

Credit Risk Transfer and Financial Sector Performance*

Wolf Wagner[†] and Ian W. Marsh[‡]

Abstract

In this paper we study the impact of credit risk transfer (CRT) on the stability and the efficiency of a financial system in a model with endogenous intermediation and production. Our analysis suggests that with respect to CRT, the individual incentives of the agents in the economy are generally aligned with social incentives. Hence, CRT does not pose a systematic challenge to the functioning of the financial system and is generally welfare enhancing. However, we identify issues that should be addressed by the regulatory authorities in order to minimize the potential costs of CRT. These include: ensuring the development of new methods of CRT that allow risk to be more perfectly transferred, setting regulatory standards that reflect differences in the social cost of instability in the banking and insurance sector; promoting CRT instruments that are not detrimental to the monitoring incentives of banks.

JEL classifications: E44, G21, G22, G28

Keywords: credit risk transfer, financial stability and efficiency, regulation

*We thank Alan Morrison, Norvald Instefjord, Demosthenes Tambakis, and seminar participants at Cambridge University for helpful comments.

[†]Corresponding author. Cambridge Endowment for Research in Finance (CERF), Judge Institute for Management, Trumpington Street, Cambridge University, Cambridge CB2 1AG, UK. Phone: +44-1223-760585, Fax: +44-1223-339701, Email: w.wagner@cerf.cam.ac.uk

[‡]CERF, Cass Business School and CEPR, Email: i.marsh@cerf.cam.ac.uk

1 Introduction

Because of its inherent illiquidity, a commercial bank loan was typically held on the bank's balance sheet until maturity or default. A few risk management techniques were available. For example, a small loan sale market existed, and banks could securitize part of their portfolios (although this was largely limited to non-commercial loans such as mortgages and credit card receivables). More recently, the introduction of credit derivative markets has given banks a new credit risk management tool. These markets have shown a rapid development. First traded in 1996, the outstanding value of credit derivatives reached US\$ 2,000bn in 2002, equivalent to some seven percent of total outstanding bank credit.¹

This significant growth in credit risk transfer has received the attention of policy makers worldwide. Most supranational and several national supervisors have issued reports on the topic (e.g., BBA 2002, BIS 2003, FSA 2002, IAIS 2003, IMF 2002, OECD 2002). These reports are rather similar in tone. On one hand, they emphasize that credit risk transfer (CRT) brings about benefits, particularly diversification gains. On the other hand, one can read a common concern that CRT may cause problems for the stability of the financial sector, for example by increasing the fragility of the risk buyer or through increased risk-taking by banks. Most of the arguments, however, are made on an informal basis due to the lack of theoretical analysis of these issues in the literature.

This paper is a first step towards filling this gap. The objective is to provide a framework for analyzing the impact of CRT on the stability and efficiency of the financial system and to provide an initial analysis of the likely impact of CRT. There are various ways in which CRT can affect the functioning of the financial system. To begin with, there is an obvious diversification effect on stability: CRT allows banks to diversify their risks better, which should help financial stability. CRT may also lead to the transferring of risk from banks to other, less fragile, financial players, such as insurance companies. The possibility of transferring risks, however, may induce banks to take up new risks through an increase in their lending activities. Risk taking by banks may also go up because banks' incentives to monitor firms may be reduced by CRT (which partly passes on the benefits from firms' success to other parties), thus possibly making loans more risky. Furthermore, by allowing banks to diversify, CRT may increase lending and thus lead to a more efficient allocation of capital in the economy. The downside, as already mentioned, is that

¹Sources: BBA (2002), BIS (2003) and FitchRatings (2003).

CRT may reduce efficiency because it worsens asymmetric information problems (by reducing banks' alignment with the firms' interest), which may create inefficiencies due to inefficiently low screening and monitoring. Finally, CRT may have an impact on the efficiency of the financial system in total by changing the relative merits of bank financing versus market financing (and thus intermediation): it reduces the cost of bank financing by allowing banks to diversify but may also reduce the benefits from bank financing (the certification effect) by lowering banks' incentives to monitor.

In our analysis, we abstract from a number of potentially important problems that characterize current (immature) CRT markets and which may affect stability and efficiency, such as the concentration of risks due to a low number market participants, a lack of transparency (making it more difficult to evaluate risk exposures), or mispricing because risk buyers may not fully understand the nature of risks involved. Rather we take a prospective view and focus on where CRT initiated by fully rational agents leads us when market imperfections become small, and compare this outcome with the socially desirable one.

To this end, we first develop a simple and tractable model of a financial system which allows us to study the above mentioned effects of CRT. In this financial system firms have access to both bank and market financing. Firms face a trade-off between market and bank financing because of an asymmetric information problem arising from firms' moral hazard. Bank financing helps to reduce the asymmetric information problem and gives firms 'certification' on the market. However, banks are fragile and need to be compensated for taking up credit risk, thus making bank financing more costly.² Stability of the financial system is determined by the (endogenous) riskiness of the portfolios of the banking sector and other financial institutions, while its efficiency depends on the ability of the financial sector to channel funds from households to firms.

We then study the impact of CRT in this model. We begin by considering *benchmark* CRT, i.e., CRT in a world with (almost) no imperfections in CRT markets. We find that CRT unambiguously raises welfare, increasing both efficiency and stability of the financial system. Efficiency increases because CRT enables banks to reduce the risk premium required on loans, thus increasing production in the economy. Stability increases because, first, risk is shifted out of the banking sector into a (assumed to be non-fragile) sector. While banks also increase their lending activities, this additional risk is transferred into the non-bank sector as well. In the case of no imperfec-

²Our model yields an endogenous bank-market financing choice that resembles the analysis in Besanko and Kanatas (1993).

tions in CRT, the economy achieves its first best and bank become pure 'originator-distributor' of loans.

Subsequently, we study perturbations of the benchmark case. First, we acknowledge that CRT may increase the fragility of the risk buyer. We find that CRT is still efficiency enhancing but that it may reduce stability, regardless whether CRT takes place within the banking sector or across sectors. We derive a condition for cross-sectoral CRT to be stability improving and argue that this condition may hold in practice. We also address optimal regulation and find that it should encourage regulatory arbitrage across sectors. This is because regulatory arbitrage is necessary to provide sufficient incentives for financial institutions to choose a (socially) optimal allocation of risk across sectors.

Second, we study asymmetric information problems between risk buyer and seller. When CRT makes banks' less interested in the prosperity of firms, this may result in less monitoring. We find that, as a consequence, efficiency in the economy can fall³ but an increased riskiness of loans (due to less monitoring) does not challenge stability since the additional portfolio risk will be spread.

Finally, we consider CRT instruments that are ineffective, in the sense that they do not allow for a complete shedding of risk (for example, because of counterparty or legal risks). As a result, there will be limits to the extent to which risk can be transferred, which obviously reduces the benefits from CRT. However, we show that stability can actually fall due to ineffective CRT.⁴

The remainder of the paper is organized as follows. In the next section we present our model of the economy. In Section 3 we analyze the benchmark case of CRT in the (near)-absence of market imperfections. Section 4 deals with the impact of CRT on the fragility of the risk buyers. Reduced monitoring incentives and ineffectiveness of CRT instruments are analyzed in Section 5. The final section summarizes and presents the policy implications.

³This confirms results in Morrison (2003). However, in contrast to Morrison we find that these efficiency losses do not necessarily cause disintermediation in the banking sector.

⁴We have started to consider the implications of CRT for systemic risk in the financial sector by allowing for an impact of CRT on the likelihood of counterparty failure. Preliminary results suggest that as long as banks internalize the impact of their transactions on counterparty risk, CRT is still stability enhancing.

2 The Model

Consider a two-period, one good, production economy. There are two types of agents in the economy: entrepreneurs (E) and investors (I). Entrepreneurs operate firms with production technology

$$y(k, e, \eta_i) = f(k, e)\eta_i \quad (1)$$

where k denotes capital, e entrepreneurial effort and η_i is a (firm-specific) productivity shock (whenever it does not create confusion we suppress the firm-specific index i). Capital depreciates completely in the production process. Exercising effort causes private costs $c(e)$ to the entrepreneur. Entrepreneurs do not have a capital endowment and live only in the first period.⁵

We make the following (standard) assumptions about the production function $f(k, e)$ and the cost function $c(e)$ to ensure interior solutions: $f_k > 0$ and $f_e > 0$ (both inputs are productive), $f_{kk} < 0$ and $f_{ee} < 0$ (there are decreasing marginal returns), $f(k, 0) = \lim_{e \downarrow 0} f(k, e) = 0$ (effort is essential in the production process), $c_e > 0$ (exercising effort is costly) and $c_{ee} \geq 0$ (marginal costs of effort are non-decreasing). Furthermore we assume $f_{ek} > 0$ and $f_{ke} > 0$ (capital and effort are complements in production). For some of the results in the paper we need to assume more structure, in particular that $f = k^\alpha e^{1-\alpha}$ and $c_{ee} = 0$.

