Norwegian School of Economics

Bergen, Fall 2016

NHH

Investing in Peer-to-Peer Lending: Risk and Return

An empirical analysis of the risk-return relationship in the crowdlending market

Peder Solberg Hjermann

Espen Dyve Jørgensen

Supervisor: Einar Bakke

Master thesis in Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

This paper applies credit-risk pricing theory from literature on bonds to price loans in the peer-to-peer (P2P) lending market. The purpose is to study the risk-adjusted return for investors investing in P2P loans. Historical data from the loan book of the P2P lending platform Lending Club, are used to attain estimates for risk in the P2P sector. These estimates are put into an extended version of the CIR univariate (ECIR) model, which in turn is used to price the loans issued at the Lending Club platform in 2015.

The main finding is that all of the the interest rates set by Lending Club are higher than the theoretical interest rates from the ECIR model. This means that Lending Club compensates investors more than the ECIR model suggests, given the risk they have taken on. Furthermore, the spread between the two increases as the riskiness of the loan increases. We have considered plausible explanations for the difference in interest rates, and find that it seems to be connected to perception of risk and market inefficiency. Still, there are reasons to believe that the difference in real and modelled interest rates indicates the excistence of excess returns for inverstors in the P2P lending market. However, further research on P2P lending will be nescesary to fully understand the risk-return relationship for the investors.

Preface

This thesis represents the completion of our Master of Science in Financial Economics and Master of Science in Business Analysis and Performance Management, at the Norwegian School of Economics (NHH).

Our interest in financial technology (FinTech) has increased with the insight obtained from finance courses at NHH, combined with the industry's increased focus on FinTech. Thus, we decided that we wanted to base our thesis on this topic.

When working on which aspect of FinTech to write about, our most important criteria was that it should be relevant, quantifiable and somewhat unchartered when it comes to academic research. After reading a number of papers, news articles and blogs on FinTech, we went on to discuss with some experts on the area and landed on the investors return from P2P lending as our topic.

In the process of specifying our topic, we got the opportunity to discuss - and receive helpful insight - from Chief Digital Officer in Skandiabanken, Christoffer Hærnes. He was the one who pointed us in the direction of P2P lending as a relevant and interesting topic for a master thesis. Further, we contacted Sebastian Harung, CEO and co-founder of the Swedish P2P lending platform Kameo, to discuss which aspect of P2P lending we should explore. We want to give special thanks to both Mr. Hærnes and Mr. Harung for inspiring us to write about P2P lending.

Lastly, we would like to thank our supervisor Einar Bakke for the support throughout the process, and for providing useful guidance and feedback.

Sincerely,

Peder Hjermann and Espen Jørgensen

Bergen, December 2016

Table of Contents

1. Introduction

As several industries experience the evolution of information technologies, the roles of the market participants' changes. According to Allen and Santomero (2001), this trend has been particularly evident when it comes to the intermediaries in financial markets. As pointed out by Berger and Gleisner (2009), the emergence of new competitors such as peer-to-peer (P2P) lending platforms, has created uncertainty about the earlier undisputed future of traditional intermediaries in the financial sector. Furthermore, Sarkar, Butler and Steinfield (2006) states that the development of electronic marketplaces seems to have made the markets more efficient. It follows that P2P lending, as a recent example of these electronic marketplaces, has the potential to increase the efficiency for both borrowers and lenders in the lending market.

We will study the investors' - by which we mean the lenders' - returns when investing in P2P loans. Further, we will apply well-tested and established credit-risk pricing theory on this relatively new intermediary. In this way, we want to contribute to the literature regarding the investment perspective of P2P lending.

As P2P lending is quite new, we will start by introducing the concept. P2P lending is also called crowdlending or loan-based crowdfunding. It is an open call, essentially through the internet, for the provision of financial resources in form of a loan, where the total loan amount is provided through the sum of several individual contributions. Investors lend borrowers money for one to five years in exchange for a cash flow of monthly interest payments and/or instalments. Details regarding the processes and agreements may vary between platforms and loans.

The P2P lending process starts when potential borrowers apply for a loan through a platform such as Lending Club, whereupon Lending Club starts the screening of the applicant. If the screening process leads to an approval of the application, Lending Club assigns the loan a credit rating between A and G. The given rating reflects the risk, where borrowers receiving an A is the group with the lowest risk. Before offering the loan to its investors, the platform decides on the interest rate that investors will receive. Note that Lending Club acts only as a facilitator matching borrowers and investors, and that the loans issued are never on the platform's balance sheet. Investors can invest in a fraction of a loan (called a note) and, by spreading their investment over several different notes, diversify their portfolio and reduce the severity in case of default. As the borrowers repays the loan according to the loan agreement, investors will receive the interest/instalments directly to their accounts. Investors can choose to manually purchase notes, or to use the platform's Automated Investing System which automatically purchases available notes that meet the investors pre-set criteria (Lendingclub.com, 2016). As the concept of P2P lending is becoming clearer, we will study what the online platforms brings to borrowers and investors in terms of value.

Even though our focus is investors' returns, it is also relevant to comment on the borrowers' perspective. For the borrowers, Lending Club offers an easy online application, low fixed interest rates and annuity payments. As the SMEs (small and medium sized enterprises) traditionally has struggled to access financing, the P2P lending market represents a new opportunity for these businesses (Wehinger, 2014). It seems intuitive that easier access to funding should gain the SMEs, but what about those offering the financing, the investors?

Lending Club projects that investors will have annual returns between 5 and 8 %. The platform also claims that 99.8% of the investors investing in 100 loans or more, earn positive returns. Other selling points are the monthly cash flow where investors receive 3 to 6% of their total investment back, the simplicity of the process and the fact that it is easy to diversify across a large number of loans. Nevertheless, the most important factor for an investor is likely to be the return, given by the interest rate on the P2P loan. Academic research has attempted to study whether these interest rate are sufficient to compensate the investors for the risk they are taking on, but there is still significant uncertainty related to the positive and negative sides of investing in P2P loans (Roig Hernando, 2016).

In this paper, we will not attempt to address all the different aspects related to investing in P2P lending, but we will focus on applying theory on credit-risk pricing to P2P loans. Through this, we hope to be able to comment on the risk-adjusted return when investing in P2P loans.

2. Theory

In the following section, we will review theory and literature relevant for pricing P2P loans.

2.1 Credit-Risk Price Modelling

This section reviews theory on credit-risk pricing, with the objective of choosing the best fit model to the P2P lending market.

There are mainly two types of credit pricing models. The first is the structural approach, which uses firm-specific information and prices corporate debt through option pricing theory as presented by Black and Scholes (1973) and Merton (1974). As Lando (2004) pointed out, these papers were the first to provide a connection between economic pricing theory and statistical default models. The structural approach offers economic intuition, but is often difficult to implement because of the many firm specific variables the model requires. The second approach does not require knowledge on firm-specific variables, instead the estimates are based on exogenous variables which makes it easier to implement on real data. This approach is known as reduced form models.

2.1.1 Choice of Approach

The academic environment differs when it comes to the preferred approach. The main difference is whether we estimate default as a result of firm specific factors, or estimate default as an exogenously determined variable. Supporters of reduced form models argue that we rarely experience perfect information, and that the most realistic way of modelling the real markets are by applying the reduced form approach (Jarrow & Protter, 2004). This argument is relevant when applying credit risk models to the P2P lending market. In P2P business loans the lenders are often unlisted SMEs, with little or none financial information available. It is not a market where we can expect to find the necessary information about each individual borrower required to successfully implement the structural approach. In addition, it can be difficult to implement the structural approach as the issuer's assets and liabilities might not be traded in financial markets, which in turn makes the estimation of relevant parameters difficult. The structural approach also requires that corporate debt is valued at the same time as every other senior liability of the firm, which makes calculations demanding. Further, Jarrow and Protter (2004) argue that the choice of approach should be determined by the purpose of the model, and that the reduced form approach is the correct one for pricing, as prices are determined by the market based on the information available.

The reasoning above states that the choice of approach should be based on the information known to the modeller. In addition, we know that default can far from always be explained by comparing assets to liabilities (Duffie and Singleton, 1999). Thus, we will apply a reduced form approach to find the theoretical price of the P2P loans issued by Lending Club.

2.1.2 Reduced Form Models

After the introduction by Jarrow and Turnbull in 1992, reduced form models have been studied by Jarrow and Turnbull (1995) and Duffie and Singleton (1999) among others. Since then, the reduced form approach has become popular due to its simplicity and applicableness. When applying reduced form models, we start by solving the theoretical equation for all the endogenous variables to get a reduced form equation. The reduced form equation can then be estimated, when assuming estimates for the exogenously given variables.

The formulation of the reduced form model applied in this paper stems from Cox, Ingersoll and Ross (1995). It is a theory of the term structure of interest rates based on the Vasicek model (Vasicek, 1977), where the market risk is the only factor driving interest rates (Cox et al, 1985). The pricing model applied is an extended version of the CIR Univariate Model from the paper 'Pricing bonds and bond options with default risk' (Barone et al, 1998), which in turn follows the lines of the model by Duffie and Singleton (1999) and Duffie and Huang (1996). The mathematical formulation of the model will be presented in section 4. Barone et al explicitly states that the model should be applicable to all defaultable debt instruments, which we believe includes P2P lending. For the remainder of this paper, the chosen model will be referred to as "ECIR". As the information about the loans in Lending Club's loan book is limited, the ECIR model is well suited for a study of the theoretical risk-adjusted price versus the actual price observed at the Lending Club platform. The price observed at the Lending Club platform is hereby referred to as the "LC price".

The model in this paper is quite similar to the ones that are used to price risk-free bonds. Using the model for assessing the theoretical price of P2P loans can be justified by the similarity of traits between bonds and P2P loans presented in the following section.

