EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Jean-Edouard Colliard

Optimal supervisory architecture and financial integration in a banking union

1786 / April 2015

Abstract

Both in the United States and in the Euro area, bank supervision is the joint responsibility of local and central/federal supervisors. I study how such a system can optimally balance the lower inspection costs of local supervisors with the ability of the central level to internalize cross-border or interstate externalities. The model predicts that centralised supervision endogenously increases market integration and cross-border externalities, further strengthening the need for centralised supervision. This complementarity implies that, for some parameterizations of the model, the economy can be trapped in a local supervision equilibrium with low supervision and integration. In such a case, the forward-looking introduction of a centralized supervisory architecture achieves a superior equilibrium.

JEL Classification Number: D53, G21, G28, G33, G38, L51.
Keywords: banking union, single supervisory mechanism, bank supervision, financial integration, regulatory federalism.

Non-technical summary

This theory paper studies the architecture of bank supervision in supranational or federal contexts. It develops a model of bank supervision where national supervisors have a comparative advantage in the on-site supervision of banks, but may also be overly lenient due to incentives to neglect e.g. the effects of bank problems at home on foreign countries. A supranational supervisor has better incentives but tries to avoid the costs of on-site supervision by using off-site monitoring to get information about the banks and the local supervisor's behaviour.

The analysis relies on a simple banking supervision model with three building blocks: (i) leveraged banks sometimes receive negative information about a pool of loans, but due to limited liability banks have too few incentives to liquidate their loans; (ii) through on-site examinations, a local supervisor learns the same information and can force the liquidation of these loans. A local supervisor may be too lenient for several reasons, for instance if the banks' losses are partly borne by foreign creditors; (iii) a central supervisor takes into account the losses to foreign creditors, but is less proficient at examining banks on-site than the local supervisor, and he can use a less informative and more costly offsite monitoring technology.

Three supervisory architectures are possible: delegation of supervision to the local level, centralization of supervision (the central supervisor performs the on-site examinations), and a joint architecture in which the central supervisor monitors the bank offsite, and inspects only if necessary.

The first part of the paper derives the optimal supervisory architecture depending on a given bank's characteristics. Three characteristics matter: (i) Aligned objectives: supervisors at both levels have more similar scopes and objectives when supervising banks with less cross-border activities; (ii) Specificity: banks that have more specific local characteristics (local products, rare language, complex legal framework...) are more costly for the central supervisor to inspect on-site; (iii) Opacity: when banks are more opaque it is more difficult to rely on off-site monitoring.

The second part of the paper endogenizes the behaviour of banks and foreign investors. I show that centralizing supervision promotes market integration, in the sense that foreign

investors lend more to local banks, which can then lend more to local firms. For this same reason, however, the incentives of local supervisors worsen, as a higher proportion of potential losses are now borne by foreigners. Because of this, supervision may have to be further centralized.

Banks that are less specific and have larger cross-border externalities are better supervised at the central level, as the costs of central supervision are lower and the benefits larger. A relatively transparent bank, one mostly engaged in simple lending activities for instance, is easier to monitor offsite and may be left to the national level, with the central supervisor keeping the right to intervene if he receives signals that his intervention is necessary. On the contrary, a bank engaged in more opaque activities can be monitored either at the local or at the central level, but a joint architecture is typically not optimal.

This analysis has several implications for the European Single Supervisory Mechanism.

First, it gives theoretically grounded criteria for deciding which banks should be supervised by the ECB, by a national supervisory agency, or by both. Large international banks typically fit the case in which centralizing supervision is optimal, as their international activities both imply high externalities and a lower informational advantage of the national supervisor. Taking into account the complexity or opacity of these banks would allow to further identify large banks that could be supervised jointly.

Second, since market conditions should react to the imposition of a more centralized supervisory architecture with more integration, the perimeter of banks supervised by the ECB should at least be flexible enough to respond to changing market conditions. Moreover, I show that a bank or a banking system can be trapped in an equilibrium where cross-border activities are small because supervision is local, and supervision is left at the local level because cross-border activities are small. The decision to allocate a bank to the central supervisor should be forward-looking and take into account that centralized supervision may itself trigger more financial integration and thus more need for such a supervision.

1 Introduction

Following the increasing integration of banking systems in the past decades, the boundaries of a bank's activity can stretch significantly further than the mandate of its supervisor. The problem has been particularly acute in Europe, where the "supervisory failings" of a system based on national supervision led the European Commission to propose the creation of a "Single Supervisory Mechanism" (SSM) for European banks in September 2012.[1] Frictions related to the supervisory architecture are also documented in the United States, where banks alternatively supervised by State and Federal supervisors appear to face forbearance by the former, as shown by Agarwal *et al.* (2012). The perception of supervisory failings also led to the suppression of the Office of Thrift Supervision (OTS) in 2011.

This paper offers a framework to answer two questions: how best to organize bank supervision in a federal context, and how does a banking system react to such large shocks on supervisory architecture as the creation of a central supervisor. Both questions are related, as market integration and supervisory architecture react to each other. While they are both important from a normative point of view, understanding the impact of supervisory changes on the banking sector also allows to derive empirical implications, that can be tested using recent reforms.

Banks in the model can have problematic loans that they try to "evergreen" instead of liquidating them. Supervisors conduct on-site inspections to learn the quality of the loans and can order a liquidation. When taking this decision, local supervisors do not take into account the cross-border (or interstate) externalities of a bank's distress, and are typically too forbearant. This is anticipated by the bank's foreign creditors, who are more reluctant to lend to the bank. Inadequate supervision thus leads to a friction in the cross-border allocation of capital.

The model makes two contributions to our understanding of the architecture of bank supervision:

First, I derive the optimal supervisory architecture, allowing for the use of both a central

[1]The term "failings" is used in the *Proposal for a Council regulation conferring specific tasks on the European Central Bank concerning policies relating to the prudential supervision of credit institutions*, European Commission, 12.09.12, p.2. The actual regulation was adopted on 15.10.13. Both texts are available on the Commission's website: http://ec.europa.eu/internal_market/finances/banking-union/.

and a local supervisor.[2] Three types of architectures observed in practice can be optimal: (i) centralized, in which case the local supervisor plays no role; (ii) delegated to the local level, which can lead to forbearance; or (iii) joint, so that inspections are conducted either by the local or the central supervisor depending on the information available about the bank.

Which arrangement is optimal depends on three dimensions. The first two are the cross-border (or interstate) activities of the bank, and the regional specificity of its assets. With important cross-border activities, central inspections lead to better decisions than local inspections because the central supervisor internalizes the payoffs to foreign agents, while the local supervisor would neglect them. But when the bank's assets are very specific to a country or region, the local supervisor has an informational advantage over the central supervisor, making central inspections more difficult and costly. The third dimension is the bank's opacity: if it is easy to obtain information about the bank off-site, the most cost-efficient solution is to have the central supervisor inspecting only if the observed figures point towards a situation in which the local supervisor is likely to be too forbearant.

Second, I show that financial integration and centralized supervision reinforce each other and derive the normative and empirical implications of this result. Imagine a bank in a region with a deficit of deposits that relies on funding from foreign agents in a surplus region. Supervisory forbearance at the local level implies that it is risky for foreign investors to lend to the bank. Centralizing supervision reduces this risk, so that the bank can borrow more from foreign investors. But, as a result, a higher proportion of the bank's losses are now borne by foreign investors, which worsens the local supervisor's incentives. It is then necessary to reinforce central supervision even more.[3]

This complementarity can generate multiple equilibria: a bank with few foreign creditors is best left to local supervisors, and for this reason few foreign investors lend to the bank. Centralizing the supervision of this bank generates more lending by foreigners, making central supervision necessary. The economy can be trapped in an equilibrium with both too little central supervision and too little market integration, when a superior equilibrium is

[2] It will be convenient to refer to the central supervisor with the female pronoun "she" and to the local supervisor with the male pronoun "he".

[3] A parallel can be made with the endogeneity of optimal currency areas, studied in Frankel and Rose (1998).

achievable. The opposite situation of a suboptimal equilibrium with too much supervision is also possible.

While the paper derives a number of normative implications about the organization of bank supervision, the second part also delivers empirical implications on the impact of changes in supervisory architecture on banks. The main mechanism is that banks supervised at the central level should increase their cross-border activities and find it easier to borrow from foreign sources than comparable banks supervised at the local level. This effect should be reflected in the rates at which banks can borrow on the interbank market, the composition of their balance sheet, or their profitability. In the European case, the criteria determining whether a bank falls under central supervision or not are public, making it possible to compare banks just above the threshold triggering a centralized supervision with those just below. I discuss in more details in section 5.2 how these implications could be tested on European or on U.S. data.

The model is flexible and allows for a number of additional applications and robustness checks, that are developed in a supplementary Internet Appendix[4]. The complementarity of centralized supervision and market integration can be generalized to a bank with foreign assets and foreign shareholders instead of only foreign creditors. I also show that banks with market power can strategically choose their balance sheet so as to avoid being supervised from the center or, more surprisingly, to make sure that they fall under the central supervisor's umbrella. Finally, as the optimal supervisory architecture derived in the first part of the paper takes supervisory incentives as given, it is easy to enrich the model with alternative microfoundations for these incentives.

The end of this section reviews the related literature. Section 2 introduces the assumptions of the model. Section 3 solves for the optimal supervisory architecture, taking the structure of the banking system as given. This structure is then endogeneized in section 4 that focuses on the interplay between supervision and market forces, and is followed by a general discussion of the model and the conclusion.

[4]Available at `https://sites.google.com/site/jecolliardengl/Supervision-Appendix.pdf`.

A first strand of the literature to which this paper is closely connected concerns the supervision of multinational banks. My modeling of supervision is close to Calzolari and Loranth (2011), who focus on the impact of the legal form of a multinational bank, and Mailath and Mester (1994). Beck, Todorov, and Wagner (2013) show that the predictions on supervisory forbearance of this modeling approach are supported by European data and discuss the implications for the European banking union. Another related paper is Holthausen and Ronde (2004), who study how different national supervisors of multinational banks can voluntarily exchange information despite their conflicting objectives.

My setup differs from this literature as I look at a "vertical" problem of cooperation between a central and a local supervisor instead of the "horizontal" interaction between equal national supervisors, and focus on the interaction between centralized supervision and market integration. This vertical dimension has been explored in several recent empirical papers (discussed in 5) that study the behavior of different bank supervisors at State and Federal level in the United States and identify systematic differences in their behavior towards supervised banks. As is discussed in the text, my results are in line with these findings and give additional testable implications.

Second, the paper is related to the literature on competition and coordination among regulators. Note that both in the United States and in the European Union regulation is already integrated. The European Banking Authority in particular is responsible for establishing a European Single Rulebook guaranteeing a uniform regulation across the European Union. Dell'Ariccia and Marquez (2006) study the incentives of financially integrated countries to voluntarily form regulatory unions. However, recent experience shows that regulatory unions may be to some extent illusory if *supervision* is still fragmented, so that the rules are not applied uniformly, or even correctly. My analysis complements their paper by studying the relation between common supervision and financial integration.

Other papers in this literature look at coordination problems between different banking regulators, either from different countries or with different objectives. Acharya (2003) for instance studies the competition between closure policies in two countries when capital adequacy regulation is already coordinated, and shows that this coordination worsens the regulatory race to the bottom. Hardy and Nieto (2011) show that common supervision alleviates the coordination problem between national deposit insurers. Kahn and Santos (2005) offer a different perspective by studying the interaction of regulators with different objectives (supervisor, deposit insurer, lender of last resort) depending on whether they are coordinated or not.

A few recent papers study the European banking union specifically. Zoican and Gornicka (2014) and Foarta (2014) focus on the impact of bail-outs and recapitalizations, thus complementing this paper's analysis of common supervision with a study of bank resolution. Beck and Wagner (2013) also study common supervision and focus on the different problem, inspired by the literature on fiscal federalism, that different regions have diverse preferences regarding financial stability. See also Vives (2001) for an early discussion of the challenges for financial integration and common regulation in Europe.

In terms of modeling, this paper follows the literature on the supervision of multinational banks in considering a static environment, whereas the theoretical literature on supervision has mostly considered dynamic environments, to focus on the timing of supervisory interventions. See for instance Merton (1978), King and O'Brien (1991), Decamps, Rochet, and Roger (2004). Forbearance in my model can still be interpreted as delayed action, but the supervisor cannot use intertemporal effects on a bank's profit to achieve better outcomes.