The production shock η has an idiosyncratic (firm-specific) component ϕ_i and an aggregate component ϕ_w . We assume that $\eta = 1 + \phi_i + \phi_w$ with $E[\phi_i] = E[\phi_w] = 0$, and thus we have $E[\eta] = 1$. The ϕ_i 's are identically and independently distributed across firms. We denote the variances of the shocks with $\sigma_i^2 = \text{var}(\phi_i)$, $\sigma_w^2 = \text{var}(\phi_w)$ and $\sigma^2 = \text{var}(\eta) = \sigma_i^2 + \sigma_w^2$.

Investors are risk-neutral and are endowed with capital but have no production opportunities. They have a storing technology that secures them a (gross) return of 1. There are no assumptions on the ratio of investors to entrepreneurs but it is assumed that the total amount of capital held by investors is sufficiently large to ensure that the economy is never capital-constrained, i.e., all worthwhile investments can be financed. Besides investing in the storing technology, investors can invest in banks (B) and nonbank financial institutions (NB).⁶ Bs and NBs in turn invest in the firms. While NBs have a pure channeling function, banks have additionally a monitoring technology. This technology enables banks to (costlessly) observe entrepreneurial effort. However,

⁵This assumption serves to simplify the analysis by ensuring that the entrepreneur sells his firm completely.

⁶Nonbank financial institution is a collective term for all channels through which funds can be directed from households to firms (except banks). Most of the time we think of them as being insurance companies but one can also interpret them as markets.

they cannot communicate an entrepreneur's effort choice to NBs.

The timing in the economy is as follows. At $t = 0$, entrepreneurs decide on the production inputs in their firm and on the share of their firm sold to Bs and NBs, and consume the revenues. Bs and NBs raise capital from the investors to finance purchases of the firms. At $t = 1$, the productivity shock realizes, production takes place and investors consume the proceeds of their investment.

More formally, at $t = 0$, after having decided upon k and e , an entrepreneur sells claims to his firm's output to NBs and Bs.⁷ We denote with b the fraction of the firm's output sold to Bs ($1 - b$ denotes then the fraction sold to NBs). The entrepreneur's optimization problem can then be written as

$$\begin{aligned} \max_{e,k,b} U(e, k, b) &= V_B(w_B) + V_N(w_N) - c(e), \text{ s.t.} & (2) \\ (i) \quad w_B &= bf(k, e) \\ (ii) \quad w_N &= (1 - b)f(k, \tilde{e}(k, b)) \end{aligned}$$

where the entrepreneur's utility $U(e, k, b)$ consists of consumption (equal to the revenues from selling the firm $V_B(w_B) + V_N(w_N)$ minus the cost of effort $c(e)$). $V_B(w_B)$ [$V_N(w_N)$] denotes the revenue from selling a fraction b [$1 - b$] of the firm to Bs [NBs]. The crucial difference between selling to the Bs and to NBs is that when selling to a B, the entrepreneur is remunerated according to the actual effort chosen e (since Bs can observe his effort choice), while when selling to NBs his remuneration will be according to the NBs expectation of his effort choice $\tilde{e}(k, b)$ based on the entrepreneur's choice of k and b . Since NBs are rational, they will correctly anticipate the entrepreneur's effort choice in equilibrium, i.e., $\tilde{e}(k, b)$ solves

$$\begin{aligned} \tilde{e}(k, b) &= \arg \max_e V_B(w_B) + V_N(w_N) - c(e), \text{ s.t.} & (3) \\ (i) \quad w_B &= bf(k, e) \\ (ii) \quad w_N &= (1 - b)f(k, \tilde{e}(k, b)) \end{aligned}$$

Next we turn to the optimization problem for investors and financial institutions. Because of their outside option (the storing technology), investors require a return on capital of at least 1 (in expectation). Competitiveness of households then implies that the required expected return

⁷Hence, it is implicitly assumed that the only contracts available in the economy are equity contracts. Allowing for more general contracts is not expected to change the main results as long as they do not allow a full separation of firm risk and incentives to monitor.

on capital for Bs and NBs is 1. Assume that there is no CRT. Then, Bs' (gross) return on their investment is

$$x_B = w_B \eta \quad (4)$$

In absence of any fragility in the financial sector this would imply that the required (expected) return on investing in firm's output is 1 (since the financial sector is assumed to be competitive as well), in particular: $V_B(w_B) = w_B$ and $V_N(w_N) = w_N$. However, it is assumed that Bs become bankrupt at $t = 1$ if the loss in their portfolio exceeds a certain threshold (consistent with a value-at-risk constraint). If bankrupt, they incur an exogenous cost (for example, because the portfolio cannot be sold at a fair price). It can then be shown that for a normally distributed portfolio, the expected loss from bankruptcy for a bank can be written as $\frac{\alpha_B}{2} \sigma_B^2$, where σ_B^2 is the variance of Bs portfolio and α_B is a parameter that measures the expected cost of a unit of bank risk (see Danielsson and Zigrand, 2003). Assuming that there is a one-to-one relationship between firms and Bs (i.e., each firm is financed through one bank only and this bank does not finance other firms), the expected value of the bank portfolio is $w_B - \frac{\alpha}{2} \sigma_B^2$ with $\sigma_B^2 = b^2 f(k, e)^2 \sigma^2$.⁸ Since investors have to be compensated for the expected losses due to bankruptcy we have

$$V_B(w_B) = w_B - \frac{\alpha_B}{2} \sigma_B^2 \quad (5)$$

Hence, Bs will behave as if they were risk-averse: their return profile is identical to CARA utility with an (absolute) risk-aversion parameter of α . Similarly, fragility in the NB-sector implies

$$V_N(w_N) = w_N - \frac{\alpha_N}{2} \sigma_N^2 \quad (6)$$

Besides causing costs to the financial institution itself, bankruptcy also causes social costs (for example due to bank runs, failure of payment system, counterparty risk). These costs are assumed to be borne by all investors in the economy and simply reduce the return on their investments. These costs are not internalized by the financial institution and thus create a scope for regulation. Analogous to (5) and (6) we denote the expected social cost of a B failing and a NB failing by $\beta_B \sigma_B$ and $\beta_N \sigma_N^2$, where β_B and β_N measure the social cost of a bankruptcy (expressed in terms of reduced returns on investment). Since ex-ante all Bs are identical and all NBs are identical, we can write the expected social cost from bankruptcies in the financial sector (the social cost of

⁸The assumption of a one-to-one relationship simplifies the analysis in that respect that the risk premium required for loans only depends on the amount of firm risk held by the bank.

financial instability) as⁹

$$I = \beta_B \sigma_B^2 + \beta_N \sigma_N^2 \quad (7)$$

The following definition completes the description of the economy.

Definition 1 *An equilibrium allocation in the economy is given by a triple of actions (k^*, b^*, e^*) for each entrepreneur, such that (k^*, b^*, e^*) solve (2), with $\tilde{e}(k, b), V_B(w_B), V_N(w_N)$ given by (3), (5) and (6).*

Corollary 1 *(Expected) welfare in the economy is determined by I (decreasing) and U (increasing).*

Corollary 1 follows because an investors' welfare (=expected consumption) depends only on the expected social costs in the economy (since they earn an expected return of 1 on their investments in absence of instability), while entrepreneurs' welfare is solely determined by U as defined in (2). We will refer to I as the instability of the financial system and to U as the efficiency of the financial system (since improvement in U can only arise from efficiency gains in the financial sector that are passed on to the entrepreneurs).

Lemma 1 *The FOCs for e, k, b in (2) are*

$$e^* : V'_B b f_e = c_e \quad (8)$$

$$k^* : [bV'_B + (1-b)V'_N] f_k + (1-b)V'_N f_e \frac{\partial \tilde{e}}{\partial k} = 1 \quad (9)$$

$$b^* : (1-b)V'_N f_e \frac{\partial \tilde{e}}{\partial b} = (V'_N - V'_B) f \quad (10)$$

Proof. Straightforward from (2).

Lemma 2 *Increases in bank financing b and/or capital k increase the equilibrium effort choice, i.e., $\partial \tilde{e} / \partial b > 0$ and $\partial \tilde{e} / \partial k > 0$.*

Proof. See appendix.

Equilibrium allocations in our economy are generally not first best. First, this is because agents' optimization ignores the social cost of instability (definition 1). Moreover, there is a second type of inefficiency arising in production. This can be seen from the FOCs of the entrepreneur's

⁹A crucial assumption in (7) is that the social cost of a (single) institution failing does not depend on whether other Bs and NBs are failing.

problem (which have been written such that marginal benefits on the left hand side are equated with marginal costs on the right).

