Bubble Bank

Osadchiy, Maksim and Sidorov, Alexander

Corporate Finance Bank (Russia, Moscow), Novosibirsk State University and Sobolev Institute of Mathematics (Russia, Novosibirsk)

22 July 2019

Bubble Bank vs Goody Bank: Structural Model of Credit Risk *

Maksim Osadchiy[†], Alexander Sidorov[‡]

November 5, 2019

Abstract

The paper studies a Vasicek-Merton model of a bank. The model allows us to specify the sets of exogenous parameters, which generate bubbles in the credit market, or contribute to formation of stable banks with self-restrictive behavior that do not require regulatory intervention. An assessment of capital structure and probability of default is carried out in terms of exogenous factors.

Key words: Banking microeconomics, Credit bubble, Probability of default, Capital adequacy ratio

JEL codes: G21, G28, G32, G33

Introduction

The paper studies an initiation of credit bubbles and ability of banks to be "goody", i.e., self-restraint from unlimited credit expansion without regulatory intervention.

The motivation of the paper is to include the relatively realistic Vasicek loss density function into a microeconomic model of a bank, taking into account the correlation of the assets of borrowers to assess the risk of default of the bank, as well as to study its capital structure. However, in the paper we will consider the more general form of the loss density function, not limited to the Vasicek function only.

Our work is based on four stylized facts related to the field of credit risks:

(1) Many banks tend to blow bubbles - to carry out dangerous credit expansion. They turn into "black holes" - fast-growing banks with unobservable negative capital. This problem is especially acute in developing countries with poor quality of banking regulation and supervision.

*We thank Henry Penikas for useful comments and suggestions, and Dirk Tasche for a kind assistance in justification of Lemma. Authors are thankful to participants of the seminar in Sobolev Institute of Mathematics for useful discussion.Authors are thankful to participants of the seminar in Sobolev Institute of Mathematics for useful discussion. This study has been funded by Russian Foundation for Basic Research, Project No. 18-010-00728.

[†]Corporate Finance Bank (Russia, Moscow). Email: OsadchyMS@cfb.ru

[‡]Novosibirsk State University and Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences (Russia, Novosibirsk) Email: alex.v.sidorov@gmail.com

(2) If credit expansion covers the entire banking sector, a credit boom begins. If it is not stopped in time, it will inevitably end in a credit crisis. History knows hundreds of examples of such crises (see, e.g., [17, pp. 344-347]).

(3) Credit crises lead to high social costs – unemployment, impoverishment, social instability. To avoid such severe consequences, the regulators limit the expansion of banks by micro-prudential policies, in particular, limiting capital adequacy ratio (CAR) from below.

(4) Instead of maximally using a capital and increasing risk assets up to the level close to the capital adequacy requirements imposed by the regulator, many banks far exceed CAR regulatory requirements. CAR has a wide spread [4, Graph I.13]: many banks are under-capitalized, and many banks are overcapitalized.

The key research questions of the paper: (1) do banks capable to self-restraint without the intervention of the regulator? (2) what conditions lead to the formation of credit bubbles? (3) why there is a wide spread of CAR?

To answer the questions, consider a hypothetical bank operating without external regulation. We build a simple microeconomic model of the bank that takes into account only credit risk. The model is constructed on the base of the Merton model (the structural model of credit risk) and the Vasicek model (the loan portfolio value model).

The bank is considered as a "Merton-Vasicek bank": the Merton firm's assets following the geometric Brownian motion are replaced with a portfolio consisting of loans to Merton firms. Loss of this portfolio under some assumptions is described by the Vasicek loss distribution. The main advantage of the Vasicek model is taking into account the effect of borrowers' assets correlation on the credit risk of the bank.

The Merton model considers liabilities of a firm as a portfolio of risk-free zero-coupon bond and a short put-option written on the assets of the firm. The option generates loss. Hence it makes possible to assess liabilities of the firm using Black-Scholes theory of option valuation.

It turned out that in the absence of liquidation costs a banker has no incentive to stop credit expansion. Liquidation costs make the incentive, and if its level is sufficient, the bank becomes "goody" – self-restraint and stable.

Similarly to the Merton model, the Merton-Vasicek model allows to assess the probability of the bank default. Maximization of the expected final capital of the Merton-Vasicek bank allows to determine the capital structure of the bank. Within the framework of the model, we construct the exogenous parameters zones where the bank is capable to be self-restraint, and where the bubble inflates – the unlimited expansion takes place. We study the structure of these areas. We study the dependence of the probability of default of the bank and its CAR on the parameters.

Using the model, we describe the mechanism of CAR choice with comparative statics of the banker's decision and of the probability of the bank's default with respect to the model parameters: the interest rates of attraction and allocation of resources, the correlation of borrowers' assets and other factors.

The microeconomic model of the bank can serve as a useful tool for prediction of regulatory intervention consequences. Also it can be used for prediction of banks' reaction on exogenous parameters shocks.

The paper is organized as follows. Section 1 presents the Merton-Vasicek model. In Section 2, we consider the banker's maximization problem, which allows to determine the equilibrium value of CAR, and study the comparative statics of equilibrium characteristics both analytically and using computer simulation. Section 3 is devoted to a visual graphical classification of decisions on the main parameters. The most important result is the determination of compliance with the requirements of Basel III. The main results and conclusions of the work are formulated in the Conclusion. A list of notations and abbreviations is placed at the beginning of Appendix.

The related literature

There are two main types of credit risk modeling: structural and reduced form (intensity) models. The aim of structural approach is to provide a relationship between default risk and capital structure, unlike the reduced form models, which consider the credit default as exogenous event driven by a stochastic process. Reduced form models do not consider the relation between default and firm value in an explicit manner. Intensity models represent the most extended type of reduced form models. In contrast to structural models, the time of default in intensity models is not determined via the value of the firm, but it is the first jump of an exogenously given jump process. The parameters governing the default hazard rate are inferred from market data.

Structural models, pioneered by Black, Scholes [6]and Merton [15], ingeniously employ modern option pricing theory in corporate debt valuation. Merton model was the first structural model and has served as the cornerstone for all other structural models, including ours. A significant extension of Merton was represented by Black and Cox in [5], who managed to relax some of the Merton's assumptions.

The next major step in generalizing the structural models was an important contribution of Leland in [13], who explicitly introduced corporate taxes and bankruptcy costs, which may be interpreted as liquidation costs. Thus, he formalized the trade-off framework and provided a way to determine both the optimal default boundary and the value-maximizing optimal capital structure. Since these classical papers the further advances of structural models in various direction. It worth to mention that firms often make their decisions in a principal agent setting, wherein managers, equity holders (borrowers), and creditors may have very different objective functions. The resulting agency problems may have significant implications for the optimal capital structure decisions and optimal contracting decisions, see the papers [9, 10]. The paper [1] develops a structural model of a financial institution that can invest in both liquid and illiquid assets dynamically, maximizing the profit of its shareholders while satisfying some regulatory constraints. It is proved that tightening the liquidity constraint adversely affects their rates of return, while preventing some large losses that occur when the portfolio is very illiquid.

The paper [2] analyses how banking firms set their capital ratios, that is, the rate of equity capital over assets. In order to study this issue, two theoretical models are developed: the "market" model for banks not affected by capital adequacy regulation, while the second one, the regulatory model, explain the behavior of banks with an optimal market ratio below a legally required regulation.

Regulation related to capital requirements is an important issue in the banking sector. One of the induces used to measure how susceptible a bank is to failure, is the capital adequacy ratio (CAR). In general, this index is calculated by dividing a measure of bank capital by an indicator of the level of bank risk. The papers [11] and [16] consider the application of stochastic optimization theory to asset and capital adequacy management in banking and construct continuous-time stochastic models for the dynamics of capital adequacy ratio established by Basel II. This ratio is obtained by dividing the bank's eligible regulatory capital (ERC) by its risk-weighted assets (RWAs) from credit, market and operational risk. The contribution of paper [7] is the construction of a stochastic dynamic model to describe the evolution of bank capital that incorporates capital gains and losses. The gains and losses are represented by loan loss reserves and the unexpected loan losses, respectively. It is studied the optimal capital management problem which maximizes the expectation of bank capital under a risk constraint on the Capital-at-Risk (CaR), where CaR is defined in terms of Value-at-Risk (VaR).

The issue of bank capital adequacy and risk management within a stochastic dynamic setting is studied in the paper [8]. An explicit risk aggregation and capital expression is provided regarding the portfolio choice and capital requirements special context. This leads to a nonlinear stochastic optimal control problem whose solution may be determined by means of dynamic programming algorithm.

Along with the great impact of structural models on the theory of the credit risk and its application, there is reasonable criticism on the prediction power of such models concerning to the pricing of corporative bonds, see, for example, [12]. On the other hand, the structural models are able to predict well the hedge ratios of corporate bonds against the equity of the underlying firm, see [18].

1 The Vasicek-Merton model

In the Merton model a firm with the assets A_t and debt D pays to a lender $\min\{A_T, D\}$ at a maturity T. If $A_T \geq D$, then the debt is paid in full – in the amount of D, and in case of default, if $A_T < D$, the debt is paid partially, in the amount of A_T. Since

$$\min\{A_T, D\} = D - (D - A_T)^+,$$

then the risky debt is equivalent to the portfolio with a risk-free zero-coupon bond with a face value D and a maturity T, and a short European put option with strike D and expiration T, written on the firm's assets. The put option generates loss to the investor. If in case of default the debt is paid in the amount of A_T, then we call such obligations the obligations of the first kind.