2.2 P2P Loans and Bonds

As previously mentioned, the ECIR model should be applicable to all debt instruments (Barone et al, 1998). To further back up this statement, we will point out the conceptual similarities between P2P loans and bonds. This will serve as additional evidence on why the literature on bond pricing should be able to enhance our knowledge on the P2P loan market.

2.2.1 Conceptual Similarities

Bonds are debt contracts between two parties. A bondholder (the investor) buys a bond from the issuer (a business) in need of financial capital. P2P loans are set up in the similar way as bonds, with a small business as the issuer and an investor buying the right to a future fixed cash flow in exchange for lending the principal amount to the business. The cash flow of a bond consists of a coupon at a given rate and an instalment, and are made periodically. Periodic payments are also made in P2P loans, also consisting of instalments and interest rate payments at a given rate. The coupon rate of bonds is effectively the interest rate for P2P loans. Note that the specific down payment plan may vary between different platforms, and that all the loans in Lending Club's loan book are annuity loans.

Bonds are often tradable in a second-hand market. This is also true for the P2P loans, although the second-hand market for P2P loans are not as developed and therefore less liquid.

2.2.2 Risk

The credit risk related to bonds applies to P2P loans as well. When investing in either a bond or a P2P loan, you risk that the borrower will not be able to pay back the periodic payments. This is known as credit risk. Financial assets have credit risk which contains of three different factors: the probability of default, the loss given default and the exposure at default (Altman, 2006). Investors choosing P2P loans are assumed to have a constant exposure at default of a 100 %, hence this factor is ignored when estimating the credit risk. There are several ways of estimating the probability of default and the loss given default. For this paper, we choose to apply constant estimates, as the Poisson process is assumed in accordance with Barone et al (1998). The Poisson process allows the historical rates within a given time interval to represent an estimate for future rates, within time intervals of the same length (Haight, 1967). The probability of default and the loss given default provides the investor with a risk measure, and as with any risk the investor should be compensated for holding that risk.

Furthermore, a bond with a long time to maturity implies a high holding-period risk. This means that the bondholder is running the risk of missing a better opportunity in the holding period, as well as increasing the probability for something negative to happen during the life-time of the bond (Berk and DeMarzo, 2014). As we have established that P2P loans are based on the same conceptual mechanisms as bonds, the same principles are likely to apply to P2P loans with different maturities.

Another relevant risk for P2P loans as well as bonds, are the so-called "callable risk". Callable risk occurs because the issuer can redeem parts of (or the entire) loan amount at any time during the loans/bonds maturity, which would reduce the expected return for the investor (Berk and DeMarzo, 2014).

2.3 Black and Scholes

With regards to the borrowers' option of paying down their loans early, we also need to consider the price of this option to quantify what we refer to as callable risk.

The Black and Scholes model for calculating the premium of an option is one of the most known option pricing models. It was developed by Fischer Black, Myron Scholes and Robert Merton and first appeared in the paper "The Pricing of Options and Corporate Liabilities" (1973). The model calculates the premium of a European option under certain assumptions. Firstly, the option is European, hence it can only be exercised at expiration. It is also assumed that there are no payments of dividends during the life of the option and the markets are said to be efficient without any form of commissions. The risk-free rate and volatility of the underlying asset are assumed to be known and constant, at the same time as the return of the underlying asset follows a normal distribution.

When applying the Black and Scholes model, it is necessary to know the following input variables: The current value of the underlying asset, the strike price at exercise, time until expiration, risk-free interest rate and volatility of the underlying asset. See section B of the appendix for the mathematical formulation.

The Black and Scholes model is interpreted in two parts, where the first part of the equation is a multiplication between the current price of the underlying asset and the effect a change in the underlying asset's price has on the call premium. The second part is the present value of paying the exercise price at the expiration date. The difference between the two parts is the call premium.

The rationale behind bonds and P2P loans are the same: a way for businesses to fund its operations and expansions. It is two very similar debt instruments, which in sum implies that we should be able to price P2P loans using the same theories about credit-risk modelling as those being used for bond pricing.

3. Data

When applying the ECIR model on the P2P lending market, we use two different datasets. The first is the loan book of Lending Club, a leading U.S. P2P lending platform. The second dataset is the constant maturity series of the U.S. treasury rate from 1962 up to March 1st, 2016.

3.1 Lending Club Loan Book

Lending Club is one of the oldest and largest P2P lending platforms, and its loan book consists of data from 2007 to the start of 2016. For reasons that we will come back to in the analysis, Lending Club are determined to be transparent. Consequently, the loan book is publicly available and can be downloaded from Lending Club's own web page (Lendingclub.com, 2016). When we downloaded the loan book it contained both business and private loans, but we have chosen to filter out the private loans. The reason is that we seek to explore the pricing of credit risk in depth, which is simpler when excluding other differences in characteristics between individuals (Lendingclub.com, 2016). Also, theory on the risk-return relationship of bonds seems to have more in common with business loans than with individual loans. Furthermore, the socio-economic issues related to the credit crunch of SMEs, suggests that the access to capital for individuals are better than for businesses (Wehinger, 2014). Thus, we find the business segment within P2P lending both more interesting and more comparable than the individual segment. All loans in the category "small business" is collected and the dataset that is left contains approximately 10 000 observations collected over 8 years.

The years of 2007, 2008 and 2009 are taken out because of the financial crisis. As the financial distress of this period hit small businesses especially hard, we expect the amount of loans, as well as the amount of defaults on these loans, to deviate from "normal". 6200 loans from the period 2010 to 2014 are used as historical statistics to obtain the different estimates needed in the model.

The loan book contains information on the loan as well as the person asking for the loan. The key elements needed in our model is listed in table 1.

Table 1: Key variables in Lending Club's loan book

Variable	Definition
ID	A unique identification number
Issue date	The month the loan was issued
Funded amount	Total value received by the borrower
Total payments	Total payments made by the borrower
Term	Length of the loan
Interest rate	The interest rate received by the investor (annualized)
Rating grade	Rating grade from A-G set by Lending Club
Loan status	Whether the loan is defaulted, late or fully paid

Notice that we only have information on the issue month of the loan as well as the present loan status. Loans can have many statuses, including "Current", "Late", "Charged Off" etc. We do not know the previous statuses a loan has had, or the specific date a change of status occurred, which means that we have to make some simplifications. A loan with a three year down payment plan has either defaulted or not during those three years, and only loans with the status "Charged Off" are counted as defaulted. The specific date of default is not available. This simplified view corresponds well with the fact that our model applies constant estimates for the default rate and loss given default.

When estimating based on historical data, the loans are grouped into rating grades. The number of observations in each rating grade can be found in the appendix, section C. As mentioned, estimates for model inputs are based on data from 2010 to 2014, and these estimates are used as input in the ECIR model to price the loans issued in 2015. We already know the prices Lending Club set on the loans they issued in 2015 (LC price), and will compare these prices to the theoretical prices provided by the chosen model.

3.2 U.S. Treasury Rate

Because our model is based on statistical models related to the term structure of interest rates, we need a dataset describing risk free rates. We use the constant maturity series of the U.S. treasury, obtained from Wharton Research Data Services (WRDS, 2016). The dataset contains interest rates with different maturities from 1962 up until march of 2016. We use the three-year U.S. treasury rate for loans with three-years maturity and five-year U.S. treasury rate for loans with five-years maturity, as we consider this the best approximation of a risk-free interest rate. The dataset contains daily observations of these rates. Figure 1 contains both the three-year and the five-year U.S. treasury rate.

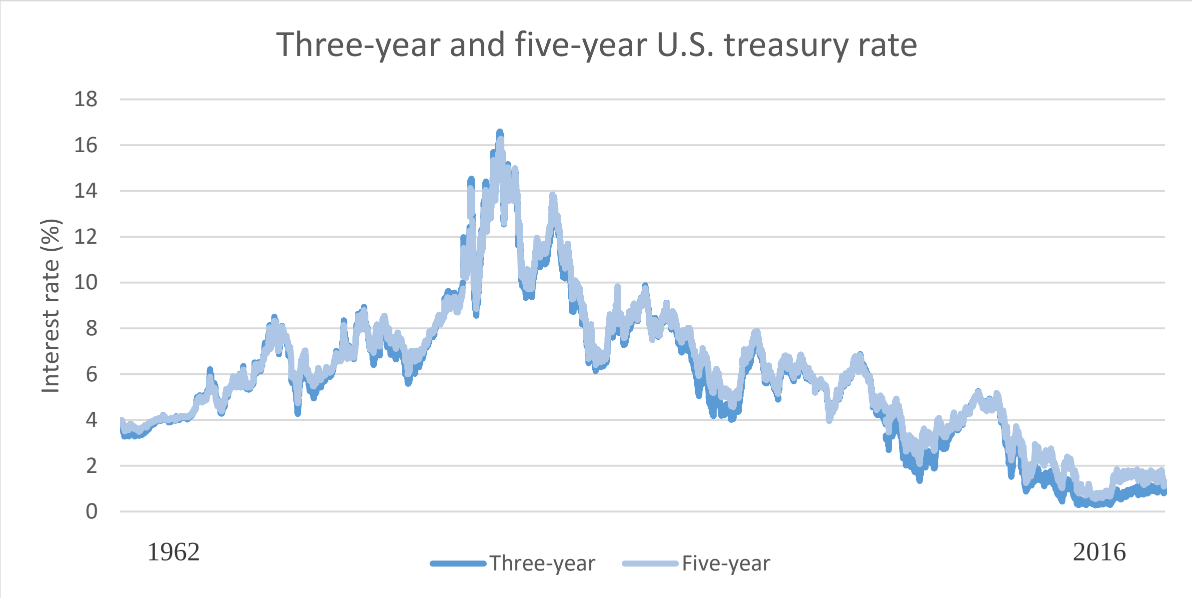


Figure 1: Three-year and five-year US treasury rate

4. Methodology

In this section, we will describe the idea behind valuation of P2P loans. The approach is based on bond valuation theory.