Equation (8) reveals that chosen effort is inefficiently low (compared to the first best effort choice, which would require $f_e = c_e$). There are two reasons for this. First, because of the fragility of the banking system, banks require a risk premium for buying firms output, i.e., $V'_B < 1$ (follows from equation 5). Second, the entrepreneur is only remunerated according to the bank's stake in the firm, adding another inefficiency when the firm is not fully financed by the bank. (9) shows that the capital choice is also not first best (which would imply $f_k = 1$ because the marginal cost of capital is foregone consumption). This is because the risk premia required by Bs and NBs imply that the average market value of the firm $bV'_B + (1 - b)V'_N$ is smaller than 1, and thus the return on capital is inefficiently low. However, there is a potentially compensating effect: a higher amount of capital invested in the firm 'signals' a higher effort choice to the NBs and thus increases the value at which the firm can be sold to NBs (second term on the left hand side of 9). The reason for the positive signalling value of capital is that an increase in k increases the productivity of effort (due to our assumption on the complementarity of inputs), thus increasing the equilibrium effort choice, which in turn increases the effort anticipated (\tilde{e}) by NBs (Lemma 2).

Equation (10) determines the extent of bank financing b in the economy. The right hand side represents the (marginal) cost of additional B-financing (compared to NB-financing): because of higher fragility in the banking sector we would generally have $V'_N - V'_B > 0$. The left hand side gives the marginal gains from additional bank financing, which again is due to a signalling effect: an increase in b increases an entrepreneur's remuneration for effort (from 8) and thus increases equilibrium effort and anticipated effort (Lemma 2). This is the bank certification effect stressed in the intermediation literature: bank financing enhances the market value of firms since markets anticipate that because of banks' monitoring there will be a more efficient effort choice by the entrepreneur. Hence, equation (10) indicates that there is a trade-off between the cost of bank financing (compared to NB-financing) and the benefits from bank financing due to the certification effect.

This concludes our exposition of the economy. In the remainder of the paper we analyze CRT by allowing Bs and NBs to trade firm risks. This changes the portfolio variances of Bs and NBs, σ_B^2 and σ_N^2 , and thus has a direct effect on stability (equation 7). Furthermore, the changes in σ_B^2 and σ_N^2 affect the risk premia required by Bs and NBs (and thus change $V_B(w_B)$ and

$V_N(w_N)$), which have an impact on entrepreneurial actions and thus the efficiency of the financial system. The model thus naturally allows us to study the impact of changes in CRT markets on the performance of the financial system.

3 Benchmark CRT

In this section we analyze CRT under ideal conditions. In particular, we assume that:

- CRT does not increase the fragility of the institution buying the risk. This is ensured in our framework by studying CRT from Bs to NBs and assuming that the NB-sector is not fragile, i.e., $\alpha_N = \beta_N = 0$ and hence

$$V_N(w_N) = w_N \tag{11}$$

$$V'_N(w_N) = 1 \tag{12}$$

This assumption is relaxed in Section 4.

- CRT does not create incentive problems between risk buyer and seller (for example, because of reputational reasons). Specifically, we assume that firms are valued in the risk transfer according to their *actual* effort level (while direct financing by the NB is based on *inferred* effort). Note that, as discussed in Section 5.1. (where we relax this assumption), this does not imply that the NB can observe the actual effort.¹⁰
- CRT is fully effective in the sense that it allows banks (in principal) to shed all firm risk from their portfolio (this also excludes leakage from the NBs back into the banking sector). This is relaxed in Section 5.2.

We model CRT as follows. After Bs have bought claims to firms (and hence entrepreneurs have set k and e), they can sell these claims to NBs at price p . Selling the claim entails a proportional cost τ on the value of the credit risk transferred, which is borne by the Bs. This cost has a broad interpretation in being related to imperfections in CRT markets (such as, for

¹⁰It can be shown that this assumption produces the same results as would arise if firms are valued in the CRT according to inferred effort but the CRT process allows the separation of the firm's risk from the firm's output CRT (as would be possible when Arrow-Debreu-securities are traded), i.e., the CRT instruments do not pose incentive problems.

example, because of transaction costs).¹¹ Denoting a bank's exposure to a firm (i.e., bank's risk taking) with g ($g = bf$ since there is one-to-one relationship between firms and banks), B's return at $t = 1$, given that a share $q \geq 0$ of $g\eta$ is sold, is

$$x_B = (1 - q)g\eta + qp - \tau pq \quad (13)$$

Since NBs are risk-neutral ($\alpha_N = 0$) and competitive, and since it is assumed that CRT is valued at actual effort, the price the bank is able to obtain is

$$p = g = g(e, k) \quad (14)$$

B's return can then be written as

$$x_B = (1 - q)g\eta + qg - \tau gq \quad (15)$$

and analogous to (5) we have the bank value of firm's output is

$$V_B = g - \tau g - \frac{\alpha}{2}(1 - q)^2 g^2 \sigma^2 \quad (16)$$

Since g is given when banks choose q , from differentiating 16 wrt. q , the optimal proportion of the bank portfolio sold is¹²

$$q^* = 1 - \frac{\tau}{\alpha g \sigma^2} \quad (17)$$

Hence, bank risk σ_B^2 is

$$\sigma_B^2 = \frac{\tau^2}{2\alpha\sigma^2} \quad (18)$$

Inserting (18) into (16) we find that the value of firm's output for the bank V_B and the marginal bank value of output V'_B are affected by the cost of CRT τ

$$V_B = (1 - \tau)g + \frac{\tau^2}{2\alpha\sigma^2} \quad (19)$$

$$V'_B = 1 - \tau \quad (20)$$

Hence, the possibility for Bs to sell risk subsequent to their investment in firms changes their required returns on investment and will thus affect entrepreneurial decisions. Substituting V'_B

¹¹It is likely that there are additional fixed costs of CRT. However, their incorporation would not yield additional insights since they will not affect marginal CRT decisions.

¹²Note that because of our definition of the cost of CRT (equation 13), q is restricted to be larger than 0. Hence, equation (17) (implicitly) imposes restrictions on the model parameters.

and V'_N (equations 20 and 11) into the entrepreneur's FOCs (8)-(10) we get

$$e^* : b(1 - \tau)f_e = c_e \quad (21)$$

$$k^* : (1 - b\tau)f_k + (1 - b)f_e\tilde{c}'(k) = 1 \quad (22)$$

$$b^* : (1 - b)f_e\tilde{c}'(b) = \tau f \quad (23)$$

Equations (21)-(23) enable us to study the impact of CRT on firm's decisions, where CRT is viewed as being caused by a reduction τ . One could simply compare the cases of no CRT ($\tau = 1$) and perfect CRT ($\tau = 0$). However, viewing τ as continuous, we are able to analyze the implications of a partial reduction in market imperfections.¹³ From now on, we will use 'increase in CRT' and 'reduction in τ ' interchangeably.

Lemma 3 *An increase in CRT increases intermediation b and effort e but does not affect investment. Banks' risk taking ($b\eta$) goes up.*

Proof. See Appendix

The intuition for the results in Lemma 3 is as follows. An increase in CRT (triggered by a reduction in τ) reduces B's exposure to firm risk which reduces the risk premium demanded and, hence, the marginal bank value of output V'_B goes up. This leads, first, to an increase in intermediation b because it reduces the marginal bank financing cost relative to NB financing $V'_N - V'_B$ (right hand side of 23). Second, it leads to a higher remuneration for effort e and thus a higher effort choice (equation 21). Third, it increases the remuneration for capital to the extent that the firm is financed by the bank (equation 22). However, there are two offsetting effects on the remuneration for capital, which come through the increase in intermediation b . First, an increase in b reduces the remuneration for capital since the marginal value of output for Bs is less than for NBs. Second, an increase in b reduces the value of the signalling effect of k , since this applies only to the share of the firm sold to NBs. In the proof of Lemma 3 it is shown that for our assumptions on the production function, all three effects on capital exactly offset each other and thus the total impact of an increase in CRT on k is zero.¹⁴ Finally, banks' risk taking goes

¹³Costs are only one type of market imperfections. In subsequent sections we focus on other imperfection arising from incentive problems between Bs and NBs and ineffective CRT.

¹⁴This independence result of capital on the firm level is not expected to extend to the total amount of capital invested in the economy if the number of projects is endogenous. This is because CRT increases the value of output V_B and thus would make undertaking a marginal project worthwhile.

up because both intermediation b and average output f goes up (the latter because effort has increased).