In the Merton model, assets A_t follow a geometric Brownian motion. Such an interpretation allows using the results of the Black-Scholes theory of options valuation to assess the liabilities and equity of the firm. We consider a modification of the Merton model: if $A_T \geq D$, then the debt is paid in full – in the amount of D, but in case of default, $A_T < D$, the debt is not paid at all. It means that the firm pays the debt in the amount of $D - D * 1_{A_T < D}$, which allows us to interpret this situation in the same way: the lender essentially buys a risk-free zero-coupon bond with a face value T and a maturity T, and at the same time shortens binary put option with strike D and expiration T, written on the firm's assets. If in case of default the debt is not paid, then we call such obligations the obligations of the second kind.

On the base of the Merton model we construct the simplest stochastic model of the bank, taking into account the credit specifics of its activities, including the diversification of credit risk. To do this, we make another change to the Merton model: now let the assets of the firm — in this case, the bank — consist exclusively of a loan portfolio. In addition, suppose that the loan portfolio consists of liabilities of Merton firms of the second kind. The loss distribution function of such a loan portfolio, if the number of firms tends to infinity, is well understood - this is the Vasicek function. The Merton model modified in this way we call the Merton-Vasicek model.

We consider two versions of the model. First, the main one - in case of default of the bank, the banker bears all the costs of bank bankruptcy. The second is the bail-out case when, in the event of a bank default, the bankruptcy costs of the bank are not borne by the banker or bank's creditors, but by a third party (e.g., the state).

Borrowers' behavior

Prior to formulate the main model of this paper, let's recall the Vasicek approach[1] [20] to modeling of the borrower's sector as a set of the Merton firms (see [15] for details) with correlated assets, which will be considered as underlying layer of our model. Let's assume that the bank provided loans for n firms. The i-th firm's assets A_i are described by the Brownian process

$$dA_i = \mu_i A_i dt + \sigma_i A_i dW_i$$

where μ_i is the mean rate of return on the assets and σ_i is the asset volatility, W_i is a standard Wiener process (Brownian motion). The liabilities of firm i are given by the single bond with face value D_i. This loan defaults if the value of the assets

$$A_i(T) = A_i(0) \exp\left(\mu_i T - \frac{1}{2}\sigma_i^2 T + \sigma_i \sqrt{T} X_i\right),$$

at the loan maturity T falls below the debt D_i, where $A_i(0) > 0$ is the initial value of assets, X_i is a standard normal variable. This implies that the probability of default of the i-th loan is

$$PD_i = \mathbb{P}(A_i(T) < D_i) = \Phi\left(\frac{\ln D_i - \ln A_i(0) - \mu_i T + \frac{1}{2}\sigma_i^2 T}{\sigma_i \sqrt{T}}\right),$$

where $\Phi(z)$ is a standard normal distribution function. In turn, the variables X_i are jointly standard normal with equal pairwise correlations ρ, and can therefore be represented as

$$X_i = \sqrt{\rho} Y + \sqrt{1-\rho} Z_i$$

where Y, Z_1, Z_2, …, Z_n are mutually independent standard normal variables. The variable Y can be interpreted as a portfolio common factor, such as an economic index. Then the term $\sqrt{\rho}Y$ is the firm's exposure to the common factor and the term $\sqrt{1-\rho}Z_i$ represents the firm specific risk.

We assume that maturity $T = 1$, all firms are identical, in particular, $PD_i = PD$ for all i, and the loss given default $LGD = 1$. Random Bernoulli variable ε_i takes two possible values: $\varepsilon_i = 1$, if loan i is defaulted (with probability PD), and $\varepsilon_i = 0$ (with probability $1 - PD$) otherwise. Then a random variable

$$\varepsilon = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i$$

characterizes the share of nonperforming loans. It is obvious that $\mathbb{E}(\varepsilon) = PD$, $\varepsilon \in [0, 1]$. The

[1]The following considerations are well-known and are adduced for completeness of the statements.

cumulative distribution function of loan losses of a very large portfolio is

$$\mathbb{P}(\varepsilon < z) = F(z; PD, \rho) = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(z) - \Phi^{-1}(PD)}{\sqrt{\rho}}\right), \tag{1.1}$$

see [20] for details. The corresponding PDF is as follows

$$f(z; PD, \rho) = \sqrt{\frac{1-\rho}{\rho}} \exp\left(\frac{1}{2}\left[(\Phi^{-1}(z))^2 - \left(\frac{\sqrt{1-\rho}\Phi^{-1}(z) - \Phi^{-1}(PD)}{\sqrt{\rho}}\right)^2\right]\right). \tag{1.2}$$

The Vasicek distribution of losses satisfies the following condition, which will be crucial for our study.

Lemma. *Let $0 < \rho < 1$, $0 < PD < 1$ and $f(z; PD, \rho)$ be the PDF of the Vasicek distribution of losses. Then the function $(1-z)f(z; PD, \rho)$ decreases with respect to z in interval $PD < z < 1$.*

Proof. See Appendix A.1.

Probability of the bank default

We study a single-period model in which a bank is created at the initial moment $t = 0$ with the initial capital $K_0 > 0$. At the same time the banker chooses the amount $D_0 \geq 0$ of the attracted deposits at the interest rate $R > 0$, and then places the borrowed and its own funds in the uniform loans of the same size at the interest rate $r > R$ before the terminal moment $t = 1$, hence, the loans are equal to $L_0 = K_0 + D_0$. The supply of loans and the demand for deposits are satisfied in full, and interest rates are exogenous parameters. At the moment $t = 1$ all loans are repaid, except for those defaulted, hence $L_1 = 0$, and all assets acquire the form of cash M_1. After the deposits are returned and interest is paid at the rate R, the bank's capital becomes equal to $K_1 = M_1 - D_1$. In framework of the model we assume that deposits can not be withdrawn prematurely, i.e., there is no liquidity risk, hence $D_1 = (1+R)D_0$.

The borrowers are considered as Merton firms with correlated assets considered in the previous subsections. The random losses of bank due to possible defaults of the firms are determined by the variable ε with the corresponding loss distribution function $F(z)$.

Given $D_0 = L_0 - K_0$, this means that

$$K_1 = (r - R - \varepsilon(1+r))L_0 + (1+R)K_0.$$

Note that the condition of the bank default

$$K_1 \leq 0 \iff \varepsilon \geq \frac{r-R}{1+r} + \frac{1+R}{1+r}\frac{K_0}{L_0}. \tag{1.3}$$

In other words, the default takes place if the loss exceeds the certain threshold, which depends on the Capital Adequacy Ratio (CAR)

$$k(L_0) = \frac{K_0}{L_0}.$$

We assume that the risk weight of loans is equal to 1.

To save space, let's introduce the following notification

$$\hat{\mathcal{E}} = \frac{r - R}{1 + r},$$

which implies

$$\frac{1 + R}{1 + r} = 1 - \hat{\mathcal{E}}$$

and let

$$\mathcal{E}(k) = \hat{\mathcal{E}} + (1 - \hat{\mathcal{E}})k.$$

Due to (1.3) $\mathcal{E}(k(L_0))$ may be interpreted as a threshold value of loss, which triggers the bank default. The value $\hat{\mathcal{E}} = \mathcal{E}(0)$ may be also interpreted as the limit loss threshold when the loan portfolio increases unrestrictedly, because $\lim\limits_{L_0 \to \infty} k(L_0) = 0$.

Now assume that the CAR k is an exogenous constant[2]. Then the probability of the bank may be calculated as follows

$$p(k) = \mathbb{P}(\varepsilon \geq \mathcal{E}(k)) = \int_{\mathcal{E}(k)}^{1} f(z)\mathrm{d}z = 1 - F(\mathcal{E}(k); PD, \rho). \tag{1.4}$$

Obviously, the probability of the bank default $p(k)$ decreases with respect to k. Moreover,

$$\mathcal{E}(k) = k + \frac{r - R}{1 + r}(1 - k),$$

which implies that $\mathcal{E}(k)$ increases with the loan rate r and decreases with the deposit rate R. Consequently, probability of the bank default decreases with r and increases with R.

2 The banker's problem

Considering the Vasicek function of losses as the main example, we generalize this approach to the general random variable ε with PDF $f(z)$ such that $(1-z)f(z)$ decreases on the interval $(PD, 1)$, where $PD = \mathbb{E}(\varepsilon)$ is the expected value of losses. Suppose that the bank's default implies the additional losses, moreover, the banker is "responsible" i.e., bears all the costs of bankruptcy. More precisely, if

[2] The endogenous optimum choice of k will be considered in the next section.

random amount of bank capital at the end of the period

$$\bar{K}_1 = (1+r)L_0(1-\varepsilon) - (1+R)D_0$$

is positive, i.e., if $\varepsilon < \mathcal{E}(k(L_0))$, then the terminal capital is equal to $\bar{K}_1$. Otherwise, in case of default i.e., $\bar{K}_1 \le 0$, the bank sells a loan portfolio with a discount $0 \le d \le 1$. The value $d = 0$ corresponds to the case when there is no additional loss, i.e., $\hat{K}_1 = \bar{K}_1$. In what follows it will be shown that assumption $d = 0$ implies, with rare exception, the formation of credit bubble. As result, the terminal capital under default is equal to

$$\hat{K}_1 = \bar{K}_1 - d(1+r)L_0(1-\varepsilon).$$

Thus, the general definition of the terminal capital is

$$K_1 = \begin{cases} \bar{K}_1, & \\ \bar{K}_1 - d(1+r)L_0(1-\varepsilon), & \text{otherwise.} \end{cases}$$