Consistent with actual supervisory arrangements, I assume that it is legally infeasible for the principal (here, the central supervisor) to use monetary transfers to control the agent's (here, the local supervisor's) incentives. While this is a classical motive for vertical integration, as in Grossman and Hart (1986), this assumption and the simplicity of the modeling would also allow to use insights from recent works on delegation games, such as Alonso and Matouschek (2008). An interesting specificity of the delegation problem studied in this paper is the interaction between delegation and market forces.

Another relevant strand of the literature studies regulatory design with lobbying. Supervisors in my model are assumed to maximize a measure of welfare, domestic or global. An alternative explanation of supervisory failings is capture by private interests. Local and central supervisors could for instance be analyzed as the internal and external auditors in Kofman and Lawarree (1993). Hiriart, Martimort, and Pouyet (2010) show, with another application in mind, that separating ex-ante monitors of risk from ex-post monitors is a powerful tool against regulatory capture. Costa Lima, Moreira, and Verdier (2012) show on the contrary some benefits of centralization. See Boyer and Ponce (2012) for an application to bank supervision and a review of the relevant literature. Martimort (1999) shows that as a regulatory agency gains information over time, it becomes less and less robust to lobbying and should be given less discretion, which is to be traded off against Section 4, which shows that centralization may have to increase over time.

2 Framework

The economy: I consider a simple intermediation model in which banks stand between borrowers and investors. All agents are price-takers and risk-neutral.

-A penniless representative *borrower* has risky projects that can either succeed or fail. He can only borrow from a bank and promise to repay $1+r$ on each unit of loan. He has access to an investment opportunity that for an investment of size L gives a gross return $1+\rho(L)$ if it is successful, where $\rho(.)$ is a strictly decreasing function. The investment fails with probability p, with $p \hookrightarrow \Phi(.)$ over $[0,1]$. In case of failure the borrower cannot repay anything. This representative borrower is price-taker and thus generates a demand for loans $q(r) = \rho^{-1}(r)$, with $q'(r) \leq 0$, $q(0) = +\infty$ and $\lim_{r \to +\infty} q(r) = 0$.

Under this specification, when the project is successful the loan generates a total surplus that can be measured by $V(r) = \int_0^{q(r)} (1+\rho(t))dt$, of which $V(r) - (1+r)L$ goes to the borrower. An investment of size L can be liquidated early, in which case the amount $(1-\ell)L$ can be recovered and represents the total surplus,[5] which is 0 in case of failure.

-*Banks* choose the quantity L of loans they want to extend to the borrower, taking the net interest rate r on these loans as given. They have no capital and a fixed quantity D of deposits. Depositors are insured and have no incentive to withdraw early, they are thus essentially passive. Their losses are covered by a deposit insurance fund. Loans in excess of D can be financed by borrowing $D_F = L - D$ from foreign investors at rate r_F. If a bank defaults, its assets are distributed pro-rata to its creditors (depositors and foreign investors).

-*Foreign investors* stand ready to lend to banks as long as their expected return on these

[5] Alternatively, the loan could need a refinancing of ℓL.

loans is higher than the return on the storage asset, normalized to 1.

Supervision and inspections: note that the liquidation value of loans is by construction not enough for banks to repay their debt, so that due to limited liability they would always choose to exert "evergreening"[6] and keep their loans. Bank supervision in this model consists in learning the probability p that the projects will succeed, and force their liquidation if this probability is too low. Either the supervisor has intervention powers himself, or his report triggers the intervention of a different player (for instance a resolution authority).

There are two risk-neutral supervisors:[7] a *local supervisor* whose aim is to maximize the welfare of local agents, and a *central supervisor* who maximizes total welfare in the economy. Denote S the state in which the bank's loans are, with $S = 1$ for successful, $S = 0$ for failed, and $S = I$ for liquidated. Global welfare in state S is denoted W_S, and local welfare $\hat{W}_S$. Define V_S, π_S and π_S^F the welfare of the borrower, the profit of banks and the profit of foreign investors, respectively. Finally, the depositors' welfare is constant but the deposit insurance fund can take losses. π_S^D denotes the joint payoff of depositors and the deposit insurance fund.

Consistent with the European case, I assume that the deposit insurance fund is "local" and its losses internalized by the local supervisor. We thus have $\hat{W}_S = V_S + \pi_S + \pi_S^D$ and $W_S = \hat{W}_S + \pi_S^F$. Assuming that the local supervisor only takes into account a fraction of the losses to the deposit insurance fund, which is closer to the United States case, is a straightforward adaptation of the model and leads to similar results. However, having the payoffs to foreign investors as the only wedge between the objective functions of both supervisors makes the exposition clearer.

Both supervisors can conduct *inspections*. Inspections cost c_0 if made by the local supervisor, and $c_0 + c$ if made by the central supervisor. c_0 is supposed to be high so that it is suboptimal to duplicate inspections, and c measures the informational advantage of the local supervisor. The supervisor who inspects the bank learns the exact value of p and decides whether the bank's loans should be liquidated. It is assumed that for informational and legal reasons it is impossible to liquidate the bank's projects without inspecting the bank first.

[6]See Peek and Rosengren (2005), as well as Albertazzi and Marchetti (2010) for recent European evidence.
[7]Extension 3 in the Internet Appendix discusses the case of risk-averse supervisors.

The central supervisor decides who should conduct the inspection. There are three different possibilities. First, the central supervisor can choose a *central architecture* under which she always inspects the bank, so that the local supervisor plays no role. Second, she can choose the symmetric *delegated architecture* in which the local supervisor always conducts inspections. The last possibility is a *joint architecture* in which who inspects the bank depends on a signal about p received by the central supervisor. The central architecture will be said to be *more centralized* than the others, and the joint architecture to be more centralized than the delegated architecture. There is no possibility for either supervisor to avoid conducting an inspection if he or she has been allocated this task.

Off-site monitoring: the central supervisor can pay *monitoring costs* $C(\lambda)$ to get a signal about p with precision λ, with $C(0) = C'(0) = 0$, $C' \geq 0, C'' \geq 0$ and $\lim_{x \to 1} C(x) = +\infty$. The signal could rely on off-site monitoring of the bank's balance sheet, stress-tests, or quarterly call reports in the U.S. case.[8]

As will be shown below, a natural feature of this model is that the central supervisor wishes to inspect the bank for *intermediate* signals about p, not for the worst ones. To analyze in closed-form which signals lead to a central inspection, I assume a particularly tractable signal structure (borrowed from Petriconi (2012)): with some probability λ, the central supervisor receives a signal s exactly equal to the true p. With probability $1 - \lambda$, s is drawn from the prior distribution Φ. The signal thus gives perfect information about the soundness of the bank with probability λ, and is noise with the opposite probability. λ thus measures the precision of the signal. Denoting $\tilde{\Phi}_s$ the cumulative distribution function of p conditional on receiving signal s, and $\tilde{\phi}_s$ the corresponding density, we have:

$$\tilde{\Phi}_s(p) = \lambda \mathcal{H}(p - s) + (1 - \lambda)\Phi(p) \quad (1)$$

$$\tilde{\phi}_s(p) = \lambda \delta(p - s) + (1 - \lambda)\phi(p) \quad (2)$$

$$\mathbb{E}(p|\sigma = s) = (1 - \lambda)\mathbb{E}(p) + \lambda s \quad (3)$$

where $\mathcal{H}(x)$ is the Heaviside step function equal to one if $x \geq 0$ and to zero otherwise, and δ

[8] See Kick and Pfingsten (2011) for evidence that on-site supervision brings additional information compared to off-site monitoring. An alternative interpretation of the signal is that it comes from market prices, as in King and O'Brien (1991).

its derivative, the Dirac function. With this signal structure, the ex-post expectation of the supervisor is simply a mixture between the prior expectation and the signal received, where the weights depend on λ, the precision of the signal.

Timeline: there are three periods in the model:

-$t = 0$: banks choose how much to borrow from foreign investors and to lend to the borrower. Competitive market equilibrium conditions determine W_I, W_1, W_0 and $\hat{W}_I, \hat{W}_1, \hat{W}_0$. The central supervisor simultaneously chooses between a delegated, centralized or joint architecture and how much to invest in the precision of the signal, λ.

-$t = 1$: p is realized. If the architecture is joint, the central supervisor receives a signal s on p and decides on who should inspect the bank. Otherwise, either the central or the local supervisor inspects, depending on the chosen architecture. In all cases, the supervisor inspecting the bank learns p and then decides whether loans should be liquidated.

-$t = 2$: if the loans were not liquidated in $t = 1$, they succeed with probability p and pay $1 + r$ to banks. The banks repay D to their depositors and $(1 + r_F)D_F$ to foreign investors. With probability $1 - p$ the loans fail, banks and foreign investors obtain a zero payoff, depositors are reimbursed by the deposit insurance fund that incurs a loss $-D$.

The timeline is summed up in the following two figures. Fig. 1 describes the timing of the overall game, whereas Fig. 2 focuses on the inspection stage, in which the decision-maker can be either the local or the central supervisor.

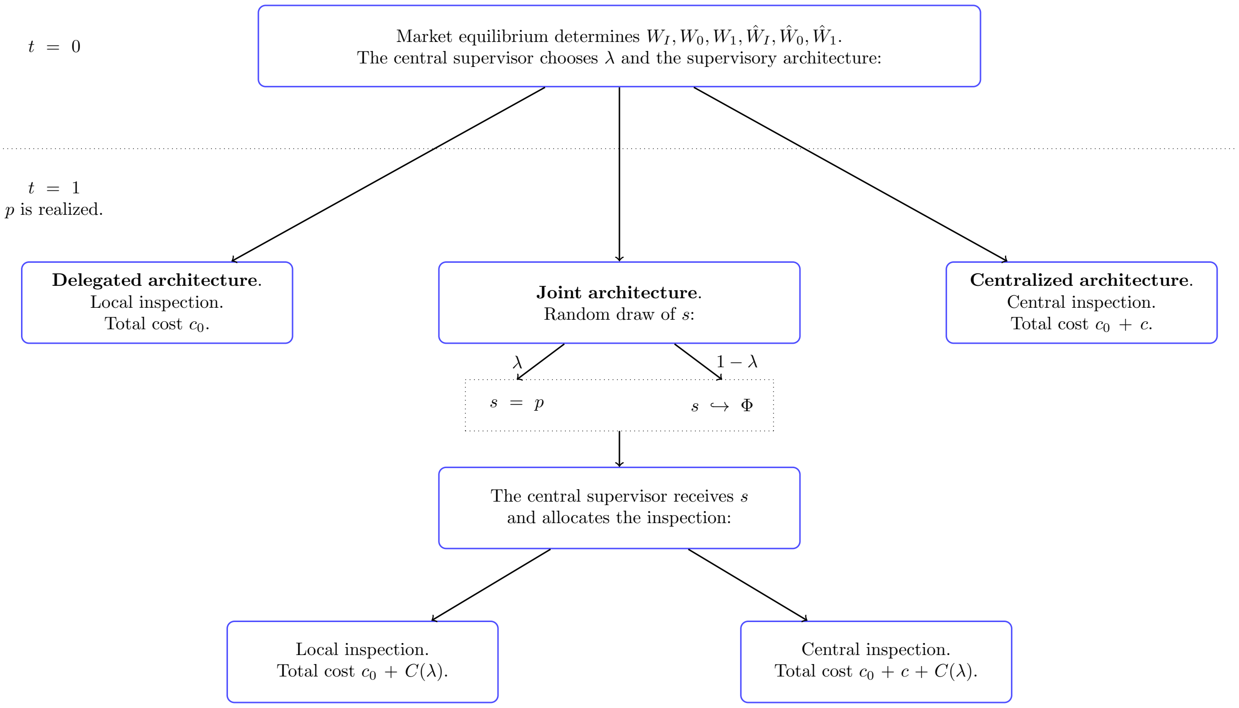

Figure 1: Timeline - Market equilibrium and choice of a supervisory architecture.