Proposition 1 *An increase in CRT unambiguously raises welfare (i.e., it increases both stability and efficiency).*

Proof. 1. Stability: From (7) and $\beta_N = 0$ we have that stability increases iff $\partial\sigma_B^2/\partial(-\tau) < 0$. $\partial\sigma_B^2/\partial(-\tau) < 0$ follows directly from (18). 2. Efficiency: We have

$$\begin{aligned} \frac{dU(b, k, e, \delta)}{d(-\tau)} &\geq \frac{\partial U(b, k, e, \delta)}{\partial(-\tau)} \\ &= -\frac{\partial U}{\partial V_B} \frac{\partial V_B(\tau)}{\partial \tau} - \frac{\partial U}{\partial V_N} V_N' f_e \frac{\partial \tilde{e}}{\partial \tau} > 0 \end{aligned}$$

where the first inequality follows because adjustments in the entrepreneur's choice variables will not make him worse off; the second inequality follows from $\partial V_B(\tau)/\partial \tau < 0$ (from 19) and $\partial \tilde{e}/\partial \tau \leq 0$ (from Lemma 3).

Proposition 1 shows that, although banks take up more risk, the stability of the financial system increases. In principal, there are two effects of increased CRT on the variance of banks' returns σ_B^2 (which determine stability of the financial system since $\beta_N = 0$). First, increased CRT increases q and thus directly reduces σ_B^2 . Second, bank risk taking g goes up, potentially increasing σ_B^2 . However, equation (18) shows that σ_B^2 is independent of g . This implies that any additional risk taken up by the bank is transferred to NBs, and hence the first effect prevails. The intuition for this result lies in the properties of the CARA-utility, under which the amount invested in a risky asset is constant regardless of the size of the portfolio. Proposition 1 further shows that efficiency is also increased since an increase in CRT reduces the risk premium required by banks and thus reduces the cost of bank financing for entrepreneurs. Hence, welfare rises.

Proposition 2 *For $\tau = 0$ (full CRT) the economy achieves the first best outcome.*

Proof. Efficiency: From (23) we find that the relative marginal cost of bank financing is zero, while the benefits from bank financing are larger than zero, hence $b = 1$; it then follows from (21) and (22) that $f_e = c_e$ and $f_k = 1$, implying that effort and capital are chosen efficiently. Stability: From (17) and (18) we have that $q^ = 1$ and $\sigma_B^2 = 0$ and it follows that $I = 0$.*

Proposition (2) tells us that when CRT is costless, there are no longer any inefficiencies in the economy. The reason is that monitoring incentives and risk bearing can then be separated

without cost. Banks become pure 'originator-distributors' of finance: they 'originate' all financing in the economy ($b = 1$) and thus ensure an efficient effort choice but distribute all claims to the NBs ($q = 1$). Thus, all firm risk is held by the institutions that are not fragile.

4 CRT and Fragility of the Risk Buyer

4.1 CRT within the Banking Sector

In the previous section we analyzed CRT into a sector that is not fragile. From a stability perspective, the risk was simply disappearing. In reality, however, stability is likely to be affected by increased risk taking by the risk buyer. In this section we study CRT within the banking sector, where the presence of such an effect is obvious. We continue to assume that there is no fragility in the NB-sector (this assumption will be relaxed in the next subsection). Since banks are ex-ante identical, gains from CRT within the banking sector can only arise from diversification of idiosyncratic firm risk. We therefore directly consider trade in the idiosyncratic risk component.

Assume to this end that B's portfolio at $t = 1$ is given by

$$x_B = g\eta - qg(1 + \phi_i) + qp - \tau pq \quad (24)$$

where q refers to the extent to which idiosyncratic risk is sold. This portfolio could, for example, be the result of banks selling their claims on firms to a financial institution which pools the claims and sells them back to the banks. Assuming that there are a large number of firms, by the law of large numbers the pool bought back will only contain aggregate risk. Thus the operation effectively sheds idiosyncratic risk from banks' portfolios.

Since idiosyncratic risk is not priced, we have that $p = g(e)$ in (24) and analogous to the previous section the optimal q can be derived as

$$q^* = 1 - \frac{\tau}{\alpha g \sigma_i^2} \quad (25)$$

which implies a portfolio variance σ_B^2 of

$$\sigma_B^2 = \frac{\tau^2}{2\alpha\sigma_i^2} + g^2\sigma_w^2 \quad (26)$$

Lemma 4 $g'(V'_B)/g < -1$.

Proof. See Appendix.

Proposition 3 *Increased CRT within the banking sector increases efficiency but will eventually reduce stability.*

Proof. Efficiency: analogous to the proof of Proposition 1. Stability: Total derivative of σ_B^2 wrt. τ is

$$\begin{aligned}\frac{d\sigma_B^2}{d\tau} &= \frac{\tau}{\alpha\sigma_i^2} + 2gg'(V'_B)\frac{\partial V'_B}{\partial\tau} \\ &< \frac{2\tau}{\alpha} - 2g^2\end{aligned}\tag{27}$$

where Lemma 4 has been used to substitute for $g'(V'_B)$. If τ becomes sufficiently small, (27) will be negative.

The efficiency result is obvious from the discussion of Proposition 1: Increased CRT reduces the risk premium required by banks and thus directly increases entrepreneur's utility. The impact on stability is less straightforward and generally ambiguous. As equation (26) shows, bank risk does not depend on the individual risk component of the initial portfolio $g^2\sigma_i^2$ (which is completely shed as in the benchmark case) but does depend on the aggregate risk component $g^2\sigma_w^2$ (since aggregate risk cannot be shed within the banking sector). Hence, if additional risk taking by banks is sufficiently large, the impact on banks' portfolio risk may be enough to outweigh the stabilizing impact of the diversification of idiosyncratic risk. Proposition 3 also states that as τ decreases, stability eventually falls. Intuitively, this is because the impact of diversification on stability becomes smaller as τ decreases (and eventually becomes zero), while additional risk taking does not fall below a certain (positive) level (as Lemma 4 shows).

4.2 CRT across Financial Sectors

We consider now CRT between the B and NB-sector and allow for fragility in the NB-sector: $\alpha_N > 0$, $\beta_N > 0$. Our model for CRT is similar to the previous subsection: banks sell claims to firm's output to a financial institution which pools all claims but sells them now to NBs (which are thus buying only aggregate risk). Denoting with q_B the fraction of a bank's portfolio sold, analogous to the benchmark case we have Bs portfolio at $t = 1$ given by

$$x_B = (1 - q_B)g\eta + pq_B - \tau gq_B\tag{28}$$

NBs portfolio consists of their initial holdings of firms (only consisting of the aggregate component, since idiosyncratic firm risk is completely shared across NBs), the claims to firms bought in the

CRT minus its price

$$\begin{aligned}
x_N &= (1-b)f(1+\phi_w) + g(1+\phi_w)q_N - pq_N \\
&= g\left[\left(\frac{1-b}{b} + q_N\right)(1+\phi_w)\right] - pq_N
\end{aligned} \tag{29}$$

where q_N refers to the amount of aggregate firm risk bought (expressed as a fraction of a bank's portfolio).

Lemma 5 *Equilibrium CRT is given by*

$$q = q_B = q_N = \frac{\alpha_B\sigma^2 - \alpha_N\sigma_w^2\frac{1-b}{b} - \tau/g}{\alpha_N\sigma_w^2 + \alpha_B\sigma^2} \tag{30}$$

and resulting B and NB-risk is

$$\sigma_B^2 = \left(\frac{\alpha_N\sigma_w^2 f + \tau}{\alpha_N\sigma_w^2 + \alpha_B\sigma^2}\right)^2 \sigma^2 \tag{31}$$

$$\sigma_N^2 = \left(\frac{\alpha_B\sigma^2 f - \tau}{\alpha_N\sigma_w^2 + \alpha_B\sigma^2}\right)^2 \sigma_w^2 \tag{32}$$

Proof. See Appendix.

Proposition 4 *An increase in CRT raises intermediation b (if $\alpha_B > 0$).*

Proof. See Appendix.

The impact of increased CRT on b depends on the marginal bank financing costs relative to market financing: $(\partial V_B/\partial f)/(\partial V_N/\partial f)$ (follows from the FOC for b , equation 10). As shown in the proof of Proposition 4, independent of the effective risk aversion parameters α_B and α_N , relative marginal bank financing costs are reduced by CRT because it reduces banks' exposure to idiosyncratic risk.

For the remainder of this section, we simplify the analysis by assuming that k and e are constant.

Proposition 5 *Increased CRT (i) improves stability for all τ if and only if*

$$\alpha_N/\alpha_B > \beta_N/\beta_B \tag{33}$$

(ii) always increases efficiency.

Proof. See Appendix.

Proposition 5 reveals that for increased CRT to be stability-improving, it is not sufficient that risk flows from the socially more fragile sector into the less fragile sector (we interpret α and β as private and social fragility, respectively). To demonstrate this, assume that $\beta_B > \beta_N$ and $\alpha_B > \alpha_N$ (the B-sector is socially and privately more fragile than the NB sector). Since $\alpha_B > \alpha_N$, there is a tendency for risk to flow into the NB-sector (depending on the relative initial risk positions in both sectors, given by b), however, as can be easily checked condition (33) is not necessarily fulfilled. The reason is that because instability in the financial system is additive in the risks in each sector (equation 7), minimizing instability is similar to standard portfolio optimization with two uncorrelated assets: the variance minimizing combination of both assets involves investment in both assets (though relatively more is invested in the less risky asset).