The banker's problem is to maximize the expected value of the terminal capital $\mathbb{E}(K_1)$ under condition $D_0 \ge 0 \iff L_0 \ge K_0$. These expectation includes the expected additional loss in case of default, provided that $d > 0$, otherwise, the objective function coincides with

$$\mathbb{E}(\bar{K}_1) = (r - R - PD(1+r))L_0 + (1+R)K_0 = (1+r)\left[(\hat{\mathcal{E}} - PD)L_0 + (1-\hat{\mathcal{E}})K_0\right].$$

In case of $d > 0$, the expected terminal capital is equal to

$$\mathbb{E}(K_1) = (1+r)\left[(\hat{\mathcal{E}} - PD)L_0 + (1-\hat{\mathcal{E}})K_0 - d \cdot L_0 \int_{\mathcal{E}(k(L_0))}^{1} (1-z)f(z)dz\right].$$

Given $f(z) = F'(z)$ is a PDF of the random losses ε, we may interpret the function

$$\mathrm{ret}(z) \equiv (1-z)f(z)$$

as a weighted share of the returned loans and let

$$\mathrm{Ret}(x) = \int_x^1 \mathrm{ret}(z)\mathrm{d}z.$$

The function $\mathrm{Ret}(x)$ is decreasing, because $\mathrm{Ret}'(x) = -\mathrm{ret}(x)$, and satisfies $\mathrm{Ret}(0) = 1 - PD$, $\mathrm{Ret}(1) =$

0. Therefore, the expected terminal capital $\mathbb{E}(K_1) = (1+r)U(L_0)$, where

$$U(L_0) = (1-\hat{\mathcal{E}})K_0 + (\hat{\mathcal{E}} - PD)L_0 - d \cdot L_0 \mathrm{Ret}(\mathcal{E}(k(L_0))), \tag{2.1}$$

is the reduced objective function.

Now the original banker's problem is equivalent to

$$\max U(L_0) \text{ s.t. } L_0 \geq K_0.$$

Theorem 1. *Let $\hat{\mathcal{E}} > PD$ and the weighted share of returns $ret(z)$ is strictly decreasing on the interval $PD < z < 1$, then $U''(L_0) < 0$ for all $L_0 \geq K_0$. Moreover, if inequality*

$$d > \frac{\hat{\mathcal{E}} - PD}{Ret(\hat{\mathcal{E}})} \tag{2.2}$$

holds, then there exists unique solution of the banker's problem

$$L_0^* = \frac{K_0}{k^*},\ D_0^* = \frac{1-k^*}{k^*}K_0 \tag{2.3}$$

where $k^* \in (0,1)$ is an equilibrium CAR, defined as the unique solution of equation

$$FOC: \ \hat{\mathcal{E}} - PD - d \cdot \Big(Ret(\mathcal{E}(k)) + (1-\hat{\mathcal{E}})k \cdot ret(\mathcal{E}(k))\Big) = 0. \tag{2.4}$$

Proof. See Appendix A.2.

Theorem 1 implies that depending on the parameter's relation, there may be two possible cases, which cause different types of the banker's behavior.

Case 1. Let the discount d be sufficiently small, e.g., $d = 0$, so that condition (2.2) is violated. This implies an unrestricted increasing of objective function $U(L_0)$ when $L_0 \to +\infty$, which means that the banker has incentives for unrestricted expansion[3] of the loan portfolio L_0.

To prevent this negative trend, the regulator restricts lending by imposing the condition

$$K_0/L_0 \geq \hat{k} \iff L_0 \leq L_0(\hat{k}) = \frac{K_0}{\hat{k}}. \tag{2.5}$$

for the exogenously given CAR $\hat{k}$. It is obvious that in this case, the modified banker problem with additional constraint (2.5) has the solution $L_0^* = L_0(\hat{k})$.

[3] The same holds when banker is "irresponsible", i.e., does not want to pay his/her liabilities in case of default, i.e., $\hat{K}_1 = 0$.

Case 2. Assume that the discount d is sufficiently large to satisfy the condition (2.2) and, therefore, there exists solution $k^* \in (0,1)$ of (2.4), which endogenizes CAR. This situation can be interpreted as if the banker imposes self-restrain $L_0 \leq K_0/k^*$, which is active in the optimum, i.e., $L_0^* = K_0/k^* > K_0$.

Remark 1. The first order condition (2.4) may be interpreted in terms of the gains-losses as follows. Choosing the amount of the loans portfolio L_0 the banker obtains the expected total gains $(1-\hat{\mathcal{E}})K_0 + (\hat{\mathcal{E}} - PD)L_0$, thus the marginal gains from the further credit expansion are

$$\hat{\mathcal{E}} - PD = \frac{\tilde{r} - R}{1+r}.$$

On the other hand, the term

$$\mathrm{Ret}(\mathcal{E}(k(L_0))) = \int_{\mathcal{E}(k(L_0))}^{1} \mathrm{ret}(z)\mathrm{d}z = \mathbb{E}(1-\varepsilon|\ \varepsilon > \mathcal{E}(k(L_0)))$$

characterizes the share of the the loans returns *in case of the bank default.* Note that the size of the loan portfolio affects both the basis L_0 and the share of returns $\mathbb{E}(1-\varepsilon|\ \varepsilon > \mathcal{E}(k(L_0)))$, then the total marginal losses in case of the bank default are sum of effects: the marginal losses from increasing of basis L_0 are equal to

$$d \cdot \mathrm{Ret}(\mathcal{E}(k(L_0))),$$

while the losses from change of share are equal to

$$d \cdot L_0 \left(-\mathrm{ret}(\mathcal{E}(k(L_0))) \frac{\mathrm{d}}{\mathrm{d}L_0} \mathcal{E}(k(L_0)) \right) = d \cdot (1-\hat{\mathcal{E}}) k(L_0) \mathrm{ret}(\mathcal{E}(k(L_0))).$$

Finally, the gross marginal losses are equal to

$$d \cdot \left(\mathrm{Ret}(\mathcal{E}(k(L_0))) + (1-\hat{\mathcal{E}}) k(L_0) \mathrm{ret}(\mathcal{E}(k(L_0))) \right),$$

Thus, the FOC 2.4 is equivalent to coincidence of the marginal gains and the marginal losses.

2.1 Comparative statics of equilibrium

Now we study how the equilibrium reacts to the changes of the parameters d, R and r.

Proposition 1. *The signs of partial derivatives of the equilibrium values of k^*, L_0^*, D_0^* with respect to d, R, and r are as follows:*

$$\frac{\partial k^*}{\partial d} > 0, \ \frac{\partial k^*}{\partial R} > 0, \ \frac{\partial k^*}{\partial r} < 0$$

$$\frac{\partial L_0^*}{\partial d} < 0, \ \frac{\partial L_0^*}{\partial R} < 0, \ \frac{\partial L_0^*}{\partial r} > 0$$

$$\frac{\partial D_0^*}{\partial d} < 0, \ \frac{\partial D_0^*}{\partial R} < 0, \ \frac{\partial D_0^*}{\partial r} > 0$$

Proof. See Appendix A.3.

These results comply with intuitive expectations. For example, increasing of discount d suppress the banker's activity, forcing to reduce the loan portfolio L_0^* and the attraction of deposits. As expected, an increasing in the deposit interest rate R reduces the demand of deposit, while increasing in the loan interest rate r increases supply of loans, etc.

CAR dependence on the correlation

Now we focus on the case of Vasicek distribution of the loan losses (1.1), which is characterized by two parameters — ρ and PD — the borrower's asset correlation and the probability of borrower's asset default, respectively. From intuitive point of view, the larger is correlation ρ, the more restrictive banking policy is required. In other words, $k^*(\rho)$ should be increasing function, but it is not clear, whether the presented model catches this effect? The analytical way, like in Proposition 1, failed due to very tedious calculations, thus, the Figures 1a and 1b show the series of computer simulations. Figure 1a shows the curves, corresponding to the fixed discount $d = 0.5$ and three values of the default probability $PD = 0.06, \ 0.07, \ 0.08$. Similarly, Figure 1b shows the curves, corresponding to the fixed probability $PD = 0.075$ and three values of discount $d = 0.3, \ 0.5, \ 0.75$. As we see, increasing in both d and PD shifts the curves upwards. An interpretation of this effect is quite natural. Increasing in both cases implies the risk of default and/or the associated losses, which forces the "responsible" banker to be more safe and conservative. Note that Figure 1b agrees with Proposition 1 statement on $\frac{\partial k^*}{\partial d} > 0$.

The lack of solution in neighborhood of $\rho = 0$ and $\rho = 1$ is result of violation of the solvability condition (2.2). The direct calculations show that for all ρ sufficiently close to 0 or 1 the fraction $\frac{\hat{\mathcal{E}} - PD}{\mathrm{Ret}(\hat{\mathcal{E}})}$ exceeds $d = 0.5$. The values $k^* = 0$, i.e., the bottom points of the "arcs", correspond to the threshold values of ρ, that satisfy the identity

$$\frac{\hat{\mathcal{E}} - PD}{\mathrm{Ret}(\hat{\mathcal{E}}; PD, \rho)} = d.$$

The fact that "arc" of the plot starts from "bottom" point ρ_1 and ends at "bottom" point ρ_2, i.e., $k^*(\rho_1) = k^*(\rho_2) = 0$, means that the finite solution L_0^* of the bankers' problem exists for all $\rho_1 < \rho < \rho_2$. On the other hand, inequalities $\rho < \rho_1$ and $\rho > \rho_2$ imply the "bubble" $L_0 \to \infty$, which may be associated with $k^* = 0$. Therefore, we can extend the function $k^*(\rho)$ on the "non-existence" areas $\rho < \rho_1$ and $\rho > \rho_2$ putting $k^*(\rho) = 0$.

