# Credit Scoring with Ego-Network Data

## Stanley Sewe[1] , Philip Ngare[2], Patrick Weke[2]

[1]Department of Mathematics, Pan African University Institute of Basic Sciences, Technology and Innovation, Juja, Kenya
[2]School of Mathematics, University of Nairobi, Nairobi, Kenya
Email: stanleysewe@gmail.com

## Abstract

This article investigates a stochastic filtering problem whereby the borrower's hidden credit quality is estimated using ego-network signals. The hidden credit quality process is modeled as a mean reverting Ornstein-Ulehnbeck process. The lender observes the borrower's behavior modeled as a continuous time diffusion process. The drift of the diffusion process is driven by the hidden credit quality. At discrete fixed times, the lender gets ego-network signals from the borrower and the borrower's direct friends. The observation filtration thus contains continuous time borrower data augmented with discrete time ego-network signals. Combining the continuous time observation data and ego-network information, we derive filter equations for the hidden process and the properties of the conditional variance. Further, we study the asymptotic properties of the conditional variance when the frequency of arrival of ego-network signals is increased.

## Keywords

Stochastic Filtering, Bayesian Updating, Credit Scoring, Filtration, Ego-Network

## 1. Introduction

In this article we propose a filtering technique that uses ego-network signals to estimate a hidden process. Consider a financial market with a single lender and borrowers, who are represented by nodes in a dynamic social network. For a particular borrower, let the process $X_t$ modeled as a mean reverting Ornstein-Ulehnbeck process capture her[1] true credit quality. On account of the information asymmetry between the borrower and the lender, the lender is unable to directly observe $X_t$. However through interactions with the borrower, the lender gets to observe a continuous time process $Y_t$, modeled as a linear

---

[1]In this article we refer to the borrower as female and the lender as male.

diffusion process. $X_t$ drives the drift of $Y_t$. This is a continuous time linear state space model with $X_t$ being the state process and $Y_t$ the observation process. Kalman-Bucy filtering can thus be used to obtain the "optimal" estimate of $X_t$ in the mean square sense.

We assume that network ties are based on homophily. Homophily [1], is the idea that individuals with similar characteristics are likely to be friends than individuals with different characteristics. Thus social network ties are based on closeness in credit type: the probability that two individuals will create/maintain a network tie between them is proportional to the distance between their credit types. The probability of a network tie formation/termination is conditional on the parties meeting. The meeting process is modeled as a random event whose probability is a deterministic function of the population size of the network, large population leading to sparse networks. Individuals know their credit type and can also observe the credit type of their direct friends (alters) in the network. The social network is thus modeled as a dynamic latent space network. The lender's view of the network is restricted to ego-network signals of borrowers at fixed discrete times. At times $0 = t_0 < t_1 < \cdots < t_{N-1} < T$, the lender observes the particular borrower's ego network and receives unbiased signals related to her credit quality and the credit quality of her alters. Thus at the information times $t_k$, $k = 0,1,\cdots,N-1$, the lender gets to observe the vector $\boldsymbol{Z}_k$ constituting the unbiased signals of the credit quality of the borrower and her alters. The dimension of the vector is a function of the actual degree (number of alters) of the borrower at the time $t_k$.

Our model proposes the inclusion of the ego-network signals $\boldsymbol{Z}_k$ into the filtering of the process $X_t$. The lender's observation filtration is augmented by the filtration generated by $\boldsymbol{Z}_k$ at discrete time points. In the proposed model, Bayesian updating at times $t_k$ is used to incorporate the information from the ego-network signals into the estimation of $X_t$. We note that by the Gaussian nature of the processes $X_t, Y_t$ and $\boldsymbol{Z}_k$ and the formulation of the ego-network likelihood, the updated estimate of $X_t$ at times $t_k$ remains Gaussian and we derive explicit results for its mean and variance. We also derive results showing that the inclusion of the signals $\boldsymbol{Z}_k$ leads to lower conditional variance for the filtered process. By introducing the meeting probability in network tie formation, our model is an extension of [2], where the conditional expected number of friends was treated as a constant. Further, we study the asymptotic behavior of the conditional variance as the frequency of network information arrivals $N \to \infty$. Increasing the frequency of network information arrival times leads to clearer signals, and in the limit as $N \to \infty$, we get to the full information scenario.

There exists several studies on the statistical modeling of social network. Some of the models proposed in these studies include the coevolution model of [3] whereby the authors proposed a continuous time network model. The nodal attributes modeled as Markov chains influence the formation of network ties, which in turn influence the transition probabilities of the nodal attributes. In [4],

the authors proposed a static latent space network model, where the nodal attributes are in a low dimensional Euclidean space, and these attributes influence the formation of network ties. The static model has been extended severally to the time varying case by among others, [5] who proposed a directed dynamic latent space model. For a review of the recent studies on latent space network models, see e.g. [6].