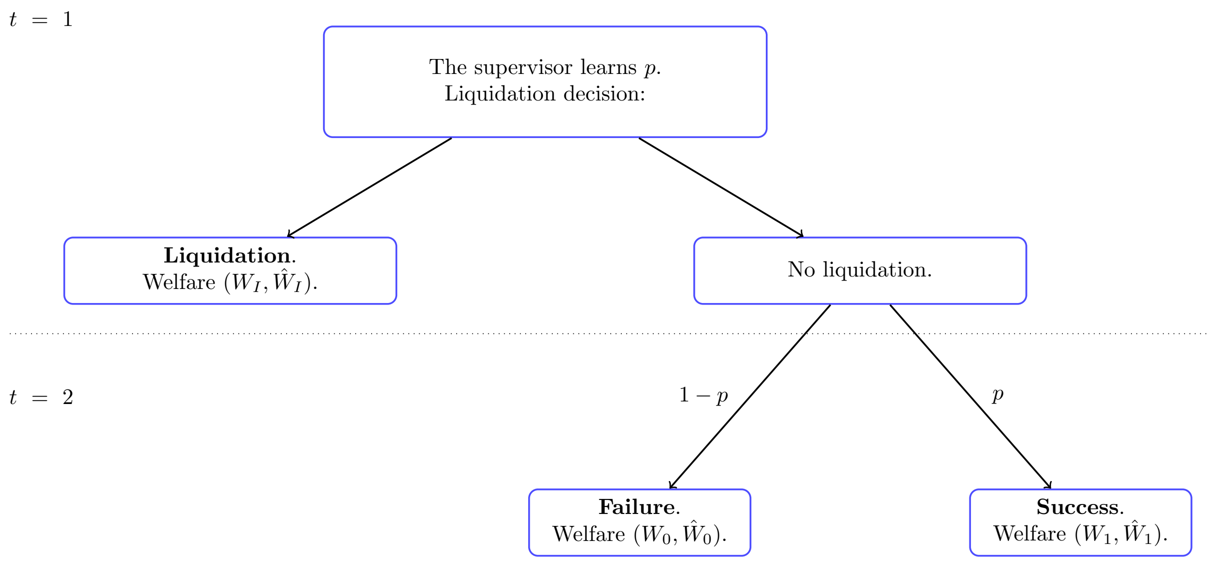

Figure 2: Timeline - Inspection stage.

Empirical counterparts: the model can deliver three different supervisory architectures with empirical counterparts in "federal" or two-layered supervisory systems, the two main examples being the United States, and more recently the Euro area.

In the Euro area, the supervision of banks under the SSM relies on a partition of the banking system into two groups. The "most significant credit institutions" are supervised directly by the European Central Bank (ECB). Although the national supervisors play a role in the inspections, the ultimate decision lies with the ECB, which corresponds to a centralized architecture. While the on-site supervision of the other banks is left to the national competent authorities, the ECB can decide at any time to take responsibility. If this decision is taken based on the ECB's information on the bank's soundness at a particular point in time, this corresponds exactly to the joint architecture. Presumably, some banks are of such small significance that the ECB would never take responsibility, in which case we have a delegated architecture.

In the United States, almost all State-chartered commercial banks are supervised both by a State supervisor and a Federal supervisor, either the Fed or the FDIC. The frequency of inspections by the Federal supervisor depends on the significance of the bank, and certain signals on its soundness such as call reports or CAMELS ratings can trigger the intervention of a Federal supervisor. This arrangement corresponds to a joint architecture, admitting as a special case the centralized architecture when the Federal supervisor always inspects the bank with maximum frequency.

Discussion: the model relies on two separate sets of assumptions. First, the assumptions on the economy are meant to model a situation in which a region with financing needs relies heavily on capital flows from outside, so that there is an imbalance in the geographic composition of the bank's assets and liabilities. This imbalance is responsible for the wedge between the two supervisors.[9] Extension 2 in the Internet Appendix studies a more general setup in which differences in endowments between two regions generate such a pattern.

Second, the assumptions on supervision rationalize that the three architectures observed empirically can be optimal. Notice that these assumptions are independent from the assumptions on the economy itself. The model only requires that there is some wedge between local and global welfare. Because of this wedge, there is a scope for having a central supervisor inspecting the bank and taking the liquidation decision maximizing global welfare. Bank in-

[9] See Beck, Todorov, and Wagner (2013) for a general analysis of how geographic imbalances between assets and liabilities can generate forbearance or on the contrary over-intervention at the local level.

spections are typically very intensive in labor however, and a local supervisor can represent a less costly alternative, in particular if he has a better knowledge of local banks. Centralizing thus involves a trade-off between better-aligned incentives and a cheaper access to information. Moreover, and this is crucial to obtain the joint architecture which is often observed in reality, it can be optimal to allocate the same bank to different supervisors depending on the information available on the bank at a particular point in time. The purpose of a joint architecture is thus to acquire less costly off-site information and conduct a central inspection only if necessary, thereby saving on inspection costs.

Notice that there are two agency problems in this economy: one between the banks and the local supervisor, and one between the local supervisor and the central supervisor. An important assumption of the model is that the second problem cannot be solved by writing contracts involving monetary transfers between the local and the central supervisor. For legal as well as political reasons, it is difficult in practice to commit to imposing a penalty on a local supervisory agency that would have failed to adequately supervise its banks, the problem is thus one of "optimal delegation" [Holmstrom (1977)]. This is a realistic assumption that allows to study supervisory architectures comparable to the ones observed in practice. In terms of normative implications, allowing for more complex contracts would theoretically have a positive impact on welfare, but would also require a thorough overhaul of current supervisory frameworks.

c is interpreted as an informational advantage of the local supervisor and is exogenously given, but one can imagine that the central supervisor will get experience over time and reduce the informational gap. This learning-by-doing effect would reinforce the multiplicity of equilibria studied in section 4. Moreover, if ultimately this gap can be completely closed, c becomes equal to zero and local supervision is necessary only as a temporary solution. An alternative interpretation of the same parameter, perhaps particularly relevant for the United States, is that local supervisors simply rely on less staff and represent a cheaper inspection technology, so that a joint architecture relies on two technologies with different costs and benefits, which can be optimal.

3 Optimal delegation of supervisory powers

This section solves for the optimal supervisory architecture, taking the structure of the banks' assets and liabilities and thus the values of the Ws and $\hat{W}$s as given.

3.1 Centralization and delegation

Proceeding by backwards induction, we can first solve the optimal decision taken by a supervisor in the inspection stage (Fig. 2). Knowing the value of p, the central supervisor can get W_I if she liquidates the loans, against an expected welfare of $pW_1 + (1-p)W_0$ if she does not. Liquidation thus occurs when p is lower than the *first-best intervention threshold* defined by:

$$p^{**} = \frac{W_I - W_0}{W_1 - W_0} \tag{4}$$

The reasoning is the same for the local supervisor, who intervenes whenever p is higher than p^* defined by:

$$p^{*} = \frac{\hat{W}_I - \hat{W}_0}{\hat{W}_1 - \hat{W}_0} \tag{5}$$

In general we have $p^* \neq p^{**}$. Depending on the exact wedge between both supervisors, the local supervisor may either intervene too often in the bank or not enough. As will be shown in section 4, we always obtain $p^* < p^{**}$ in this model, so that the local supervisor exerts supervisory forbearance.[10] This is due to the local supervisor neglecting the losses borne by foreign investors when the bank's projects fail. The quantity $p^{**} - p^*$ measures the intensity of the *conflict of objectives* between the local supervisor and the general interest.

The results in this section do not depend on a particular microfoundation of p^* and p^{**} and are kept general on purpose, so that the properties of the optimal regulatory architecture do not depend on a particular theory of the political economy frictions that drive a wedge between the two levels of supervision. The only loss of generality is the assumption that $p^{**} < p^*$. It is straightforward to adapt the reasoning to the opposite case $p^{**} > p^*$, at the cost of an additional assumption that the local supervisor cannot choose to liquidate the

[10] Forbearance is always suboptimal in this model. See Morrison and White (2013) for a different view, relying on the idea that forbearance can be an optimal answer to reputation concerns.

loans on his own if he does not inspect the bank. Otherwise the liquidation decision would in effect always be taken by the local supervisor. In the case $p^* < p^{**}$ this distinction is irrelevant because the local supervisor is always more forebearant than the central one.

We can now look at the central supervisor's choice to allocate inspections to the local or to the central level, taking the signal s and its precision λ as given. If she pays the cost c_0+c, the central supervisor will intervene if she finds that $p \leq p^{**}$. If she does not inspect the bank, she anticipates that the local supervisor intervenes if and only if $p \leq p^*$. Comparing the total welfare in both cases gives us that the central supervisor pays the cost $c_0 + c$ upon receiving signal s if and only if:

$$\int_0^{p^{**}} W_I\tilde{\phi}_s(p)dp+\int_{p^{**}}^1 [pW_1+(1-p)W_0]\tilde{\phi}_s(p)dp-c \geq \int_0^{p^*} W_I\tilde{\phi}_s(p)dp+\int_{p^*}^1 [pW_1+(1-p)W_0]\tilde{\phi}_s(p)dp \tag{6}$$

Using the definition of p^{**}, this condition can be rewritten as:

$$(W_1 - W_0)\int_{p^*}^{p^{**}} (p^{**} - p)\tilde{\phi}_s(p)dp \geq c \tag{7}$$

The integral in (7) is the average difference between p^{**} and the actual p in the interval $[p^*, p^{**}]$, conditionally on receiving signal s. $[p^*, p^{**}]$ is a region of conflicting objectives, in which the central supervisor would like to overrule the local supervisor's decision, the more so when the difference between p^{**} and p is higher. Instead, when $p < p^*$ both the central and the local supervisor want to intervene, while when $p > p^{**}$ both want not to intervene. The particular signal structure chosen makes it straightforward to reexpress equation (7) explicitly, using (2). It is first useful to define:

$$\bar{B}(p^*, p^{**}) = \int_{p^*}^{p^{**}} (p^{**} - p)\phi(p)dp \tag{8}$$

The expected benefit from inspecting upon receiving signal s with precision λ is then:

$$B(\lambda, s) = \begin{cases} (W_1 - W_0)\left[(1-\lambda)\bar{B}(p^*, p^{**}) + \lambda(p^{**} - s)\right] & \text{if } s \in [p^*, p^{**}] \\ (W_1 - W_0)(1-\lambda)\bar{B}(p^*, p^{**}) & \text{if } s \notin [p^*, p^{**}] \end{cases} \tag{9}$$

Lemma 1. *For a given* λ*:*

1. A centralized architecture is optimal if and only if $c < B(\lambda, 0)$.

2. A delegated architecture is optimal if and only if $c > B(\lambda, p^*)$.

3. When $c \in [B(\lambda, 0), B(\lambda, p^*)]$, a joint architecture is optimal and there exists $\bar{s}(\lambda) \in [p^*, p^{**}]$ such that the central supervisor inspects the bank if $s \in [p^*, \bar{s}(\lambda)]$, and the local supervisor inspects the bank otherwise.

The proof is in the Appendix A.1. $\bar{s}(\lambda)$ will be called the *inspection threshold.* A higher threshold implies a higher probability that the central supervisor inspects the bank. Whenever $c \in [B(\lambda, 0), B(\lambda, p^*)]$, we have:

$$\bar{s}(\lambda) = p^{**} - \frac{c}{\lambda(W_1 - W_0)} + \frac{1-\lambda}{\lambda}\bar{B}(p^*, p^{**}) \tag{10}$$

Lemma 1 is quite intuitive. When the central supervisor receives a signal s outside $[p^*, p^{**}]$ there is probably no conflict of objectives. If she nonetheless inspects, she would do so *a fortiori* for a signal inside the interval. Conversely, among all the signals she can receive inside $[p^*, p^{**}]$, the most favorable to inspection is s close to but above p^*: a conflict of objectives is likely, and since p is probably low not liquidating is very costly. If upon receiving this signal the supervisor does not inspect, then she never does.

Lemma 1 defines three regions in the $(\lambda, c, p^*, p^{**})$ space, where either a centralized, a delegated or a joint architecture are optimal. Figure 3 gives a 3D plot of these regions for a given p^{**}.[11] The next proposition shows how one moves from one region to the next when changing the parameters of the model:

Proposition 1. *1. An increase in c can trigger a switch to a less centralized architecture.*

2. A higher p^{**} or a lower p^* can trigger a switch to a more centralized architecture.

3. When a joint architecture is optimal, the inspection threshold increases in p^{**} and decreases in p^*. It also increases in λ if $c > (W_1 - W_0)\bar{B}(p^*, p^{**})$, and decreases otherwise.

[11] The parameters used for the figure are $\Phi(x) = x$, $W_1 = 1.05, W_I = 0.8, W_0 = 0$, so that $p^{**} = 0.76$.