4.1 Valuation

Consider a bond lasting from time t to time T, with m different time intervals Δt. At the end of the bond's maturity at time T, it pays the amount X. The value of such a bond today would be P^D. Also, the bond has a default risk connected to it. This implies that there is uncertainty attached to whether the investor will receive the promised cash flows. We assume that the default risk remains constant over the time interval (Δt), and occurs as an exogenously given hazard rate h. If the default occurs, a fraction L of the promised cash flows are lost (loss given default). The possibility of a default in each period, multiplied with the loss given default will provide the default premium η as shown in equation (1).

$$\eta = Lh\Delta t \tag{1}$$

General valuation theory uses the expected cash flows discounted at the risk-free rate r, as the current value of a bond. In the case of defaultable claims, such as defaultable bonds and P2P loans, these expected cash flows must be adjusted for the possibility of default. The present value (P^D) of a defaultable bond with only one payment X at time T would be as shown in equation (2).

$$P^D = X[1 - h\Delta \mathrm{tL}]\frac{1}{1 + r\Delta t} \tag{2}$$

Equation (2) describes the formula for pricing bonds using binomial trees. The formula is applied by calculating the value of each node back in time, until you arrive at the node representing the time today. See Barone et al (1998) for an example of the binomial tree in this setting. Instead of multiplying the amount paid with the default probability and loss given default, we adjust the risk-free interest rate (r) to a risk-adjusted interest rate R. By using the risk-adjusted interest rate (R), the formula in equation (2) can be written as equation (3):

$$P^D = X\frac{1}{1 + R\Delta t} \tag{3}$$

Where the risk-adjusted interest rate (R) is given by equation (4), expressed with the risk-free interest rate (r) and the default premium (η):

$$\mathrm{R} = r + \eta \tag{4}$$

As shown in equation (4), it is necessary to add a default premium (η) to the risk-free rate (r) when pricing defaultable claims.

So far, the formulas have considered zero-coupon bonds. When introducing coupons, the price of a bond (${P^D}_C$) will be as shown in equation (5).

$$P_C^D = \sum_{j=1}^{m} P_j^D \, a_j \tag{5}$$

The price at each period ${P^D}_j$ represents the price at the *j-th* cash flow a_j, where the cash flow (a_j) consists of both coupon and capital instalment. ${P^D}_j$ can be viewed as a discounting factor, discounting each future payment to the present value.

A coupon bond can be viewed as a series of strip bonds. A strip bond is a bond where all coupons are divided up and sold separately, as if every coupon were its own zero-coupon bond. Based on this thought we can view a coupon bond as the sum of all its strip bonds. As shown in equation (5), the price of the coupon bond is the sum of the individual prices of the imagined zero-coupon bonds, multiplied with the associated cash flow.

Based on the previously mentioned conceptual similarities between bonds and P2P loans, we argue that the bond valuation methodology in this section should also be applicable to P2P loans. Moving on, we will apply the valuation methodology in a pricing model.

4.2 An Extended Version of the CIR Univariate Model

In this section, we will apply the Extended version of the CIR univariate (ECIR) model from "Pricing Bonds and Bond Options with Default Risk" (Barone et al, 1998).

The model that follows is based on the CIR model (Cox, Ingersoll and Ross, 1985):

$$\mathrm{dr} = k(\theta - r)dt + \sigma\sqrt{r}\,dz \tag{6}$$

The CIR model describes the development of interest rates. It states that interest rates are driven by a single market risk factor and that the rate follows mean reversion. Mean reversion results in an equilibrium value where the rates will return to in the long-term perspective. The first part of equation (6), $k(\theta - r)dt$, describes the instantaneous change in the interest rate at the current time. Parameter k represents the speed of mean reversion, which tells us how fast the interest rate will drift towards the long-term equilibrium. This long-term equilibrium is represented by the expected long-term interest rate θ. As previously mentioned, r represents the short-term instantaneous interest rate. The second part of equation (6), $\sigma\sqrt{r}\,dz$, consists of the volatility of interest rates (σ^2) together with a stochastic process (dz) modelling the random market risk factor. The square root of the short-term instantaneous interest rate (r) is added to avoid negative interest rates.

As both valuation theory and the ECIR model has been presented it is possible to combine the two. In their original paper from 1985, Cox, Ingersoll and Ross presented the fundamental equation for the price of a discount bond as shown in equation (7).

$$P_j^D + P_r^D[k\theta - (k+\lambda)r] + \frac{1}{2}P_{rr}^D\sigma^2 \mathrm{r} - \mathrm{R}P^D = 0 \tag{7}$$

Further, this partial differential equation was subject to a derivation to obtain the price of a defaultable zero-coupon bond. The solution to the derivation of equation (7) is given in equation (8). Notice that for the following set of equations, equation (8) is the formula of interest as this is the pricing formula. Equations (9) to (13) serves as auxiliary calculations for an easier read, and the output from these equations are not easily interpreted alone. For equation (8), $P^D(R,t;s)$ is the value at time t of a defaultable zero-coupon bond with face value and maturity at time s.

$$P^D(R,t;s) = A(t;s)e^{-\eta(s-t)-B(t;s)r} \tag{8}$$

Where

$$\mathrm{A(t;s)} = \left[\frac{2\gamma e^{\left(\frac{1}{2}\right)(\gamma+\beta)(s-t)}}{W(t;s)}\right]^{\zeta} \tag{9}$$

$$\mathrm{B(t;s)} = \frac{2\left[e^{\gamma(s-t)} - 1\right]}{W(t;s)} \tag{10}$$

$$W(t;s) = (\gamma+\beta)\left[e^{\gamma(s-t)} - 1\right] + 2\gamma \tag{11}$$

$$\zeta = \frac{\gamma+\beta}{\sigma^2}R_\infty \tag{12}$$

$$\gamma = \sqrt{\beta^2 + 2\sigma^2} \tag{13}$$

Equation (8) prices a defaultable zero-coupon bond, but we want to price a defaultable coupon bond. In accordance with Barone et al (1998) we view the price of a coupon bond as a linear combination of defaultable zero-coupon bonds. The rationale behind this assumption is as explained in the paragraph on strip bonds in section 4.1. We consider a model where the value of a bond at time *t*, with default risk and maturity at time *T*, is given by P_C^D(R, t; T) as shown in equation (14). The price of such a bond would then be given by

$$P_C^D(R,t;T) = \sum_{i=1}^{m} a_j \, P^D(R,t;t_i) \tag{14}$$

Note that P_C^D (R, t; T) in equation (14) is identical to equation (5) from the last section. Once more, the P^D functions as a discount factor. Multiplied with the cash flow at the given point in time, it provides a present value of the future payment (a_j).

The general formula presented in equation (14) describes the pricing of a defaultable asset with payment *X* at time *T* (Barone et al, 1998). Bonds are one example of such an asset, and so are P2P loans. Supported by all the earlier mentioned arguments, it follows that the model should be applicable to price P2P loans in the same way as it is used to price bonds.

As all the elements of the original equation (14) has been accounted for, the missing factors are the inputs. The inputs needed are the j-th cash flow (a_j), the default premium (η), the instantaneous short-term interest rate (r), the β, which is mainly connected to the slope of the interest rate curve, the long-term average interest rate (θ), the asymptotic long-term interest rate (R_∞) and the volatility of interest rates (σ^2).

In the next section, we will explain how we estimate the different input variables, keeping in mind that the goal is to apply the model to the P2P market.

4.3 Inputs

4.3.1 Annuity

The first input is the cash flow (a_j), which consists of coupon payments and instalments. To make the down-payment plan as easy and comprehensible as possible, Lending Club uses a constant monthly payment (annuity). Note that Lending Club charges a 1 % fee of the monthly payment as a service fee in addition to an origination fee. Because this paper focus on the investor perspective, the origination fee can be disregarded as it is paid directly by the borrower and does not affect the investor's cash flow. The 1 % service fee is subtracted from every payment (annuity) the borrower makes to the investor, thus it reduces the investors expected cash flows in relation to the loan. When calculating the monthly payment, we use the following annuity formula:

$$PV_{Ann} = a_j\left(\frac{1-(1+i)^{-m}}{i}\right) \quad (15)$$

where i is the interest rate of the loan and PV_{Ann} is the funded amount.

4.3.2 Default Premium

Moving on, the second input is the default premium (η). It consists of the default rate (h) and the loss given default (L), both estimated on historical data. As previously explained the model assumes that default is an unpredictable event that follows a given rate, in accordance with the assumption of a Poisson process. The default rate in one period of time is therefore assumed to be the same as the default rate in any other given period of time, assuming that the length of the time interval of the two periods are the same. Furthermore, the observed historical default rates are used as an estimate for future defaults. Note that there is no information in Lending Club's loan book on historical default rates for each company. Therefore, the rating grades provided by Lending club on each company is used to group loans with similar chance of default. The default rates are then estimated for each rating class and each maturity, by using historical monthly averages.

The investor's actual net cash flow divided by the net expected cash flows, provides a fraction called the loss given default (L). The estimates are done within each rating class and each maturity, thinking that firm-specific variables and the length of the loan could affect the ability to pay back the loan. Multiplied with the default rate (h), the loss given default (L) provides the default premium (η).

4.3.3 Interest Rates

To estimate the risk-free rate (r), we use the constant maturity series of U.S. treasuries (CMT). As the loan book from Lending Club only provides the month in which the loan has been issued, we apply the monthly average of both three and five year U.S. treasuries as the risk-free rate input in the model.