Existing studies on the mathematical modeling of consumer credit risk include [7] where the authors proposed a continuous time model of a borrower's credit type. By modeling the credit type as a jump diffusion process and applying the Option pricing theory, the authors were able to derive explicit formulation of the borrower's default probability. Consumer credit risk modelling is mainly focused on credit scoring, the use of statistical models to aid in credit granting decisions. Common techniques used for credit scoring include linear discriminant analysis, logistic regression, bayesian classifiers, random forest and finite Markov chains, see e.g. [8] for a review. In recent times, Hidden Markov models (HMM) have been applied for credit scoring e.g. [9] who compared the performance of HMM and logistic regression in the classification of customers and evaluation of the probability of default. [10] modeled the consumer's credit rating as a discrete time Markov chain process upon incorporating a latent variable which captures the prevailing economic conditions. In [11], the author proposed a credit scoring model whereby the borrower's hidden credit type modeled as a discrete time Markov chain is learned through observing network related variables including reputation, trust and distrust. Proposing a static credit scoring model, [12] used ego-network signals to update the lender's belief of the borrower's unobserved credit type modeled as a Gaussian random variable. For a review of the application of social network data to consumer credit risk modeling , see [13].

In [14], the authors augmented the observation filtration with discrete time expert opinion to estimate the hidden Gaussian process driving the drift of the stock price. The model is an extension of the Black-Litterman model of [15] to the continuous time case. The authors in [16] estimated the unobserved drift parameter on an observation filtration initially enlarged with some anticipative information perturbed by independent noise. Text book treatment of stochastic filtering includes [17], [18] and [19].

The article is organized as follows. In Section 2, the credit risk and dynamic network models are presented. Results on stochastic filtering are presented in Section 3. In Section 4, the properties of conditional variances are derived in details under the various information setups. Brief numerical results are presented in Section 5, whilst Section 6 concludes.

## 2. The Model Setup

Consider a filtered probability space $\left(\Omega, \mathcal{A}, \mathbb{A}, \mathbb{P}\right)$ with $\mathbb{A} = \left(\mathcal{A}_t\right)_{t \geq 0}$ satisfying the usual conditions of right continuity and completeness. All processes are as-

sumed to be $\mathbb{A}$ adapted.

### Borrower's Behavioral Dynamics

The borrower's hidden credit quality process $X_t$ is modeled as a mean-reverting Ornstein-Ulehnbeck process defined as

$$dX_t = \mu(\delta - X_t)dt + \gamma dB_t \tag{1}$$

$$X_0 = x$$

where $\delta \in \mathbb{R}, \mu > 0, \gamma > 0$ are constants and $B_t$ is a Brownian motion. $x \sim \mathcal{N}(m_0, v_0)$, $\mathbb{A}$ measurable and is independent of $B$. Thus $X_t$ is a Gaussian process with the mean and variance given by

$$\delta + e^{-\mu t}(m_0 - \delta) \tag{2}$$

$$\frac{\gamma^2}{2\mu} + e^{-2\mu t}\left(v_0 - \frac{\gamma^2}{2\mu}\right) \tag{3}$$

respectively. The hidden credit quality $X_t$ drives the drift of the borrower's observed behavioral dynamics which is modeled as a diffusion process defined as

$$dY_t = \alpha X_t dt + \sigma dW_t \tag{4}$$

The parameters $\alpha, \sigma > 0$ are assumed to be constants and $W_t$ is a $\mathbb{A}$ adapted one dimensional Brownian motion. $W_t$ and $X_t$ are assumed to be independent.

### Network Dynamics

Let $\mathcal{Z}$ be the population of a society, such that individuals are represented as nodes in a dynamic network. Each individual in the population is assumed to have an independent and time varying credit quality $X_{it}$ modeled as a Gaussian process. When a pair of individuals $i$ and $j$ get the opportunity to meet, they may decide to create, terminate or continue a network tie by mutual consent. Thus network tie formation and termination are conditioned on the meeting probability $\upsilon = \dfrac{1}{\mathcal{Z}(\mathcal{Z}-1)}$. Modeling the meeting probability as a function of population size captures network sparseness, which is a property observed in real life social networks. Thus the meeting probability reduces with increased number of individuals in the population. Assuming an undirected network *i.e.* $\mathcal{Y}_{ij}(t) = \mathcal{Y}_{ji}(t)$, then we let

$$\mathcal{Y}_{ij}(t) \mid \pi_{ij}(t) \sim \text{Bern}(\pi_{ij}(t)) \tag{5}$$

for every $i \neq j$ and $t \geq 0$ with

$$\pi_{ij}(t) = e^{-\frac{1}{2}(X_{it} - X_{jt})^2} \tag{6}$$

Hence the network ties $\mathcal{Y}_{ij}(t) \in \{0,1\}$ are independent Bernoulli random variables conditioned on the nodal attributes $X_t$. The network tie formation probability $\pi_{ij}(t) \in (0,1)$ is modeled as a probit link function. Conditional on the meeting process, existence of a network tie between individual $i$ and $j$ at time $t$ is a function of the Euclidean distance between their respective credit types.

The network model captures homophily, since shorter distance between credit types leads to higher probability of network tie formation. The model assumes zero cost incurred on network tie formation or termination.

Define $G_t$ as the graph of friendship ties in the society at time $t$. The set of borrower $i$'s direct friends (alters) at time $t$, known as her ego-network is defined as $g_t^1 = \{ij \mid ij \in G_t\}$. For a particular borrower $i$, we consider her hidden credit quality process $X_t$ and observed behavioral score $Y_t$.

