

Can We Prove a Bank Guilty of Creating Systemic Risk? A Minority Report*

Jon Danielsson Kevin R. James Marcela Valenzuela
Ilknur Zer

November 2015

Forthcoming in *Journal of Money, Credit and Banking*

Abstract

Since increasing a bank's capital requirement to improve the stability of the financial system imposes costs upon the bank, a regulator should ideally be able to prove beyond a reasonable doubt that banks classified as systemically risky really do create systemic risk before subjecting them to this capital punishment. Evaluating the performance of two leading systemic risk models, we show that estimation error alone prevents the reliable identification of the most systemically risky banks. We conclude that it will be a considerable challenge to develop a riskometer that is both sound and reliable enough to provide an adequate foundation for macroprudential policy.

Keywords: Systemic risk, macroprudential policy, financial stability, risk management

*Corresponding author Kevin R. James, k.james1@lse.ac.uk. We thank Philippe Mueller, Robert DeYoung, an anonymous referee, and seminar participants at the LSE, the Bank of England's Centre for Central Banking Studies, the Reserve Bank of India, and the Central Bank of Colombia. We also thank the Economic and Social Research Council (UK) [grant number: ES/K002309/1] for its financial support and the AXA Research Fund for its financial support provided via the LSE Financial Market Groups research program on risk management and regulation of financial institutions. Valenzuela acknowledges the support of Fondecyt Project No. 11140541 and Instituto Milenio MIPP IS130002. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. Earlier versions of this paper have been distributed under the title: "Dealing with Systemic Risk When We Measure It Badly".

†Dr. Jon Danielsson, *Department of Finance and the Systemic Risk Centre, London School of Economics* (j.danielsson@lse.ac.uk). Dr. Kevin R. James, *the Financial Markets Group and the Systemic Risk Centre, London School of Economics* (k.james1@lse.ac.uk). Dr. Marcela Valenzuela, *University of Chile (DII)* (mvalenzuela@dii.uchile.cl). Dr. Ilknur Zer, *the Federal Reserve Board* (ilknur.zerboudet@frb.gov).

1 Introduction

In the film *Minority Report* the PreCrime unit of the DC Police uses psychics to detect people who are about to commit crimes and then locks them up before they do so. As it happens, Andrew Crockett's original vision of macroprudential policy (which we call Targeted Macroprudential Policy or *TMP*) is envisioned to work in more or less the same way (Crockett, 2000). The crime in this case is to cause a financial crisis; the perpetrators are banks and other financial firms (henceforth banks) pursuing systemically risky strategies; the task of the PreCrime units of banking supervisors is to detect these banks and require them to increase the proportion of costly capital in their balance sheets so that they do not create a crisis by becoming stressed or failing. The crucial role of the psychics is played by algorithms or *riskometers* that empirically implement a theoretical measure of how much systemic risk each bank creates. In the movie the government shuts down PreCrime as it turns out that the psychics' visions are not accurate enough to justify punishing people on the basis of the information they provide. In this paper we examine the limits of riskometer performance to see if riskometers can be reliable enough to provide a sound foundation for *TMP*. We are, to the best of our knowledge, the first to analyze the relationship between the reliability of systemic risk measurements and macroprudential policy.¹

In our analysis we assume that a regulator acts to minimize the sum of the expected social cost of a crisis and the social cost of bank capital. The regulator is endowed with a riskometer that it can use to get a risk reading for each bank, with the reading equal to either *Guilty* (beyond a reasonable doubt of creating systemic risk) or *Safe*. The regulator chooses between two policy options: i) the *TMP* option in which it requires only banks with a *Guilty* reading to operate with a high level of capital (the level appropriate for a systemically risky bank); and ii) a Blanket Macroprudential Policy or *BMP* option in which the regulator requires all banks to operate with a high level of capital. The *BMP*

¹Our analysis draws upon previous critiques of VaR by Berkowitz and O'Brien (2002); Danielsson (2002, 2008); Hendricks (1996).

option eliminates crisis risk, but it does so inefficiently as both actually risky banks and safe banks must operate with high levels of costly capital. The *TMP* option conserves on costly bank capital, but at the cost of increased crisis risk: banks that do create systemic risk but cannot be shown beyond a reasonable doubt to do so can operate with a low level of capital. We define a riskometer’s reliability to be equal to the proportion of risky banks that get a *Guilty* reading when using that riskometer. The relative advantage of the *TMP* option increases with the reliability of the regulator’s riskometer, and the *TMP* option is always the optimal choice if the regulator has a perfectly reliable riskometer.

