ON THE OPTIMAL DESIGN OF RISK RETENTION IN SECURITISATION

Metin Kaptan*

Abstract

This paper examines the optimal design of retention in securitisation, in order to maximize welfare of screening per unit of retention, assuming that screening is costly and that the bank intends to securitise its loans. In contrast to the focus of previous literature on tranche retention, we deviate from the constitutional mechanisms of tranche retention to present a pareto-optimal method of tranche retention. Unlike the current ad-hoc-regulations, we derive the optimal design of retention from a utility maximization problem. We show that the level of retention per tranche should be dependent on the rate of credit default, i.e. the higher the rate of default, the higher the optimal rate of retention required to provide an incentive to screen carefully. From this approach, it follows that the rate of retention per tranche should be higher, the higher the position within the ranking order of subordination. Accordingly, the efficiency of tranche retention can be enhanced, reducing the level of retention required to maintain a given level of screening-effort. This retention design entails a recovery of the bank’s equity capital, thereby increasing liquidity and lending capacities.

Keywords: Securitization, Screening Incentives, Retention Requirements, Moral Hazard

JEL Classification: D82, G14, G21, G28

* Department of Economics, University of Muenster, Germany
Email: kaptan@uni-muenster.de

1 Introduction

Securitisation, the transformation of illiquid bank loans and other financial assets into liquid, tradable securities, grew tremendously over the period 2004-2007, and declined rapidly thereafter. The annual amount declined from over $3.5 trillion to just over $2 trillion in 2008 (Fender and Mitchell, 2009). This decline in the volume of securitisation reflects a loss of investor trust in the instrument of securitisation, a consequence of malpractice both before and during the crisis.

The technique of securitisation allows banks to transfer default risk to the capital markets. This instrument entails various benefits, such as increased liquidity and lending capacities, and a cost reduction of lending (Geithner, 2011). However, it simultaneously creates a moral hazard problem, which is likely to arise “when individuals engage in risk sharing under conditions such that their privately taken actions affect the probability distribution of the outcome” (Hölmlstrom, 1979). Purnanandam (2011) has shown empirically, that banks with high involvement in the so-called originator-to-distribute market did not devote sufficient resources to screening their borrowers. In other words, banks have little incentive to screen borrowers carefully if they intend to securitise the default risk, and once the risk has been transferred, they have no incentive to monitor loans to reduce the probability of credit default. Thus, moral hazard refers here to the tendency towards a low incentive to screen borrower solvency.

The debate on solutions to the problem of moral hazard can be traced back to Arrow (1963), after which it developed into a broad strand of economic literature, specific to the field of insurance economics. Economists have mainly discussed the three following solutions to the problem of moral hazard: (i) “incomplete coverage against loss”, (ii) “observation by the insurer of the care taken to prevent loss” (Shavell, 1979) and (iii) “reputational concerns” (Stiglitz and Weiss, 1981). Within the framework of securitisation transactions, screening-effort is unobservable, because it is too time-consuming to be economically viable. The recent crisis has shown that reputational concerns can overcome the moral hazard problem only to some extent. Thus, in this paper, we focus on incomplete coverage against losses, i.e. the originating bank retains so-called “skin in the game”. The originating bank should retain some risk associated with the securitised credit portfolio. The share of risk held by the originating bank provides, ceteris paribus, an incentive to prevent losses and to screen effectively, so that the interests of the investors and the originator are at least partially aligned (Franke and Krahnen, 2008).

In fact, before and during the crisis, originating banks typically retained the first loss piece of their
transactions, similar to a deductible in insurance contracts. “By construction, the first loss piece fully absorbs all [credit] default losses up to its notional amount” (Franke and Krahnen, 2008). However, the notional amount of the first loss piece was so small that it did not provide an incentive to pursue sufficient screening. In addition, a first loss retention generally does not provide an incentive to prevent losses that exceed the notional amount of that first loss piece.

In order to address the deficiencies that contributed to the global financial crisis, the U.S. and European legislative authorities passed laws that require a risk retention of no less than 5 percent, which exceeds the traditional level. Accordingly the United States passed the Dodd-Frank Wall Street Reform and Consumer Protection Act (“Dodd-Frank Act”) in July 2010 (Kiff and Kissner, 2011). The amendments to the securitisation market include “greater transparency for investors, measures to mitigate conflicts of interest at credit ratings agencies, and [a required] […] credit risk retention” of no less than 5 percent (Geithner, 2011). In May 2009, the European Capital Requirements Directive was approved and came into force in January 2011. This reform introduces risk retention requirements and intensifies due diligence obligations. “In December 2010, the Committee of European Bank Supervisors (CEBS) issued final guidelines with respect to the application of Article 122a” (Geithner, 2011).