### Lender's information

The lender observes in continuous time the process $Y_t$ denoting the borrower's behavior. Further, at discrete fixed times $0 = t_0 < t_1 < \cdots < t_{N-1} < T, N \in \mathbb{N}$ the lender observes the borrower's ego network and receives signals from her and her alters. Let the vector $\mathbf{Z}_k = \{Z_{ik}, Z_{jk} \mid ij \in g_t^1\}$ denote the ego-network signals received by the lender at times $t_k, k = 0, 1, \cdots, N$, comprising the borrower's own signal $Z_{ik} = X_{it_k} + \sqrt{\Lambda}_k \varepsilon_{ik}$ and the signals from her alters $Z_{jk} = X_{jt_k} + \sqrt{\Lambda}_k \varepsilon_{jk}$. The variables $\varepsilon_{jk} \sim \mathcal{N}(0,1)$ are i.i.d across individuals with $\mathbb{E}(\varepsilon_{js}\varepsilon_{lu}) = 0$ for $j \neq l$, $s \neq u$. Thus the lender receives noisy but unbiased signals upon observing the borrower's ego-network at time $t_k$.

The information available to the lender can thus be represented by the following filtrations

$$\mathbb{F}^Y = \left(\mathcal{F}_t^Y\right)_{t \in [0,T]} \text{ with } \mathcal{F}_t^Y \text{ generated by } \{Y_s, s \leq t\}$$

$$\mathbb{F}^Z = \left(\mathcal{F}_t^Z\right)_{t \in [0,T]} \text{ with } \mathcal{F}_t^Z \text{ generated by } \{Y_s, s \leq t, Z_k, t_k \leq t\}$$

$$\mathbb{F}^O = \left(\mathcal{F}_t^O\right)_{t \in [0,T]} \text{ with } \mathcal{F}_t^O \text{ generated by } \{Z_k, t_k \leq t\}$$

$\mathbb{F}^Y$ corresponds to the continuous time behavioral information only, $\mathbb{F}^O$ consists of the ego-network signals received at discrete times whilst $\mathbb{F}^Z$ is the combination of behavioral information and the ego-network signals. We assume that the $\sigma$-algebras $\mathcal{F}_t^Y$ and $\mathcal{F}_t^Z$ are augmented with the $\mathbb{P}$ null sets. Note that for each $t \geq 0$, $\mathcal{F}_t^Z = \mathcal{F}_t^Y \vee \mathcal{F}_t^O$.

## 3. Stochastic Filtering

The focus of stochastic filtering is to estimate the hidden stochastic process $X_t$ based on observations up to time $t$. Let $\hat{X}_t^H$ be the projection of the process $X_t$ onto the observed filtration $\mathbb{F}^H, H \in \{O, Y, Z\}$ *i.e.* $\hat{X}_t^H = \mathbb{E}\left(X_t \mid \mathcal{F}_t^H\right)$. In this section, we derive explicit results for the filtering equations and the conditional variance of the hidden process $X_t$.

### Behavioral Observations

When the lender's observation $\sigma$-algebra is $\mathcal{F}_t^Y$, *i.e* when the lender does not receive any ego-network signals, we are in the realm of the classical Kalman-Bucy filter, see e.g. [18] and [19]. This is since the state and observation equations constitute a linear Gaussian state space model. Let $\hat{X}_t^Y = \mathbb{E}\left(X_t \mid \mathcal{F}_t^Y\right)$ and $\lambda_t^Y = \mathbb{E}\left[\left(X_t - \hat{X}_t^Y\right)^2 \mid \mathcal{F}_t^Y\right]$ be the conditional mean and variance respec-

tively in the $\sigma$-algebra $\mathcal{F}_t^Y$.

The dynamics of $\hat{X}_t^Y$ is given by the following SDE

$$d\hat{X}_t^Y = \left( \mu\left(\delta - \hat{X}_t^Y\right) - \alpha^2\sigma^{-2}\lambda_t^Y \hat{X}_t^Y \right)dt + \alpha\sigma^{-2}\lambda_t^Y dY_t, \hat{X}_0^Y = m_0 \tag{7}$$

whilst the dynamics of $\lambda_t^Y$ is given by the deterministic ODE

$$\frac{d\lambda_t^Y}{dt} = -2\mu\lambda_t^Y + \gamma^2 - \alpha^2\sigma^{-2}\left(\lambda_t^Y\right)^2, \lambda_0^Y = v_0 \tag{8}$$

Equation (8) is the well known Riccati equation, a deterministic equation whose unique solution is given as

$$\lambda_t^Y = \frac{-\mu\sigma^2}{\alpha^2} + C_0 \frac{C_1 + C_2 e^{-2\frac{\alpha^2}{\sigma^2}C_0 t}}{C_1 - C_2 e^{-2\frac{\alpha^2}{\sigma^2}C_0 t}} \tag{9}$$

given that the initial value is $\lambda_0^Y = v_0$. In Equation (9), $C_0 = \frac{\sigma}{\alpha}\sqrt{\frac{\mu^2\sigma^2}{\alpha^2} + \gamma^2}$,

$C_1 = v_0 + C_0 + \frac{\mu\sigma^2}{\alpha^2}$ and $C_2 = v_0 - C_0 + \frac{\mu\sigma^2}{\alpha^2}$ (see e.g. [20]).