We investigate the practicality of *TMP* by first estimating a lower bound for how reliable a riskometer needs to be in order for *TMP* to dominate *BMP* and then estimating an upper bound for how reliable riskometers can be by examining riskometer reliability in a best case scenario. Based upon recent Basel Committee on Banking Supervision (2010) estimates of the cost of a crisis, the level of capital needed to eliminate crisis risk, and the cost of bank capital, we find that the *TMP* option dominates the *BMP* option if the reliability of the regulator’s riskometer exceeds 75% (in our base specification).² Our estimates of riskometer reliability in a best case scenario fall far short of the required 75% level. Consequently, we think that it will be a considerable challenge to develop a riskometer that is reliable enough to provide a sound foundation for *TMP*.

We use the language of crime and punishment when discussing *TMP* since a bank designated as systemically risky definitely suffers a punishment—the higher capital requirement (and other regulatory interventions) that such a designation entails places it at a considerable competitive disadvantage relative to its commercial rivals that are not classified as systemically risky. Indeed, the insurance company MetLife is (at the time of this writing) suing the FSOC over its designation as “systemic”, precisely because it would “put the company on an uneven playing field with its competitors”.³ So, given that *TMP*

²While we use the crisis cost and capital cost figures from the Basel Committee for concreteness, we note that these figures are very similar to those found in other studies on this topic such as de Ramon et al. (2012); Kragh-Sorensen (2012).

³“MetLife to mount legal challenge to systemic risk label”, Financial Times, 13 January 2015.

involves imposing a substantial private cost upon firms, we think that *TMP* will only be sustainable if a regulator can convincingly show (that is, show beyond a reasonable doubt) that the institutions it charges with creating systemic risk really are guilty of creating systemic risk. Since *BMP* does not place any given bank at a disadvantage relative to its peers, we assume that a regulator can pursue a *BMP* policy without having to prove that every bank is guilty of creating systemic risk.

We assume that banks can be ranked by the amount of systemic risk they create, and that a bank is systemically risky if it is in the upper tail (e.g., the top decile or the top quartile) of the bank risk distribution. A technically perfect riskometer (one that implements the true systemic risk model in the correct way) estimated with a sufficiently large sample would provide a *Guilty* reading for all systemically risky banks and a *Safe* reading for all other banks. In reality riskometer readings will not be perfectly accurate due to technical imperfections and estimation risk. There is now a substantial research effort aimed at improving riskometer design (see, for example, Bisias et al. (2012)'s survey of 31 systemic risk models or the ECB's (2012) report on the progress of its Macroprudential Research Network), and we assume that this research effort has been a success and that each of the riskometers we examine is technically perfect. We focus our analysis entirely upon the impact of estimation risk, which even a perfect riskometer will suffer from.

To construct a best case scenario for riskometer reliability, we take steps to minimize estimation risk by analyzing the reliability of two relatively simple (but legitimate) riskometers which we implement in the most straight-forward way that we can devise. The riskometers we choose are the Marginal Expected Shortfall (MES) measure developed by Acharya et al. (2010) and the CoVaR measure developed by Adrian and Brunnermeier (2008, 2011). These two riskometers are among the most widely cited and analyzed systemic risk measures, so they are undoubtedly legitimate riskometers (see, for example, Benoit et al., 2013; Danielsson et al., 2015; Idier et al., 2011; Zhang et al., 2015). Again, as we assume away any error due to design imperfections, any inherent design imperfec-

tions in the riskometers themselves and/or any simplifying assumptions that we make when implementing them that reduce their economic validity will in no way adversely affect our measure of riskometer reliability.

To evaluate riskometer reliability, we obtain daily stock returns for US financial firms for the period 1974–2012. In our base specification we use a sample of the largest 200 firms and rolling 5 year estimation windows. We find that MES has an average reliability of 39% and that CoVaR has an average reliability of 6%.⁴ Each of the riskometers we examine thus falls far short of the level of reliability that *TMP* requires due to estimation error alone. It follows that *TMP* needs a riskometer that is both technically superior to the two relatively simple riskometers that we analyze here while simultaneously being far less subject to estimation risk.