What is the general intuition behind condition (33)? Ignoring the cost of CRT for the moment, when the ratio of private and social fragilities are identical across sectors (equation 33 with an equality sign), the private sector achieves a risk allocation across sectors that is also socially desirable. However, if the the ratio of private to social fragility is larger in the NB-sector (condition 33), the NB-sector behaves 'too risk-averse' from a social perspective, with the result that there is too much risk in the B-sector. A reduction in τ , reducing the cost of CRT, increases CRT into the NB-sector (equation 30) and is thus socially desirable.

Is condition (33) fulfilled in reality? The answer is not obvious. It is realistic to assume that $\beta_B > \beta_N$ and $\alpha_B \geq \alpha_N$, but, as mentioned above, this does not imply that condition (33) is fulfilled. However, given the large fragility of the banking sector, β_B may be substantially larger than β_N , while the effective risk aversion in both sectors, α_B and α_N may not differ substantially. Hence it is not implausible to conjecture that condition (33) is met.

The above discussion suggests that increased CRT is only welfare improving because the relative private fragilities are not leading to a socially efficient outcome in the first place. This suggest scope for regulation. Regulators, through capital requirements and other regulatory instruments, can in practice effectively set parameters α_B and α_N (keep in mind that α implies a certain value at risk). Proposition 6 sheds light on what determines the optimal ratio of α 's across sectors, $r = \alpha_N^R/\alpha_B^R$, in terms of maximizing stability.

Proposition 6 *For fixed output f , optimal (stability maximizing) regulation r is given by*

$$r^* = \beta_N/\beta_B - \frac{\tau(\sigma_w^2\beta_N + \sigma^2\beta_B)}{f} \quad (34)$$

Hence, optimal regulation generally encourages regulatory arbitrage ($\alpha_N/\alpha_B < 0$) if $\beta_N < \beta_B$.

Furthermore, optimal regulatory arbitrage exceeds the relative externalities across sectors ($\alpha_N/\alpha_B < \beta_N/\beta_B$)

Proof. See appendix.

Hence, as our discussion above suggested, in absence of CRT costs ($\tau = 0$), $r^* = \beta_N/\beta_B$, i.e., optimal regulation sets the ratio of private fragility costs equal to the social costs. However, if CRT is costly and $r^* = \beta_N/\beta_B$, CRT is inefficiently low from a social perspective since banks have to bear the costs of CRT and thus transfer less. Optimal, stability maximizing regulation therefore sets r in excess of the ratio of social fragilities across sectors in order to compensate for this underincentive to transfer credit risk.

What can be said about the relative merits of cross-sectoral CRT compared to CRT within the banking sector? Our analysis suggest that cross-sectoral CRT may be preferable in several respects. First, it allows for the transfer of risk into the less fragile NB-sector. Second, it permits, in contrast to CRT among banks, for the complete shedding of additional risks taken on by banks. Finally, diversification gains are higher across sectors than within sectors.¹⁵

5 Imperfections in CRT Instruments

5.1 Incentive Problems between Risk Buyer and Seller

So far it has been assumed that CRT does not cause incentive problems between Bs and NBs. However, when buying firm risk from banks, NBs may face the same problems as when buying from the entrepreneur directly, in that they cannot directly observe the entrepreneur's effort choice. Hence, there may be an incentive for banks to agree on a lower effort choice with the entrepreneur because the bank does not benefit from effort exercised on that part of the firm it sells to NBs.¹⁶

In practice there are several effects that mitigate this problem and effectively justify the assumption made earlier that firms are in the CRT valued according to actual effort. For example, banks' reputation may ensure that banks transfer only firms which operate efficiently. Additionally, CRT instruments may be designed to reduce possible incentive problems.¹⁷ Furthermore,

¹⁵Our setup understates the gains from cross-sectoral diversification. This is because typically Bs and NBs have exposures to different types of risks.

¹⁶A broader discussion on incentive problems caused by CRT can be found in Kiff et. al. (2003).

¹⁷See for example Gorton and Pennacchi (1995[15]), DeMarzo and Duffie (1999) and Duffie and Zhou (2001).

the risk buyer may be partially able to take over monitoring from the bank. However, these effects will generally not eliminate the problem.

In this section we study the impact of these additional incentive problems caused by CRT. Morrison (2003) has shown that CRT can reduce welfare and lead to disintermediation by undermining banks' role as an intermediary that mitigates asymmetric information problems in the economy. There may also be an impact on stability. When banks can transfer firm risk, they may have less incentives to monitor firms, and as a consequence, the riskiness of firms may increase.

We deliberately choose a setup that maximizes the (adverse) impact of CRT on incentive problems by assuming that there are no mitigating effects as described above. Consequently, a bank benefits from entrepreneurial effort only according to its stake in the firm *after* CRT (because the price at which the firm is transferred to NBs does not depend on the actual effort choice).¹⁸ It is further assumed that banks cannot commit to retain a certain stake in the firm after CRT. This assumption is plausible because credit derivatives are largely traded over-the-counter, which makes it difficult to verify the extent to which a bank remains exposed to firm risk after CRT.

The price at which a firm's output can be transferred in the CRT depends now on the inferred effort choice $\tilde{e}(k, r)$. Denote with $\tilde{g} = g(\tilde{e}(b, r))$ the bank's claim to output based on inferred output and let $g = g(e)$ continue to denote the bank's actual claim to output. Analogous to the benchmark case ($\alpha_N = \beta_N = 0$) case we have

$$p = \tilde{g} \tag{35}$$

Banks' portfolio is given by

$$x_B = (1 - q)g(e) + q\tilde{g} - \tau\tilde{g}q \tag{36}$$

from which we can derive

$$q^* = 1 - \frac{g - (1 - \tau)\tilde{g}}{\alpha g \sigma^2} \tag{37}$$

Resulting bank risk and bank value of output are

$$\sigma_B^2 = \frac{(g - (1 - \tau)\tilde{g})^2}{\alpha g^2 \sigma^2} \tag{38}$$

$$V_B(g, \tilde{g}, \tau) = (1 - \tau)\tilde{g} + \frac{g^2/2 - (1 - \tau)\tilde{g}g + 1/2(1 - \tau)^2\tilde{g}^2}{\alpha g^2 \sigma^2} \tag{39}$$

The use of CRT instruments may even improve incentives, see Arping (2003).

¹⁸It is straightforward to generalize the analysis by studying intermediate cases, i.e., cases between the benchmark case and the analysis in this section.

Lemma 6 $\partial V_B/\partial g = (1 - \tau)(1 - q)$ and $\partial V_B/\partial \tilde{g} = (1 - \tau)q$ after setting $g = \tilde{g}$.

Proof: See Appendix.

Lemma 7 The FOCs for e, k, b are

$$e^* : b(1 - q)(1 - \tau)f_e = c_e \quad (40)$$

$$k^* : (1 - b\tau)f_k + [1 - b + q(1 - \tau)b]f_e\tilde{e}'(k) = 1 \quad (41)$$

$$b^* : [1 - b + q(1 - \tau)b]f_e\tilde{e}'(b) = \tau f \quad (42)$$

with $q = q(\tau)$ given by (37)

Proof. see Appendix.

The difference in the FOC for e compared to the benchmark case (equation 21) is the $(1 - q)$ term on the LHS of (40). Since effort does not affect the price at which the bank transfers claims to the NB sector, the entrepreneur will only be remunerated according to the (anticipated) bank's stake in the firm *after* CRT: $b(1 - q\tau)$. Hence, effort choice is lower than in the benchmark case (for a given τ and thus q). The FOC for k (equation 41) changes in comparison to the benchmark case because signalling becomes more valuable than in the benchmark case since a higher inferred effort choice does now also increase the bank's value of output. This is because an increase in the inferred output increases the price at which banks can transfer firm output, which is represented by the additional term $q(1 - \tau)b$ in equation (41). Similarly, the FOC for b also changes because of this increased value of signalling.

Proposition 7 An increase in CRT (i) enhances financial stability, (ii) eventually reduces efficiency, (iii) increases intermediation.