**Behavioral Observations and Network Information**

This is the case of most interest in the study. The lender's observation $\sigma$-algebra is $\mathcal{F}_t^Z = \mathcal{F}_t^Y \vee \mathcal{F}_t^O$ being the augmentation of $\mathcal{F}_t^Y$ with discrete time ego-network signals. Since the lender's observation of the network is restricted to borrower $i$'s ego network, at each time $t \geq 0$, with no other additional borrower information, an individual's credit quality is assumed to have the distribution $X_{jt} \sim \mathcal{N}\left(0, q^{-1}\right)$. The lender uses the assumed density for all other individuals $X_{jt}$ who are alters to borrower $X_{it}$. The following lemma gives the expected degree (number of direct friends) conditional on the borrower's true credit type.

**Lemma 1**

*At each time $t \geq 0$, conditional on the meeting process and the borrower's credit quality $X_t$, the expected number of friends $\mathbb{E}\left(\eta_t \mid X_t\right)$ is given as*

$$\frac{1}{\mathcal{Z}-1}\sqrt{\frac{q}{q+1}}e^{-\frac{q}{2(q+1)}X_t^2}$$

*Proof.*

Conditioned on the borrower's true credit type $X_t$ and the meeting process, the probability of having a network tie with any other individual is

$$\int_{-\infty}^{\infty} e^{-\frac{(X_t - s)^2}{2}}\sqrt{\frac{q}{2\pi}}e^{-q\frac{s^2}{2}}ds = \sqrt{\frac{q}{q+1}}e^{-\frac{q}{2(q+1)}X_t^2} \tag{10}$$

Thus the conditional expected number of friends $\mathbb{E}\left(\eta_t \mid X_t\right)$ is given by

$$\mathbb{E}\left(\eta_t \mid X_t\right) = \mathcal{Z}v\sqrt{\frac{q}{q+1}}e^{-\frac{q}{2(q+1)}X_t^2} = \frac{1}{\mathcal{Z}-1}\sqrt{\frac{q}{q+1}}e^{-\frac{q}{2(q+1)}X_t^2} \tag{11}$$

**Proposition 1.**

*For any $k = 0, 1, \cdots, N-1$ and $t > 0$, let $p_k = \Lambda_k^{-1}$ denote the precision of*

the ego-network signals at time $t_k$. Further define the variable

$$\theta_k^Z = \frac{(p_k+q+1)}{\lambda_{tk-}^Z \left( p_k(p_k+q+1)+\eta_k(p_k+q)\right)+(p_k+q+1)} \tag{12}$$

Then it holds that:

1) For any $t \in [t_k, t_{k+1})$, the filtered estimate $\hat{X}_t^Z$ is Gaussian with the dynamics

$$d\hat{X}_t^Z = \left( \mu\left(\delta-\hat{X}_t^Z\right)-\alpha^2\sigma^{-2}\lambda_t^Z\hat{X}_t^Z \right)dt+\alpha\sigma^{-2}\lambda_t^Z dY_t \tag{13}$$

whilst the equation of the conditional variance is given as

$$\lambda_t^Z = \frac{-\mu\sigma^2}{\alpha^2}+C_0\frac{C_{1k}+C_{2k}e^{-2\frac{\alpha^2}{\sigma^2}C_0 t}}{C_{1k}-C_{2k}e^{-2\frac{\alpha^2}{\sigma^2}C_0 t}} \tag{14}$$

with initial values $\hat{X}_{tk}^Z$ and $\lambda_{tk}^Z$. $C_0$ is same as in Equation (9) whilst $C_{1k}=\lambda_{tk}^Z+C_0+\frac{\mu\sigma^2}{\alpha^2}$ and $C_{2k}=\lambda_{tk}^Z-C_0+\frac{\mu\sigma^2}{\alpha^2}$.

2) At information date $t_k$, $\hat{X}_{t,k}^Z$ is Gaussian. The mean $\hat{X}_{t,k}^Z$ and variance $\lambda_{t,k}^Z$ are updated from their respective values before the arrival of ego-network signals $t_{k-}$ to

$$\hat{X}_{t,k}^Z = \theta_k^Z\lambda_{tk-}^Z\left( \frac{\hat{X}_{t,k-}^Z}{\lambda_{tk-}^Z}+p_k Z_{ik}+\frac{p_k}{p_k+q+1}\sum_{j\in Z}Z_{jk} \right) \tag{15}$$

and variance

$$\theta_k^Z\lambda_{tk-}^Z = \frac{\lambda_{tk-}^Z(p_k+q+1)}{\lambda_{tk-}^Z\left( p_k(p_k+q+1)+\eta_k(p_k+q)\right)+(p_k+q+1)} \tag{16}$$

Proof.

1) Between two information dates, $t \in [t_k, t_{k+1})$, there is no new arrival of ego-network signals. The lender's $\sigma$-algebra is defined as $\mathcal{F}_t^Z = \mathcal{F}_{t,k}^Z \vee \sigma\{Y_s, t_k < s \le t\}$. Thus we revert to the classical Kalman-Bucy filtering situation with the respective initial values for the conditional mean and variances given as $\hat{X}_{tk}^Z$ and $\lambda_{tk}^Z$. For this case, the formulations for conditional mean and variance follow closely from Equations (7) and (8).