The next input we calculate is the β, which is a variable mainly connected to the slope of the interest rate curve. To calculate the β this paper follows the approach of Barone and Risa (1994):

$$\beta = k + \lambda \tag{16}$$

The market risk premium is represented by λ. This is a factor that has been extensively researched in financial literature, but there is still no universal answer to the size of the market risk premium. Throughout this paper the market risk estimate will be 5 % in accordance with Koller (Inc, Koller and Goedhart, 2015). The estimate of mean reversion speed (k) is difficult to calculate. An easier understandable representation of the speed is found using half-life ($H_{1/2}$) of the interest rate. $H_{1/2}$ represents the time it takes for the interest rate to return halfway to the mean (θ). Using historical data from the constant maturity series of US treasuries, it is possible to provide an estimate of $H_{1/2}$. The half-life rate is given by:

$$\mathrm{r}\left(H_{\frac{1}{2}}\right) = \frac{r_0 - \theta}{2} + \theta \tag{17}$$

Calculating all the current rates and their respective half-life rates, we estimate the average time it takes from r_0 to $r(H_{1/2})$. With the estimates for half-life we can calculate the speed of mean reversion (k) with the following formula (Meucci, 2009).

$$H_{1/2} = \frac{\log(2)}{k} \tag{18}$$

Together with market risk premium (λ) the speed of mean reversion (k) will provide the estimate for β. In addition, the long-term expected interest rate (θ) is calculated as the daily average of three and five year U.S. interest rates from 1962-2016. The asymptotic mean of long term risk-free rate (R_∞), is calculated as shown in equation (19), in accordance with Barone and Risa (1994):

$$R_\infty = 2k\theta/(\gamma + \beta) \tag{19}$$

The volatility of the risk-free rate (σ^2) has been calculated through quadratic variation, using the difference between the average rate and the daily observations of the three and five year U.S. treasury rates.

5. Results

In this section, we will present the results from the input calculations as well as from the application of the ECIR model. We will also explain the output variables where necessary.

5.1 Default Premium

Firstly, we will present the default premium we estimated based on data for the various rating grades from 2010-2014. This default premium is used as an estimate for the future default premium η. Table 2 shows the results of the estimation for three-year loans, while table 3 shows the results of the estimation for five-year loans.

Table 2: Default premium, three-year loans

Rating grade	Default rate	Loss given default	Default premium
A	1.96 %	58.44 %	1.14 %
B	2.59 %	56.91 %	1.47 %
C	2.50 %	57.33 %	1.43 %
D	3.06 %	57.13 %	1.75 %
E	4.08 %	59.75 %	2.44 %
F	3.94 %	63.86 %	2.51 %
G	4.96 %	69.22 %	3.44 %

Table 3: Default premium, five-year loans

Rating grade	Default rate	Loss given default	Default premium
A	3.85 %	48.25 %	1.86 %
B	3.97 %	60.11 %	2.38 %
C	4.60 %	66.94 %	3.08 %
D	3.79 %	68.02 %	2.58 %
E	4.17 %	70.03 %	2.92 %
F	4.61 %	72.32 %	3.33 %
G	4.91 %	71.74 %	3.52 %

The default rate and default premium is given annually in table 2 and 3. The calculation looks at the total number of loans that has been issued and total number of defaulted loans in the period 2010-2014. Notice that the loan book is updated constantly and that the only information known publicly is the issue date and the current status. We do not know when each loan was "charged off" (defaulted). Because of this, we assume an even distribution which means that there is an equal amount of defaults each month.

We did expect the "Default rate" and "Loss given default" to increase as the rating grades moved towards G. Table 2 and 3 shows that this is generally what we found. However, due to a low number of observations when dividing on both rating grades and maturities, the trend is not true for all the rating grades. The trend of increasing rates with decreasing rating grades was more evident when we looked at all the observations without dividing by maturity.

Despite some random results due to a small dataset, the main results are as expected: the default premium increases as we move towards lower rating grades.

5.2 Interest Rates

When it comes to the monthly average of three- and five-year U.S. treasuries, the following input variables is used throughout the paper. See section 4.3.3 for explanations regarding the auxiliary calculation.[a]

Table 4: Input variables connected to the term structure of interest rates

	β	**σ^2**	**γ**	**ϴ**	**R∞**
Three years	0.13	0.0180	0.2301	0.0578	0.0257[b]
Five years	0.12	0.0599	0.3662	0.0605	0.0174

5.3 Model Output

In this section the ECIR model is applied to the P2P business loans in Lending Club's loan book. We present some examples of the model-given price levels relative to the prices observed at the Lending Club platform (LC price). From the model price, we derive an implicit interest rate that corresponds to the interest rate set by Lending Club. We also present the spread between the two. Furthermore, the results are presented separately for loans with

[a] The inputs are the speed of mean reversion (k), the market risk premium (λ), the β, which is mainly connected to the slope of the interest rate curve, the instantaneous short-term interest rate(r), the volatility of interest rates (σ^2), the γ as is given by equation (13), the long-term average interest rate (θ) and the asymptotic interest rate (R_∞).

[b] Notice that the asymptotic interest rate (R_∞) is higher for the three-year loan compared to the five-year loan. We have no intuitive explanation for this finding, but mathematically it seems to be connected to a higher estimated speed of mean reversion (k) for the three-year rate.

maturity of three and five years, as it is expected that the length of the loan affects the risk and thus the price of the loan. Note that all the loans in Lending Club's loan book has a maturity of either three or five years. Examples of the output from the ECIR model can be seen in table 4 and 5.

Table 5: Examples of loans with three-year maturity

ID	Rating grade	Model price	LC price	LC ratio	Implicit rate	LC rate	Spread
354795	C	$7 790.31	$6 800.00	0.8729	4.06 %	13.33 %	9.27 %
382607	B	$8 631.80	$8 000.00	0.9268	4.05 %	9.17 %	5.12 %
737150	C	$34 468.20	$30 000.00	0.8704	4.19 %	13.67 %	9.48 %
36300829	F	$8 807.41	$6 825.00	0.7749	5.05 %	22.99 %	17.94 %
36370598	E	$6 232.23	$5 025.00	0.8063	4.98 %	19.99 %	15.01 %
36411294	D	$42 144.53	$35 000.00	0.8305	4.32 %	17.14 %	12.82 %
36733219	F	$15 453.29	$11 975.00	0.7749	5.05 %	22.99 %	17.94 %
36783346	D	$16 677.53	$14 075.00	0.8440	4.32 %	15.99 %	11.67 %
37197490	C	$6 571.95	$5 600.00	0.8521	4.02 %	14.99 %	10.97 %
37227469	E	$14 324.82	$11 550.00	0.8063	3.80 %	19.99 %	16.19 %

Table 6: Examples of loans with five-year maturity

ID	Rating grade	Model price	LC price	LC ratio	Implicit rate	LC rate	Spread
16111205	E	$33 309.83	$24 000.00	0.7205	5.46 %	19.99 %	14.53 %
29053621	G	$36 275.29	$22 400.00	0.6175	6.16 %	28.49 %	22.33 %
36803880	D	$44 622.75	$35 000.00	0.7844	5.07 %	15.59 %	10.52 %
37601806	D	$31 873.68	$25 000.00	0.7844	5.07 %	15.59 %	10.52 %
37620413	E	$49 602.25	$35 000.00	0.7056	5.47 %	20.99 %	15.52 %
37660697	C	$24 445.98	$20 000.00	0.8181	5.65 %	14.31 %	8.66 %
37661609	E	$47 814.35	$35 000.00	0.7320	5.47 %	19.24 %	13.77 %
37731496	C	$29 334.98	$24 000.00	0.8181	5.65 %	14.31 %	8.66 %
37821555	F	$45 942.01	$30 225.00	0.6579	5.94 %	24.99 %	19.05 %
37831593	B	$16 561.27	$14 000.00	0.8454	4.84 %	11.99 %	7.15 %

Table 5 and 6 presents some examples of loans and how they are priced by the model and by Lending Club. The first three columns are descriptive information regarding the loan. The ID column consists of the unique "ID" assigned to each loan, "Interest rate" is the interest rate which the borrower pays to the lender (investor) to compensate for the risk and "Rating grade" is the rating dedicated to the borrower by Lending Club. The output of the model is a "Model price" and are to be seen relative to the price set by Lending Club (LC price). The "LC ratio" is given as the LC price divided by the model price.

The next section will explain the "LC price" and "LC ratio" terms in depth. "Implicit Rate" is the implicit, monthly interest rate paid by the borrowers to the investors. The implicit interest rate is found by using the annuity formula and the solver tool in Excel. We compare this to the "LC rate", which is the rate that Lending Club presents for each loan to potential investors. "Spread" represents the spread between the implicit interest rate from the ECIR model and the interest rate set by Lending Club.

5.3.1 Price vs. Interest Rates

The results from the ECIR model provides the "Model price" given in table 5 and 6, which is equal to the present value of all future, risk-adjusted cash flows. This theoretical price can be compared to the "LC price", which is the amount the investor pays to the borrower for the promised future cash flow (the size of the loan). Even though the LC price variable is not set directly by Lending Club, the price is set implicitly through Lending Club's choice of an interest rate and down payment plan.

In table 5 and 6 we present the price of the loan and the interest rates, which could both be derived from the other. As we are studying returns on loans, it is more intuitive to refer to the interest rate that the borrower needs to pay to the investor. However, due to practical considerations, we do not calculate the implicit interest rate for each of the 6000 loans as we did for the examples in table 5 and 6. The LC ratio is more convenient to calculate for all the loans. Therefore, we will be using both "price" and "interest rates" when discussing the loans.