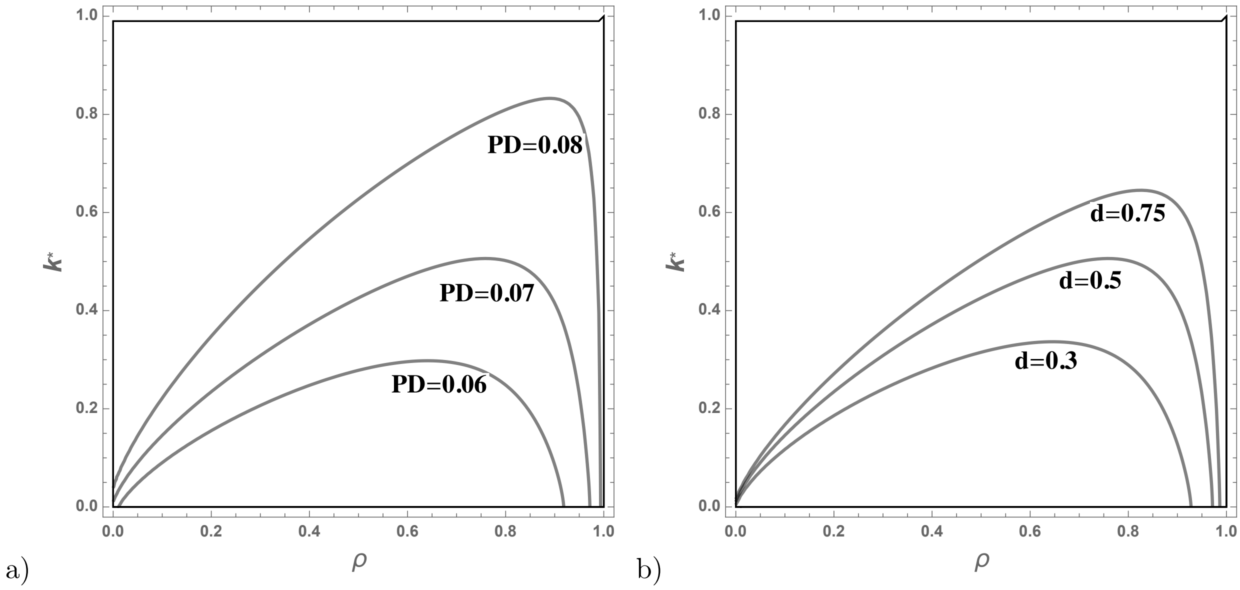

Figure 1: CAR k^* as a function of ρ, $R = 0.05$, $r = 0.15$, a) $d = 0.5$, b) $PD = 0.07$

2.2 Endogenous probability of the bank's default

The considered above optimum asset liability management is based on the risk-neutral behavior, targeted to maximize the expected terminal capital $\mathbb{E}(K_1)$, which is nominally greater than initial capital K_0, due to Theorem 1. However, the risk of default persists even if the management decisions are optimal. Due to (1.4) the probability of the bank's default is equal to

$$p = \mathbb{P}(\varepsilon \geq \mathcal{E}(k^*)) = 1 - F(\mathcal{E}(k^*)),$$

where k^* is solution of equation (2.4). Function $1 - F(\mathcal{E}(k))$ strictly decreases with respect to k, therefore, Proposition 1 implies that

$$\frac{\partial p}{\partial d} < 0, \ \frac{\partial p}{\partial R} < 0, \ \frac{\partial p}{\partial r} > 0,$$

which is quite intuitive. Note that the equilibrium values of k^* must be positive, therefore, the feasible values of the bank's probability of default satisfy inequality $p = 1 - F(\mathcal{E}(k^*)) < 1 - F(\hat{\mathcal{E}})$.

Focusing on the Vasicek distribution of losses, we can consider the comparative statics of probability p with respect to specific parameters — the correlation ρ and the probability of borrower's default PD. Unfortunately, the analytic study of this question is problematic. The Figure 2 shows the result of the computer simulations with $d = 0.25$, $r = 0.15$, $R = 0.05$, $PD = 0.075$.

Given the FOC

$$G(k, \rho) \equiv \hat{\mathcal{E}} - PD - d \cdot \Big(\mathrm{Ret}(\mathcal{E}(k); \rho) + (1 - \hat{\mathcal{E}})k \cdot \mathrm{ret}(\mathcal{E}(k); \rho)\Big) = 0,$$

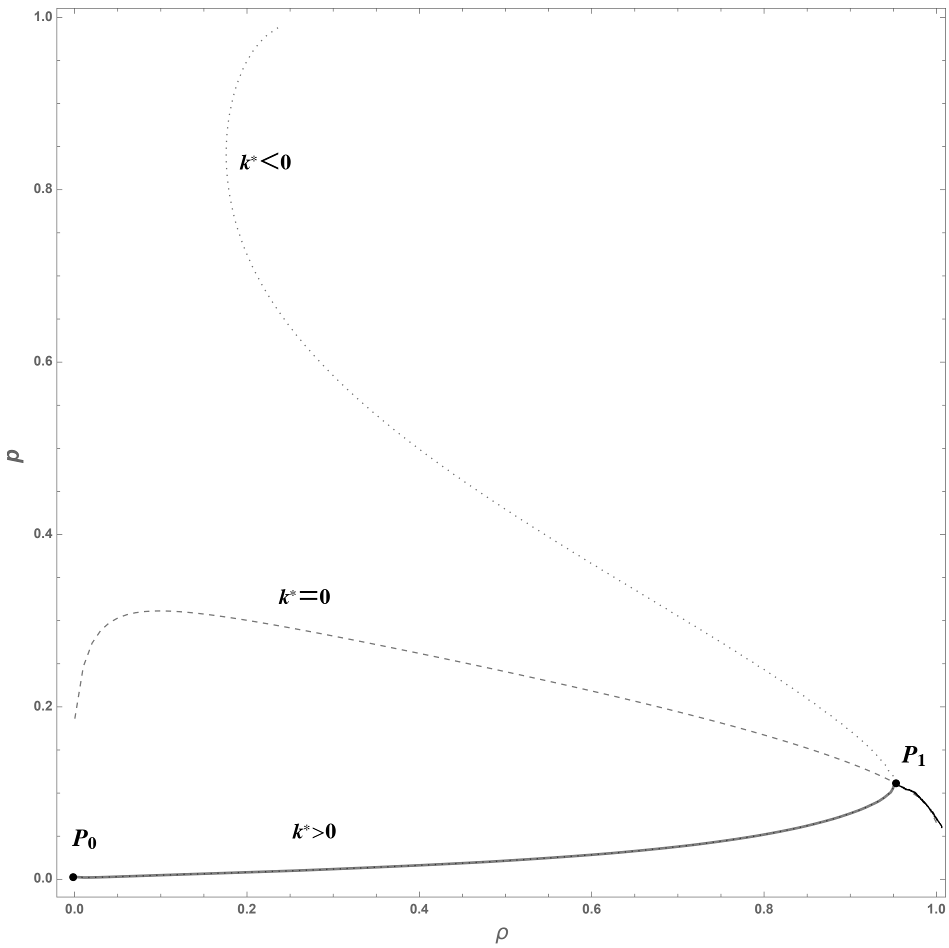

Figure 2: Probability of the banker's default as function of ρ, $d = 0.25$, $PD = 0.075$

we substitute $k = \frac{F^{-1}(1-p;\rho)-\hat{\mathcal{E}}}{1-\hat{\mathcal{E}}}$ obtaining the equation

$$H(\rho, p) = G\left(\frac{F^{-1}(1-p;\rho) - \hat{\mathcal{E}}}{1-\hat{\mathcal{E}}}, \rho\right) = 0,$$

which determines the implicit function $p(\rho)$. The set of all solutions (ρ, p) of this equation contains the "fictive" roots, violating the feasibility condition

$$k^* > 0 \iff p(\rho) < 1 - F(\hat{\mathcal{E}}, \rho).$$

To screen the fictive solutions we draw the delimiting border

$$k^* = 0 \iff p(\rho) = 1 - F(\hat{\mathcal{E}}, \rho),$$

which is depicted by dashed curve on Figure 2. The solid curve P_0P_1 is a set of all feasible solution (ρ, p), satisfying both $H(\rho, p) = 0$ and $p(\rho) < 1 - F(\hat{\mathcal{E}}, \rho) \iff k^* > 0$. The points (ρ, p) of the pointed curve above the border $p = 1 - F(\hat{\mathcal{E}}, \rho)$, satisfying $H(\rho, p) = 0$ and $p(\rho) > 1 - F(\hat{\mathcal{E}}, \rho) \iff k^* < 0$, are non-feasible.

As for definition of the default probability for ρ rightward to P_1, let's to recall that in these cases the banker can not impose the self-restriction at some finite amount of the loan portfolio, which implies $L_0 \to \infty \iff k(L_0) \to 0$. Thus we may define the function $k(\rho)$ as follows

$$k^*(\rho) = 0 \Rightarrow p(\rho) = 1 - F^{-1}\left(\hat{\mathcal{E}}; \rho\right),$$

i.e., the continuation of the probability of the bank default belongs to the delimiting curve $p = 1 - F(\hat{\mathcal{E}}, \rho)$.