Our analysis rests upon two key assumptions. First, we assume that capital is socially costly. While there is some debate on this point (see, for example Admati and Hellwig, 2013), the consensus in the policy community is that increasing bank capital requirements adversely affects economic activity. The cost of crisis and cost of capital numbers we use in our analysis reflect this consensus. Second, we assume that a regulator pursuing the *TMP* option will need to use a formal systemic risk model (that is, a riskometer) to identify systemically risky banks. In theory, regulators could classify banks as systemic or non-systemic by using intuitive judgement informed by a systemic risk model instead, and in this scenario the reliability of the regulator’s riskometer may not be of central importance. However, as regulators cannot directly observe a bank’s contribution to systemic risk and will not experience many crises, it is difficult to see how they could form a reliable model-free view of a bank’s contribution to systemic risk (let alone a model-free view that could withstand independent challenge from an objecting bank).⁵

⁴We consider a variety of alternative specifications in which we vary the proportion of banks that are systemically risky, the length of the estimation window, and the number of banks in the sample. The choice of specification does not materially alter our reliability results.

⁵The potential for external challenge of regulatory decisions has been heightened by the fact that most regulators are under formal or informal requirements to undertake a cost-benefit analysis of policy decisions. For example, the act authorizing the creation of the UK’s prudential supervisor the PRA requires it to show that “A burden or restriction which is imposed ...should be proportionate to the

Riskometers will therefore be central to any *TMP* effort, and this in turn implies that an effective *TMP* requires a reliable riskometer.

We organize our analysis as follows. In the next section, we analyze the relationship between macroprudential policy and riskometer reliability. Given the reliability threshold this analysis yields, we describe the empirical methodology in Section 3. In Section 4, we present the riskometers we use in our analysis and evaluate their performance. Conclusions follow.

2 Macroprudential Policy and Riskometer Reliability

In this section we derive the relationship between the cost of a crisis, the social cost of bank capital, riskometer reliability, and a regulator's optimal policy choice.

We assume that the regulator oversees a continuum of banks. A proportion ψ of these banks are risky and the remainder are safe. Risky banks create systemic risk and safe banks do not. Banks are otherwise identical. All banks prefer to operate with a privately optimal low level of capital, and safe banks with low capital never cause a crisis. However, a risky bank that operates with low capital may set in train a series of events that cause a crisis if it becomes stressed or fails. A risky bank can eliminate the chance that it will cause a crisis by operating with a high level of capital.

If all risky banks operate with low capital, the expected cost of a crisis in each period is χ . The social cost of requiring all banks (risky and safe) to operate with high capital is κ . We assume that the probability of a crisis is small and hence that the expected cost of a crisis is an approximately linear function of the proportion of risky banks operating with low capital.

benefits" (Financial Services and Markets Act of 2012, Chapter 3, Section 3B(1)(B)). Similarly, due to a series of recent court decisions, US banking supervisors now operate under stringent cost-benefit requirements (Rose and Walker, 2013).

The regulator chooses between policy options to minimize the expected total social cost of financial crises. This cost equals the sum of the expected cost of a crisis itself and the social cost of any additional bank capital that the regulator requires banks to hold to reduce the probability of a crisis. The regulator is also endowed with a riskometer that it can use to obtain a systemic risk reading for each bank. The riskometer provides a reading of either *Guilty* or *Safe*, with a *Guilty* reading meaning that the bank in question is guilty beyond a reasonable doubt of creating systemic risk and a *Safe* reading meaning that the bank is not guilty beyond a reasonable doubt of creating systemic risk. We assume that a proportion ϕ of risky banks gets a *Guilty* reading, and that all remaining banks get a *Safe* reading. We therefore set the riskometer's reliability equal to ϕ .

The regulator chooses between a Targeted Macroprudential Policy (*TMP*) option and a Blanket Macroprudential Policy (*BMP*) option. Under the *TMP* option, the regulator uses its riskometer to get a risk reading for each bank and requires only those with a *Guilty* reading to operate with a high level of capital. All other banks can operate with the privately optimal low level of capital. Under the *BMP* option, the regulator requires all banks to operate with the high level of capital appropriate for a systemically risky bank.