Before we move on to the analysis, we specify the relationship between interest rates and LC price. From the investor's point of view, higher interest rates mean higher returns. At the same time, a lower price for a given expected cash flow, also means higher returns. An investor will therefore prefer an interest rate which is higher at Lending Club, compared to the model given interest rate. The same lender will prefer an LC price lower than the model given price for a given cash flow. As we continue our analysis from the investor's perspective, high interest rates and low prices at the Lending Club platform are favourable for the investors.

In the upcoming section, we will present three main findings from the application of the ECIR model on loans from Lending Club's loan book.

5.3.2 Returns Higher at Lending Club Compared to the ECIR Model

Table 7 and 8 displays the average LC interest rate, which is an average for each rating grade. To compare these rates with implied interest rates from the ECIR model, we added all the loans in each rating grade into rating grade portfolios. We find the implied interest rate by solving numerically for the yield of a standard annuity. This is an implied rate for the entire rating grade portfolio, which serves as an approximation of the average implied interest rates of each rating grade. The "Interest rate spread" returns the difference between the interest rate Lending Club assigns each portfolio, and the interest rate the ECIR model assigns each portfolio.

Table 7: Comparison of interest rates, three-year loans

Rating class	Average LC interest rate	Implied interest rate	Interest rate spread
A	7.32 %	3.80 %	3.52 %
B	10.39 %	4.14 %	6.25 %
C	13.48 %	4.05 %	9.43 %
D	16.82 %	4.37 %	12.46 %
E	19.28 %	4.98 %	14.31 %
F	23.74 %	5.03 %	18.71 %
G	26.69 %	6.10 %	20.59 %

Table 8: Comparison of interest rates, five-year loans

Rating class	Average LC interest rate	Implied interest rate	Interest rate spread
A	7.91 %	4.29 %	3.62 %
B	10.89 %	4.87 %	6.03 %
C	13.67 %	5.65 %	8.02 %
D	16.80 %	5.04 %	11.76 %
E	19.54 %	5.50 %	14.04 %
F	23.80 %	5.96 %	17.84 %
G	26.95 %	6.19 %	20.76 %

The first important finding from running the ECIR model is that investing in any of Lending Club's P2P business loans yields a return that exceeds the risk-adjusted return projected by the model. In other words, the interest rates from the ECIR model are lower than the interest rates set by Lending Club, for every loan on the platform. The fact that every loan receives a lower rate than the interest set by Lending Club, can be seen through the descriptive statistics found in section A of the appendix. Note that we did expect the rates generated from the ECIR model to be lower than the rates observed at Lending Club, as we recognise that the model is a simplification of reality with assumptions such as no-arbitrage and perfect markets. Even though we expected to find a premium relative to the theoretical rates, we did not know if parts of the spread could represent excess returns for the investors. To address this, we need to study the explanations for the spread. We will return to this in the analysis.

5.3.3 Variations Across Rating Grades

The second important finding is the pattern of the interest rate spread in table 7 and table 8. We find that the interests rate spread increases as the rating grades becomes lower, for both maturities. This implies that the amount of which the observed return on the loans exceeds the theoretical return, increases as the loans becomes riskier. This is seen in table 9 where the interest rate spreads from table 7 and 8 are presented side by side.

Table 9; Interest rate spread increases with decreasing rating grades

Rating class	Interest rate spread (three-year)	Interest rate spread (five-year)
A	3.52 %	3.62 %
B	6.25 %	6.03 %
C	9.43 %	8.02 %
D	12.46 %	11.76 %
E	14.31 %	14.04 %
F	18.71 %	17.84 %
G	20.59 %	20.76 %

5.3.4 Variations Across Maturites

The third important finding from the model output is that as the maturity of the loans increases, the relative spread between the theoretical prices and the observed prices increases. This is demonstrated by comparing the average LC ratio from three- and five-year loans. The LC ratio decreases as the maturity increases for all rating grades, as seen in figure 10. This implies that the difference between the LC price and the model price increases, as the maturity goes from three to five years. Remember that the LC ratio is given as $\frac{LC\ price}{Model\ price}$, an LC ratio of 1 would indicate that the difference in the prices are 0. Any LC ratio between 0 and 1, indicates that the model price is higher than the LC price. The lower the LC ratio becomes, the bigger is the relative difference between the LC price and the model price. Table 10 shows how the average LC ratio is higher for loans with five-year maturity compared to loans with three-year maturity, for every rating grade.

Table 10: Change in average LC ratio as the maturity increases

Rating grade	Average LC ratio (three-years)	Average LC ratio (five-years)
A	0.9491	0.9168
B	0.9117	0.8677
C	0.8716	0.8307
D	0.8349	0.7647
E	0.8150	0.7281
F	0.7679	0.6747
G	0.7480	0.6368

The previous sections have presented the three main findings: Firstly, there is a spread between the return currently found at the Lending Club platform, and the theoretical return predicted by the ECIR model. Secondly, this spread increases as the rating of the loans decreases. Thirdly, the difference in return increases as the maturity of the loans increases.

In the next section, we will analyse the possible reasons behind the observed spread, to understand whether parts of the spread could be explained by the existence of excess returns in the P2P lending market.

6. Analysis

To analyse the results from the application of the ECIR model, embedded financial and economic theory will be used in an attempt to explain the output and to assess how Lending Club's P2P loans are priced relative to the credit-risk pricing model. We will also conduct a sensitivity analysis, discuss the robustness of the model and comment on the potential implications of our findings.

The spread between Lending Club's prices and the ECIR model is the main finding. Note that we expected to find a spread of some sort, although the size of the spread was larger than expected. To study the risk-return relationship for investors we will analyse the reasons behind the spread, and evaluate if these reasons are likely to explain the whole spread. If not, it could indicate that investors are in fact earning excess returns when investing in P2P loans. In a perfect market where the ECIR model is good a way of pricing the risk of P2P loans, Lending Club's prices would equal the theoretical price from the ECIR model and there should be no excess returns. However, the finding of a spread does not necessarily imply that the ECIR model does not fit the P2P lending market, but rather that the P2P lending market is inefficient.

In theory, the high interest rates set by Lending Club should attract investors, but also make the borrowers go somewhere else for financing. The problem with this reasoning is the lack of opportunities for the borrowers when it comes to alternative ways of financing. Many of the borrowers in the P2P lending market are in the market because they are struggling to access financing in the traditional markets. Furthermore, stricter regulations are costly for banks and forces them to set aside more capital and demand more collateral from borrowers, this leads to many SMEs experiencing a credit squeeze. With no other place to go, the borrowers are left with the option of giving up on their business, or accepting an expensive P2P loan.

With this inefficiency in mind, we will now present possible explanations for the three main findings: the existence of a spread and the increase of the spread with both decreasing rating and increasing maturity.

6.1 Explanations of the Spread

The higher interest rates relative to the ECIR model might be due to the way Lending Club perceives risk. The interest rate of a loan determines the return, which is the investor's compensation for the risk of investing in the loan. The riskier a loan is, the higher the interest rate has to be, to compensate the investor. If Lending Club and the ECIR model differs when it comes to the perception of a loans riskiness, the interest rate of the loan will also differ between the two. We suspect that asymmetric information and callable risk could help explain the difference in perception of risk, and we will address these two in the sections below.

6.1.1 Asymmetric Information

The differences in perceived risk could be a consequence of asymmetric information. As Lending Club conducts the screening of the lender and prices the loan accordingly, an investor does not know if all the risk found has been incorporated in the price. A possible explanation for the spread may be that Lending Club compensates the investors for the uncertainty by adding a premium in addition to the interest rates predicted by the ECIR model. Lending Club tries to mitigate the problem of asymmetric information by having their loan book publicly available. Still, the platform has been accused of selling loans on false premises and manipulating rating scores over the last year, which has led to the Lending Club stock-price collapsing and the CEO's resignation (Forbes, 2016). This leads us to the problem of moral hazard: the investors know that Lending Club does not have any of the loans on their balance sheet, and that Lending Club's incentives when setting interest rates are tied to attracting investors (maximize fees) and not to avoid defaults. As moral hazard means more uncertainty for the investors, they would demand higher interest rates when investing in P2P loans.

It seems evident that the concerns regarding asymmetric information and moral hazard are relevant when it comes to the P2P lending industry.

6.1.2 Callable Risk

The differences in the perception of risk could also be explained by the ECIR model not considering other elements of risk. For example, we do know that the ECIR model does not contain estimates on callable risk or liquidity risk. Callable risk is a concern because the borrowers at Lending Club can pay down the entire loan at any time during the loans maturity, which would reduce the expected return for the investor. An example would be a start-up company turning to Lending Club, as the lack of collateral and credit history made things difficult with traditional banks. If the start-up company succeeds and start to make profits quickly, they would have the incentive to pay down their relatively expensive Lending Club loan and lend money at a lower interest rate somewhere else. P2P lending investors having bought the loan would miss the interest rates from the remaining periods of the maturity, which leads us to believe that this risk should be considered when pricing the loans. Could the callable risk explain the spread between the Lending Club interest rate and the theoretical interest rate from the ECIR model?

The opportunity the borrowers have to pay down the loan at any time, is equivalent to an American call option where the borrower is short in the loan and long in the call option. It follows that the investor is long in the loan and short in the option, and that the value of the loan will be reduced (from the investor's perspective) because of the call option.

To estimate the value of the option, we will apply the Black and Scholes model for European call options. We know that the borrowers can pay down the loan at any time during the loan period, but we have no data on how many actually do this. For simplicity, we will look at the risk of early fulfilment as a European call option and assume a time to exercising. Because an American option will always be valued at least as high as a European call option, the latter can be viewed as a "minimum value" of the call option. We believe that this reflects the risk in a good way, given the uncertainty related to the information availability. See the appendix, section B for explanations surrounding the parameters of the Black and Scholes model.