3 Parametric zoning by the solution types

The main aim of the present section is to visualize the various types of equilibria in terms of the model primitives. First, assume that the deposit interest rate R, and CDF $F(z)$ for the loan losses ε are given and its PDF $f(z)$ satisfies the condition $\mathrm{ret}(z)$ decreases for all $PD < z < 1$. Consider the set $\mathcal{S}$ of feasible points $r > R$, $0 \le d \le 1$ of the parameter plane (r, d). With any point of this set we associate specific type of equilibrium, which corresponds to the whole set of parameters, including the given ones. Figure 3 shows two examples of such zoning of $\mathcal{S}$ for the Vasicek distribution of losses, which is characterized by two additional parameters, ρ and PD.

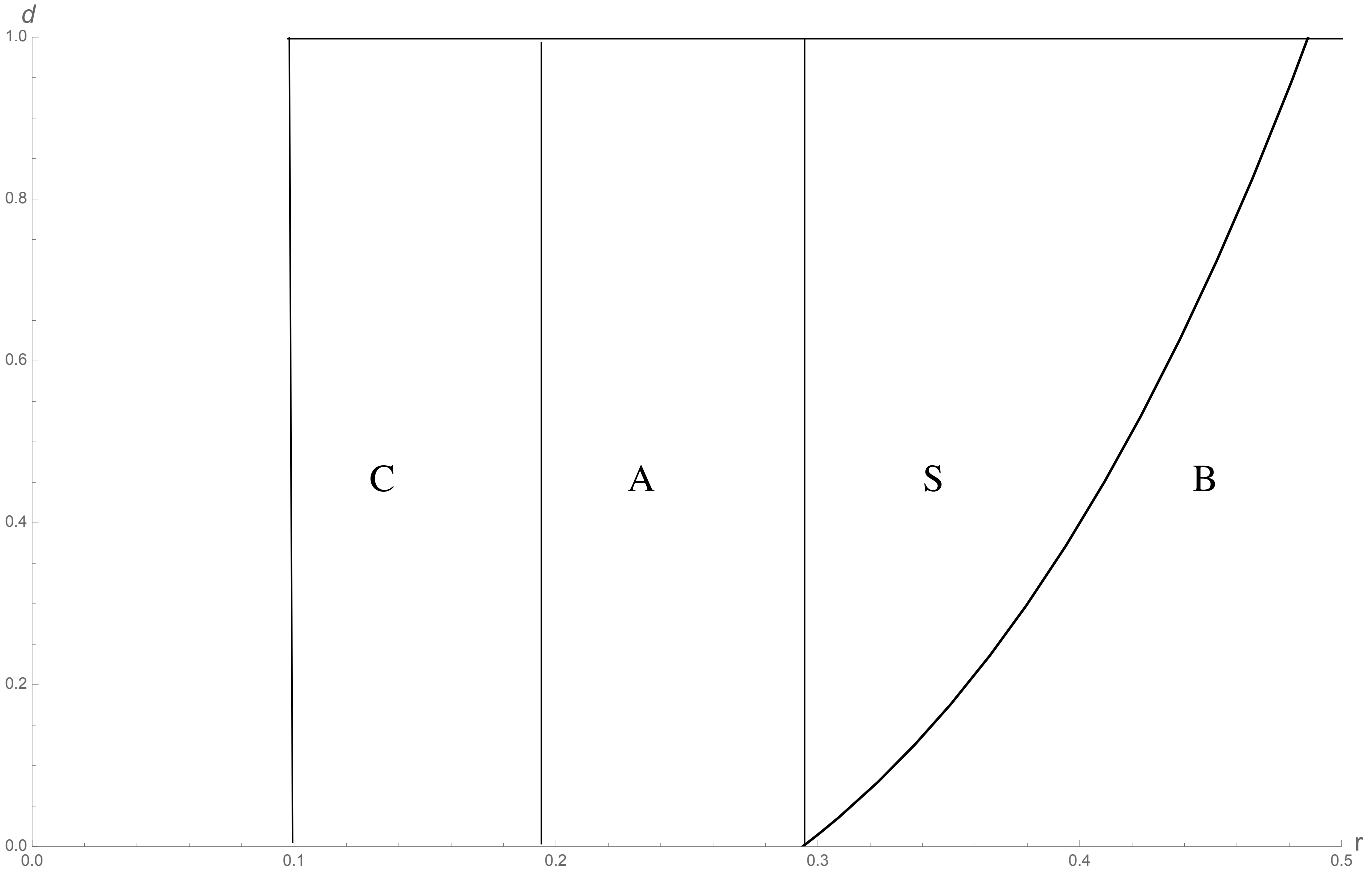

Figure 3: Zoning of parameters (r, d) for the Vasicek function: $R = 0.1$, $PD = 0.04$, $\rho = 0.1$

There are three areas in parameters space, which may be described as follows.

I. Bubble area B corresponds to the unrestricted credit expansion. It consists of points $(r, d) \in \mathcal{S}$, that violate condition (2.2).

II. Self-Restrained area S corresponds to case when the bank attracts deposits and places funds to the loan portfolio of the limited size. It consists of points $(r, d) \in \mathcal{S}$ that satisfy conditions (2.2) and $\tilde{r} > R$

III. Autarchy area A consists of points $(r, d) \in \mathcal{S}$ that satisfy the inequality $\tilde{r} < R$, $0 \leq d \leq 1$, which means that condition (2.2) trivially holds and the banker's optimum solution is degenerate: the bank does not attracts deposits, i.e., $D_0 = 0$, while the the loan portfolio $L_0 = K_0 > 0$.

Remark 2. For any given positive value of discount $d_+ > 0$, no matter how small is it, we obtain the nonempty intersection of the line $d = d_+$ with all three areas. If $d = 0$, the Self-Constrained area **S** vanishes and we obtain only two generic cases —Bubble area **B** and Autarchy area **A**.

The main result of this subsection is that the shapes of this zoning does not depend, on choice of distribution function

Proposition 2. *The structure of areas* **B**, **S**, **A** *is persisting.*

Proof. See Appendix A.4.

3.1 The Basel III requirements

The Basel III requires that the probability of the bank's default

$$p = 1 - F(\mathcal{E}(k^*))$$

must not exceed 0.001, which implies the inequality

$$k^* \geq \bar{k} = \frac{\mathrm{VaR}_{99.9} - \hat{\mathcal{E}}}{1 - \hat{\mathcal{E}}}.$$

where $\mathrm{VaR}_{99.9} = F^{-1}(0.999)$.

The Basel III analysis uses the Vasicek loan losses distribution (1.1), therefore,

$$\mathrm{VaR}_{99.9} = F^{-1}(0.999; PD, \rho) = \Phi\left(\sqrt{\frac{\rho}{1-\rho}}\Phi^{-1}(0.999) + \sqrt{\frac{1}{1-\rho}}\Phi^{-1}(PD)\right), \tag{3.1}$$

which allows to calculate the corresponding required value of CAR. Now we are going to identify sets of the bank parameters d, r, R, PD, ρ which guarantee that the banker complies voluntarily with Basel III requirements, or, on the contrary, the external regulation is needed. Substituting $\mathrm{VaR}_{99.9} = \mathcal{E}(\bar{k})$ into equation (2.4), we can determine the minimum value of discount d_B, as a function of r, guaranteeing the precise discharge of Basel III requirements, as follows

$$d_B(r) = \frac{\hat{\mathcal{E}} - PD}{\mathrm{Ret}(\mathrm{VaR}_{99.9}) + (\mathrm{VaR}_{99.9} - \hat{\mathcal{E}})\mathrm{ret}(\mathrm{VaR}_{99.9})}. \tag{3.2}$$

Let parameters R, PD, ρ be given, consider the curve $d = d_B(r)$ in the parameter plane (r, d). Obviously it starts from point $d = 0$, $r = \frac{PD+R}{1-PD}$, moreover, function $d_B(r)$ strictly increases, because function $\hat{\mathcal{E}} = \frac{r-R}{1+r}$ is increasing with respect to r. To illustrate this division, consider the following example with $R = 0.1$, $PD = 0.04$, $\rho = 0.2$, presented on Figure 4. The dashed "Basel curve" $d = d_B(r)$ divides area $\mathbf{S}$ into two sub-areas: $\mathbf{S}_A$, where Basel III requirements are violated, and $\mathbf{S}_B$, where they are complied.