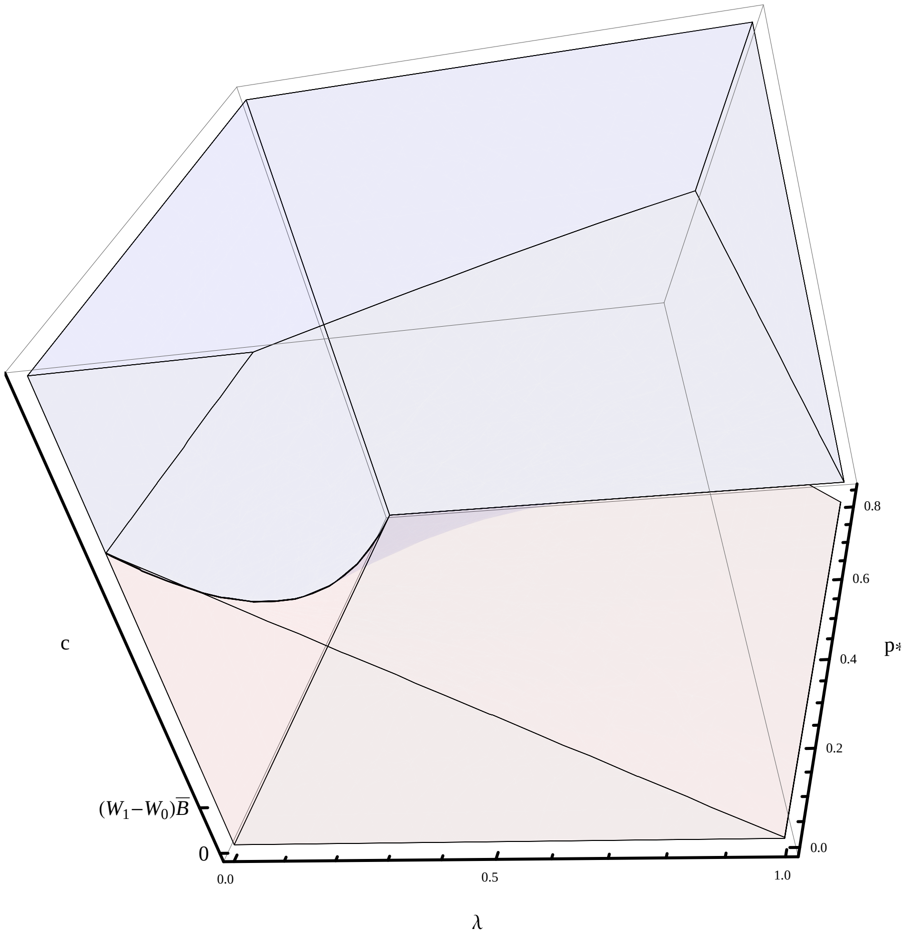

Figure 3: Regions with optimal centralized (red), delegated (blue) and joint architectures (transparent) depending on p^*, λ and c.

Point 1 follows directly from Lemma 1, points 2 and 3 are proven in the Appendix A.1. Intuitively, the central supervisor always inspects more when her information disadvantage c is lower or when the conflict of objectives with the local supervisor, as measured by the wedge between p^{**} and p^*, is higher.

The role of λ is more subtle. When the cost c is low, a supervisor with very imprecise information wants to inspect banks on-site as this is not very costly and delegating would be risky. If the signal is more precise, she will sometimes get signals outside $[p^*, p^{**}]$ giving a very high probability that the local supervisor will take the first-best decision, and it is then optimal to delegate. More information thus allows the central supervisor to take less risks when delegating to the local supervisor.

Conversely, if c is high, a supervisor with imprecise information fully delegates to the local supervisor because there is a risk that a costly inspection is unnecessary, even if the signal points towards a conflict of objectives. With a higher precision, the central supervisor is more confident that inspecting will be useful if the signal belongs to $[p^*, p^{**}]$, so that it is sometimes optimal to inspect. More information allows to take less risks when inspecting.

3.2 Investment in monitoring

We can now solve for the optimal level of investment by the central supervisor in the monitoring technology. The supervisor's problem is to maximize in λ the expected benefits from supervision minus the investment costs. Denoting $\mathcal{B}(\lambda)$ this quantity, we have:

$$\begin{aligned}\mathcal{B}(\lambda) &= \mathbb{E}(\max(B(\lambda,s)-c,0)) - C(\lambda) \\ &= \begin{cases} (W_1-W_0)\bar{B}(p^*,p^{**}) - c - C(\lambda) & \text{if } B(\lambda,0) > c \\ -C(\lambda) & \text{if } B(\lambda,p^*) < c \\ \int_{p^*}^{\bar{s}(\lambda)}(B(\lambda,s)-c)\phi(s)ds - C(\lambda) & \text{otherwise} \end{cases}\end{aligned}$$

If the chosen architecture is not joint then surely the supervisor chooses $\lambda = 0$, as acquiring a signal is costly and no signal affects her choice. If $\lambda = 0$ then either c is high and the solution is to delegate, or it is low and the solution is to centralize. Finally, as getting an infinitely precise signal is by assumption too costly, $\lambda = 1$ cannot be optimal. Hence for a given c we have to compare the benefit obtained with $\lambda = 0$ and the benefit obtained with an interior solution satisfying the following first-order condition:

$$C'(\lambda) = (W_1-W_0)\int_{p^*}^{\bar{s}(\lambda)}\left(p^{**} - s - \bar{B}(p^*,p^{**})\right)\phi(s)ds \tag{11}$$

Proposition 2. *There exist $\underline{c}$ and $\overline{c}$ with $\overline{c} > (W_1-W_0)\bar{B}(p^*,p^{**}) > \underline{c}$ such that:*

1. If $c < \underline{c}$ the centralized architecture is optimal, with $\lambda = 0$.

2. If $c > \overline{c}$ the delegated architecture is optimal, with $\lambda = 0$.

3. If $c \in [\underline{c}, \overline{c}]$ the joint architecture is optimal, with $\lambda > 0$. The optimal λ is increasing in c for $c \in [\underline{c}, (W_1-W_0)\bar{B}(p^*,p^{**})]$ and decreasing in c for $c \in [(W_1-W_0)\bar{B}(p^*,p^{**}), \overline{c}]$.

See the Appendix A.2 for the complete proof. Figure 4 shows the optimal architecture and the optimal λ as a function of c, as well as the probability that the central supervisor inspects and the probability of a liquidation.[12] For low inspection costs the best option is to fully centralize and not invest in a more precise signal. As costs increase, inspection is more costly and investing in a more precise signal is a way to save on these costs, hence λ is increasing in c. As the threshold $(W_1 - W_0)\bar{B}(p^*, p^{**})$ is reached, inspecting is so costly that the default option without any signal would be to fully delegate. Investing in a better signal is a way to make sure that the central supervisor will still inspect when it's worth it. When $\bar{c}$ is reached the supervisor decides to fully delegate and thus does not invest in a better signal.

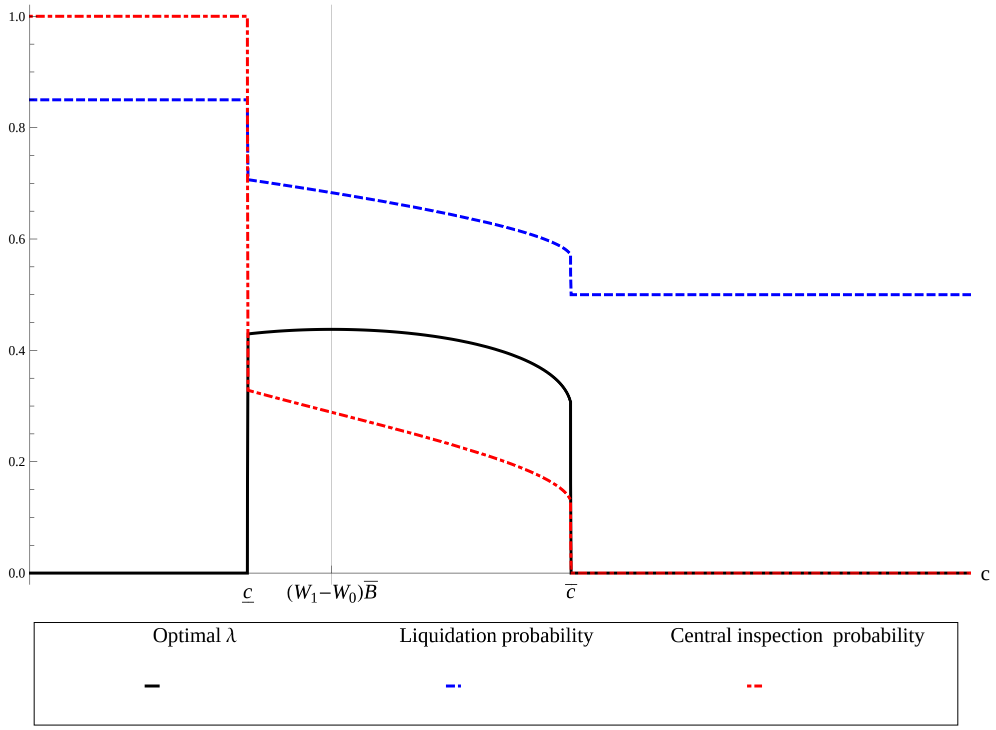

Figure 4: Optimal architecture and λ as a function of c, and implied probabilities of inspection and liquidation.

Considering equation (11), it is immediate that near an interior solution a cost function with a higher marginal cost gives a lower optimal λ and a lower expected benefit:

Corollary 1. *If the cost function C can be written as $C(\lambda) = \gamma\tilde{C}(\lambda)$ and a joint architecture is optimal, an increase in γ leads either to a joint architecture with a lower λ, or to a switch to*

[12] The cost function used to produce the graph is $C(\lambda) = 0.05\lambda^2$.

a delegated architecture if $c > (W_1 - W_0)\bar{B}(p^*, p^{**})$*, or to a switch to a centralized architecture if* $c < (W_1 - W_0)\bar{B}(p^*, p^{**})$.

Higher monitoring costs thus lead to an extreme solution, either centralization or delegation. Finally, we can look at the impact of p^* and p^{**} on the optimal architecture:

Corollary 2. *1. A higher* p^{**} *or* $W_1 - W_0$ *can trigger a switch to a more centralized architecture.*
2. If $\Phi(p^{**}) - \Phi(p^*) < \frac{1}{2}$*, a lower* p^* *can trigger a switch to a more centralized architecture.*

See the Appendix A.3 for the proof. In general, increasing the conflict of objectives implies a move towards regions with more centralization. The impact of p^* is not straightforward: decreasing p^* worsens the conflict of objectives and should encourage acquiring more information. At the same time, since the interval $[p^*, p^{**}]$ is wider, even a less precise signal in this interval would be enough to inspect, which goes in the other direction. The sufficient condition given for the first effect to dominate is actually quite mild, as it means that the probability of p being in the conflict of objectives region is less than one half.

4 Market response to supervision

We have seen how the optimal supervisory architecture depends on the welfare measures W and $\hat{W}$. By reexpressing these measures directly as functions of market prices and quantities, I first derive the best response of the central supervisor to the market outcome. Conversely, I then solve for how private agents react to the supervisory architecture. Combining both gives us the equilibrium of the game in $t = 0$, in which the supervisory architecture is an optimal reaction to the structure of the banking system, and banks' lending and borrowing optimally react to the supervisory architecture.

4.1 Supervisory forbearance

We can first endogeneize the values of the Ws and $\hat{W}$s. If the inspection leads to state I, the proceeds of the liquidated projects $(1 - \ell)L$ are shared between the foreign investors and the depositors (thus reducing the losses of the deposit insurance fund). Denoting α the share accruing to depositors, we have $V_I = 0, \pi_I = 0, \pi_I^D = \alpha L(1-\ell), \pi_I^F = (1-\alpha)L(1-\ell)$, $W_I = (1-\ell)L$ and $\hat{W}_I = \alpha L(1-\ell)$. As the proceeds are distributed pro-rata, we have:

$$\alpha = \frac{D}{D + D_F(1+r_F)} \tag{12}$$

If the projects were not liquidated and are successful, debts are reimbursed and deposits are given back to the depositors, so that we have $V_1 = V(r) - (1+r)L, \pi_1 = (1+r)L - D_F(1+r_F) - D$, $\pi_1^D = D$, $\pi_1^F = D_F(1+r_F)$ and thus $\hat{W}_1 = V(r) - D_F(1+r_F), W_1 = V(r)$. If the projects fail instead, the deposit insurer gives D to depositors and foreign investors get nothing, hence $V_0 = 0, \pi_0 = 0, \pi_0^D = 0, \pi_0^F = 0$ and $\hat{W}_0 = W_0 = 0$.