2.1 When is Macroprudential Policy Optimal?

If the regulator chooses the *BMP* option, the expected cost of a crisis is 0 but all banks operate with high capital. The cost of this policy (C_{BMP}) is then

$$C_{BMP} = \kappa. \tag{1}$$

If the regulator instead pursues the *TMP* option, all safe banks and the risky banks that receive a *Safe* reading operate with low capital at no social cost and only the risky banks with a *Guilty* reading need to operate with a high level of capital. Unfortunately, the

$1 - \phi$ proportion of risky banks that obtain a *Safe* reading may also cause a financial crisis. It follows that the cost of this policy (C_{TMP}) is

$$C_{TMP} = (1 - \phi)\chi + \phi\psi\kappa. \quad (2)$$

where ϕ is the riskometer reliability, χ is the expected cost of a crisis in each period, ψ is the proportion of risky banks, and κ is the social cost of requiring all banks (risky and safe) to operate with high capital.

The *TMP* option dominates the *BMP* option if

$$C_{TMP} = (1 - \phi)\chi + \phi\psi\kappa \leq \kappa = C_{BMP}. \quad (3)$$

It follows that the regulator prefers the *TMP* option if the reliability of his riskometer ϕ exceeds

$$\phi \geq \frac{\chi - \kappa}{\chi - \psi\kappa}. \quad (4)$$

Denote the value of ϕ such that the regulator just prefers the *TMP* option by ϕ^* .

2.2 How Reliable Does a Riskometer Need to Be?

We can use equation (4) to calculate how reliable a riskometer needs to be in order for a regulator to find the *TMP* option optimal given values for the expected cost of a crisis and the social cost of capital (together with an assumption on the proportion of banks that are risky). Happily, a recent report by the Basel Committee on Banking Supervision (2010)—henceforth BCBS10—estimates precisely these parameters. We base our estimates of ϕ^* upon the BCBS10 analysis.

When all banks operate with a low level of capital, the key driver of the expected cost of a crisis (χ) is the extent to which a crisis has long-lasting effects beyond the immediate shock. As the evidence reviewed in BCBS10 suggests that financial crises do have longer

lasting effects, we set χ in our base case equal to the expected reduction in output from a crisis assuming a moderate longer term effect. In this case, χ equals 2.64% of output/year. We also calculate ϕ^* under a High χ case and a Low χ case. In the High χ case, we assume that a crisis has a large long term effect ($\chi = 6.62\%$ of output/year) and in the Low χ case we assume that a crisis has no long term effects ($\chi = 0.8\%$ of output per year).⁶

BCBS10 finds that requiring all banks to operate with a capital ratio of 15% will essentially reduce the probability of a crisis to 0. So, we set the High level of capital to 15%. Reviewing the evidence on bank capital requirements and bank activity, BCBS10 finds that increasing the capital ratio of the banking system as a whole to 15% will reduce output by 0.72% of output/year. In our base case, we therefore set $\kappa = 0.72\%$ of output/year. To get an idea of how the regulator's optimal policy choice will vary with the cost of bank capital, we also consider a High κ and a Low κ case. In the High κ case we increase our base κ by 50% to 1.08% of output/year, and in the Low κ case we decrease base κ by 50% to 0.36% of output/year.

We set ψ equal to 10%. Changing ψ to 1%, 5%, or 25% has only a minimal effect upon ϕ^* , so we do not consider the impact of variations in ψ further.⁷ Intuitively, ϕ^* is insensitive to changes in ψ because the marginal impact of this change on the total social cost of capital under the *TMP* option can be offset by a small change in the expected cost of a crisis (which is a linear function of ϕ).

We calculate ϕ^* for each χ/κ combination in Table 1. In our base case, we find that the reliability of the regulator's riskometer must equal at least 75% to make it worthwhile for the regulator to choose the *TMP* option.

[Table 1 approximately here]

The minimum reliability requirement for the regulator's riskometer: i) increases as the expected cost of a crisis increases (as the cost of allowing a risky bank to operate with

⁶These figures are from BCBS10, Table 8.

⁷For example, in our base case, ϕ^* falls from 78% to 73% as ψ falls from 25% to 1%.

low capital increases); and ii) decreases as the social cost of bank capital increases (the relative benefit of *TMP* increases as the cost of requiring all banks to operate with high capital increases). However, setting the expected cost of a crisis equal to its base case value and increasing the social cost of bank capital by 50% decreases ϕ^* to only 62%. The minimum level of reliability that a riskometer must achieve to make *TMP* optimal falls substantially below that of our base case only if one assumes that the expected cost of a crisis is low.⁸

The expected cost of a crisis is low if a crisis has no long lasting effects upon economic performance. In light of recent events, we think that it would be rash to base one's policy choice upon the assumption that financial crises do not have long-lasting effects. Consequently, we think a regulator must have a riskometer with a reliability of least 75% in order for *TMP* to be a sensible choice.⁹