The "Basel friendly" combination of parameters admits an arbitrary value of discount d, while the loan interest rates should not be too large. The existence of area $\mathbf{S}_B$ may explain the paradoxical dispersion of the observed values of CAR. Indeed, even if the values of parameters r, R, PD, and ρ are the same for all bank, the dispersion of parameter d still persists. This parameter may be idiosyncratic for every bank, reflecting the difference in reputation, or, possibly, the difference in the risk aversion

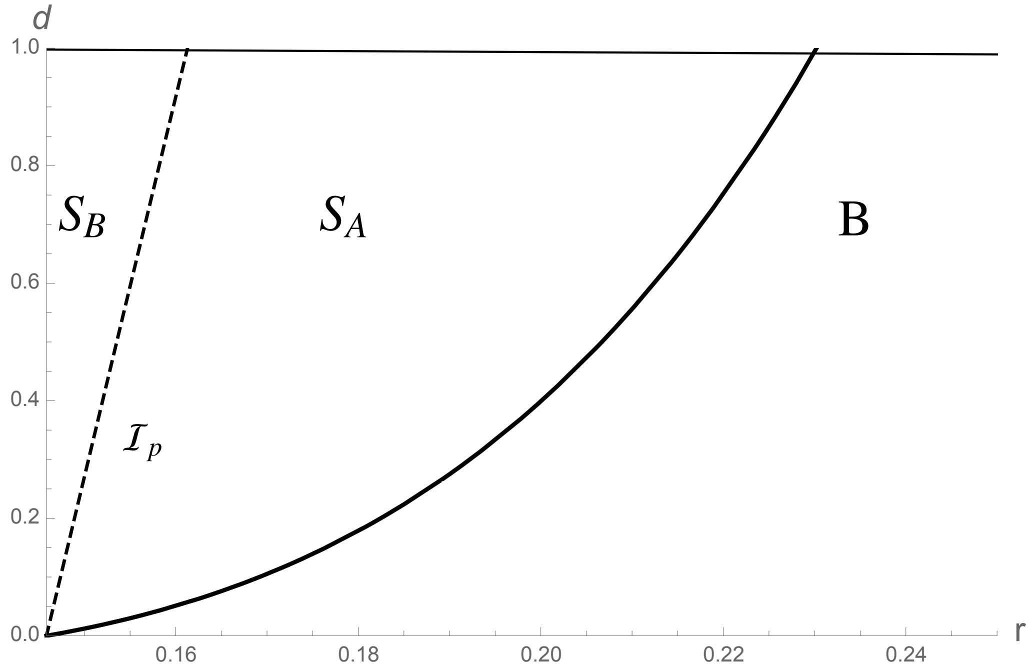

Figure 4: Dichotomy "self-restriction – external restriction"

of banker, because this parameter is applied to the decision making for the not yet defaulted bank.

3.2 Generalized Basel and the equiprobability curves

The curve dividing area **S** into two subareas in Figure 5 was determined by specific Basel III requirement. Let's generalize this approach considering an arbitrary value of the bank's default probability p as a parameter and determining the equibrobability curve $\mathcal{I}_p$ associated with the value of p, as a set of pairs (r, d), which generate the equilibrium with the probability of default equal to p, provided that the rest of the model parameters, including CDF $F(z)$, are given. We also keep the assumption on decreasing of the function $\mathrm{ret}(z)$ on the interval $PD < z < 1$

Theorem 2. *The assemblage of curves $\mathcal{I}_p$ is characterized by the following properties:*

1. All curves $\mathcal{I}_p$ associated with different probabilities p start from the same point $r = \frac{R+PD}{1-PD}$, $d = 0$.

2. If probability of the bank's default p converges to zero, then the curves $\mathcal{I}_p$ converge to the vertical line $r = \frac{R+PD}{1-PD}$, $0 \leq d \leq 1$.

3. The assemblage of curves $\mathcal{I}_p$ for all $p < 1 - F(PD)$ fills the whole Self-Constrained area **S**, *moreover, for $p < p'$ the curve $\mathcal{I}_p$ resides leftward and above the curve $\mathcal{I}_{p'}$.*

4. For all sufficiently small p the equiprobability curve $\mathcal{I}_p$ does not intersect the border of the areas **B** *and* **S** *for $0 < d \leq 1$.*

Proof. See Appendix A.5.

Figure 5 illustrates the three possible cases of the equiprobability curves described in Theorem 2

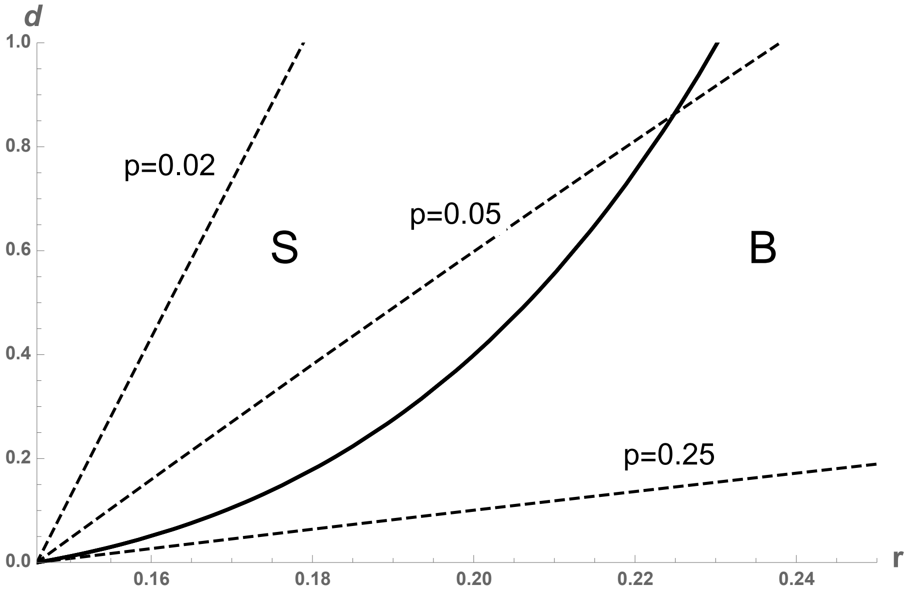

Figure 5: Equiprobability curves $\mathcal{I}_p$ for $p = 0.02,\ 0.05,\ 0.25.$

for the Vasicek distribution of losses with $PD = 0.04$, $\rho = 0.2$, and $R = 0.1$. Solid curve is the border between Self-Constrained and Bubble areas, while three dashed lines are the equiprobability curves $\mathcal{I}_p$ associated with three values of the bank probability of default: for the very large probability $p = 0.25$ the curve $\mathcal{I}_p$ leave the Self-Constrained area immediately, for the intermediate value $p = 0.05$ the curve $\mathcal{I}_p$ intersect the border, while the small probability of the bank default $p = 0.02$ generates the curve $\mathcal{I}_p$ intersecting the line $d = 1$.

4 Conclusion

The banking is one of the most over-regulated and over-supervised industries, and the pressure on banks continues to grow. A natural question arises: can banks do without a regulator - at least in some aspects of their activities that are now under strict regulation and supervision? For example, can banks limit their credit expansion on their own, without intervention of a regulator? To answer this question, we built a simple microeconomic model of a bank with one stochastic factor – the share of nonperforming loans. It turns out that if the close-out sale of a loan portfolio in case of a bank default is discountless, then the banker has no incentives to limit its credit expansion, even despite the prospect of huge losses incurring. This means that in this case, banking cannot do without a regulator, only the state can restrict the credit expansion.

The situation changes drastically, when we assume that in the event of a bank failure, its loan portfolio is sold with discount. In this case, when certain limitations on the model parameters are satisfied, an endogenous restriction of credit expansion arises. Unlike external restrictions that banks have learned to successfully circumvent, these restrictions are internal, and deceiving oneself is usually not beneficial. However, from the point of view of the regulator, which evaluates the result in terms of CAR, the level of bank self-restraint may seem inappropriate, for example, if the ratio is too low. In this paper we derive the conditions of the existence and uniqueness of equilibrium, which have the clear economic interpretation and appropriate for both analytical and numerical study.

It is shown that with sufficiently weak and natural restrictions on the loss distribution function, the parameter space is divided into 3 non-empty zones in which one of the three possible outcomes is realized: B ("Bubble") - there are no bounded solutions (we get the outcome similar to the linear model with zero discount); S ("Self-Constrained") with limited solutions; and, finally, A - autarchy solutions - deposits are not attracted, loans are placed only at own expense. In addition, a more subtle identification of compliance with the requirements established by Basel III in the area S was carried out. A natural dispersion of exogenous parameters, in the first place, the discount d, implies the observed dispersion of CAR between banks.

References

[1] Astic, F., and Tourin, A. (2014), Optimal bank management under capital and liquidity constraints, Journal of Financial Engineering, vol. 1, no. 03, 1450022 (21 pages)

[2] Barrios, V.E., and Blanco, J.M., (2003), The effectiveness of bank capital adequacy regulation: A theoretical and empirical approach, Journal of Banking & Finance, 27:1935–1958

[3] BIS (2019), IRB approach: risk weight functions – January

[4] BIS (2019), BIS Annual Economic Report 2019 - June

[5] Black F, Cox J.C. (1976), Valuing corporate securities: some effects of bond indenture provisions, J. Finance, 31:351–67

[6] Black F, Scholes M. (1973), The pricing of options and corporate liabilities. J. Polit. Econ. 81:637–54

[7] T. Bosch, J. Mukuddem-Petersen, M.P. Mulaudzi and M.A. Petersen (2008), Optimal Capital Management in Banking, Proceedings of the World Congress on Engineering 2008 Vol II WCE 2008, July 2 - 4, London, U.K.

[8] F. Chakroun and F. Abid (2016), Capital adequacy and risk management in banking industry, Applied Stochastic Models in Business and Industry, vol. 32:113-132.

[9] DeMarzo P, Fishman M. (2007), Agency and optimal investment dynamics. Rev. Financ. Stud. 20:151–88

[10] DeMarzo P, Sannikov Y. (2006), Optimal security design and dynamic capital structure in a continuous-time agency model. J. Finance 61:2681–724

[11] Fouche, C.H., Mukuddem-Petersen J. and Petersen M.A. (2006), Continuous-timestochastic modelling of capital adequacy ratios for banks. Applied Stochastic Models in Business and Industry, 22(1):41-71.

[12] Huang, J. and M. Huang (2012), How much of the corporate-treasury yield spread is due to credit risk?, Review of Asset Pricing Studies 2.2:153-202.

[13] Leland H. (1994), Corporate debt value, bond covenants, and optimal capital structure. J. Finance 49:1213–52

[14] Leland, H., (2004), Predictions of default probabilities in structural models of debt. Journal of Investment Management, 2:5–20.