Proof. (i) Stability: immediate with (38) after setting $\tilde{g} = g$. (ii) Efficiency: For $\tau \rightarrow 0$ we have from (17) that $q \rightarrow 1$. Hence, we get from (40) that $e \rightarrow 0$. From $\lim_{e \downarrow 0} f(k, e) = 0$ (assumption that effort is necessary for production), we obtain that $f \rightarrow 0$ and hence $U \rightarrow 0$. (iii) Intermediation: For $f(k, e) = k^\alpha e^{1-\alpha}$ we have $f_e\tilde{e}'(b)/f = \bar{c}/b$ where \bar{c} is a positive constant (see Appendix, proof for Lemma 3). Inserting in (42) and solving for b gives $b = \bar{c}/(\tau + \bar{c}(1 - q(1 - \tau)))$. Hence $b'(\tau) < 0$ since $\bar{c} > 0$.

Increased CRT increases stability for the same reason as in the benchmark case. In fact, the equation for σ_B^2 (equation 38) becomes identical to (17) for $g = \tilde{g}$. Thus, bank risk does not depend on additional risk taking (and hence there is only the direct stability increasing impact

of the risk transfer). This general insight suggests that even if firms become more risky due to the reduced effort choice (which is not modelled here), this will not impact bank risk since any additional risk will be shed by banks.

Due to the impact of CRT on banks' incentives, there are two counteracting effects on efficiency. As in the benchmark case, increased CRT reduces risk premia and thus the financing costs, which enhances efficiency. However, CRT now reduces banks' participation in the firm. This lowers the entrepreneur's remuneration for effort and increases the inefficiency in the effort choice (in terms of the financial intermediation literature, the certification value of bank financing is reduced). The net effect is generally ambiguous. However, Proposition 7 shows that efficiency eventually decreases. This is because as τ becomes small ($\tau \rightarrow 0$), banks will transfer the whole firm ($q \rightarrow 1$), thus the entrepreneur will not be remunerated for effort at all (the left hand side of equation 40 becomes zero) and his effort choice goes to zero. Since effort is assumed to be very productive at small levels, this process reduces efficiency.¹⁹

As in the benchmark case, a reduction in τ reduces the relative cost of bank financing (right hand side of equation 42). On top of that, the signalling value of b increases because an increase in τ increases the stake in the firm ultimately held by the market and thus the importance of inferred effort. As a consequence, perhaps surprisingly, although efficiency is eventually reduced, this does not cause disintermediation, i.e., firms do not substitute bank for market financing.

This result is in contrast to Morrison (2003), who finds that CRT tends to cause disintermediation. We conjecture that the difference in results obtains because in Morrison's model banks incur a fixed cost of monitoring and that the effort choice is discrete (to monitor or not to monitor). Hence, if the (absolute) certification value of bank financing becomes small because of reduced monitoring incentive due to CRT, it is no longer worthwhile incurring this monitoring cost and certification value becomes zero. Firms will then be completely market financed. In our model, there are no fixed costs and what determines the degree of bank financing is the marginal signalling value of bank financing (LHS of equation 42), i.e., the marginal certification value. As described above, the latter increases due to increased CRT. On top of that, CRT makes bank financing less costly.

In reality, there are likely to be both fixed and variable elements in monitoring. Banks may choose not to monitor at all if their exposure to risk is small (fixed effect), however, once above a

¹⁹It should be kept in mind that our setup is extreme in that it maximizes the incentive problems caused by CRT.

certain threshold, the incentive to monitor will generally depend on the size of the exposure. For a hybrid model with fixed and flexible monitoring elements we would expect that intermediation increases up to a certain point (because of the effects described in our model) but eventually jumps to zero (when incurring the fixed monitoring costs is no longer worthwhile).

The fact that even though increased CRT causes inefficiency it may not reduce intermediation suggests a potential problem: since their business is not undermined (and may even be strengthened) by this inefficiency, banks may lack the incentives to develop CRT instruments that minimize incentive problems. As a consequence, welfare increasing innovations may not take place and regulation may be called for.

5.2 Ineffective CRT Instruments

In this subsection we relax the assumption that CRT allows banks to shed risks completely. There are good reasons why a complete shedding of risks is unrealistic. In practice, CRT is subject to several imperfections such as basis risk, counterparty risk and legal risk (arising from uncertainty about the definition of credit events). CRT may also be incomplete in a broader sense. Transferring risk into the NB sector may increase the likelihood of failure in the latter. Typically, a bank will have linkages with the NB-sector besides CRT. A failure in the NB sector may then leak back into the banking sector. As an example, consider an insurance company that has a credit line with a bank. In the case of a credit event, the insurance company may draw upon the credit line. Thus, (liquidity) risk is ultimately still allocated in the banking sector.

Such incompleteness in CRT can be formalized by recognizing that CRT instruments cannot provide a full hedge against banks' portfolio risk. To this end, we generalize our model of CRT from the benchmark case as follows. Assume that a CRT instrument for a specific firm has a pay-off $g\eta_h$. Denote with $\rho = \text{cor}(\eta, \eta_h)$ its correlation with the bank's portfolio. Without loss of generality assume that $\rho > 0$ (i.e., the bank has to sell the asset to hedge its portfolio) and assume furthermore $\sigma_h^2 = \sigma^2$. The benchmark case arises then in the case $\rho = 1$.

Banks' portfolio is then

$$x_B = g\eta - qg\eta_h - \tau gq \quad (43)$$

and from the FOC for q we find

$$q = \rho - \frac{\tau}{\alpha g \sigma^2} \quad (44)$$

Hence, resulting bank risk is

$$\sigma_B^2 = \frac{\tau^2}{2\alpha\sigma^2} + g^2(1 - \rho^2)\sigma^2 \quad (45)$$

Proposition 8 *If CRT instruments are ineffective ($\rho < 1$), increased CRT will eventually reduce stability.*

Proof. Analogous to Proposition 3, second part.

The impact of increased CRT on stability is generally ambiguous when $\rho < 1$. The reason is exactly the same as for CRT within the banking sector (Proposition 3): additional risk cannot be completely shed, here because the instruments do not *allow* for complete shedding. This can be seen by looking at equation (45): bank risk depends on bank risk taking g . A similar conclusion is reached by Instefjord (2003) who studies optimal credit risk hedging by a bank in continuous time. He concludes that risk taking may go up when the elasticity of loans is large enough. Unfortunately, there is no explicit solution to his hedging problem and only limited insights can be gained into what determines the trade-off between the stability enhancing and stability reducing impacts of CRT. Equation (45) reveals that this trade-off depends crucially on the hedging ability of CRT instruments. The more effective instruments are ($\rho \uparrow$), the more additional risk taking can be spread. In the case of $\rho = 1$, all risks can be spread and stability is unambiguously improved.

6 Summary and Policy Implications

This paper has analyzed the impact of CRT on the stability and the efficiency of the financial sector. Regarding stability, our results suggest that with respect to CRT, individuals' incentives are generally aligned with regulators' incentives. Although banks ignore the social cost of a failure of their institution, due to their risk-aversion they have an incentive to diversify and to shift risk out of the banking sector, which tends to increase stability. Improving the opportunities for CRT strengthens these incentives and is consequently socially desirable. Banks' risk aversion ensures also that any additional risk they take up as a consequence of CRT (because of increased firm financing or an increase in the riskiness of loans due to reduced monitoring incentives) are shed.

CRT also tends to enhance the efficiency of the financial system. The diversification it brings about reduces the risk premia required by financial institutions for financing firms, which lowers firms' financing costs and increases output in the economy. There is, however, a potential caveat:

CRT can reduce the certification value of bank financing because it may lower the bank's incentives to monitor.

In sum, our analysis does not indicate a systematic threat to the functioning of the financial system through CRT. The main problems of CRT seem thus to be limited to the immaturity of CRT markets (which may cause risk concentration, misevaluation of risks, lack of transparency etc.). Since CRT undoubtedly has benefits, this suggests that public policy, rather than restricting CRT, should aim to promote CRT markets (and thus overcome its imperfections) and, if necessary, direct CRT. Our analysis has highlighted several ways in which public policy can ensure that the benefits from CRT are maximized:

- Regulatory arbitrage across sectors should be encouraged in order to better align the incentives for CRT in the financial sector with the social incentives to increase the stability of the financial sector.
- CRT out of the banking sector should be preferred over CRT within the banking sector because it directs risks into a less fragile sector, brings about higher diversification benefits and allows for a better shedding of additional risks taken on by banks.
- Regulators should ensure that CRT instruments allow banks to *effectively* shed risk, for example by encouraging standards for the settlement of legal disputes of credit events or by discouraging the taking of counterparty risks. An effective shedding of risk is obviously beneficial for stability, simply because it transfers more risk. As our analysis has shown, however, it is *crucial* for the overall impact on stability in that it ensures that additional risks taken on by banks can be shed.
- Regulators should promote instruments that minimize asymmetric information problems in the CRT transfer. This is because banks do not necessarily have incentives to reduce efficiency losses arising from incentive problems between risk buyer and seller.