3 A Method to Measure Riskometer Reliability

To measure riskometer reliability, we begin by assuming that the regulator oversees a banking system consisting of Q banks, numbered B_1, \dots, B_Q . Each bank B_q will create a threat to the ability of the financial system to function if it becomes stressed or fails, and we assume that banks can be ordered (from highest to lowest) by the severity of the threat that they pose. Denote B_q 's position in this ordering by $\Theta_{q, True}$. We assume that a proportion ψ of these banks create a severe threat to the ability of the financial system to function and so classify these banks as systemically risky. Hence, a bank is systemically risky if $\Theta_{q, True} \leq \psi Q$.

⁸In this case, the expected cost of a crisis assuming all banks operate with low capital is approximately equal to the social cost of requiring all banks to operate with a high level of bank capital, implying that the *BMP* option does not lower the expected total social cost of a financial crisis. The *TMP* option will therefore dominate the *BMP* option if it leads to even a minimal decrease in the probability of crisis, which it will do even if the regulator's riskometer is very unreliable.

⁹While we base our analysis upon a 15% capital ratio, using the BCBS10 parameters for the expected cost of a crisis and the expected cost of bank capital for capital ratios of 12% to 14% imply that $\phi^* \geq 75\%$. Our 15% capital ratio assumption is therefore a conservative one.

The regulator cannot directly observe $\Theta_{q, True}$ and so cannot know with certainty which banks are guilty of creating systemic risk. Instead, the regulator must use a riskometer ρ and a sample S to estimate the probability that a bank is guilty of creating systemic risk. Denote this probability by $\pi_{q, \rho, S}$, with

$$\pi_{q, \rho, S} = \text{Prob} \left[\Theta_{q, True} \leq \psi Q \right] \Bigg|_S. \quad (5)$$

We use an indicator variable G to identify banks that are guilty *beyond a reasonable doubt* of creating systemic risk. Given the substantial penalty that a bank suffers when a regulator designates it as systemic, we assume that a bank is guilty beyond a reasonable doubt of creating system risk if $\pi_{q, \rho, S} \geq 0.9$.¹⁰ It follows that

$$G_{q, \rho, S} = \begin{cases} 1 & \text{if } \pi_{q, \rho, S} \geq 0.9, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

A riskometer's reliability ϕ then equals

$$\phi = \frac{\sum_q G_{q, \rho, S}}{\psi Q}. \quad (7)$$

For simplicity, we assume here that there are no false positive findings of guilt (this assumption will create an upward bias in our measure of riskometer reliability).

A riskometer's reliability can suffer due to technical imperfections and/or estimation risk. A riskometer is technically imperfect if the position of a given bank in the list of banks sorted by systemic risk as measured by a given riskometer in the limiting case of an infinite amount of data ($\Theta_{q, \rho, Limit}$) is not equal to $\Theta_{q, True}$. A riskometer is subject to estimation risk if the position of a given bank in the list of banks sorted by systemic risk

¹⁰McCauliff (1982) undertook a survey of Federal Judges in the US, and found that majority of them believe that "beyond a reasonable doubt" means: with a probability of 0.9 or higher. Weinstein and Dewsbury (2007) consider this topic as well and argue that "beyond a reasonable doubt" should mean: with probability 0.95 or higher.

as measured by a given riskometer for a given sample S ($\Theta_{q,\rho,S}$) is not necessarily equal to $\Theta_{q,\rho,Limit}$. All riskometers will suffer from estimation risk as the constantly evolving nature of both individual banks and the financial system mean that the amount of data available to estimate a riskometer reading for a given bank at a given time is necessarily limited.

As we are seeking to find an upper bound on how reliable a riskometer can be in the best case scenario, we assume that each riskometer we examine is technically perfect, that is:

$$\Theta_{q,\rho,Limit} = \Theta_{q,True}. \quad (8)$$

We focus our analysis upon the impact of estimation risk on bank guilt probabilities and so on riskometer reliability.

3.1 Estimating Guilt Probabilities

We estimate each bank's guilt probability by using a bootstrap consisting of T trials, numbered $\tau = 1, \dots, T$. In each trial we: i) draw a bootstrap sample from our original data (we discuss this process below); ii) use a riskometer ρ to obtain each bank q 's systemic risk score, which we denote by $\omega_{q,\rho,\tau}$; and iii) sort banks by their risk scores (from highest to lowest). Denote the position of B_q in this list by $\theta_{q,\rho,\tau}$.