[15] Merton, R.C., (1974), On the pricing of corporate debt: the risk structure of interest rates. J. Finance, 29:449–70

[16] J. Mukuddem-Petersen and M.A. Petersen (2008), Optimizing Asset and Capital Adequacy Management in Banking, J. Optim. Theory Appl. 137: 205–230

[17] Reinhart, C. M. and Rogoff, K. S. (2009) This Time Is Different: Eight Centuries of Financial Folly, Princeton University Press

[18] Schaefer S.M, Strebulaev I. (2008), Structural models of credit risk are useful: evidence from hedge ratios on corporate bonds. J. Financ. Econ. 90:1–19

[19] Vasicek O., (1987), Probability of Loss on Loan Portfolio, KMV Corporation

[20] Vasicek O., (2002), The Distribution of Loan Portfolio Value, Risk, December, 160-162

A Appendix

Notations and abbreviations

K_t	capital
D_t	deposits
M_t	cash
L_t	loans
r	loan interest rate
R	deposit interest rate
ε	share of nonperforming loans (the portfolio percentage loss)
$PD = \mathbb{E}(\varepsilon)$	probability of default of a borrower
$\tilde{r} = r - (1+r)PD$	loan risk-adjusted interest rate
d	discount of loan nominal value in case of selling of the loan
$k(L_0) = \frac{K_0}{L_0}$	CAR (capital adequacy ratio) – capital/risk weighted assets
$\hat{\mathcal{E}} = \frac{r-R}{1+r}$	the limit threshold for the loan losses
$\mathcal{E}(k) = \hat{\mathcal{E}} + (1-\hat{\mathcal{E}})k$	the threshold for the loan losses
$\mathrm{ret}(z) = (1-z)f(z)$	the weighted share of the returned loans
$\mathrm{Ret}(x) = \int_x^1 \mathrm{ret}(y)\mathrm{d}y$	the expected returns of loans in case of the bank default
$F(x)$	CDF (cumulative density function)
$f(x) = F'(x)$	PDF (probability density function)
$\Phi(z)$	standard normal distribution
FOC	First-Order Condition
SOC	Second-Order Condition
ρ	borrower's asset correlation
$U(L_0)$	objective function
r_f	risk-free rate

A.1 Proof of Lemma

Assume first that $\rho < 1/2$ and $PD \leq 1/2$, then in this case the PDF (1.2) is unimodal with mode at

$$z_{\mathrm{mode}} = \Phi\left(\frac{\sqrt{1-\rho}}{1-2\rho}\Phi^{-1}(PD)\right),$$

(see, e.g., [20]). Moreover, $\rho < 1/2$ and $PD \leq 1/2$ imply

$$\frac{\sqrt{1-\rho}}{1-2\rho} > 1 \Rightarrow \frac{\sqrt{1-\rho}}{1-2\rho}\Phi^{-1}(PD) \leq \Phi^{-1}(PD) \Rightarrow z_{\text{mode}} \leq \Phi(\Phi^{-1}(PD)) = PD,$$

because $\Phi^{-1}(PD) \leq 0$, therefore, $f(z)$ decreases with respect to z, as well as $(1-z)f(z)$ does.

Now assume that $1 > \rho \geq 1/2$ and $PD \leq 1/2$. Substituting $x = \Phi^{-1}(z)$ we obtain the following problem: to prove that the function

$$h(x) = \sqrt{\frac{1-\rho}{\rho}} \exp\left(\frac{1}{2}\left[x^2 - \left(\frac{\sqrt{1-\rho}x - c}{\sqrt{\rho}}\right)^2\right]\right)(1-\Phi(x)) = \varphi\left(\frac{\sqrt{1-\rho}x - c}{\sqrt{\rho}}\right)\frac{\Phi(-x)}{\varphi(x)}$$

is decreasing with respect to x, where $c = \Phi^{-1}(PD) < 0$, $\varphi(z) = \Phi'(z) > 0$ is the density function of the standard normal distribution satisfying the identity $\varphi'(x) = -x\varphi(x)$. Differentiating $h(x)$ we obtain

$$h'(x) = \frac{\varphi\left(\frac{\sqrt{1-\rho}x-c}{\sqrt{\rho}}\right)}{\varphi(x)}\left[\left(\frac{c\sqrt{1-\rho}}{\rho} + \frac{2\rho-1}{\rho}x\right)\Phi(-x) - \varphi(x)\right].$$

Assume first that $x \leq 0$. Given $c = \Phi^{-1}(PD) \leq 0$, we obtain

$$\left(\frac{c\sqrt{1-\rho}}{\rho} + \frac{2\rho-1}{\rho}x\right)\Phi(-x) - \varphi(x) < 0 \Rightarrow h'(x) < 0.$$

Now let $x > 0$, then

$$\left(\frac{c\sqrt{1-\rho}}{\rho} + \frac{2\rho-1}{\rho}x\right)\Phi(-x) - \varphi(x) < x\Phi(-x) - \varphi(x) < 0,$$

due to

$$\frac{2\rho-1}{\rho} = 1 - \frac{1-\rho}{\rho} < 1$$

and $\varphi(x) > x\Phi(-x)$ for all $x \geq 0$. Indeed, $\varphi(0) > 0 = 0 \cdot \Phi(-0)$ and for all $x > 0$ the inequality

$$\varphi'(x) = -x\varphi(x) = -x\varphi(-x) > \Phi(-x) - x\varphi(-x) = (x\Phi(-x))'$$

holds, which completes the proof of this case.

Finally, assume that $PD > 1/2$, then $c = \Phi^{-1}(PD) > 0$, and $z > PD$ implies $x = \Phi^{-1}(z) > c > 0$ and, consequently,

$$\frac{c\sqrt{1-\rho}}{\rho} + \frac{2\rho-1}{\rho}x < \frac{2\rho-1+\sqrt{1-\rho}}{\rho}x = x - \frac{\sqrt{1-\rho}(1-\sqrt{1-\rho})}{\rho}x < x$$

for all $0 < \rho < 1$. The rest of proof is similar. Q.E.D.

A.2 Proof of Theorem 1

Proof. Differentiating the function (2.1), we obtain

$$U'(L_0) = \hat{\mathcal{E}} - PD - d \cdot \Big(\mathrm{Ret}(\mathcal{E}(k(L_0))) + (1 - \hat{\mathcal{E}})k(L_0)\mathrm{ret}(\mathcal{E}(k(L_0))\Big) \tag{A.1}$$

while the second derivative takes on the form

$$U''(L_0) = -d \cdot \left[-\mathrm{ret}(\mathcal{E}(k(L_0)) \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}L_0} + (1 - \hat{\mathcal{E}})\mathrm{ret}(\mathcal{E}(k(L_0)) \frac{\mathrm{d}k}{\mathrm{d}L_0} + (1 - \hat{\mathcal{E}})k(L_0)\mathrm{ret}'(\mathcal{E}(k(L_0)) \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}L_0}\right] =$$
$$= \frac{d}{L_0} \cdot (1 - \hat{\mathcal{E}})^2 k^2(L_0)\mathrm{ret}'(\mathcal{E}(k(L_0)) < 0, \tag{A.2}$$

due to

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}L_0} = (1 - \hat{\mathcal{E}})\frac{\mathrm{d}k}{\mathrm{d}L_0} = (1 - \hat{\mathcal{E}})\frac{k(L_0)}{L_0}.$$

To justify the solvability of equation $U'(L_0) = 0$ on the interval $(K_0, +\infty)$, note that

$$U'(K_0) = \hat{\mathcal{E}} - PD - d\,(1 - F(1) - PD + \mathrm{Ret}(0)) = \hat{\mathcal{E}} - PD > 0,$$

while for all sufficiently large L_0 the values of $U'(L_0)$ are negative under the conditions of Theorem. Indeed,

$$\lim_{L_0 \to \infty} U'(L_0) = \hat{\mathcal{E}} - PD - d \cdot \mathrm{Ret}(\hat{\mathcal{E}}) < 0 \iff d > \frac{\hat{\mathcal{E}} - PD}{\mathrm{Ret}(\hat{\mathcal{E}})}.$$

Given the SOC $U'' < 0$, we obtain that there exists the unique solution of the FOC $U'(L_0) = 0$. Due to (A.1) we may represent the FOC as the equation

$$\hat{\mathcal{E}} - PD - d \cdot \Big(\mathrm{Ret}(\mathcal{E}(k)) + \Big(1 - \hat{\mathcal{E}}\Big)\, k \cdot \mathrm{ret}(\mathcal{E}(k))\Big) = 0$$

of variable $k = K_0/L_0$, which sets the correspondence between solution of this equation k^* and the equilibrium size of the loan portfolio L_0^*.

Q.E.D.

A.3 Proof of Proposition 1

To simplify calculations, consider the following substitution of variables $E = \hat{\mathcal{E}} + (1 - \hat{\mathcal{E}})k$. Then the FOC (2.4) is equivalent to the equation

$$G(E) \equiv \hat{\mathcal{E}} - PD - d \cdot \left(\text{Ret}(E) + \left(E - \hat{\mathcal{E}}\right) \text{ret}(E)\right) = 0.$$

Let E^* be the solution of this equation, considered as an implicit function of all parameters. The corresponding derivative with respect to an arbitrary parameter a is as follows

$$\frac{\partial E^*}{\partial a} = -\frac{\partial G}{\partial a} \Big/ \frac{\partial G}{\partial E},$$

where

$$\frac{\partial G}{\partial E} = d \cdot \left(E - \hat{\mathcal{E}}\right) \left(f(E) - (1 - E)f'(E)\right) = -d \cdot \left(E - \hat{\mathcal{E}}\right) \text{ret}'(E) > 0,$$

because $E > \hat{\mathcal{E}}$ and $\text{ret}(E)$ is a decreasing function. Moreover,

$$\frac{\partial G}{\partial d} = -\left(\text{Ret}(E) + \left(E - \hat{\mathcal{E}}\right) \text{ret}(E)\right) < 0,$$

which implies $\frac{\partial E^*}{\partial d} > 0$.