Using (4) and (5), the intervention thresholds for both supervisors are given by:

$$p^* = \frac{\alpha(1-\ell)L}{V(r) - D_F(1+r_F)}, \ p^{**} = \frac{(1-\ell)L}{V(r)} \tag{13}$$

Using (12), the equality $D_F = L - D$ and rearranging, $p^{**} \geq p^*$ is equivalent to $V(r) \geq (1+r_F)L - r_F D$, which is true as by definition the surplus $V(r)$ is higher than $(1+r)L$ and $r \geq r_F$. The local supervisor thus exerts forbearance. When $D = L$ both $\hat{W}_1 = W_1$ and $\hat{W}_I = W_I$, so that forbearance goes to zero.

The difference between the local supervisor's behavior and the first-best comes from two effects. First, in case of intervention a fraction $1-\alpha$ of the liquidated projects goes to foreign investors instead of local agents, giving less incentives to intervene than in the first-best. Second, when loans are repaid a part of the surplus also goes to foreign investors, giving less incentives to let the banks operate. The combination of these two effects is always towards forbearance. When D is close to L, the banks almost do not need to borrow, hence local welfare is close to global welfare in all states and the local supervisor's incentives are aligned with the first-best.

4.2 Multiplicity of equilibria

We have a model of how the supervisor behaves for given interest rates and loan volumes. But if the central supervisor reacts to the current market situation, for instance by choosing more centralization, this will in turn affect the foreign investors' incentives to lend to the banks, and thus interest rates. This will lead to a further change in supervisory architecture, and so on. The outcome of such a process is an equilibrium in which both supervision and market outcomes are endogenous: the central supervisor's decision is optimal for the anticipated interest rates and loan volumes, private agents' decisions are optimal given the anticipated supervisory architecture, and anticipations are correct. Moreover, while private agents are assumed to correctly anticipate the value p^* chosen by the local supervisor, p^{**} and λ chosen by the central supervisor, as they are infinitesimally small they neglect the effect of their own behavior on these variables.

Denote $\Pr(I)$ the probability that loans are liquidated (more formally, the probability that p and s are such that they lead to liquidation), and $\mathbb{E}(p|\bar{I})$ the expectation of p conditionally on the loans not being liquidated. A bank makes an expected profit of:

$$\pi = (1 - \Pr(I))\mathbb{E}(p|\bar{I})[L(r - r_F) + r_F D] \tag{14}$$

Thus it must be the case in equilibrium that $r = r_F$, otherwise the banks would choose to borrow and lend more. r_F is determined by the investors' incentives to lend to a bank. The return on such a loan is $1 + r_F$ if the bank's projects succeed, 0 if they fail. If the bank's assets are liquidated, they are worth $(1 - \ell)L$, of which creditors recover a share $1 - \alpha$. In equilibrium, r_F must be such that foreign investors are indifferent between investing D_F at the risk-free rate 1 and lending the same amount to the banks:

$$\Pr(I)\frac{(1 + r_F)(1 - \ell)L}{D + D_F(1 + r_F)} + (1 - \Pr(I))\mathbb{E}(p|\bar{I})(1 + r_F) = 1 \tag{15}$$

The probability of liquidation depends on the supervisory architecture. With a centralized architecture we have $\Pr(I) = \Phi(p^{**})$, while with a delegated architecture $\Pr(I) = \Phi(p^*)$. Under a joint architecture liquidation is obtained if either supervisor intervenes, which gives:

$$\begin{aligned}
\Pr(I) &= \lambda\Phi(\bar{s}(\lambda)) + (1-\lambda)[\Phi(p^*) + (\Phi(p^{**}) - \Phi(p^*))(\Phi(\bar{s}(\lambda)) - \Phi(p^*))] && (16)\\
(1 - \Pr(I))\mathbb{E}(p|\bar{I}) &= \lambda\int_{\bar{s}(\lambda)}^{1} s\phi(s)ds + (1-\lambda)\int_{p^*}^{1} p\phi(p)dp && (17)\\
&- (1-\lambda)(\Phi(\bar{s}(\lambda)) - \Phi(p^*))\int_{p^*}^{p^{**}} p\phi(p)dp
\end{aligned}$$

Although the exact expressions are lengthy, the intuition is simple: conditional on some success probabilities $p \in [p^*, p^{**}]$, the probability of liquidation increases with more supervision (centralization or higher $\bar{s}(\lambda)$). We can now derive the equilibrium conditions ensuring that both market participants and supervisors react optimally to each others' actions:

Lemma 2. *In a market equilibrium the interest rate r^* satisfies:*

$$\Pr(I)\frac{(1+r^*)(1-\ell)q(r^*)}{D+(q(r^*)-D)(1+r^*)} + (1-\Pr(I))\mathbb{E}(p|\bar{I})(1+r^*) = 1 \tag{18}$$

Moreover, p^* and p^{**} satisfy (13); equations (16) and (17) are satisfied; λ and the central supervisor's behavior obey Proposition 2 and Lemma 1.

This Lemma directly follows from the previous analysis, equating r^* with r_F and imposing the equilibrium condition $L = q(r^*)$. It gives a long-run equilibrium in which demand equals supply both for loans to the borrower and to banks, the local supervisor optimally reacts to the balance sheet of banks, the central supervisor optimally invests in monitoring and inspects based on the local supervisor's behavior, and the supervisory architecture is taken into account by foreign investors when they ask for a certain interest rate. The following proposition studies how supervision and lending react to each other:

Proposition 3. *1. r is lower (more foreign lending) under the optimal architecture chosen by the central supervisor than if there is only a local supervisor.*
2. If the elasticity of demand for loans is high enough and $\Phi(1-\ell) < \frac{1}{2}$, an exogenous decrease in r (more foreign lending) favors a more centralized architecture.

The proof is in the Appendix A.4. The assumption on ℓ simply ensures that the condition given in point 3 of Corollary 2 is necessarily met. A decrease in r pushes p^* upwards because the loans are less costly to liquidate; the assumption on demand elasticity ensures that this effect is more than compensated by the increase in foreign lending. These assumptions give sufficient conditions only and are by no means necessary.

Foreign lending and centralized supervision reinforce each other: more centralized supervision increases foreign lending, more foreign lending increases the conflict of objectives, and a higher conflict of objectives increases the incentives for central supervision. If increased central supervision more than compensates the higher degree of forbearance of the local supervisor, foreign lending increases again. This complementarity between foreign lending and centralized supervision can lead to a multiplicity of equilibria. The next proposition gives sufficient conditions for the extreme cases of centralization and delegation to be both possible equilibrium outcomes:

Proposition 4. *If the elasticity of demand for loans is high enough and* $\Phi(1-\ell) < \frac{1}{2}$*, then for high enough monitoring costs* $C(\lambda)$ *there exist* c_1, c_2 *with* $c_1 < c_2$ *so that for* $c \in [c_1, c_2]$ *both an equilibrium with centralization and an equilibrium with delegation exist.*

At least for c *close to* c_1*, the equilibrium with centralization is associated to a higher global welfare than the equilibrium with delegation.*

See the Appendix A.5 for the proof. The assumption that $C(.)$ is high simplifies the problem by excluding the possibility of a joint architecture in equilibrium, but is not necessary.

According to the proposition, two very different equilibria may obtain. With a centralized architecture, foreign investors are ready to lend a high quantity at a low interest rate because the central supervisor takes into account their potential losses. Banks then have high cross-border externalities, making a centralized architecture necessary. With the same parameters but a delegated architecture, foreign investors lend less because they expect the local supervisor to be too lenient. Cross-border externalities are then low and supervision is optimally delegated to the local level.

Fig. 5 gives an example in which multiplicity is obtained.[13] On the left panel we have for different values of r the marginal gross return of one unit lent to a bank according to (18), both if liquidation is triggered by the local ($p \leq p^*$) and by the central supervisor ($p \leq p^{**}$). The points at which these returns are equal to 1 give us r_c and r_d, the equilibrium interest rates under a delegated and a centralized architecture. The right panel displays the surplus from centralizing the supervision, equal to $(W_1 - W_0)\bar{B}(p^*, p^{**})$, as a function of r. To get multiplicity, we need a cost c such that the level of r making the central supervisor indifferent between delegating and centralizing falls between r_d and r_d. This occurs in the figure for example with $c = 0.83$, such that centralization is obtained whenever $r \leq 0.1$, and delegation

[13]The parameters used are $\ell = 0.6$, $D = 1$, $q(r) = r^{-1.5}$, and p follows a Beta distribution with parameters $(9, 1)$, so that $\mathbb{E}(p) = 0.9$.

is obtained otherwise. As $r_c = 0.08$ and $r_d = 0.12$, both equilibria are possible.

The figure illustrates well the condition for multiplicity: the marginal return on a loan to a bank must be higher under the centralized architecture than under the delegated architecture so that $r_c \leq r_d$ (Point 1 of Proposition 3), and the surplus from centralizing must be decreasing in r (Point 2 of Proposition 3).

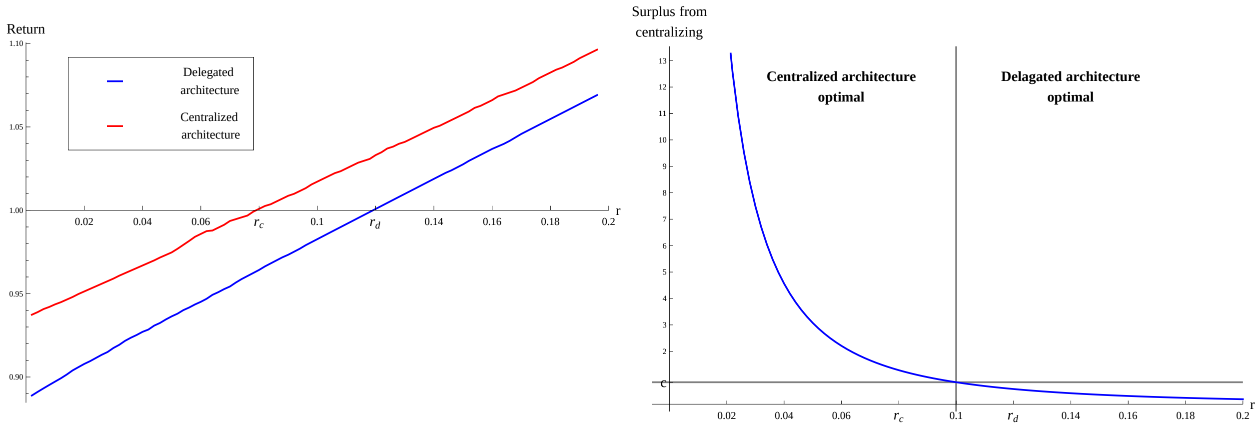

Figure 5: Marginal gross return for foreign investors (left) and surplus from centralizing supervision (right).

5 Discussion

I first review the implications of the model on which banks should be attributed to a central or a local supervisory authority, or be supervised jointly. I then focus on the additional policy and empirical implications of considering that in the long-run market integration reacts to the supervisory architecture.

5.1 Towards a supervisory typology of banks

The results of section 3 shed light on possible options for organizing banking supervision at a federal or supranational level. The optimal arrangement depends on three different dimensions:

-the **conflict of objectives** is measured in the model by $\bar{B}(p^*, p^{**})$. As shown in 4, this dimension typically depends on the impact of a bank's default on non-local agents, and more generally on its **cross-border (or interstate) activities**. Access of a bank to federal deposit insurance would be a typical indicator of potential conflicts of objectives. In addition, in its guidelines to identify G-SIBs[14] the Basel Committee on Banking Supervision (BCBS) gives indicators for interconnectedness and cross-jurisdictional activities, which can also proxy for this dimension. A last measure can be obtained by looking at past supervisory interventions, estimating how "delayed" they were and what are the determinants of the delay. Beck, Todorov, and Wagner (2013) provide such estimates for the European case.

-the **informational advantage** of the local supervisor is measured in the model by c. This parameter should be related to a bank's **national/regional specificity**. The supervision of a bank investing mostly in local assets and getting funds from domestic agents requires specialized supervisory teams with a good knowledge of the local conditions, financial products or structures, and language. In Europe, examples of such products largely specific to particular countries would be "Prêts cautionnés" (France), "Pfandbriefe" (Germany) or shares in Finnish residential housing companies. The French model of "bank insurance", driven by local tax conditions, is also unique and its future treatment by the central supervisor is the topic of much debate.