The analysis of risk transfer and moral hazard has attracted considerable attention in the literature. Keys et al. (2010) investigate whether the securitisation process reduces the incentive of banks to pursue proper screening. Chiesa (2008) examines the impact of credit risk transfer on screening-efforts and the incentive for banks to engage in credit risk transfer. Fender and Mitchell (2009) use a moral hazard model, closely related to Innes (1990), to analyse the effectiveness of different forms of retention on the originator’s level of screening-effort. The authors differentiate between the retention of a vertical slice, a first loss tranche and a mezzanine tranche. Introducing accounting frictions, Kiff and Kissner (2011) compare the efficiency of equity and mezzanine retention. They demonstrate theoretically, that different forms of risk retention result in different levels of screening-effort.

In contrast to Fender and Mitchell (2009) and Kiff and Kissner (2011) we deviate from existing regulations and derive an optimal design of retention that leads to the welfare maximizing level of screening incentives. We show that neither of the current regulations is welfare maximizing. Our retention scheme increases screening incentives, given the level of retention. Such a scheme is desirable from a social point of view, because it increases the efficiency of capital allocation. This in turn induces an increase in portfolio quality, which then decreases spreads and leads to an increase in investor demand for securities. The increase in the quality of credit portfolios may be used to reduce the required level of risk retention, which induces a release of costly equity capital and thus promotes financial stability. Furthermore, the release of equity capital leads to an increase in lending capacities, facilitating economic growth.

We use a slightly modified version of the models presented by Bender (2002) and Holmström (1979), to derive the optimal form of an incentive contract under moral hazard. We show that the level of retention per tranche should be relatively high for highly ranked tranches and relatively low for subordinated tranches. The underlying logic is that the bank should be punished for a “bad” outcome and rewarded for a “good” one.

This paper is organised as follows. Section 2 presents the current regulations. Section 3 introduces the model and Section 4 concludes and considers the policy implications.

2 Current Securitisation Regulations

The theoretical literature on risk allocation generally finds that under asymmetric information, the screening incentives decrease when risk is transferred. The real-world poor quality of the underlying assets of securitised portfolios during the financial crisis supports the theoretical literature. As a response to the financial crisis, the US and EU legislative authorities interfered through regulation, introducing tranche retention to solve the problem of moral hazard.

In April 2010, the Securities and Exchange Commission (SEC) proposed a rule obliging originators to satisfy risk retention requirements. “Under the proposed Regulation AB II regime, [the originator has to] […] retain either (i) at least 5 percent of the nominal amount of each tranche [of the portfolio securitised] […] or (ii) in the case of revolving [exposures] […] a seller’s interest of at least 5 percent of the nominal amount of the securitized exposures” (De Sear and Hwang, 2011).

The European Union considered a risk retention regime for asset backed securities (ABS), which to some extent differs in its design from AB II. On 1st January 2011, the European retention requirements came into force. The key requirement of tranche retention (see Article 122a, Capital Requirements Directive) stipulates that EU-based credit institutions investing in securitisation transactions.

---


72 Innes (1990) considers a principal-agent problem, in which a risk-neutral agent makes an unobservable effort choice and thereby influences the principal’s income under limited liability of the agent.

Note that in contrast to the US-regulation AB II, the EU Capital Requirements Directive “does not impose risk retention requirements directly on originators of securitisations […], but instead regulates the activities of
investors in asset-backed securities” (De Sear and Hwang, 2011).

74 This form of risk retention is not depicted in Figure 1, because it can be treated as equivalent to a vertical slice.


76 The terms “originator” and “bank” are used interchangeably.

77 It is assumed that, even with meticulous screening, the default rate can never be reduced to zero.

retain no less than 5 percent of the “net economic interest”. Article 122a provides the following retention options that may be chosen from (see also Figure 1):

- Random selection - originators are required to retain randomly selected exposures equal to no less than 5 percent of the nominal amount of the securitised exposures, provided that the nominal amount is no less than 100 at origination,
- First loss tranche - originators are required to retain the subordinated first loss tranche and, if necessary, other tranches of the same or a higher risk profile as those sold or transferred, so that the retention equals no less than 5 percent of the nominal amount of the securitized exposures,
- Vertical slice - originators are required to retain no less than 5 percent of the nominal value of all securitised tranches sold or transferred to investors,
- Pari passu share - originators are required to retain no less than 5 percent of the nominal value of securitised exposures in the case of securitisations of revolving exposures (Capital Requirements Directive, Article 122a).