2) On the information arrival date $t_k$, the lender receives ego-network signals $\boldsymbol{Z}_k$ and gets to update the conditional mean and variance of the filtered estimate. To incorporate the ego-network signals into the estimate, Bayesian updating is carried out since there is no time evolution from $t_{k-}$ to $t$. At time $t_{k-}$, the conditional prior distribution of $X_t$ is Gaussian and the signals received are also Gaussian. The posterior probability of the borrower's credit type $\mathbb{P}\left(X_{t_k}\mid\boldsymbol{Z}_k\right)$ is obtained by

$$\mathbb{P}\left(X_{t_k}\mid\boldsymbol{Z}_k\right) \propto \mathbb{P}\left(X_{t_k},\boldsymbol{Z}_k\right)$$

$$= \int_{-\infty}^{\infty}\mathbb{P}\left(\boldsymbol{Z}_k\mid X_{t_k},X_{j,t_k}\right)\mathbb{P}\left(X_{t_k}\right)\mathbb{P}\left(X_{j,t_k}\mid X_{t_k}\right)dX_{j,t_k}$$

$$= \int_{-\infty}^{\infty}\mathbb{P}\left(\boldsymbol{Z}_k\mid X_{t_k},X_{j,t_k}\right)\mathbb{P}\left(X_{t_k}\right)\mathbb{P}\left(X_{j,t_k}\right)dX_{j,t_k}$$

The last equality is as a result of the assumption of independence for the $X_{jtk}$.

We have $\mathbb{P}\left(X_{j,t_k}\right) \propto \prod_{ij \in g_k^1} e^{-\frac{q}{2}X_{j,t_k}^2}$ being the assumed density of any individual $X_j$ for $j \neq i$. Thus the integrand is given by

$$
\begin{aligned}
&\mathbb{P}\left(\boldsymbol{Z}_k \mid X_{t_k}, X_{j,t_k}\right)\mathbb{P}\left(X_{t_k}\right)\mathbb{P}\left(X_{jt_k}\right) \\
&= e^{-\frac{\left(X_{tk}-X_{tk-}^Z\right)^2}{2\lambda_{tk-}^Z}} \times e^{-\frac{p_k}{2}\left(Z_{ik}-X_{tk}\right)^2} \\
&\quad \times \underbrace{\prod_{ij \in g_k^1} e^{-\frac{q}{2}X_{j,t_k}^2} \times \prod_{j \in Z_k} e^{-\frac{p_k}{2}\left(Z_{jk}-X_{jtk}\right)^2} \times \prod_{ij \in g_k^1} v e^{-\frac{1}{2}\left(X_{tk}-X_{jtk}\right)^2}}_{(a)} \\
&\quad \times \underbrace{\prod_{ij \notin g_k^1}\left(1 - \frac{\mathbb{E}\left(\eta_{tk} \mid X_{tk}\right)}{\mathcal{Z}}\right)}_{(b)}
\end{aligned}
\tag{17}
$$

where

- The first term denotes the product of the conditional prior density (before the arrival of network information) and the likelihood function for the observation $Z_{it,k}$.
- (a) denotes the assumed prior density for $X_{jt}$ for $ij \in g_{t_k}^1$ times the probability that at time $t_k$ borrower $i$ is friends with the individuals $j$, $ij \in g_{t_k}^1$ within her ego-network whose signals are in $\boldsymbol{Z}_k$ and that these friends have the signals as collected in $\boldsymbol{Z}_k$
- (b) denotes the probability that at time $t_k$ borrower $i$ is not friends with anyone outside $g_{t_k}^1$.

As $\mathcal{Z} \to \infty$, $v \to 0$, then by the monotone convergence theorem and applying lemma 1 we have

$$
\prod_{ij \notin g_k^1}\left(1 - \frac{\mathbb{E}\left(\eta_{tk} \mid X_{tk}\right)}{\mathcal{Z}}\right) \to 1
\tag{18}
$$

Hence,

$$
\begin{aligned}
\mathbb{P}\left(X_{t_k} \mid \boldsymbol{Z}_k\right) \propto \int_{-\infty}^{\infty} &e^{-\frac{\left(X_{tk}-X_{tk-}^Z\right)^2}{2\lambda_{tk-}^Z}} \times e^{-\frac{p_k}{2}\left(Z_{ik}-X_{tk}\right)^2} \times \prod_{ij \in g_k^1} e^{-\frac{q}{2}X_{j,t_k}^2} \\
&\times \prod_{j \in Z_k} e^{-\frac{p_k}{2}\left(Z_{jk}-X_{jtk}\right)^2} \times \prod_{ij \in g_k^1} e^{-\frac{1}{2}\left(X_{tk}-X_{jtk}\right)^2} \mathrm{d}X_{j,t_k}
\end{aligned}
\tag{19}
$$

whereby the integrand is a product of Gaussian densities. Upon integrating out $X_{jtk}$, and matching the terms of $X_{tk}$ and $X_{tk}^2$ we obtain the posterior distribution as Gaussian with the given expectation and variance. The modeling of the network tie probability $\pi_{ij}(t)$ as a probit link function enables the elegant formulation of the posterior probability as a Gaussian.

Lastly, we consider a unique case, whereby the lender does not observe the continuous time information but only receives the discrete time ego-network signals *i.e.* when $H = O$. Thus between network information times $t \in [t_k, t_{k+1})$,

the lender receives no information. We thus have the following corollary.