-**off-site monitoring costs** are measured in the model by the function $C(\lambda)$ and determine how costly it is for the central supervisor to acquire information without an on-site inspection. It should be thought of as mostly related to a bank's **complexity and opacity**. The BCBS uses three proxies for complexity, namely the amounts of over-the-counter derivatives, level 3 assets, and trading and available-for-sale securities. Additional proxies for opacity have been used in the literature, such as the proportion of "transparent assets" or whether the bank is publicly traded or not.[15]

Proposition 2 shows that the key parameter for determining the optimal architecture is c divided by $(W_1 - W_0)\bar{B}(p^*, p^{**})$, that is the informational advantage of local supervisors relative to the conflict of objectives. With a high ratio the optimal architecture tends to delegation, but if off-site monitoring costs are low enough the central supervisor inspects the bank after low signals. With a low ratio the optimal architecture tends to centralization, but

[14]BCBS, "Global systemically important banks: assessment methodology and the additional loss absorbency requirement", November 2011.

[15]See e.g. Morgan (2002).

the central supervisor delegates inspections to the local level after high signals, to save on inspection costs. The following table gives the optimal architecture in four cases:

$c/[(W_1 - W_0)\bar{B}]$ Monitoring costs	Low	High
High	Centralized	Delegated
Low	Joint, Delegate after high signal	Joint, Centralize after low signal

While it is beyond the scope of this paper to fully flesh out this typology with an empirical analysis, it is worth noting that several proxies already exist for two of the three relevant dimensions. The last ingredient needed to operationalize the above typology is a measure of national specificities for a given bank, based for instance on the composition of its assets or the complexity of the national legal framework.

Current supervisory arrangements already favor centralized supervision when the impact of a bank's default is not restricted to the local level, corresponding to the conflict of objectives dimension. In the United States, the access of a bank to federal deposit insurance automatically implies some supervision at the federal level, the intensity of which depends in particular on the size of the bank and thus the potential losses that the deposit insurance fund might incur. Agarwal *et al.* (2012) provide evidence on supervisory forbearance at the State level, which they relate to proxies for regulatory capture and to staff characteristics. These variables can be seen as other measures of the conflict of objectives, and give guidance on whether a more centralized architecture would be more profitable. While there is no common deposit insurance in the Euro area yet, a bank that has requested assistance from the European Stability Mechanism or the European Financial Stability Facility automatically falls under SSM supervision, which obeys the same logic. The significance of a bank's cross-border activities and its size are also explicit criteria for a bank to fall under centralized supervision.

Also in the U.S. case, Rezende (2011) shows that joint examinations by State and Federal supervisors are more frequent for large and complex institutions, which corresponds well to the effect of opacity in the model. An alternative explanation developed by the author is that Federal supervisors support State supervisors with not enough staff to properly supervise a particular bank. This is also consistent with the model if one reinterprets the difference $p^{**} - p^{*}$ as coming not from externalities but from different abilities to spot problem banks. The trade-off is then between a more costly but more robust inspection technology (Federal supervisor) and a cheaper but less robust one (State supervisor).

None of the criteria used for assigning Euro area banks to central or local supervision seem directly linked to complexity and specificity, with the exception of size which is probably correlated with both, as well as with the conflict of objectives.[16] Taking into account these additional dimensions could increase the cost-effectiveness of the SSM by identifying banks that are transparent enough to be supervised under a joint architecture, or banks that are so specific that supervising them from the center is too costly or too difficult.

One last conclusion that applies to both the United States and the Euro area is that the central supervisor should not necessarily focus her attention on the banks on which she receives the worst signals, as in these cases the local supervisor takes the correct decision because potential losses in case of no intervention far outweigh the benefits.

5.2 Supervisory architecture and market integration

Policy implications: the assumption that local deposits are fixed and that banks can only get extra funding from foreign investors models a situation in which the optimal allocation of capital makes large cross-border capital flows necessary. The supervised banks can be seen as operating in a country with a savings deficit, while foreign investors live in a surplus country. This fits the European situation particularly well, also because European banks play a major role in the cross-border allocation of capital.[17]

Adequate banking supervision then has an impact on cross-border capital flows and thus on market integration. To see this, consider the extreme case in which there is no bank su-

[16] The list of criteria is available at `http://www.ecb.europa.eu/ssm/html/index.en.html`.

[17] Gilje, Loutskina, and Strahan (2013) provide recent evidence that bank branches also play an important role in financial integration in the United States.

pervisor and loans are never liquidated. Applying (18), the equilibrium interest rate satisfies:

$$(1+r)\int_0^1 p\phi(p)dp = 1 \tag{19}$$

In comparison, a welfare-maximizing interest rate would satisfy:

$$\Phi(p^{**})(1-\ell) + \int_{p^{**}}^1 (1+r)p\phi(p)dp = 1,\ p^{**} = \frac{(1-\ell)q(r)}{V(r)} \tag{20}$$

Notice that for $p \leq p^{**}$ we have by definition $1-\ell \geq p(1+r)$. This shows that for a given r the left-hand side of (20) is larger than $(1+r)\int_0^1 p\phi(p)dp$. The optimal r is thus lower than the one prevailing in the absence of supervision.

In other words, a low level of supervision acts as a friction preventing capital to flow through banks from surplus to deficit countries. The first part of Proposition 3 shows that the European SSM can indeed contribute to restoring market integration in the Euro area, which is one of the objectives stated by the European Commission.[18] However, this reassuring result comes with three caveats:

-another face of the same coin is that market integration increases cross-border externalities, and thus the conflict of objectives between local and central supervisors.[19] A conclusion for the design of the European SSM is that flexibility in the degree of centralization is valuable, because changes in supervisory architecture affect supervised agents, which changes the trade-off for the central supervisor.

-the timing assumption matters in Proposition 4. If the central supervisor could credibly commit to a specific architecture, she could always choose the best equilibrium in case of multiplicity. This may involve setting up a more centralized architecture even though *current* market conditions would not justify it (or on the contrary a delegated architecture even though the bank is currently well integrated). In other words, the optimal architecture should be forward-looking and take into account how it can affect market integration in the

[18] Morrison and White (2009) also show that harmonized supervision increases market integration, through a different mechanism: absent a "level playing field", good banks choose to be chartered in countries with a strong supervisory reputation, while banks of lower quality choose lenient countries, giving rise to distortions in the banking market.

[19] Moreover, market integration can encourage banks to specialize more, leading to asymmetric banking crises. See Fecht, Gruner, and Hartmann (2012).

long-run.[20]

-a risk is that the forward-looking players may be the banks themselves. Extension 1 in the Internet Appendix considers a bank with market power that anticipates the central supervisor's choice and can strategically decide to keep the amount it borrows from foreign creditors just below the threshold above which the bank would be allocated to the central supervisor. More surprisingly, a bank relying a lot on foreign lending can choose to remain just *above* the threshold to avoid being allocated to the local supervisor. These phenomena would be the European equivalent of U.S. banks strategically choosing their supervisor, as studied in e.g. Rosen (2003) or more recently Rezende (2014).

Empirical implications: the main implication of this section is that banks switching to a more centralized supervisory architecture should get funding at lower rates and attract more funding from foreign agents. This prediction could be tested in several ways:

The implementation of the SSM in the Euro area gives a unique opportunity to empirically study the impact of bank supervision, as it is a large change affecting differently the banks directly supervised by the ECB and the other ones. While the two groups are not similar before the implementation of the SSM, the heterogeneity of banking systems in the Euro area implies that it is possible to match at least the smaller "treated" banks with comparable non-treated banks. The treatment is not random, but the selection criteria are public and based on observable variables, so that selection biases can be controlled for. A particularly interesting selection criterion is that at least the three most significant credit institutions in each participating country shall be supervised directly by the ECB, implying that some banks may be in the treated group although they are actually relatively small and comparable to some non-treated banks in other countries. Another implication is that the treatment introduces a large discrepancy between the third and the fourth largest credit institutions in some countries, suggesting a regression discontinuity design approach, the problem being then a very small sample size.

The model predicts that treated and untreated banks should diverge over time in a predictable way. Banks directly supervised by the ECB should be able to borrow at lower rates

[20] A parallel can be made with the classical commitment problem in bank closure, which is the subject of Mailath and Mester (1994).

from foreign agents, typically on interbank markets, compared to similar banks under national supervision. CDS spreads should also fall. Extension 2 in the Internet Appendix shows that the complementarity between market integration and centralized supervision is more general than is shown in section 4:[21] treated banks should attract more foreign creditors but also have more foreign assets and more foreign shareholders. The reasoning is the same: centrally supervised banks face less incentives to manipulate the domestic/foreign composition of their different balance sheet items to ensure a more favorable supervisory treatment. As a result, national biases decrease and the cross-border activities of the treated banks should increase compared to the non-treated banks.

Such drastic and partly exogenous changes in the supervision of a bank are more difficult to find in the United States, especially because banks that switch their primary supervisor choose to do so, introducing a selection bias. A recent exception is the dissolution of the Office of Thrift Supervision in 2011, but this event does not provide a control group as by definition we cannot observe OTS supervision after this date. The rotating supervision by State and Federal supervisors used in Agarwal *et al.* (2012) offers a nice identification strategy for the impact of State vs. Federal supervision. A prediction of the model that can be tested using the same approach is that during the Federal supervisor's shift a bank should find it easier to borrow on the interbank market, especially from banks in a different State.

A difficulty is of course to get access to detailed data on bank funding. An alternative for listed banks is to look at their stock price. In the absence of the effect on bank funding, one would expect more supervision to decrease the value of the bank to shareholders, as less losses will be borne by the deposit insurance fund and the bank's loans will be liquidated more often. This negative impact could be more than compensated for banks that find it difficult to borrow on interbank markets. The stock price of banks falling under the central supervisor's umbrella can thus be expected to increase if the bank is in a region with a savings deficit, or relies more heavily on central bank funding.

[21] Ongena, Popov, and Udell (2013) study a market in which there is also substitutability: stricter regulation of multinational banks in Western Europe leads their subsidiaries in Central and Eastern Europe to lower their credit standards, in particular when the home supervisor is less efficient. A more centralized supervision would presumably lead to less lending in Eastern Europe, reducing cross-border exernalities.

6 Conclusion

This paper develops a framework to analyze optimal supervisory architectures in a federal/international context in which local supervisors have incentives to be forbearant, but also have more information about domestic banks than does a central supervisor. Whether the optimal architecture is centralized, delegated, or uses both local and central supervisors is shown to depend on three dimensions. Supervision should be more centralized when the conflict of objectives between the two supervisors is more severe, for banks with less specific assets, and banks with an opaque structure difficult to monitor from afar.

The conflict of objectives is the dimension that has attracted the most attention, but is also the most difficult to evaluate: it does not depend only on the different banks' sizes *per se* or on whether their creditors are foreign or domestic, but on who ultimately bears losses in case of default, as well as on a correct evaluation of the local supervisor's incentives. The other important dimensions of the trade-off between centralized and delegated supervision may be easier to estimate empirically, and could thus give more guidance on who should be the supervisor of a particular bank.

The optimal architecture should endogenously change due to the reaction of the banking sector to a new supervisory framework. More centralized supervision reduces market fragmentation, thus increasing cross-border externalities and making centralized supervision even more necessary. A good supervisory architecture should be flexible to accommodate evolutions in market conditions triggered by the supervisory changes themselves. Moreover, due to the complementarity between centralized supervision and market integration, it is possible to be stuck in an equilibrium in which market integration and centralization are both low, when another equilibrium in which both are high would be possible and preferable. The choice of a supervisory architecture should thus be forward-looking and anticipate how the market will react to the new supervision framework.

A Appendix

A.1 Proof of Lemma 1 and Proposition 1

Proof of the Lemma: $B(\lambda, s)$ takes the same value $B(\lambda, 0)$ for all $s \notin [p^*, p^{**}]$, which is lower than the value it takes inside the interval. If $c < B(\lambda, s)$ then a central inspection is optimal for any s, so that the centralized architecture is optimal. For $s \in [p^*, p^{**}]$, $B(\lambda, s)$ is decreasing in s and thus takes its maximal value for $s = p^*$. If $c > B(\lambda, p^*)$ then a local inspection is optimal for any s, so that a delegated architecture is optimal. For intermediate values of c the supervisor inspects only for some $s \in [p^*, p^{**}]$. $B(\lambda, s)$ is decreasing in s, higher than c in $s = p^*$ and lower in $s = p^{**}$, showing point 3.

