroeconomics of credit money.* Cambridge University Press.

Nilsson, Roland, José M Peña, Johan Björkegren, and Jesper Tegnér. 2007. Consistent feature selection for pattern recognition in polynomial time. *Journal of Machine Learning Research* 8 (Mar): 589–612.

Osborn, Denise R. 1995. Moving average detrending and the analysis of business cycles. *Oxford Bulletin of Economics and Statistics* 57 (4): 547–558.

Pollin, Robert. 1991. Two theories of money supply endogeneity: some empirical evidence. *Journal of Post Keynesian Economics* 13 (3): 366–396.

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC bioinformatics* 8 (1): 25.

Towers, G. 1939. Minutes of proceedings and evidence respecting the bank of canada. *Committee on Banking and Commerce, Ottawa: Government Printing Bureau.*

Weidlich, Wolfgang, and Günter Haag. 1983. *Concepts and models of a quantitative sociology: the dynamics of interacting populations.* Vol. 14. Springer Science & Business Media.

Werner, Richard A. 2014. Can banks individually create money out of nothing? the theories and the empirical evidence. *International Review of Financial Analysis* 36:1–19.

Wray, L Randall. 1995. Keynesian monetary theory: liquidity preference or black box horizontalism? *Journal of Economic Issues* 29 (1): 273–282.