The need for retention requirements is widely accepted, but the level and nature remain controversial. Some market participants fear that a retention rate of 5 percent is not high enough to create a sufficient incentive for banks to carefully screen borrowers. Others argue that a retention of 5 percent is too high and thus constrains lending capacities.

Instead of analysing the optimal level of risk retention, this paper aims to identify the optimal retention design. For this purpose, we deviate from the earlier presented forms of retention. The risk retention should rather be designed so that the screening incentive is maximized per unit of retention. This approach therefore maximizes the originator’s incentive to screen carefully without causing additional costs in terms of retained equity capital. The efficiency of securitisation transaction may be increased in two ways. (i) An increase in screening-effort due to optimal incentive setting, ceteris paribus, reduces the notional level of retention needed to ensure a given level of effort and thereby increases bank lending capacities. (ii) With a given percentage share of tranche retention, an increase in screening-effort entails an increase in the quality of the portfolio. This provides greater certainty among investors, thereby decreasing spreads and increasing the demand for asset backed securities.

Thus, we provide a simple incentive-based approach to enhance the market for securities by decreasing the costs and increasing the quality.

3 The Model

The model focuses on a principal-agent relationship between two utility maximizing market participants in a securitisation transaction - an originator (bank) B and an investor I. The investor is risk-averse. The bank may or may not be risk-averse.

The bank gives credit to borrowers. Borrowers i differ in default probabilities \( x_i \in \{ x; \overline{x} \} \). The default probability \( x \) cannot be influenced by borrowers, i.e. there is a purely adverse selection problem between bank and borrowers. The bank can mitigate the adverse selection problem only with the costly screening-effort \( \vartheta \). The bank securitises the credit portfolio and the size of the credit portfolio is normalized to 1. The default risk of the portfolio \( x \in \{ \overline{x}; \overline{x} \} \) is determined by the screening-effort. The level of screening-effort cannot be observed by the investor. Since screening is costly, originators are tempted not to screen, when credits are to be securitised. Accordingly, the information asymmetry originates in the unobservability of the screening-effort.

The investor’s utility function is defined by wealth only, while the originator’s utility function is defined by both wealth and effort. An increase in effort entails two opposing effects on the originator’s utility function: (i) there is a direct negative effect as a result of an increase in costs and (ii) there is an indirect positive effect due to a decrease in the probability of credit default. The latter effect only applies, if the originator retains risk in terms of tranche retention. The lower the level of risk retention, the higher is the probability that the negative effect prevails. If the level of risk retention does not exceed the expected (non-influenceable) level of credit default, the originator has no incentive to pursue proper screening, since an increase in effort only affects the utility negatively through an increase in costs. In contrast to the originator, the investor always benefits from an increase in effort, due to the decrease in credit default to be covered and because the costs are borne by the originator. Therefore, a conflict in objectives is likely to arise.

The originator’s utility-function is given by

\[
\varphi_B(y) = \varphi_B(x, \vartheta) = \varphi_B(\Omega_B - R(x) - c(\vartheta)),
\]

where \( \Omega_B \) denotes the originator’s initial wealth, \( x \) displays the credit default rate, \( R(x) \) is the level of the credit default which is covered by the

plane
originator and $\epsilon$ represents the screening-effort. Let $c(\epsilon)$ denote the costs of screening, with $c'(\epsilon) = 1$.

The initial wealth already accounts for a risk-premium paid to the investor for covering part of the default risk.

The investor’s utility function is given by

$$u_I(z) = u_0(x) = u_0(\Omega_I - (x - R(x))).$$  \(2\)

where $\Omega_I$ displays the investor’s initial wealth, including the premium for taking the default risk.

The investor benefits from an increase in effort, since it reduces the credit default rate and consequently leads to a decrease in the covered losses $(x - R(x))$, whereas screening-effort implies, ceteris paribus, that financial penalties accrue to the originator when the level of retention $R(x)$ is low. Consequently, an incentive problem arises, conditional on the level of unobservability of screening-effort. An incentive to pursue costly screening can be attained by risk retention, i.e. the investor does not cover all potential losses, $R(x) > 0$.