Proof of the Proposition: using Lemma 1, to prove point 2 we need to show that $B(\lambda, 0)$ and $B(\lambda, p^*)$ increase in p^{**} and decrease in p^*. We have:

$$\begin{aligned} B(\lambda, 0) &= (W_1 - W_0)(1 - \lambda)\bar{B}(p^*, p^{**}) \\ B(\lambda, p^*) &= (W_1 - W_0)((1 - \lambda)\bar{B}(p^*, p^{**}) + \lambda(p^{**} - p^*)) \end{aligned}$$

Differentiating $\bar{B}$ gives:

$$\frac{\partial \bar{B}}{\partial p^*} = -(p^{**} - p^*)\phi(p^*) < 0, \ \frac{\partial \bar{B}}{\partial p^{**}} = \Phi(p^{**}) - \Phi(p^*) > 0 \quad (21)$$

It is then easy to show that $B(\lambda, 0)$ and $B(\lambda, p^*)$ vary with p^* and p^{**} as mentioned above.

Point 3: that $\bar{s}(\lambda)$ increases in p^{**} and decreases in p^* follows directly from (10) and (21). To conclude the proof, we can differentiate $\bar{s}$ with respect to λ:

$$\bar{s}'(\lambda) = \frac{1}{\lambda^2(W_1 - W_0)}(c - \bar{B}(p^*, p^{**})(W_1 - W_0)) \quad (22)$$

which shows the desired result on the impact of a higher λ.

A.2 Proof of Proposition 2

Define first the expected benefit B^d for the central supervisor if he fully delegates and chooses $\lambda = 0$, and B^c the benefit if he fully centralizes and $\lambda = 0$. We have $B^d = 0$ and $B^c = (W_1 - W_0)\bar{B}(p^*, p^{**}) - c$. Notice that $\bar{B}(p^*, p^{**}) \leq p^{**} - p^*$, so that if if $c > (W_1 - W_0)(p^{**} - p^*)$ a delegated architecture is surely optimal because even a perfectly informative signal that $p = p^*$ would not trigger an inspection by the central supervisor. We thus have that $\bar{c} \leq (W_1 - W_0)(p^{**} - p^*)$. Assuming c to be lower than $(W_1 - W_0)(p^{**} - p^*)$, define B^j the maximum benefit if joint inspection is chosen. According to Lemma 1, for a given c choosing a positive λ is useful only if it is such that $c \in [B(\lambda, 0), B(\lambda, p^*)]$, which can be rewritten as $\lambda > \underline{\lambda}(c)$, with:

$$\underline{\lambda}_c = \max\left(\frac{\bar{B}(p^*, p^{**})(W_1 - W_0) - c}{\bar{B}(p^*, p^{**})(W_1 - W_0)}, \frac{c - \bar{B}(p^*, p^{**})(W_1 - W_0)}{(W_1 - W_0)(p^{**} - p^* - \bar{B}(p^*, p^{**}))}\right)$$

We then have:

$$B^j = \max_{\lambda \geq \underline{\lambda}_c} \int_{p^*}^{\bar{s}(\lambda)} ((W_1 - W_0)((1 - \lambda)\bar{B}(p^*, p^{**}) + \lambda(p^{**} - s)) - c)\phi(s)ds - C(\lambda)$$

We can now compare B^c, B^d and B^j for different values of c. In $c = 0$ we surely have $B^c > B^j > B^d$ so that a centralized architecture is optimal. As c increases, we have $\partial B^c / \partial c = -1$ whereas, using the envelope theorem, $\partial B^j / \partial c = -(\Phi(\bar{s}(\lambda)) - \Phi(p^*)) > -1$. To show that there exists $\underline{c} < (W_1 - W_0)\bar{B}(p^*, p^{**})$ such that the centralized architecture is optimal for $c < \underline{c}$ and a joint architecture is optimal for $c \in [\underline{c}, \bar{c}]$, we just need to prove that in $c = (W_1 - W_0)\bar{B}(p^*, p^{**})$ we have $B^j > B^c \geq B^d = 0$. This value of c being precisely such that $B^c = B^d = 0$, it is enough to show that $B^j > 0$.

With this value of c we have $\underline{\lambda}_c = 0$. Using the definition of B^j, choosing $\lambda = 0$ would yield $B^j = 0$. It is thus enough to show that a marginal increase in λ brings a positive benefit. The marginal increase in expected benefit when increasing λ is given by:

$$\frac{\partial \bar{s}(\lambda)}{\partial \lambda} \underbrace{(B(\lambda, \bar{s}(\lambda)) - c)}_{=0} \phi(\bar{s}(\lambda)) + \int_{p^*}^{\bar{s}(\lambda)} (p^{**} - s - \bar{B}(p^*, p^{**}))\phi(s)ds - C'(\lambda) \tag{23}$$

where the first term is null by definition of $\bar{s}(\lambda)$. Using $C'(0) = 0$, in $\lambda = 0$ this derivative is:

$$(W_1 - W_0)\bar{B}(p^*, p^{**})(1 - (\Phi(\bar{s}(\lambda)) - \Phi(p^*))) > 0$$

This shows that $B^j > B^c = B^d = 0$ in $c = (W_1 - W_0)\bar{B}(p^*, p^{**})$, so that a joint architecture is optimal for c in this neighborhood. As c increases past this point B^j decreases and ultimately becomes negative, so that the delegated architecture becomes optimal past some $\bar{c} > (W_1 - W_0)\bar{B}(p^*, p^{**})$.

Finally, we can study how the optimal λ depends on c by differentiating the marginal benefit of λ given by (23) with respect to c (notice that $\lambda = 1$ is never optimal as the cost is assumed to go to infinity). This gives us:

$$\frac{\partial \bar{s}(\lambda)}{\partial c}(p^{**} - \bar{s}(\lambda) - \bar{B}(p^*, p^{**})) = \frac{\partial \bar{s}(\lambda)}{\partial c} \times \frac{c - (W_1 - W_0)\bar{B}(p^*, p^{**})}{\lambda(W_1 - W_0)}$$

As $\frac{\partial \bar{s}(\lambda)}{\partial c} \leq 0$ this shows that the impact of c depends on whether it is above or below $(W_1 - W_0)\bar{B}(p^*, p^{**})$.

A.3 Proof of Corollary 2

Point 1. We need to show:

$$0 = \frac{\partial B^d}{\partial p^{**}} \leq \frac{\partial B^j}{\partial p^{**}} \leq \frac{\partial B^c}{\partial p^{**}} \quad (24)$$

the same inequalities must be true when differentiating with respect to $W_1 - W_0$. Let us start with p^{**}. Using the envelope theorem, we have:

$$\begin{aligned}\frac{\partial B^j}{\partial p^{**}} &= (W_1 - W_0)\left(\lambda + (1-\lambda)\frac{\partial \bar{B}(p^*, p^{**})}{\partial p^{**}}\right)(\Phi(\bar{s}(\lambda)) - \Phi(p^*)) \\ &= (W_1 - W_0)(\Phi(\bar{s}(\lambda)) - \Phi(p^*))[\lambda + (1-\lambda)(\Phi(p^{**}) - \Phi(p^*))]\end{aligned}$$

which is positive, proving the first inequality in (24). The derivative of B^c with respect to p^{**} is:

$$\frac{\partial B^c}{\partial p^{**}} = (W_1 - W_0)(\Phi(p^{**}) - \Phi(p^*))$$

The second inequality in (24) is equivalent to:

$$\Phi(p^{**}) - \Phi(p^*) > (\Phi(\bar{s}(\lambda)) - \Phi(p^*))(\lambda + (1-\lambda)(\Phi(p^{**}) - \Phi(p^*)))$$

which is true as $\Phi(p^{**}) > \Phi(\bar{s}(\lambda))$ and the second term on the right-hand side is lower than one. Following the same reasoning for the impact of $W_1 - W_0$, we have:

$$\frac{\partial B^j}{\partial(W_1 - W_0)} = \int_{p^*}^{\bar{s}(\lambda)} ((1-\lambda)\bar{B}(p^*, p^{**}) + \lambda(p^{**} - s))\phi(s)ds + \frac{\partial \bar{s}(\lambda)}{\partial(W_1 - W_0)} \times \underbrace{\partial B^j / \partial \bar{s}(\lambda)}_{=0} \geq 0$$

which shows the first inequality. We need to compare the impact of $W_1 - W_0$ on B^c and B^j:

$$\frac{\partial B^c}{\partial(W_1 - W_0)} \geq \frac{\partial B^j}{\partial(W_1 - W_0)} \Leftrightarrow \bar{B}(p^*, p^{**}) \geq \int_{p^*}^{\bar{s}(\lambda)} ((1-\lambda)\bar{B}(p^*, p^{**}) + \lambda(p^{**} - s))\phi(s)ds$$

For $\lambda = 1$ we have an equality. The right-hand side is increasing in λ, so that the inequality is always true. An increase in $W_1 - W_0$ thus favors a centralized over a joint solution.

Point 2. Following the same reasoning, the impact of p^* on B^j is given by:

$$\frac{\partial B^j}{\partial p^*} = (W_1 - W_0)(1-\lambda)\frac{\partial \bar{B}(p^*, p^{**})}{\partial p^*}(\Phi(\bar{s}(\lambda)) - \Phi(p^*)) - (B(\lambda, p^*) - c)\phi(p^*)$$

Using (21), B^j decreases in p^* while B^d is unaffected. We then have to compare this impact with the impact on B^c:

$$\frac{\partial B^c}{\partial p^*} = -(W_1 - W_0)(p^{**} - p^*)\phi(p^*)$$

$\frac{\partial B^c}{\partial p^*} < \frac{\partial B^j}{\partial p^*}$ is equivalent to:

$$(W_1 - W_0)(p^{**} - p^*)(1 - (1-\lambda)(\Phi(\bar{s}(\lambda)) - \Phi(p^*))) > (W_1 - W_0)((1-\lambda)\bar{B}(p^*, p^{**}) + \lambda(p^{**} - p^*)) - c$$

It is sufficient to show that this inequality holds for $c = 0$, in which case it simplifies to:

$$(p^{**} - p^*)(1 - (\Phi(\bar{s}(\lambda)) - \Phi(p^*))) > \bar{B}(p^*, p^{**}) \tag{25}$$

As moreover $\bar{B}(p^*, p^{**}) < (p^{**} - p^*)(\Phi(p^{**}) - \Phi(p^*))$ it is enough to have $1 > (\Phi(p^{**}) -$

$\Phi(p^*)) + (\Phi(\bar{s}(\lambda)) - \Phi(p^*))$. This is certainly true if $\Phi(p^{**}) - \Phi(p^*) < \frac{1}{2}$ as $\bar{s}(\lambda) \leq p^{**}$.

A.4 Proof of Proposition 3

Point 1. The choice of the supervisory architecture by a central supervisor increases the probability of liquidation conditional on p being in $[p^*, p^{**}]$. I first show that, all else equal, this increases the left-hand side of equation (15).

By definition, for $p \in [p^*, p^{**}]$ total welfare would be higher with liquidation but local welfare is higher with no liquidation. As the difference between the two quantities is the foreign investors' profit, for such a p this profit is higher in case of liquidation. Total foreign profit is equal to the marginal return on loans times $L - D$, which is given, hence for such a p the marginal return is also higher in case of liquidation. Finally, an increase in the probability of liquidation for a given p gives more weight in the left-hand side of (15) to the marginal return conditional on liquidation and less weight on the marginal return conditional on no liquidation, hence the expression increases.

This shows that the left-hand side of (18) increases when a central supervisor is introduced. Differentiation shows that the left-hand side is always increasing in r^*. Thus to restore equilibrium r^* has to decrease when supervision increases. Since $D_F = q(r^*) - D$, foreign lending increases.