In table 9 and 10 the "Interest rate spread" column is the original spread between the interest rate from Lending Club and the ECIR model for each rating grade as shown in section 5.3.2, while the "Adjusted interest rate spread" column is the interest rate spread after adjusting for the call option value. The "Change in spread" column shows the change to the spread when taking the callable risk into consideration.

Table 11: Spread adjusted for call option value, three-year loans

Rating class	Interest rate spread	Adjusted interest rate spread	Change in spread
A	3.52 %	2.05 %	1.47 %
B	6.25 %	4.62 %	1.63 %
C	9.43 %	7.46 %	1.97 %
D	12.46 %	10.27 %	2.19 %
E	14.31 %	12.17 %	2.13 %
F	18.71 %	16.09 %	2.62 %
G	20.59 %	18.16 %	2.44 %

Table 12: Spread adjusted for call option value, five-year loans

Rating class	Interest rate spread	Adjusted interest rate spread	Change in spread
A	3.62 %	1.32 %	2.30 %
B	6.03 %	3.76 %	2.27 %
C	8.02 %	5.88 %	2.14 %
D	11.76 %	9.07 %	2.69 %
E	14.04 %	11.30 %	2.74 %
F	17.84 %	14.96 %	2.88 %
G	20.76 %	17.72 %	3.04 %

As we can see from table 11 and 12, the spread between the interest rate from Lending Club and the ECIR model is reduced across all rating grades when deducting the value of the call option from the theoretical price. The callable risk is not considered in the ECIR model, so this result implies that this risk element may be a rather large part of the explanation for the spread (assuming that Lending Club has included callable risk in their pricing of the loans).

However, we should be careful concluding that the callable risk alone can explain this much of the spread. Remember that most of the borrowers in the P2P lending industries are companies that cannot access loans at favourable rates in traditional banks, and that to exercise the call option they need to have the cash to pay down the loan before maturity. This kind of refinancing requires that the company's credit situation has improved significantly, and it might not be realistic to assume that many borrowers will have the opportunity to do this.

Note that liquidity risk might also be added by Lending Club, because the second-hand market is far from fully developed. If Lending Club has a model that includes these two risks, we know that they will consider each loan riskier than the model does, hence it could explain why they assign the loans higher interest rates than the ECIR model predicts. The differences in perception of risk could also stem from Lending Club pricing the same amount of risk differently, in other words, an underlying difference in risk aversion.

We have now presented and addressed the differences in the perception of risk as a possible explanation for the spread, and we will move on to present other explanations.

6.1.3 Profit Premium

The spread could also be tied to a profit premium. P2P loans are the service of the Lending Club platform, and it is not realistic that Lending Club would price their loans in accordance with the no-arbitrage argument, as they would want to make a profit. Lending Club's pricing of the loans could be easier to understand when comparing to the pricing of a bond. A higher rate for a bond means a lower price, but for P2P loans the price is the amount borrowed and therefore the price is already defined. The consequence of a lower interest rate for P2P loans are therefore higher cash flows instead. Whether the price for a given cash flow is lowered, or the cash flows for a given price are raised results in the same: A more attractive loan for investors. Keeping in mind that Lending Club receives 1 % of every cash flow, a higher interest rate results in a scenario where higher cash flows means a higher absolute value of the fee to Lending Club. In addition, a higher interest rate could attract more investors, which would result in even more cash flows and fees. Thus, Lending Club's strive for higher profits could be a part of the explanation of the spread between Lending Club's interest rates and interest rates from the ECIR model.

So far in section 6.1 we have discussed possible explanations for the first finding of this paper: the high interest rates set by Lending Club relative to the comparable interest rates from the ECIR model. The reason for the spread is probably a combination of asymmetric information, callable risk and a search for profit. More in-depth information about Lending Club's credit risk model, lenders and borrowers, will be necessary to further study the explanations of the spread.

After addressing explanations for the spread, we will now direct our focus towards the two other main findings from applying the ECIR model on Lending Club's loan book from 2015. The second finding was that the spread increases with the riskiness of the loans.

6.2 Spread Increases with the Riskiness of Loans

We find that the spread between Lending Club's interest rates and the interest rates from the ECIR model increases with the risk of the loans, that is when the loans' ratings decreases. It makes sense that the interest rate of the loans with a higher rating grade lies closer to the theoretical interest rate than the riskier loans, as the added risk premium for these loans are closer to zero. As the loans become riskier, investors expect higher risk premiums. And the riskier the loans are, the more difficult it is to price the total amount of risk. The uncertainty related to both the size of the risk premium and to the pricing of that risk premium, are probable explanations for why the spread between Lending Club's interest rates and the model's interest rates increases when the loans' ratings decreases.

If a stronger risk aversion is the explanation for the higher interest rates set by Lending Club, compared to the ECIR model interest rates, the spread would increase exponentially with the risk. The fact that the spread increases for poorer rated loans could be explained by the inefficiency in the P2P lending market, as it would be even harder to access finance from other sources when your creditworthiness declines.

Regardless of the explanation, the second important insight from applying the ECIR model to the Lending Club loan book, is that the spread between the interest rates set by Lending Club and the interest rates provided by the model increases as the loans' ratings decreases.

The third major finding is that the spread also increases with the maturities of the loans.

6.3 Spread Increases with the Maturity of Loans

That the differences in interest rates between Lending Club and the ECIR model increases with the maturity of the loans, can be explained by the fact that a five-year loan entails a longer period where investors cannot reinvest the amount lent out than with a three-year loan, which means that the investors will demand to be compensated for the increased holding-period risk. Moreover, as we assumed that defaults follow a Poisson distribution, longer maturity loans imply that the accumulated probability of default increases. The investors would demand compensation for this as well.

Another aspect from theory on bond risk is that longer maturity implies a higher duration, which in turn implies higher interest rate risk. Hence, the five-year P2P loans are more sensitive to changes in risk-free rates which affects the price an investor would get if he/she were to sell the loan in a second-hand market. Note that, as the second-hand market for P2P loans are in the very early stages, this risk is probably not the most significant.

We have addressed plausible explanations for the three major findings from running the ECIR model on Lending Club’s loan book. It is difficult to quantify the effect of each of the explanations, but the analysis provides an indication on which of the explanations that are most likely to be the reasons for the spread.

When summing up, the spread between the interest rates set by Lending Club and the interest rates set by the ECIR model seems to increase as the risk of the loan increases, regardless of the reason behind the increase in risk. This implies that the perception of risk differs between the credit-pricing models used by Lending Club, and the ECIR model. It follows that an underlying difference in risk-aversion is a probable explanation for all three findings.

The fact that some risks, such as liquidity and callability, are not considered in the ECIR model, implies a logical explanation of the spread. However, the spread seems too large for these risks to explain the whole spread alone. The remaining parts of the spread could be explained by a market inefficiency, profit premium or difference in underlying risk-aversion. Note that these last explanations do not exclude the possibility of the investors, potentially, earning excess returns in the P2P lending market. If this is the case, it might be a well-founded strategy from Lending Club to attract investors who are sceptic to a new and unchartered investment like P2P lending. The aforementioned inefficiencies of the P2P lending market make this kind of strategy possible.

So far, we have looked at each loan as just one loan, which is a simplification. In reality, each loan is divided into several notes. The next section will address the possible implications the concept of notes will have for the results from section 6.

6.4 Diversification

Investors have the opportunity to diversify when investing in P2P loans. As previously mentioned, a single P2P loan gets divided into several notes so that a single investor does not have to cover the entire loan amount. At the same time, the notes enable the investors to spread their investment over several different loans, and thereby diversify their portfolio.

Diversification reduces the risk, but this effect has not been added to the ECIR model, as the model assumes that one investor covers the entire loan. However, we do not know if Lending Club has considered the diversification effect in their own model.

In a scenario where Lending Club’s model does not consider the diversification effect, both the Lending Club interest rates and the model-given interest rates should have been adjusted to a lower level. The spread between the two prices would remain the same as presented in the results section, assuming that the diversification effect would have an equal effect on both Lending Club and the model.

In a scenario where Lending Club’s model does consider the diversification effect, the ECIR model is the only model were the interest rate should be adjusted to a lower level. Which means that the spread between the interest rates set by Lending Club and the model-given interest rates should be even larger than presented in the results section.

Knowing that the diversification effect is not considered in the ECIR model, the differences in interest rates between Lending Club and the ECIR model are either as presented, or even more significant. After accounting for the various explanations, including the possibility for diversification, we find that it is not unlikely that investors in P2P lending are earning excess returns.

7. Robustness

As we have applied both a model and theory from literature on bonds on P2P loans, we find it suitable to address the robustness of the ECIR model and to perform a sensitivity analysis on the input variables.

7.1 Sensitivity Analysis

In this section, we will look at the model's sensitivity to the different input variables, to study the robustness of the model. As explained, the implied interest rates are only quantifiable through the solver tool in Excel, and cannot be easily implemented on every single loan. Therefore, we return to the concept of prices for this sensitivity analysis. Using what-if simulations, the price output is tested by changing the input variables default premium (η), interest rate volatility (σ^2), risk-free rate (r) and the β which is mainly connected to the slope of the interest rate curve. These four variables are chosen because we find that there is some uncertainty connected to the calculation of the estimates. The input estimates are changed in a setting where everything else is kept the same (ceteris paribus). For the other input variables, we believe that the previously described calculation methods are sufficiently solid.