**Corollary 1.**

*When the lender's information set is $\mathcal{F}_t^O$ we have*

- *For $t \in [t_k, t_{k+1})$ between information arrival times, the respective conditional mean and variance are given by the equations.*

$$\hat{X}_t^O = \delta + \mathrm{e}^{-\mu(t-t_k)}\left(\hat{X}_{tk}^O - \delta\right) \tag{20}$$

$$\lambda_t^O = \mathrm{e}^{-2\mu(t-t_k)}\lambda_{t_k}^O + \frac{\gamma^2}{2\mu}\left(1 - \mathrm{e}^{-2\mu(t-t_k)}\right) \tag{21}$$

- *At information date $t_k$, it holds that $X_t^O$ is Gaussian with mean and variance*

$$\hat{X}_{t,k}^0 = \theta_k^0 \lambda_{tk-}^0 \left(\frac{\hat{X}_{t,k-}^0}{\lambda_{tk-}^0} + p_k Z_{ik} + \frac{p_k}{p_k + q + 1}\sum_{j \in Z} Z_{jk}\right) \tag{22}$$

$$\theta_k^0 \lambda_{tk-}^0 = \frac{\lambda_{tk-}^O\left(p_k + q + 1\right)}{\lambda_{tk-}^O\left(p_k\left(p_k + q + 1\right) + \eta_k\left(p_k + q\right)\right) + \left(p_k + q + 1\right)} \tag{23}$$

*respectively.*

*Proof.*

For $t \in [t_k, t_{k+1})$, the proof can be found in corollary 4 of [2]. At information date $t_k$, the conditional prior distribution at time $t_{k-}$ is updated using the ego-network signals received from the vector $\mathbf{Z}_k$. Following in a similar fashion to proposition 1 (part (2)), the conditional prior density is updated to a Gaussian posterior density with the given mean and variance.

## 4. Properties of the Conditional Variance

We study the properties of the conditional variance under the various information settings discussed in Section 3. We show that inclusion of ego-network signals leads to better estimates of the hidden process $X_t$. A key result within this section is proposition 2 where we show that increasing the frequency of network information arrival times leads to the full information case in the limit as $N \to \infty$ The following lemma shows that the ego-network signals improves the lender's estimate of the credit quality.

**Lemma 2.**

For $H \in \{Y, O\}$, $t \in [0, T]$ and $k = 0, 1, \cdots, N-1$

$$\lambda_t^Z \leq \lambda_t^H \tag{24}$$

*Proof.*

The proof similar to proposition 6 of [2], where a detailed proof is available.

The following proposition shows that as we increase the frequency of arrivals of ego-network information *i.e.* as $N \to \infty$ then the variances $\lambda_t^Z, \lambda_t^O$ tends to zero. It is an adaptation of the asymptotic result of [14].

**Proposition 2.**

*Let $\left\{\pi^N = \left(t_k^N\right)_{k=0}^{N-1}\right\}_{N \geq 1}$ be a refining sequence of partitions of the interval*

$[0,T]$ such that information dates are retained i.e. for $N \leq N'$ then $\left(\pi^N\right) \subset \left(\pi^{N'}\right)$. Let $\Delta_N = \max_{k=1,\cdots,N}\left\{t_k^N - t_{k-1}^N\right\}$ be the mesh size. Further, let $\left(\Lambda_k^N\right)_{k=0,\cdots,N-1}$ be a sequence of corresponding variances at information times $t_k^N$. Assume that there exists a constant $\overline{\Lambda} > 0$ such that $\Lambda_k^N \leq \overline{\Lambda}$ for all $k = 0,1,\cdots,N-1$ and all $N \in \mathbb{N}$. Then it holds that for all $t \in (0,T]$, the conditional variances $\lambda_t^{ON}$ and $\lambda_t^{ZN}$ tend to 0 as $N \to \infty$ and $\Delta_N \to 0$.

Proof.

From lemma 2, since $0 \leq \lambda_t^Z \leq \lambda_t^0$ we need only prove the assertion for $\lambda_t^{O,N}$. Further, we can assume ego-network information with constant variances i.e. $\Lambda_k = \overline{\Lambda}$ for all $k = 0,1,\cdots,N-1$. This assumption generalizes the proof even for the case where $\Lambda_k < \overline{\Lambda}$. For ease of notation we write $t_k$ instead of $t_k^{(N)}$ noting the dependence on $N$. For any $k = 0,1,\cdots,N-1$ and any $t \in [t_k, t_{k+1})$ we know that $\lambda_t^{0,N}$ is given by

$$\lambda_t^{0,N} = e^{-2\mu(t-t_k)}\lambda_{t_k}^O + \frac{\gamma^2}{2\mu}\left(1 - e^{-2\mu(t-t_k)}\right) \tag{25}$$

where $\lambda_{t_k}^O = \theta_k^0 \lambda_{t_k-}^O$ with

$$\theta_k^0 = \frac{(p_k + q + 1)}{\lambda_{tk-}^O\left(p_k\left(p_k + q + 1\right) + \eta_k\left(p_k + q\right)\right) + \left(p_k + q + 1\right)} \tag{26}$$

Since $\left(1 - e^{-2\mu(t-t_k)}\right) \leq 2\mu(t - t_k) \leq 2\mu\Delta_N$ it follows that

$$\lambda_t^{0,N} \leq \theta_k^0 \lambda_{tk-}^0 + \gamma^2\Delta_N \tag{27}$$

We iterate this inequality for all $l \leq k$ and denote $\overline{\theta}_k^0 = \max_{l \in (0,1,\cdots,k)}\left\{\theta_l^0\right\}$. This yields for $\lambda_{0-}^N = v_0$ and $t \in [t_k, t_{k+1})$

$$\lambda_t^0 \leq v_0\left(\overline{\theta}_k^0\right)^{k+1} + \gamma^2\Delta_N\sum_{l=0}^{k}\left(\overline{\theta}_k^0\right)^l \leq v_0\left(\overline{\theta}_k^0\right)^{k+1} + \frac{\gamma^2\Delta_N}{1 - \overline{\theta}_k^0} \tag{28}$$