References

- [1] Arping, S.: 2002, 'Playing Hardball: Relationship Banking in the Age of Credit Derivatives'. *mimeo, University of Amsterdam*.
- [2] Bank of England: December 2001, 'Risk Transfer Between Banks, Insurance Companies and Capital Market: An Overview'.
- [3] Basel Committee on Banking Supervision: January 2003, 'Credit Risk Transfer'.
- [4] Besanko, D. and G. Kanatas: 1993, 'Credit Market Equilibrium with Bank Monitoring and Moral Hazard'. *Review of Financial Studies* **6**, 213–232.
- [5] Boot, A. W. A. and A. V. Thakor: 1997, 'Financial System Architecture'. *Review of Financial Studies* **10**, 693–733.
- [6] British Bankers Association: 2002, 'Credit Derivatives Report 2001/2002'.
- [7] Danielsson, J. and J. P. Zigrand: 2003, 'What Happens When You Regulate Risk? Evidence from a Simple Equilibrium Model'. *working paper, London School of Economics*.
- [8] DeMarzo, P. and D. Duffie: 1999, 'A Liquidity-Based Model of Security Design'. *Econometrica* **67**, 65–100.
- [9] Diamond, D. W.: 1984, 'Financial Intermediation and Delegated Monitoring'. *Review of Economic Studies* **51**, 393–414.
- [10] Diamond, D. W.: 1991, 'Monitoring and Reputation: The Choice Between Bank Loans and Directly Placed Debt'. *Journal of Political Economy* **99**, 689–721.
- [11] Duffee, G. R. and C. Zhou: 2001, 'Credit Derivatives in Banking: Useful Tools for Managing Risk?'. *Journal of Monetary Economics* **48**, 25–54.
- [12] Financial Stability Authority: May 2002, 'Cross-Sector Risk Transfer'.
- [13] FitchRatings: September 2003, 'Global Credit Derivatives: A Qualified Success'.
- [14] Gorton, G.: 1988, 'Banking Panics and Business Cycles'. *Oxford economic papers* **40**, 751–781.

- [15] Gorton, G. and G. Pennacchi: 1995, 'Banks and Loan Sales: Marketing Non-Marketable Assets'. *Journal of Monetary Economics* **35**, 389–411.
- [16] IMF: March 2002, 'Global Financial Stability Report'.
- [17] Instefjord, N.: 2003, 'Risk and Hedging: Do Credit Derivatives Increase Bank Risk?'. *working paper, Birkbeck College*.
- [18] International Association of Insurance Supervisors: March 2003, 'Credit Risk Transfer Between Insurance, Banking and Other Financial Sectors'.
- [19] Kiff, J., F.-L. Michaud, and J. Mitchell: June 2003, 'An Analytical Review of Credit Risk Transfer Instruments'. *Banque de France FSR* pp. 106–131.
- [20] Morrison, A. D.: 2003, 'Credit Derivatives, Disintermediation and Investment Decisions'. *forthcoming Journal of Business*.
- [21] OECD: June 2002, 'Risk Transfer Mechanisms: Converging Insurance, Credit and Capital Markets'.
- [22] Santomero, A. M. and J. J. Trester: 1998, 'Financial Innovation and Risk Taking'. *Journal of Economic Behavior and Organization* **35**, 25–37.
- [23] The Economist: August 16th 2003, 'Special Report on Bank Lending'.

Appendix

Proof of Lemma 2 Partial differentiation of (8) wrt. b (assuming $c_{ee} = 0$) gives $f_e + b f_{ee} \frac{\partial \tilde{e}}{\partial b} = 0$.

Hence

$$\frac{\partial \tilde{e}}{\partial b} = \frac{f_e}{b(-f_{ee})} > 0 \quad (46)$$

Partial differentiation of (8) wrt k gives $b f_{ek} + b f_{ee} \frac{\partial \tilde{e}}{\partial k} = 0$, hence

$$\frac{\partial \tilde{e}}{\partial k} = \frac{f_{ek}}{-f_{ee}} > 0 \quad (47)$$

Proof of Lemma 3 Assume $c_{ee} = 0$ and $f(k, e) = k^\alpha e^{1-\alpha}$.

1. $b'(\tau) < 0$: Inserting (46) in (10) and rearranging for f gives

$$\frac{1 - b f_e}{b} \frac{f_e}{f(-f_{ee})} = \tau \quad (48)$$

for $f(k, e) = k^\alpha e^{1-\alpha}$ we have

$$\frac{f_e}{f} \frac{f_e}{(-f_{ee})} = \bar{c} = \frac{1 - \alpha}{\alpha} \quad (49)$$

inserting in (48) gives

$$b = \frac{\bar{c}}{\bar{c} + \tau} \quad (50)$$

and hence

$$b'(\tau) < 0 \quad (51)$$

2. $e'(\tau) < 0$: Inserting (50) in (21) gives

$$\frac{\bar{c}(1 - \tau)}{\bar{c} + \tau} f_e = c_e \quad (52)$$

hence

$$e'(\tau) = -\frac{1 + c}{(c + \tau)(1 - \tau)} \frac{f_e}{-f_{ee}} > 0 \quad (53)$$

3. $k'(\tau) = 0$: From (47) and $f = k^\alpha e^{1-\alpha}$ we get

$$f_e \tilde{e}'(k) = f_e \frac{f_{ek}}{-f_{ee}} = \frac{1 - \alpha}{\alpha} f_k = \bar{c} f_k$$

inserting in (22) gives

$$[1 - \tau b + (1 - b)\bar{c}] f_k = 1$$

and using (50) to substitute \bar{c} gives

$$\begin{aligned} [1 - \tau b + (1 - b)\frac{b\tau}{1 - b}]f_k &= 1 \\ f_k &= 1 \end{aligned}$$

hence

$$k'(\tau) = 0$$

4. $g'(\tau) < 0$ ($g = bf$): follows directly from $b'(\tau) < 0$, $e'(\tau) < 0$ and $k'(\tau) = 0$

Proof of Lemma 4 The total derivative of g wrt τ is given by

$$g'(\tau) = b'(\tau)f + bf_e e'(\tau) + bf_k k'(\tau)$$

Since $k'(\tau) = 0$ we get after dividing by g

$$\frac{g'(\tau)}{g} = \frac{b'(\tau)}{b} + \frac{f_e}{f} e'(\tau) \quad (54)$$

From (50) we have

$$\frac{b'(\tau)}{b} = \frac{-1}{\bar{c} + \tau}$$

and using (53) we get

$$\frac{f_e}{f} e'(\tau) = -\frac{(1 + c)c}{(c + \tau)(1 - \tau)}$$

Hence

$$\begin{aligned} \frac{g'(\tau)}{g} &= \frac{-1}{\bar{c} + \tau} - \frac{(1 + c)c}{(c + \tau)(1 - \tau)} \\ &= -\frac{(1 - \tau) + (1 + c)c}{(c + \tau)(1 - \tau)} \end{aligned} \quad (55)$$

Differentiating (55) wrt τ gives

$$\partial \frac{g'(\tau)}{g} / \partial \tau < 0$$

Hence

$$g'(\tau)/g \leq g'(\tau = 0)/g = -(1 + c) < -1$$

Proof of Lemma 5 From (28) and (29) we have for Bs and NBs portfolio volatility

$$\sigma_B^2 = (1 - q_B)^2 \sigma^2 g^2 \quad (56)$$

$$\sigma_N^2 = \left(\frac{1 - b}{b} + q_N\right)^2 g^2 \sigma_w^2 \quad (57)$$

The FOC for q_B and q_N are then (after arranging for p/g)

$$p/g = 1 - \alpha_B g[(1 - q_B)\sigma^2] + \tau \quad (58)$$

$$p/g = 1 - \alpha_N g\left(\frac{1-b}{b} - q_N\right)\sigma_w^2 \quad (59)$$

Equilibrium in CRT markets requires $q_B = q_N$ ($=: q$). Using (58) and (59) to substitute p and solving for q gives

$$q = \frac{\alpha_B \sigma^2 - \alpha_N \sigma_w^2 \frac{1-b}{b} - \tau/g}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2} \quad (60)$$

Inserting into (56) and (57) yields

$$\sigma_B^2 = \left(\frac{\alpha_N \sigma_w^2 f + \tau}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2}\right)^2 \sigma^2 \quad (61)$$

$$\sigma_N^2 = \left(\frac{\alpha_B \sigma^2 f - \tau}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2}\right)^2 \sigma_w^2 \quad (62)$$

Proof of Proposition 4 Inserting (60),(61) and (62) into V_B and V_N and differentiating wrt to f we obtain

$$\frac{\partial V_B}{\partial f} = \left(1 - \frac{\alpha_B \sigma^2 - \alpha_N \sigma_w^2 \frac{1-b}{b} - \tau/g}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2} \tau\right) - \alpha_B \left(\frac{\alpha_N \sigma_w^2 f + \tau}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2}\right) \sigma^2 \quad (63)$$

$$\frac{\partial V_N}{\partial f} = 1 - \alpha_N \alpha_B \sigma^2 \left(\frac{\alpha_B \sigma^2 f - \tau}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2}\right)^2 \sigma_w^2 \quad (64)$$