The model is most sensitive to changes in the default premium, although it is not very sensitive to any of the four chosen variables. Even when doubling the default premium, the change in model price is too low to bring the model price below the price observed at the Lending Club platform. Changes connected to the risk-free rate are even less severe, this results in changes in the LC ratio of only 0.01. These changes are small relative to the existing size of the spread which means that the conclusion of the result does not change significantly: that the P2P loans issued at the Lending Club platform are priced lower than the ECIR model predicts. No matter which of the three major inputs that are changed, the result is still a model price that is higher than the LC price. This result is exemplified in the appendix, section D. The sensitivity analysis supports the main result from the previous sections, even though the size of the spread between Lending Club and the ECIR model are mildly affected by input changes.

7.2 Fitting the Model

In addition to the sensitivity analysis we want to study what the default rate has to be for the model price to equal the LC price. If we find that this default rate is unrealistically high, it is an argument in favour of our previous findings: that Lending Club sets their prices lower than the prices suggested by the ECIR model. All input variables except the default premium is kept constant, including the model price relative to the LC price (LC ratio = 1).

Table 13 shows the results from fitting the model to three-year loans, segmented by rating grades. Default rate is the estimate of default rate used throughout this paper, while Implied default rate is the result as we solve for interest rate with LC ratio equal to 1. Implied default rate with the call premium deducted is the same as implied default rate, only after we have taken the call option into account. A corresponding table to the five-year loans can be found in section E of the appendix, as the analysis below applies to both maturities.

Table 13: Fitting the model, three-year loans

Rating grade	Default rate	Implied default rate	Implied default rate with the call premium deducted
A	1.96 %	8.25 %	5.50 %
B	2.59 %	14.17 %	10.86 %
C	2.50 %	19.72 %	15.53 %
D	3.06 %	25.96 %	21.03 %
E	4.08 %	29.14 %	24.40 %
F	3.94 %	34.60 %	28.73 %
G	4.96 %	36.41 %	31.10 %

We find that the increase in default premium is quite large. It indicates that the default rate at the Lending Club platform must be significantly higher than the ECIR model (and Lending Club's websites) suggests, for the output price from the model to equal Lending Club's price. However, we make a simplification by only counting the status "Charged Off" as default. Had statuses such as "late" or "in grace period" been counted as default as well, the default rates could have been higher.

Looking at the loan book of Lending Club, there are not enough loans with the statuses of "late" or "in grace period" to increase the default rate as much as the difference between the historical default rate and the implied default rate in table 13 implies. This is also true when taking the call premium into account. There is a possibility that the default premiums are higher than the estimates used throughout this paper, but the estimates are not likely to be high enough for the ECIR model price to equal the LC price. Therefore, fitting of the model supports the previous results from this paper: there seems to be a considerable spread between the two prices.

7.3 Criticism

It follows from the sensitivity analysis that there are reasons to believe that the output from the model are robust with regards to uncertainty about inputs. This were supported when fitting the model to test the quality of the results. But is the ECIR model a good way to price P2P loans? We have presented several arguments for this earlier in the paper, but realise that a deeper analysis of Lending Club's credit rating model could be helpful to fully understand the reasons behind the spread.

Also, the fact that P2P lending is quite new entails several aspects of uncertainty related to the limited time window of the data. This makes it difficult to state the correct price of P2P loans, and thus difficult to conclude on the correctness of the LC price.

Lastly, we have only looked at the business loans and only at the Lending Club platform. For future research, it would be interesting to also study other platforms and market segments.

All things considered, we still believe that the assumptions, simplifications and estimations necessary to perform the analysis of this paper has been accounted for and controlled, and that the paper is a relevant contribution to the literature on P2P lending.

8. Conclusion

We started by looking at what P2P Lending is and how it works. As a new type of intermediary in the financial sector we were particularly interested in the opportunities for the investors. We wanted to get a deeper understanding of the returns an investor receives and the risk that comes with investing in P2P loans, to shed some light on a new and uncharted participant in the financial services industry.

To study the returns on investing in P2P lending, we applied a reduced form model approach. The chosen model was set as a benchmark for pricing of risk, assuming a no-arbitrage argument. The prices and interest rates observed at the Lending Club platform were then analysed relative to the model, and we used the comparison to address investor compensation.

We found that the observed returns were higher than what the model suggested was necessary to compensate an investor for the risk they had taken on. Further, we found this to be true for all the loans in Lending Club's loan book. As the model is a simplification of reality with rather strict assumptions, we expected this result. However, the size of the spread and the possible explanations were interesting, as this could indicate whether investors were earning excess returns. Furthermore, the compensation investors received increased relative to the model-given benchmark as the loans got riskier. This were true both when the rating grade of the borrower decreased, and when the maturity of the loans increased.

The reasons for the difference in interest rates are difficult to conclude upon, but several possibilities have been addressed. These includes that Lending Club perceives the risk itself (or the investors' attitude towards risk) differently than to the ECIR model does, because of aspects such as assymetric information, callable risk and the possibility for diversificaiton. Inefficiency in the P2P lending market or Lending Club's search for profits could also help explain the spread.

As the academic research on P2P lending are pretty much non-existing and the information availability is rather scarce, it is problematic to calculate the net effect of the various risk elements involved. More extensive research and more information on Lending Club's credit model is required to conclude on one thing or another. However, we have proposed and applied a model, and analysed accompanying aspects, for studying the risk-return relationship when investing in P2P loans. Because P2P lending is new and unknown to most investors, it seems plausible that Lending Club intentionally sets the interest rates quite high in order to attract investors. Especially since we have pointed out inefficienies in the market that would allow them to execute such a strategy. In sum, we find that it is not unlikely that investors investing in P2P loans has been earning excess returns, taken into account the risk they have taken on.

Appendix

A. Descriptive Statistics

Table: Descriptive statistics for loans with three-year maturity

Rating grade	Average LC ratio	Max LC ratio	Min LC ratio
A	0.9491	0.9791	0.9360
B	0.9117	0.9416	0.8899
C	0.8716	0.8915	0.8521
D	0.8349	0.8530	0.8171
E	0.8150	0.8285	0.7846
F	0.7679	0.7874	0.7464
G	0.7480	0.7568	0.7258

Table: Descriptive statistics for loans with five-year maturity

Rating grade	Average LC ratio	Max LC ratio	Min LC ratio
A	0.9168	0.9168	0.9168
B	0.8677	0.9202	0.8453
C	0.8307	0.8631	0.8095
D	0.7647	0.7885	0.7377
E	0.7281	0.7494	0.6911
F	0.6747	0.7001	0.6458
G	0.6368	0.6513	0.6116

In the descriptive statistics, we present the results for each of the seven rating classes, separating the loans by maturity of three and five years. The “Average LC ratio” is the average LC ratio within each rating class. “Max”/”Min” is the maximum/minimum value of the LC ratio observed in the given rating class. Notice how all of the “Max LC ratio” are less than 1, which tells us that all of the prices set by Lending Club are lower than what the model suggests.

B. Black and Scholes model

The Black and Scholes model has been presented theoretically earlier in this paper. The following section will present the mathematical formulation of the same model.

$$C = SN(d_1) - N(d_2)Ke^{-rt} \tag{1}$$

Where

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}} \tag{2}$$

$$d_2 = d_1 - \sigma\sqrt{t} \tag{3}$$

The *C* represents the premium of the option, which in this case is a European call option. *S* is the current value of the underlying asset while *K* represents the strike price. The risk-free rate is represented by *r*, while the volatility of the underlying asset is the σ^2. N is the cumulative standard normal distribution, and *ln* represents the natural logarithm.

When applying the Black and Scholes model we need to know the following parameters: Time until expiration, risk-free interest rate, the volatility of the underlying asset, the current value of the underlying asset and the strike price at exercise. We will now present how we estimate these inputs in our application of the Black and Scholes Model on the P2P loans. The quantification of the parameters is not shown as they are dependent loan specific data, hence the parameters are in many cases individually adapted to each loan.

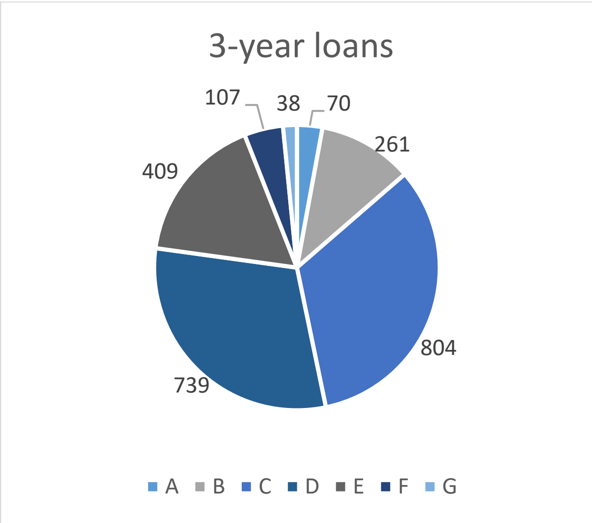
The exercise date of the option could be any time during the loans maturity, but we must assume an exercise date since the model only prices European options. For simplicity, we assume a time to exercise halfway through the maturity of the loan. This means 1.5 years for the three-year loans and 2.5 years for the five-year loans.

Both the risk-free rate and the volatility of the underlying asset has been calculated when we applied the ECIR model. The same estimates will be used for the Black and Scholes model as well. This means that the three-year loans use estimates based on the monthly risk free rates from the three-year U.S. treasury rate, while the five-year loans use estimates from the five-year U.S. treasury rate. Monthly averages for the risk-free rate is calculated and applied to each loan depending on which month the different loans were issued. The volatility of the underlying asset is estimated through quadratic variation on the two U.S. treasury rates.