Let $u \in (0,T]$, $\varepsilon > 0$ and $\overline{p} = \overline{\Lambda}^{-1}$. We desire to show that $\lambda_u^0 < \varepsilon$ for a suitably chosen $N$. Define $k_N$ to be the index for which $u \in [t_{kN}, t_{kN+1})$. Suppose that for all $N_0$ there exists a $N \geq N_0$ such that

$$\min\left\{\lambda_{t_0-}^{0,N}, \lambda_{t_1-}^{0,N}, \cdots, \lambda_{t_{k_N-}}^{0,N}\right\} \geq \varepsilon/2 \tag{29}$$

Then we have that

$$\overline{\theta}_k^{0,N} \leq \frac{1}{\min\left\{\lambda_{tk-}^{O,N}\right\}\left(p_k\left(p_k + 1\right)\right) + 1} \leq \frac{2}{\overline{p}\varepsilon + 2}$$

where the minimum is over all $\lambda_{tk-}^{O,N}$. This yields (with one iteration less)

$$\lambda_u^{0,N} \leq v_0\left(\frac{2}{\varepsilon\overline{p} + 2}\right)^{k_N} + \gamma^2\Delta_N\left(\frac{\varepsilon\overline{p} + 2}{\varepsilon\overline{p}}\right) \tag{30}$$

Note that in this case $\overline{\theta}_k^{0,N} < 1$ and $k_N \to \infty$ as $N \to \infty$. Thus for all $N \geq N_0$, $\lambda_u^{0,N} \to 0$. Thus we can choose $N_0$ such that $\lambda_u^{0,N} < \varepsilon/2$ which is a contradiction of the assumption in Equation (29).

Thus there exists a $N_0$ such that for all $N \geq N_0$ there exists an index set $j_N \leq k_N$ with $\lambda_{t_{jN-}}^{0,N} < \varepsilon/2$. For each such $N$ we choose $t_{jN-} \leq t_{k_N-}$ as the last

information arrival time before $t_{k_N-}$ such that $\lambda_{t_{j_N-}} < \varepsilon/2$. In the case that $j_N = k_N$, then for a suitably large $N$ from Equation (27) implies that $\lambda_u^0 \le \varepsilon$.

For the case when $j_N < k_N$ for $k = j_N+1, \cdots, k_N$ we have that $\lambda_{t_{k_N-}}^{0,N} \le \varepsilon/2$ and $\overline{\theta}_k^{0,N} \le \dfrac{2}{\overline{p}\varepsilon+2}$. We can choose a suitable $N_1 \ge N_0$ such that

$$\gamma^2 \Delta_N \left( \frac{\varepsilon \overline{p} + 2}{\varepsilon \overline{p}} \right) < \varepsilon/2.$$ An iteration similar to Equation (28) starting from $j_N$

with initial value $\lambda_{t_{j_N-}}^{0,N}$ yields that

$$\lambda_u^{0,N} < \varepsilon$$

for all $N \ge N_1$ as desired.

## 5. Numerical Results

In this section we provide a brief illustration of our findings on the properties of the conditional variance. We assume that the ego-network signals $Z_k$ arrive at equidistant time points $t_k$, $k = 0, 1, \cdots, N-1$. We simulate the processes $X_t, Y_t$ and $Z_k$ using the parameter values in **Table 1** with $q = 0$.

To illustrate the impact of the number of alters on the borrower's conditional variances $\lambda_t^O$ and $\lambda_t^Z$, **Figure 1** plots a comparison between the variances of $\lambda_t^O$ and $\lambda_t^Z$ with and with no friends. The left panel plots a comparison of the variances for the case when the number of friends is constant at zero and five respectively. In the right panel, the number of friends is modeled as a Poisson random variable with parameter $\rho = 5$. In both plots, the conditional variances for the case when there exists friends' data in $Z_k$ is lower as compared to the case with zero friends. Besides, the perturbations in the conditional variance with randomly varying number of friends are well depicted in panel 2 of the plot (see **Figure 1**).
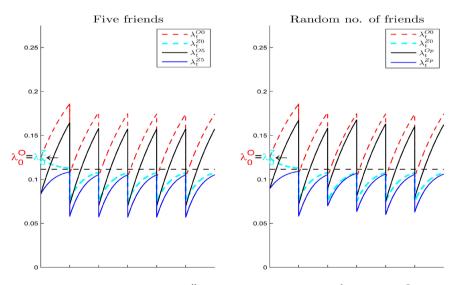


**Figure 1.** Left: plot of variances $\lambda_t^H$ when no. of friends $\eta_t = 0$ and $\eta_t = 5$ Right: plot of variances $\lambda_t^H$ when no. of friends $\eta_t = 0$ and $\eta_t = \rho$.