**The maximization problem**

Since screening-effort cannot be observed, the originator will choose a level of effort, such that the marginal costs of an additional unit of effort will equal the marginal benefit. Accounting for this condition, the retention rate $R(x)$ must be such that the originator chooses a screening-effort which simultaneously maximizes total utility.

We apply the First-Order Approach (FOA) to solve this problem\(^78\):
\[
\max_{R(x), \varepsilon} L = \left( \int_{x} u_{g}\{\Omega_{g} - R(x) - c(\varepsilon)\} f(x, \varepsilon) \, dx \right) + \mu \left( \int_{x} u_{i}\{\Omega_{i} - (x - R(x))\} f(x, \varepsilon) \, dx - u_{i}^{r} \right) + \lambda \left( \int_{x} U(x, \varepsilon) f_{g}(x, \varepsilon) \, dx - c'(\varepsilon) \right) \int_{x} U(x, \varepsilon) f(x, \varepsilon) \, dx \right) \quad (3)
\]

with
\[
Y(x, \varepsilon) = \Omega_{g} - R(x) - c(\varepsilon).
\]

Let \( \mu \) and \( \lambda \) denote the Lagrangian multipliers. In the spirit of Holmström (1979), the necessary condition for optimality is replaced by the first-order constraint. MLRP and CDFC are sufficient conditions for the FOA to be valid\textsuperscript{79}. The first-order constraint reflects the fact that at the optimum, the marginal costs of screening-effort equal its marginal benefit.

A point-wise optimization of the Lagrangian with respect to \( R(x) \) yields:

\[
-u_{g}'(Y) f(x, \varepsilon) + \mu u_{i}'(Z) f(x, \varepsilon) + \lambda \left[ -u_{g}'(Y) f_{g}(x, \varepsilon) + c'(\varepsilon) u_{g}'(Y) f(x, \varepsilon) \right] = 0. \quad (4)
\]

Rewriting equation (4) gives:

\[
\frac{u_{i}'}{u_{g}'}(Z) + \frac{\lambda c'(\varepsilon)}{\mu} = \frac{1}{\mu} + \frac{\lambda f_{g}(x, \varepsilon)}{\mu f(x, \varepsilon)} \quad (5)
\]

The optimal shape of the retention function

The optimal shape of the retention function

Differentiating equation (5) with respect to \( x \) yields:

\[
\frac{u_{i}'}{u_{g}'} \left( \frac{dR(x)}{dx} - 1 \right) - \frac{u_{i}''}{u_{g}'} \left( \frac{-dR(x)}{dx} \right) + \frac{\lambda c'(\varepsilon)}{\mu} \frac{\partial \frac{u_{i}'}{u_{g}'}}{\partial x} = \frac{\lambda f_{g}}{\mu} \quad (6)
\]

with \( \frac{dZ(x)}{dx} = dR(x)/dx \) and \( dY(x)/dx = -dR(x)/dx \).

Factoring out \( u_{i}'/u_{g}' \) gives:

\[
\frac{u_{i}'}{u_{g}'} \left[ u_{i}' \left( \frac{dR(x)}{dx} - 1 \right) - u_{g}' \left( \frac{-dR(x)}{dx} \right) \right] + \frac{\lambda c'(\varepsilon)}{\mu} \frac{\partial \frac{u_{i}'}{u_{g}'}}{\partial x} = \frac{\lambda f_{g}}{\mu} \quad (7)
\]

We can simplify equation (7) by using the Arrow-Pratt Measure, \( \gamma_{1} = -u_{i}'/u_{i} \), with \( i = B, I \):

\[
\frac{u_{i}'}{u_{g}'} \left( -\gamma_{1} \frac{dR(x)}{dx} + \gamma_{2} - \gamma_{2} \frac{dR(x)}{dx} \right) - \frac{\lambda c'(\varepsilon)}{\mu} \frac{\partial \gamma_{1}}{\partial x} = \frac{\lambda f_{g}}{\mu} \quad (8)
\]

Extending the quotient \( \frac{\partial \gamma_{1}}{\partial x} \) with \( \frac{\partial \gamma_{1}}{\partial x} = \frac{\partial \gamma_{1}}{\partial y} \frac{\partial y}{\partial x} \) yields:

\[
\frac{u_{i}'}{u_{g}'} \left( -\eta_{1} \frac{dR(x)}{dx} + \eta_{1} - \eta_{2} \frac{dR(x)}{dx} \right) + \frac{\lambda c'(\varepsilon)}{\mu} \frac{\partial \gamma_{1}}{\partial x} \frac{dR(x)}{dx} = \frac{\lambda f_{g}}{\mu} \quad (9)
\]