The underlying asset is the P2P loan and the current value of this asset could be represented by the model price as calculated in the ECIR model. As previously mentioned, this model price is the discounted value of the future cash flows, which theoretically would represent the price of the asset. However, as we assume that the option can be exercised halfway into the loans maturity, the options only affect half of the loan, hence only half of the future cash flows are used to estimate the current value of the underlying asset.

The exercise price is estimated by taking the “Funded amount” variable from Lending Club’s loan book (the size of the loan) and taking away all down payments until time of exercise. The interest payments of the annuity cash flows are not withdrawn, only the part of the cash flow that are considered a down payment. The borrowers can at any time pay down the outstanding amount of the loan, which means that the exercise price naturally would equal the outstanding amount of the loan.

C. Number of loans divided by maturity and rating class.



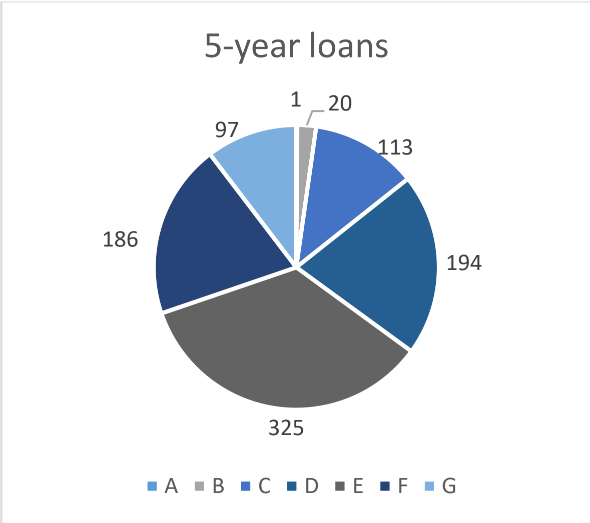


Figure: Number of loans per rating grade

D. Sensitivity Analysis Examples

Examples of how the price reacts to changes in the default premium is presented in the figure.

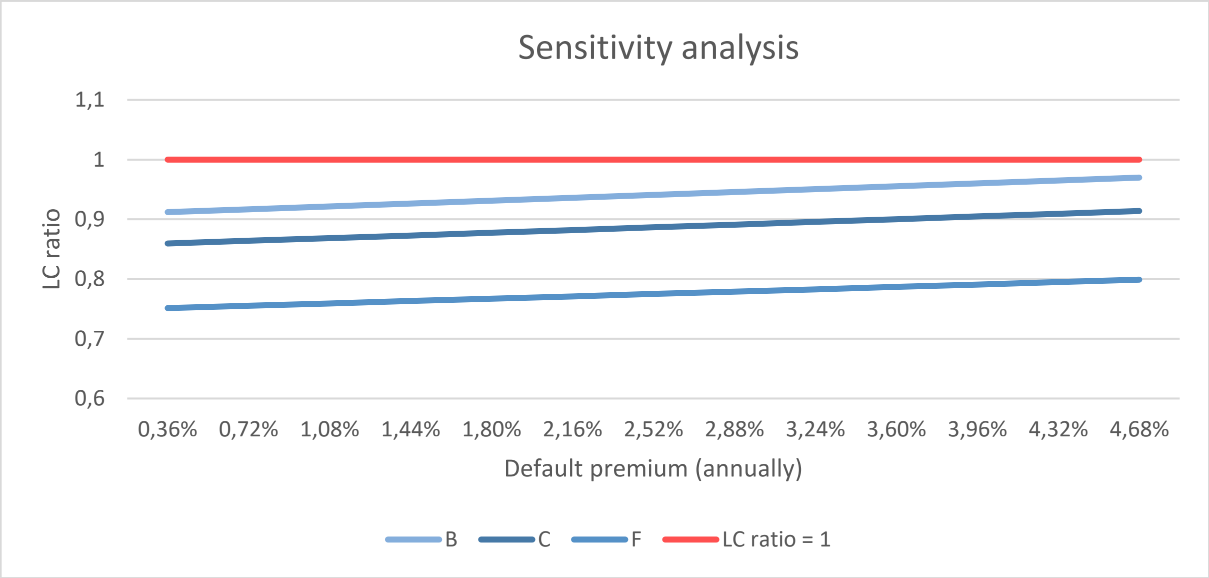


Figure: Changing the default premium

The horizontal axis represents various default premiums, while the vertical axis represents the LC ratios. An LC ratio of 1, indicates that the LC price and the model price are the same. Scenarios where the LC ratio rises above 1, means that the previous conclusion of higher interest rates at Lending Club, changes. As we can see from the figure above, the slopes of the lines are not very steep. This means that the loans are not that sensitive to small changes in the default premium. Most of the loans are estimated with an LC ratio less than 0.9, which means that even a scenario where the default premium doubles will not push the LC ratio above the benchmark.

E. Fitting the Model

Table: Fitting the model, five-year loans

Rating grade	Default rate	Implied default rate	Implied default rate with the call premium deducted
A	3.85 %	10.32 %	6.04 %
B	3.97 %	12.58 %	8.94 %
C	4.60 %	14.83 %	11.55 %
D	3.79 %	18.49 %	14.08 %
E	4.17 %	21.25 %	16.60 %
F	4.61 %	25.47 %	20.28 %
G	4.91 %	29.31 %	23.38 %

The fitting of the model was exemplified through the three-year loans in section 7.2 The table above shows the same calculation for the five-year loans.

References

Allen, F. and Santomero, A. (2001). What do financial intermediaries do? *Journal of Banking & Finance*, 25(2), pp.271-294.

Altman, E. (2006). Credit Risk and the Link between Default and Recovery Rates. *CFA Institute Conference Proceedings Quarterly*, 23(3), pp.34-43.

Barone, E. and Risa, S. (1994). Valuation of Floaters and Options on Floaters under Special Repo Rates. *SSRN Electronic Journal*.

Barone, E., Barone-Adesi, G. and Castagna, A. (1998). Pricing Bonds and Bond Options with Default Risk. *SSRN Electronic Journal*.

Berger, S. and Gleisner, F. (2009). Emergence of Financial Intermediaries in Electronic Markets: The Case of Online P2P Lending. *Bus Res*, 2(1), pp.39-65.

Berk, J. and DeMarzo, P. (2014). *Corporate finance*. 3rd ed. Boston, Mass.: Pearson.

Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3), pp.637-654.

Brown, K., Harlow, W. and Tinic, S. (1988). Risk aversion, uncertain information, and market efficiency. *Journal of Financial Economics*, 22(2), pp.355-385.

BROWN, S. and DYBVIG, P. (1986). The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates. *The Journal of Finance*, 41(3), pp.617-630.

Cox, J., Ingersoll, J. and Ross, S. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2), p.385.

Dietsch, M. and Petey, J. (2002). The credit risk in SME loans portfolios: Modeling issues, pricing, and capital requirements. *Journal of Banking & Finance*, 26(2-3), pp.303-322.

Duffie, D. and Huang, M. (1996). Swap Rates and Credit Quality. *The Journal of Finance*, 51(3), p.921.

Duffie, D. and Singleton, K. (1999). Modeling Term Structures of Defaultable Bonds. *Rev. Financ. Stud.*, 12(4), pp.687-720.

Financial Industry Regulatory Authority, (2016). *Understanding Bond Risk | FINRA.org*. [online] Finra.org. Available at: http://www.finra.org/investors/understanding-bond-risk [Accessed 18 Oct. 2016].

Forbes, (2016). *Forbes Welcome*. [online] Forbes.com. Available at: http://www.forbes.com/sites/rogeraitken/2016/07/22/lending-club-scandal-provokes-major-uk-peer-to-peer-investigation/#37ec5f21fc3d [Accessed 28 Nov. 2016].

Haight, F. (1967). *Handbook of the Poisson distribution*. New York: Wiley.

Hotchkiss, E. and Ronen, T. (2002). The Informational Efficiency of the Corporate Bond Market: An Intraday Analysis. *SSRN Electronic Journal*.

Inc, M., Koller, T. and Goedhart, M. (2015). *Valuation*. Hoboken: Wiley.

Jarrow, R. and Protter, P. (2004). Structural Versus Reduced Form Models: A New Information Based Perspective. *Journal of Investment Management*, Vol. 2(No. 2, (May 2004), pp.pp. 1-10.

Jarrow, R. and Turnbull, S. (1995). Pricing Derivatives on Financial Securities Subject to Credit Risk. *The Journal of Finance*, 50(1), p.53.

Lando, D. (2004). *Credit risk modeling*. Princeton, NJ: Princeton University Press.

Lendingclub.com. (2016). *Earn Solid Returns - Lending Club*. [online] Available at: https://www.lendingclub.com/site/investing/solid-returns [Accessed 19 Sep. 2016].

Lendingclub.com. (2016). *Lending Club Statistics | Lending Club*. [online] Available at: https://www.lendingclub.com/info/download-data.action [Accessed 9 Nov. 2016].

Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, 29(2), p.449.

Meucci, A. (2009). Review of Statistical Arbitrage, Cointegration, and Multivariate Ornstein-Uhlenbeck. *SSRN Electronic Journal*.

Roig Hernando, J. (2016). Crowdfunding: The collaborative economy for channelling institutional and household savings. *Research in International Business and Finance*, 38, pp.326-337.

Sarkar, M., Butler, B. and Steinfield, C. (2006). Intermediaries and Cybermediaries: Sarkar, Butler and Steinfield. *Journal of Computer-Mediated Communication*, 1(3), pp.0-0.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), pp.177-188.

Wehinger, G. (2014). SMEs and the credit crunch. *OECD Journal: Financial Market Trends*, 2013(2), pp.115-148.

Wehinger, G. (2014). SMEs and the credit crunch. *OECD Journal: Financial Market Trends*, 2013(2), pp.115-148.

Wharton Research Data Services, (2016).