**Table 1.** Model parameter values.

| | | | |
|---|---|---|---|
| Mean reversion speed $\delta$ | 0.05 | Drift $\alpha$ | 2 |
| Return's volatility $\sigma$ | 0.3 | Volatility $\gamma$ | 1 |
| Mean reversion level $\mu$ | 2.0 | Network variance $\Lambda$ | 0.04 |

## 6. Conclusions

In this article, we have presented stochastic filtering results whereby the hidden credit quality process, modeled as an Ornstein-Ulehnbeck equation drives the drift process of the borrower's observed behavior score. We have formulated a latent space network model such that the ego-network signals received at discrete fixed times are incorporated into the credit quality filtering by way of Bayesian updating. Modeling of network tie probability using the probit link function enabled the elegant formulation of the conditional posterior density of the hidden process. We have presented explicit results for the conditional mean and conditional variance under the various information setups. Further, we have presented asymptotic properties of the conditional variance when the frequency of the network information arrival times is increased.

The results in this article thus present a theoretical justification of including ego-network data in credit scoring, for a network model based on credit type homophily. Future studies may consider network models whereby the network ties capture the strength or frequency of interaction between the nodes, instead of binary network ties.

## Acknowledgements

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] McPherson, M., Smith-Lovin, L. and Cook, J.M. (2001) Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, **27**, 415-444.
https://doi.org/10.1146/annurev.soc.27.1.415

[2] Sewe, S., Ngare, P. and Weke, P. (2019) Dynamic Credit Quality Evaluation with Social Network Data *Journal of Applied Mathematics*, **2019**, Article ID: 8350464.
https://doi.org/10.1155/2019/8350464

[3] Snijders, T., Steglich, C. and Schweinberger, M. (2017) Modeling the Coevolution of Networks and Behavior. In: van Montfort, K., Oud, J. and Satorra, A., Eds., *Longitudinal Models in the Behavioral and Related Sciences*, Routledge, London, 41-71.
https://doi.org/10.4324/9781315091655-3

[4] Hoff, P.D., Raftery, A.E. and Handcock, M.S. (2002) Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, **97**, 1090-1098. https://doi.org/10.1198/016214502388618906

[5] Sewell, D.K. and Chen, Y. (2015) Latent Space Models for Dynamic Networks. *Journal of the American Statistical Association*, **110**, 1646-1657. https://doi.org/10.1080/01621459.2014.988214

[6] Kim, B., Lee, K., Xue, L. and Niu, X. (2017) A Review of Dynamic Network Models with Latent Variables. ArXiv: 1711.10421.

[7] De Andrade, F.W.M. and Thomas, L.C. (2007) Structural Models in Consumer Credit. *European Journal of Operational Research*, **183**, 1569-1581. https://doi.org/10.1016/j.ejor.2006.07.049

[8] Crook, J.N., Edelman, D.B. and Thomas, L.C. (2007) Recent Developments in Consumer Credit Risk Assessment. *European Journal of Operational Research*, **183**, 1447-1465. https://doi.org/10.1016/j.ejor.2006.09.100

[9] Oguz, H.T. and Gurgen, F.S. (2008) Credit Risk Analysis Using Hidden Markov Model. *2008 23rd International Symposium on Computer and Information Sciences*, Istanbul, 27-29 October 2008, 1-5. https://doi.org/10.1109/ISCIS.2008.4717932

[10] Malik, M. and Thomas, L.C. (2012) Transition Matrix Models of Consumer Credit Ratings. *International Journal of Forecasting*, **28**, 261-272. https://doi.org/10.1016/j.ijforecast.2011.01.007

[11] Bundi, D.N. (2016) Social Network Analysis for Credit Risk Modeling. Unpublished Doctoral Dissertation, University of Nairobi, Kenya.

[12] Wei, Y., Yildirim, P., Van den Bulte, C. and Dellarocas, C. (2015) Credit Scoring with Social Network Data. *Marketing Science*, **35**, 201-340. https://doi.org/10.1287/mksc.2015.0949

[13] Francis, E., Blumenstock, J. and Robinson, J. (2017) Digital Credit: A Snapshot of the Current Landscape and Open Research Questions. CEGA White Paper.

[14] Gabih, A., Kondakji, H., Sass, J. and Wunderlich, R. (2014) Expert Opinions and Logarithmic Utility Maximization in a Market with Gaussian Drift. *Communications on Stochastic Analysis*, **8**, No. 1. https://doi.org/10.31390/cosa.8.1.03

[15] Black, F. and Litterman, R. (1992) Global Portfolio Optimization. *Financial Analysts Journal*, **48**, 28-43. https://doi.org/10.2469/faj.v48.n5.28

[16] Danilova, A., Monoyios, M. and Ng, A. (2010) Optimal Investment with Inside Information and Parameter Uncertainty. *Mathematics and Financial Economics*, **3**, 13-38. https://doi.org/10.1007/s11579-010-0025-y

[17] Elliott, R.J., Aggoun, L. and Moore, J.B. (2008) Hidden Markov Models: Estimation and Control. Springer Science & Business Media, Berlin, Heidelberg.

[18] Cohen, S.N. and Elliott, R.J. (2015) Stochastic Calculus and Applications. Springer, Switzerland. https://doi.org/10.1007/978-1-4939-2867-5

[19] Bain, A. and Crisan, D. (2009) Fundamentals of Stochastic Filtering. Springer, Switzerland. https://doi.org/10.1007/978-0-387-76896-0

[20] Lakner, P. (1998) Optimal Trading Strategy for an Investor: The Case of Partial Information. *Stochastic Processes and Their Applications*, **76**, 77-97. https://doi.org/10.1016/S0304-4149(98)00032-5