Differentiating (63) and (64) wrt τ gives

$$\frac{\partial}{\partial \tau} \frac{\partial V_B}{\partial f} = -\frac{2\alpha_B \sigma^2 - \alpha_N \sigma_w^2 \frac{1-b}{b} - 2\tau/g}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2} = -2q - \frac{\alpha_N \sigma_w^2 \frac{1-b}{b}}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2} < 0 \quad (65)$$

$$\frac{\partial}{\partial \tau} \frac{\partial V_N}{\partial f} = \alpha_N \alpha_B \sigma^2 \left(\frac{1}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2}\right)^2 \sigma_w^2 \geq 0 \quad (66)$$

Dividing the FOC for b (equation 10) by $\partial V_N / \partial f$ yields

$$\begin{aligned} (1-b)f_e \frac{\partial \tilde{c}}{\partial b} &= \left(1 - \frac{\partial V_B}{\partial f} / \frac{\partial V_N}{\partial f}\right) f \\ \frac{1-b}{b} \bar{c} &= 1 - \frac{\partial V_B}{\partial f} / \frac{\partial V_N}{\partial f} \end{aligned} \quad (67)$$

where $f_e \frac{\partial \tilde{c}}{\partial b} / f = \bar{c} / b$ (equation 49) has been used. With (65) and (66) we have that the RHS of (67) is increasing in τ , hence $b'(\tau) < 0$.

Proof of Proposition 5 (i) stability "if part": $\beta_B / \alpha_B > \beta_N / \alpha_N$. Differentiating (7) wrt. to τ (keeping f constant) gives

$$\begin{aligned} \frac{\partial I}{\partial \tau} &= \left(\frac{\alpha_N \sigma_w^2 + \tau/f}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2}\right) \left(\frac{1/f}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2}\right) \sigma^2 \beta_B - \left(\frac{\alpha_B \sigma^2 - \tau/f}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2}\right) \frac{1/f}{\alpha_N \sigma_w^2 + \alpha_B \sigma^2} \sigma_w^2 \beta_N \\ &= (\alpha_N \beta_B - \alpha_B \beta_N) \sigma^2 \sigma_w^2 + \tau/f (\sigma^2 \beta_B + \sigma_w^2 \beta_N) \end{aligned} \quad (68)$$

Since $\beta_B/\alpha_B > \beta_N/\alpha_N$, it follows that

$$\frac{\partial I}{\partial \tau} > \tau/f(\sigma^2\beta_B + \sigma_w^2\beta_N) \geq 0$$

thus $\partial I/\partial \tau > 0$ for all τ . "only if part": $\partial I/\partial \tau > 0$ for all τ . In particular, for $\tau = 0$ we have

$$(\alpha_N\beta_B - \alpha_B\beta_N)\sigma^2\sigma_w^2 > 0$$

from which follows that $\beta_B/\alpha_B > \beta_N/\alpha_N$. (ii) efficiency: proof analogous Proposition 1,

Proof of Proposition 6 Substituting σ_B^2 and σ_N^2 in I (equation 7) using (61) and (62) and substituting α_N using $r = \alpha_N^R/\alpha_B^R$ gives

$$I = \left(\frac{r\alpha_B\sigma_w^2 + b\tau/g}{r\alpha_B\sigma_w^2 + \alpha_B\sigma^2}\right)^2\sigma^2\beta_B + \left(\frac{\alpha_B\sigma^2 - b\tau/g}{r\alpha_B\sigma_w^2 + \alpha_B\sigma^2}\right)^2\sigma_w^2\beta_N \quad (69)$$

The FOC for r is (for constant f)

$$\begin{aligned} 0 &= 2\left(\frac{r\alpha_B\sigma_w^2 + b\tau/g}{r\alpha_B\sigma_w^2 + \alpha_B\sigma^2}\right)\frac{\alpha_B\sigma_w^2(r\alpha_B\sigma_w^2 + \alpha_B\sigma^2) - \alpha_B\sigma_w^2(r\alpha_B\sigma_w^2 + b\tau/g)}{(r\alpha_B\sigma_w^2 + \alpha_B\sigma^2)^2}\sigma^2\beta_B + \\ &= (r\alpha_B\sigma_w^2 + b\tau/g)\sigma^2\beta_B - (\alpha_B\sigma^2 - b\tau/g)\sigma_w^2\beta_N \end{aligned} \quad (70)$$

Solving for r gives

$$\begin{aligned} r &= \frac{(\alpha_B\sigma^2 - b\tau/g)\sigma_w^2\beta_N - b\tau/g\sigma^2\beta_B}{\alpha_B\sigma_w^2\sigma^2\beta_B} \\ &= \beta_N/\beta_B - \frac{\tau(\sigma_w^2\beta_N + \sigma^2\beta_B)}{f} \end{aligned} \quad (71)$$

Proof of Lemma 6 $\partial V_B/\partial g$: The partial derivative of (39) wrt g is

$$\begin{aligned} \frac{\partial V_B}{\partial g} &= \frac{(g - (1 - \tau)\tilde{g})\alpha g^2\sigma^2 - 2\alpha g\sigma^2(g^2/2 - (1 - \tau)\tilde{g}g + 1/2(1 - \tau)^2\tilde{g}^2)}{\alpha^2 g^4 \sigma^4} \\ &= \frac{(1 - (1 - \tau)) - 2(1/2 - (1 - \tau) + 1/2(1 - \tau)^2)}{\alpha g \sigma^2} \end{aligned} \quad (72)$$

$$= \frac{\tau - \tau^2}{\alpha g \sigma^2} = (1 - \tau)\frac{\tau}{\alpha g \sigma^2} = (1 - \tau)(1 - q) \quad (73)$$

where $g = \tilde{g}$ has been used to obtain (72) and (17) to obtain (73)

$\partial V_B/\partial \tilde{g}$: The partial derivative of (39) wrt \tilde{g} is

$$\begin{aligned} \frac{\partial V_B}{\partial \tilde{g}} &= 1 - \tau + \frac{-(1 - \tau)g + (1 - \tau)^2\tilde{g}}{\alpha g^2 \sigma^2} \\ &= 1 - \tau + \frac{-(1 - \tau) + (1 - \tau)^2}{\alpha g \sigma^2} \end{aligned} \quad (74)$$

$$= (1 - \tau)q \quad (75)$$

where $g = \tilde{g}$ has been used to obtain (74) and (17) to obtain (75)

Proof of Lemma 7 FOC for e (equation 40). From (39), $V'_N(w_N) = 1$ and $w_N = \frac{1-b}{b}\tilde{g}$ we have

$$\begin{aligned} dU/de &= dV_N/de + dV_B/de - c'(e) \\ &= \frac{\partial V_B}{\partial g} \frac{\partial g}{\partial e} - c'(e) = (1-q)(1-\tau)bf_e - c'(e) \end{aligned}$$

FOC for k (equation 41).

$$\begin{aligned} \frac{dU}{dk} &= \frac{dV_B}{dk} + \frac{dV_N}{dk} = \frac{\partial V_B}{\partial g} \frac{\partial g}{\partial k} + \frac{\partial V_B}{\partial \tilde{g}} \frac{\partial \tilde{g}}{\partial k} + \frac{\partial V_B}{\partial \tilde{g}} \frac{\partial \tilde{g}}{\partial \tilde{e}} \tilde{c}'(k) + \frac{\partial V_N}{\partial k} + \frac{\partial V_N}{\partial \tilde{e}} \tilde{c}'(k) \\ &= [(1-\tau)b + (1-b)]f_k + [(1-\tau)qb + (1-b)]f_e \tilde{c}'(k) \end{aligned}$$

where $f = \tilde{f}$ has been used. FOC for b (equation 42).

$$\begin{aligned} \frac{dU}{db} &= \frac{dV_B}{db} + \frac{dV_N}{db} = \frac{\partial V_B}{\partial g} \frac{\partial g}{\partial b} + \frac{\partial V_B}{\partial \tilde{g}} \frac{\partial \tilde{g}}{\partial b} + \frac{\partial V_B}{\partial \tilde{g}} \frac{\partial \tilde{g}}{\partial \tilde{e}} \tilde{c}'(b) + \frac{\partial V_N}{\partial b} + \frac{\partial V_N}{\partial \tilde{e}} \tilde{c}'(b) \\ &= (1-\tau)f - f + [(1-\tau)qb + (1-b)]f_e \tilde{c}'(b) \\ &= -\tau f + [(1-\tau)qb + (1-b)]f_e \tilde{c}'(b) \end{aligned}$$