Point 2. As $\Phi(1-\ell) < \frac{1}{2}$ and $p^{**} < 1 - \ell$, both points of Corollary 2 apply. An increase in r has an impact on p^*, p^{**} and $W_1 - W_0$. If we show that an increase in r leads to a higher p^*, a lower p^{**} and a lower $W_1 - W_0$, point 2 will be directly implied by Corollary 2. We have $W_1 - W_0 = V(r)$ and $V'(r) = q'(r)(1+r) \leq 0$. For p^{**}:

$$\frac{\partial p^{**}}{\partial r} = (1-\ell)q'(r)\frac{V(r) - q(r)(1+r)}{V(r)^2} \leq 0.$$

The last thing to show is that p^* increases in r. Using (12), we have:

$$p^* = \frac{D(1-\ell)q(r)}{V(r)[q(r)(1+r) - rD] - [q(r)(1+r) - rD]^2}$$

and, using a log-derivative:

$$\frac{\partial \ln p^*}{\partial r} = \frac{q'(r)}{q(r)} - \frac{q'(r)(1+r)+q(r)-D}{q(r)(1+r)-rD} + \frac{q(r)-D}{V(r)-q(r)(1+r)+rD}$$

Denoting $\varepsilon = \frac{-(1+r)q'(r)}{q(r)}$ the elasticity of the demand for loans to r, we get:

$$\frac{\partial \ln p^*}{\partial r} \geq 0 \Leftrightarrow \varepsilon \frac{rD}{(1+r)[q(r)(1+r)-rD]} \geq \frac{(q(r)-D)(V(r)-2[q(r)(1+r)-rD])}{[q(r)(1+r)-rD](V(r)-q(r)(1+r)+rD)}$$

The right-hand side is lower than if we had $[q(r)(1+r)-rD]$ instead of $2[q(r)(1+r)-rD]$. It is thus sufficient to have the left-hand side higher than this upper bound on the value of the right-hand side, which simplifies to:

$$\varepsilon \geq \frac{(1+r)q(r)-D}{rD} - 1$$

This condition is satisfied if the elasticity ε of the demand for loans to r is high enough.

A.5 Proof of Proposition 4

Let us first assume that there exist two equilibrium interest rates r_c and r_d corresponding to an equilibrium with centralization and an equilibrium with delegation, respectively, and denote $p^*(r_c), p^*(r_d), p^{**}(r_c), p^{**}(r_d)$ the intervention thresholds (notice in particular that r_c and r_d do not depend on the cost c, as the central supervisor is supposed to intervene always or never). We check that equilibrium conditions for same parameters equilibrium conditions can be met in both cases. The market equilibrium condition (18) gives us:

$$\begin{aligned}
\Phi(p^*(r_d))\frac{(1+r_d)(1-\ell)q(r_d)}{D+(q(r_d)-D)(1+r_d)} + (1+r_d)\int_{p^*(r_d)}^{1} p\phi(p)dp &= 1 \\
\Phi(p^{**}(r_c))\frac{(1+r_c)(1-\ell)q(r_c)}{D+(q(r_c)-D)(1+r_c)} + (1+r_c)\int_{p^*(r_c)}^{1} p\phi(p)dp &= 1
\end{aligned}$$

As shown in A.4, under the assumption of high elasticity p^* is increasing in r. Moreover p^{**} is decreasing in r, and $p^{**} = p^*$ if $q(r) = D$. This implies that for any r_c and r_d such that $q(r_c) > D$ and $q(r_d) > D$ we have $p^{**}(r_c) \geq p^{**}(r_d)$. Then Proposition 3 implies that $r_c \leq r_d$.

From this we deduce that $p^*(r_c) \leq p^*(r_d)$ and $p^{**}(r_c) \geq p^{**}(r_d)$.

When for any λ the cost $C(\lambda)$ is made arbitrarily high, the central supervisor only chooses between centralization, which brings $V(r)\bar{B}(p^*, p^{**}) - c$, and delegation, which brings zero. To have both equilibria as possible outcomes we need:

$$V(r_c) \int_{p^*(r_c)}^{p^{**}(r_c)} (p^{**}(r_c) - p)\phi(p)dp \geq c \geq V(r_d) \int_{p^*(r_d)}^{p^{**}(r_d)} (p^{**}(r_d) - p)\phi(p)dp \tag{26}$$

The integral on the left is greater than the integral on the right due the comparison between $p^*(r_c)$ and $p^*(r_d)$ on the one hand, and $p^{**}(r_c)$ and $p^{**}(r_d)$ on the other hand. As $r_c \leq r_d$ we also have $V(r_c) \geq V(r_d)$. Thus the left-hand side in equation (26) is higher than the right-hand side, hence there are intermediate values of c such that both equilibria can be obtained.

Let us now show that the equilibrium with centralization is associated with a higher global welfare than the equilibrium with delegation when c is close enough to the lower bound for multiplicity to obtain. Welfare under delegation or centralization being continuous in c, it is sufficient to consider the case in which c is equal to the lower bound:

$$c = c_1 = V(r_d) \int_{p^*(r_d)}^{p^{**}(r_d)} (p^{**}(r_d) - p)\phi(p)dp$$

Let us denote W_c the global welfare under the centralized equilibrium for this particular value of c, and W_d the global welfare in the delegation equilibrium. We can write W_c as:

$$W_c = \int_0^1 \max((1-\ell)q(r_c), pV(r_c))\phi(p)dp - c_1$$

By definition of c_1, welfare W_d obtained in the delegation equilibrium is also equal to welfare if supervision were centralized, but at interest rate r_d instead of r_c:

$$W_d = \int_0^1 \max((1-\ell)q(r_d), pV(r_d))\phi(p)dp - c_1$$

Given that $r_c \leq r_d$, $q(r_c) \geq q(r_d)$ and $V(r_c) \geq V(r_d)$. Hence welfare is equal or higher in the centralized equilibrium for any value of p. This shows that $W_c \geq W_d$.

A similar reasoning can be applied for $c = c_2$, where at rate r_c the central supervisor is indifferent between the centralized and delegated architectures, so that we can compare the outcome of the delegated architecture with r_c and with r_d. There are two opposite effects however: for a given p^* we still have that welfare is higher with r_c than with r_d. However, $p^*(r_c) < p^*(r_d)$ so that the local supervisor's incentives are worse for the lower interest rate. Either equilibrium may be superior in terms of welfare, depending on which effect is stronger.

References

Acharya, V. (2003): "Is the International Convergence of Capital Adequacy Regulation Desirable?," *Journal of Finance*, 58(6). 7

Agarwal, S., D. Lucca, A. Seru, and F. Trebbi (2014): "Inconsistent Regulators: Evidence from Banking," *Quarterly Journal of Economics*, 129(2). 4, 29, 33

Albertazzi, U., and D. J. Marchetti (2010): "Credit supply, flight to quality and evergreening: an analysis of bank-firm relationships after Lehman," Working paper. 10

Alonso, R., and N. Matouschek (2008): "Optimal Delegation," *Review of Economic Studies*, 75(1). 8

Beck, T., R. Todorov, and W. Wagner (2013): "Supervising cross-border banks: theory, evidence and policy," *Economic Policy*, 28(73). 7, 14, 28

Beck, T., and W. Wagner (2013): "Supranational Supervision - How Much and for Whom?," Working paper. 8

Boyer, P. C., and J. Ponce (2012): "Regulatory capture and banking supervision reform," *Journal of Financial Stability*, 8(3), 206–217. 9

Calzolari, G., and G. Loranth (2011): "Regulation of multinational banks: A theoretical inquiry," *Journal of Financial Intermediation*, 20(2). 7

Costa Lima, R., H. Moreira, and T. Verdier (2012): "Centralized decision making against informed lobbying," Working paper. 9

Decamps, J.-P., J.-C. Rochet, and B. Roger (2004): "The three pillars of Basel II: optimizing the mix," *Journal of Financial Intermediation*, 13(2), 132–155. 8

Dell'Ariccia, G., and R. Marquez (2006): "Competition among regulators and credit market integration," *Journal of Financial Economics*, 79(2), 401–430. 7

Fecht, F., H. P. Gruner, and P. Hartmann (2012): "Financial integration, specialization, and systemic risk," *Journal of International Economics*, 88(1), 150–161. 31

Foarta, O. (2014): “The Limits to Partial Banking Unions: A Political Economy Approach,” Working paper. 8

Frankel, J. A., and A. K. Rose (1998): “The Endogeneity of the Optimum Currency Area Criteria,” *Economic Journal*, 108(449), 1009–25. 5

Gilje, E., E. Loutskina, and P. Strahan (2013): “Exporting Liquidity: Brank Banking and Financial Integration,” Working paper. 30

Grossman, S. J., and O. D. Hart (1986): “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,” *Journal of Political Economy*, 94(4), 691–719. 8

Hardy, D., and M. Nieto (2011): “Cross-border coordination of prudential supervision and deposit guarantees,” *Journal of Financial Stability*, 7. 7

Hiriart, Y., D. Martimort, and J. Pouyet (2010): “The public management of risk: Separating ex ante and ex post monitors,” *Journal of Public Economics*, 94. 8

Holmstrom, B. (1977): *On Incentives and Control in Organizations*. PhD. Thesis, Stanford University. 15

Holthausen, C., and T. Ronde (2004): “Cooperation in International Banking Supervision,” Working paper. 7

Kahn, C., and J. Santos (2005): “Allocating bank regulatory powers: Lender of last resort, deposit insurance and supervision,” *European Economic Review*, 49. 8

Kick, T., and A. Pfingsten (2011): “The importance of qualitative risk assessment in banking supervision before and during the crisis,” Working paper. 11

King, K. K., and J. M. O’Brien (1991): “Market-based, risk-adjusted examination schedules for depository institutions,” *Jounal of Banking and Finance*, 15(4-5), 955–974. 8, 11

Kofman, F., and J. Lawarree (1993): “Collusion in Hierarchical Agency,” *Econometrica*, 61(3). 8

Mailath, G., and L. Mester (1994): “A Positive Analysis of Bank Closure,” *Journal of Financial Intermediation*, 3. 7, 32

Martimort, D. (1999): “The Life Cycle of Regulatory Agencies: Dynamic Capture and Transaction Costs,” *Review of Economic Studies*, 66(4). 9

Merton, R. C. (1978): “On the Cost of Deposit Insurance When There Are Surveillance Costs,” *The Journal of Business*, 51(3), 439–52. 8

Morgan, D. P. (2002): “Rating Banks: Risk and Uncertainty in an Opaque Industry,” *The American Economic Review*, 92(4), pp. 874–888. 28

Morrison, A., and L. White (2009): "Level Playing Fields in International Financial Regulation," *Journal of Finance*, 64(3). 31

Morrison, A. D., and L. White (2013): "Reputational contagion and optimal regulatory forbearance," *Journal of Financial Economics*, 110(3), 642 – 658. 16

Ongena, S., A. Popov, and G. F. Udell (2013): "When the cat's away the mice will play: Does regulation at home affect bank risk-taking abroad?," *Journal of Financial Economics*, 108(2). 33

Peek, J., and E. S. Rosengren (2005): "Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan," *American Economic Review*, 95(4), 1144–1166. 10

Petriconi, S. (2012): "Bank Competition, Information Choice and Inefficient Lending Booms," Working paper. 11

Rezende, M. (2011): "How Do Joint Supervisors Examine Financial Institutions? The Case of State Banks," Working paper. 29

——— (2014): "The Effects of Bank Charter Switching on Supervisory Ratings," Working paper. 32

Rosen, R. J. (2003): "Is Three a Crowd? Competition among Regulators in Banking," *Journal of Money, Credit and Banking*, 35(6), 967–98. 32

Vives, X. (2001): "Restructuring Financial Regulation in the European Monetary Union," *Journal of Financial Services Research*, 19(1), 57–82. 8

Zoican, M., and L. Gornicka (2014): "Banking Union Optimal Design under Moral Hazard," Working paper. 8

Acknowledgements
I am grateful to Elena Carletti, Dean Corbae, Hans Degryse, Giovanni Di Iasio, Co-Pierre Georg, Peter Hoffmann, Charles Kahn, Julian Kolm, Myron Kwast, Perrin Lefebvre, Olivier Loisel, Gyongyi Loranth, David Marques-Ibanez, Marcelo Rezende, Roberto Savona, Bernd Schwaab, Alexandros Vardoulakis, Larry Wall, Marius Zoican, seminar participants at the European Central Bank, the Deutsche Bundesbank, the Paris School of Economics, the University of Mannheim, the University of Zurich and HEC Paris, and participants to the EUI Conference on Macroeconomic Stability, Banking Supervision and Financial Regulation, the 2014 IBEFA-ASSA meetings, the 5th Financial Stability Conference in Tilburg, the 2013 Bocconi-Caren Conference, the 2013 CREDIT Conference, the 2013 Conference on Banks and Governments in Globalised Financial Markets and the 2013 EEA Congress for helpful comments and suggestions. This paper won the 1st SUERF/Unicredit & Universities Foundation Research Prize. The views expressed in this paper are the author's and do not necessarily reflect those of the European Central Bank or the Eurosystem.

Jean-Edouard Colliard
European Central Bank, Financial Research; e-mail: colliard@hec.fr

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Internet www.ecb.europa.eu

ISSN 1725-2806 (online)
ISBN 978-92-899-1599-1
DOI 10.2866/449351
EU catalogue number QB-AR-15-026-EN-N