Now let $a = R$, given

$$\hat{\mathcal{E}} = \frac{r - R}{1 + r}$$

we obtain

$$\frac{\partial G}{\partial R} = -\frac{1}{1 + r} - d \cdot \frac{\text{ret}(E)}{1 + r} < 0,$$

which implies $\frac{\partial E^*}{\partial R} > 0$. Furthermore, the inequality

$$\frac{\partial G}{\partial r} = \frac{1 + R}{(1 + r)^2} + d \cdot \frac{(1 + R)\text{ret}(E)}{(1 + r)^2} > 0$$

implies $\frac{\partial E^*}{\partial r} < 0$.

Given $E^* = (1 - \hat{\mathcal{E}})k^* + \hat{\mathcal{E}}$ and $\hat{\mathcal{E}} = \frac{r-R}{1+r}$, we obtain that

$$k^* = \frac{(1 + r)E^* - (r - R)}{1 + R}, \ L_0^* = \frac{(1 + R)K_0}{(1 + r)E^* - (r - R)}, \ D_0^* = L_0^* - K_0,$$

therefore,

$$\frac{\partial k^*}{\partial d} > 0, \frac{\partial L_0^*}{\partial d} < 0, \frac{\partial D_0^*}{\partial d} < 0$$

Moreover,

$$\frac{\partial k^*}{\partial R} = -\frac{1+r}{(1+R)^2}E^* + \frac{1+r}{1+R}\frac{\partial E^*}{\partial R} + \frac{1+r}{(1+R)^2} = \frac{1+r}{(1+R)^2}(1-E^*) + \frac{1+r}{1+R}\frac{\partial E^*}{\partial R} > 0,$$

$$\frac{\partial L_0^*}{\partial R} = \frac{\partial D_0^*}{\partial R} = \frac{K_0\left((1+r)E^* - (r-R)\right) - (1+R)K_0\left((1+r)\frac{\partial E^*}{\partial R}+1\right)}{\left((1+r)E^*-(r-R)\right)^2} =$$
$$= -\frac{(1+r)K_0\left((1-E^*) + (1+R)\frac{\partial E^*}{\partial R}\right)}{\left((1+r)E^*-(r-R)\right)^2} < 0,$$

because $E^* < 1$, $\frac{\partial E^*}{\partial R} > 0$.

Finally,

$$\frac{\partial k^*}{\partial r} = \frac{1}{1+R}\left[(1+r)\frac{\partial E^*}{\partial r} - (1-E^*)\right] < 0$$

$$\frac{\partial L_0^*}{\partial r} = \frac{\partial D_0^*}{\partial r} = -\frac{(1+R)K_0}{\left((1+r)E^*-(r-R)\right)^2}\left[E^* - 1 + (1+r)\frac{\partial E^*}{\partial r}\right] > 0,$$

because of $E^* < 1$ and $\frac{\partial E^*}{\partial r} < 0$. Q.E.D.

A.4 Proof of Proposition 2

The statement about area **A** is obvious. The rest is to show the robustness of shapes of areas **B** and **S**. Note that the function

$$d_S(r) = \frac{\hat{\mathcal{E}}(r) - PD}{\mathrm{Ret}(\hat{\mathcal{E}}(r))},$$

where $\hat{\mathcal{E}}(r) = \frac{r-R}{1+r}$, satisfies the following conditions:

1. $d_S\left(\frac{PD+R}{1-PD}\right) = 0,$

2. $d_S\left(r\right)$ strictly increases for all $r > \frac{PD+R}{1-PD}$.

The first statement is obvious due to definition of $d_S\left(r\right)$. Then, representing the function $d_N(r)$ as follows

$$d_S(r) = \frac{\tilde{r}-R}{1+r}\cdot\frac{1}{\mathrm{Ret}\left(\frac{r-R}{1+r}\right)},$$

and given the functions $\frac{\tilde{r}-R}{1+r}$, $\frac{r-R}{1+r}$ are positive and strictly increasing with respect to r we obtain that the function $d_0(r)$ is also strictly increasing. Finally, the function $d_S\left(r\right)$ is unrestrictedly increasing with $r \to \infty$, because

$$\frac{\tilde{r}-R}{1+r} \to 1 - PD,\ \frac{r-R}{1+r} \to 1.$$

Therefore, its graph intersects the line $d = 1$ in finite point $r_S > \frac{PD+R}{1-PD}$, which determines the base of the curvilinear triangle **S**. Q.E.D.

A.5 Proof of Theorem 2

Consider the probability of the bank's default p as a parameter with possible values from the interval $[0, 1]$. Formula (1.4) implies that for any given p, the equation $p = 1 - F(E)$ determines the value $E_p = F^{-1}(1-p)$. This means that the equiprobability curve $\mathcal{I}_p$ is determined by equation

$$\hat{\mathcal{E}}(r) - PD - d \cdot \Big(\mathrm{Ret}(E_p) + \Big(E_p - \hat{\mathcal{E}}(r)\Big)\,\mathrm{ret}(E_p)\Big) = 0,$$

or, equivalently,

$$d = d_p(r) \equiv \frac{\hat{\mathcal{E}}(r) - PD}{\mathrm{Ret}(E_p) + \Big(E_p - \hat{\mathcal{E}}(r)\Big)\,\mathrm{ret}(E_p)}, \tag{A.3}$$

where $\hat{\mathcal{E}}(r) = \frac{r-R}{1+r}$. It is obvious that $d_p\left(\frac{PD+R}{1-PD}\right) = 0$ for all p, which completes the first statement of the theorem. Moreover, $p \to 0$ implies $E_p \to 1$, therefore, $\lim\limits_{p\to 0} d_p(r) = +\infty$ for any $r > \frac{R+PD}{1-PD} \iff \hat{\mathcal{E}}(r) - PD > 0$, which completes the second statement.

Recall that the border of areas **S** and **B** is determined by the function

$$d_S\,(r) = \frac{\hat{\mathcal{E}}(r) - PD}{\mathrm{Ret}(\hat{\mathcal{E}}(r))}.$$

Calculating and comparing the derivatives of $d_S(r)$ and $d_p(r)$ at the starting point $r_0 = \frac{R+PD}{1-PD}$

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}r} d_S(r_0) &= \frac{(1-PD)^2}{(1+R)\mathrm{Ret}(PD)},\\
\frac{\mathrm{d}}{\mathrm{d}r} d_p(r_0) &= \frac{(1-PD)^2}{(1+R)\left(\mathrm{Ret}(E_p) + (E_p - PD)\mathrm{ret}(E_p)\right)},
\end{aligned}$$

we obtain that

$$\frac{\mathrm{d}}{\mathrm{d}r} d_p(r_0) > \frac{\mathrm{d}}{\mathrm{d}r} d_S(r_0) \iff \mathrm{Ret}(PD) > \mathrm{Ret}(E_p) + (E_p - PD)\mathrm{ret}(E_p).$$

Consider the function

$$G(x) = \mathrm{Ret}(x) + (x - PD)\mathrm{ret}(x),$$

which obviously satisfies $G(PD) = \mathrm{Ret}(PD)$. Moreover,

$$G'(x) = (x - PD)\mathrm{ret}'(x) < 0$$

for all $x > PD$. This implies that

$$x = E_p = F^{-1}(1-p) > PD \iff p < 1 - F(PD)$$

is necessary and sufficient condition for the curve $\mathcal{I}_p$ to belong the area **S**, at least in some neighborhood of r_0.

Let's determine the point of intersection of the equiprobability curve $\mathcal{I}_p$ with the border of areas **S** and **B** from the following equation

$$d_S(r) = d_P(r) \iff \mathrm{Ret}(\hat{\mathcal{E}}(r)) = \mathrm{Ret}(E_p) + \Big(E_p - \hat{\mathcal{E}}(r)\Big)\,\mathrm{ret}(E_p).$$

The unique solution $r(p)$ of this equation is determined by identity

$$E_p = \hat{\mathcal{E}}(r(p)) \iff r(p) = \frac{E_p + R}{1 - E_p} = \frac{F^{-1}(1-p) + R}{1 - F^{-1}(1-p)} > r_0 = \frac{PD + R}{1 - PD}$$

because $F^{-1}(1-p) > PD$. Note that this point of intersection is actual only in case

$$d_S(r(p)) = d_p(r(p)) \leq 1,$$

otherwise, the equiprobability curve intersects the line $d = 1$ instead of $d_S(r)$. This happens if and only if

$$d_S(r(p)) > 1 \iff E_p - PD - \mathrm{Ret}(E_p) > 0.$$

Note that the function

$$H(x) = x - PD - \mathrm{Ret}(x)$$

for $x \geq PD$ satisfies the following conditions: $H(PD) < 0$, $H(1) = 1 - PD > 0$, and $H'(x) = 1 + (1-x)f(x) > 0$. This implies that there is $x^* \in (PD, 1)$ such that for all $x > x^*$ the function $H(x) > 0$, which is equivalent to

$$p < 1 - F(x^*) \Rightarrow E_p - PD - \mathrm{Ret}(E_p) > 0.$$

Q.E.D.