Solving for \( \frac{dR(x)}{dx} \) yields:

\[
VIRTUSINTERPERS
\]

54
For simplicity and without changing the main result, constant absolute risk aversion (respectively risk neutrality) is assumed, i.e. \( \frac{\partial R}{\partial x} = 0 \). Hence, equation (10) reduces to:

\[
\frac{dR(x)}{dx} = \frac{\eta \left( \frac{u_1}{u_B} \right) - \frac{\lambda}{\mu} \left( \frac{\partial f(x)}{\partial x} \right)}{(R_g + \eta \left( \frac{u_1}{u_B} \right)}.
\]

Equation (11) indicates that the shape of the function \( R(x) \) is determined mainly by two factors, (i) the relationship between the originator’s and the investor’s risk-profile and (ii) the likelihood ratio.

The higher the investor’s absolute risk aversion \( \eta \) in relation to the originator’s risk-attitude \( R_g \), the steeper the function. In other words, the more risk-averse the investor, the higher is the necessary rate of risk retention.

The likelihood ratio reflects the impact of an increase in screening-effort on the density function. The likelihood ratio decreases monotonically in \( x \), because an increase in screening-effort induces a decrease in the probability of high default rates and an increase in the probability of low default rates, i.e. \( \frac{\partial f(x) \cdot c(x)}{\partial x} < 0 \). With \( \lambda > 0 \), the slope of the retention function increases progressively.

Equation (11) indicates that the level of retention should be increasing in \( x \), i.e. the higher the default rate, the higher the relative share of default carried by the originator. In other words, the level of retention per tranche should be relatively low for low default rates and relatively high for high default rates, i.e. the level of retention should increase for higher ranked tranches. Intuitively, we could argue that the bank will be punished for a bad outcome and rewarded for a good outcome, since the outcome reflects the level of screening-effort. Accordingly, we increase the incentive for banks to prevent high default rates.

Figure 2. Optimal retention function

Applied to the structure of a securitised credit portfolio, this means that the originator should retain a share of each tranche that might be affected by a credit default and whose probability of default can be influenced by screening. The higher the default rate \( x \), the higher the proportional share carried by the originator, i.e. the originator should retain a relatively small share of low ranked tranches and an increasing
share, the higher the position within the ranking order of subordination. Accordingly, the originator is punished disproportionately severely for bad outcomes (high default rates) and rewarded excessively for good outcomes (low default rates). Thus, the originator has an increased incentive to screen carefully in order to avoid high default rates. Figure 3 depicts an example of an optimal retention design. The retention is marked in grey. The thin

Figure 3. Optimal design of retention

Concluding Remarks

It is common wisdom that the transfer of risk induces moral hazard. As has been demonstrated dramatically during the financial crisis, the securitisation of loan portfolios severely reduces and even eliminates the originator’s screening incentives. Consequently, the quality of securities deteriorates, so that investors demand for an extremely high risk premium or alternatively leave the market for securities. As shown by Akerlof (1970), markets may even collapse due to informational asymmetries. In order to prevent an overall breakdown of the market for securities, the partial retention of risk has emerged as the most effective solution. Risk retention entails an alignment of originator and investor incentives.

In the aftermath of the financial crisis, policymakers have decreed that originator’s must retain 5 percent of the portfolio, thereby creating an incentive for originators to screen carefully. Nevertheless, policy makers should not neglect the fact that retention design has a strong impact on its effectiveness. This is of particular importance, because bank equity capital is limited, so that lending capacities are affected directly. Thus, retention should be minimized, but no less than necessary to ensure an optimal screening level. This can only be achieved by an incentive-maximizing retention structure.

The current ad-hoc-regulation does not fulfill this requirement and hence is not optimal. This paper provides an approach which demonstrates that the design of retention should be modified. Policymakers should design risk retention so as to maximize the incentive for careful screening, while minimizing the costs. Therefore, the rate of retention should be relatively high for high-rank tranches and lower for subordinate tranches, as to motivate originators to prevent high default rates. The first loss tranche, which is so small that it may not even be influenced by excessive screening-efforts, should be securitised. In short, the originator should only retain the risk of tranches within the bounds of influenceable credit default, and the level of retention should increase in credit default rates in order to prevent the originator
for lax screening. Thus, with less risk retention, the same amount of screening-effort could be generated. In this manner, the bank’s equity capital will be discharged and lending capacity will increase.

References