We create an indicator variable g to denote the banks that are systemically risky based upon the evidence of a given trial, with

$$g_{q,\rho,\tau} = \begin{cases} 1 & \text{if } \theta_{q,\rho,\tau} \leq \psi Q, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

The probability that a bank is guilty of creating systemic risk is then

$$\pi_{q,\rho,S} = \frac{\sum_{\tau} g_{q,\rho,\tau}}{T}. \quad (10)$$

Our sample will contain both time series and cross-sectional relationships. To capture both aspects of the data, we estimate guilt probabilities using a stationary bootstrap consisting of 1000 trials. Following the Politis and Romano (1991) algorithm, we set the block-size in our bootstrap to 40 observations to capture the time series features of the data. As a sensitivity check, we experimented with block sizes of 20, 30, and 50 observations and found that our choice of block-size did not materially affect our results.

4 Estimating Riskometer Reliability

To establish a plausible upper bound on riskometer reliability, we apply the method we describe in the last section to measure the reliability of two relatively simple riskometers (that is, riskometers that are just complex enough to be considered true *systemic* risk measures). We then use these riskometers to estimate individual bank guilt probabilities, and we use these guilt probabilities in turn to measure riskometer reliability. We focus our analysis on two relatively simple riskometers as it is likely that estimation risk will increase with riskometer complexity (as more complex riskometers place greater demands upon the data). So, by looking at simple riskometers, we can find a lower bound for the impact of estimation risk. Since we have already assumed away the impact of technical imperfections, it follows our analysis will provide a best case scenario for riskometer reliability.

4.1 Riskometers

The two riskometers we examine are Acharya et al. (2010)’s MES and Adrian and Brunnermeier (2008)’s CoVaR. Both MES and CoVaR are true systemic risk measures in that a bank’s risk reading is a function of the relationship between the bank’s performance and that of the market as a whole or the financial system rather than the bank’s performance in isolation. Furthermore, each of these riskometers can be implemented in a very “bare bones” fashion that captures the bank/wider system relationship in a straight-forward manner. Surveying the systemic risk measurement literature, we at least could not find less complex systemic risk measures than these.¹¹

4.1.1 The MES Riskometer

The premise of MES is that a bank creates systemic risk if it performs poorly at the same time that the economy as a whole performs poorly. This idea leads to the MES riskometer, in which a bank’s systemic risk score is equal to its expected return conditional upon the market performing poorly. Acharya et al. (2010) operationalize MES as follows: $\omega_{q,MES}$ is equal to the average return for B_q on days when the market return is at or below its 5% value-at-risk (VaR_M^{05}) value. Denote the set of days for which $R_M < \text{VaR}_M^{05}$ by Z and B_q ’s returns on these days by $R_{B_q,Z}$. Then,

$$\omega_{q,MES} = \text{Mean} [R_{B_q,Z}] . \quad (11)$$

4.1.2 The CoVaR Riskometer

Adrian and Brunnermeier’s (2008, 2011) CoVaR is based upon the idea that a bank creates systemic risk when stress at the bank coincides with stress in the financial system

¹¹Of course, the basic versions of MES and CoVaR that we examine have been subject to numerous refinements that aim to improve their technical quality (see, for example, Adrian and Brunnermeier (2011) and Lopez-Espinoza et al. (2012) for CoVaR and Brownlees and Engle (2015) and Acharya et al. (2012) for MES). But, as we are not evaluating riskometers on the basis of their technical merits, we do not analyze the more elaborate versions of MES or CoVaR here.

as a whole. This idea is implemented in the riskometer ΔCoVaR , which measures a bank's systemic risk score by looking at the change in financial system risk conditional upon the bank being under stress.

The measure of financial system risk is VaR_F^{01} , which is the 1% VaR on the value-weighted portfolio of financial firms. To see how this measure of risk changes as a given bank becomes stressed, suppose that B_q is under stress when $R_{B_q} = \text{VaR}_{B_q}^{01}$. B_q 's ΔCoVaR then equals

$$\omega_{q,\Delta\text{CoVaR}} = \text{VaR}_F^{01} |_{R_{B_q}=\text{VaR}_{B_q}^{01}} - \text{VaR}_F^{01} |_{R_{B_q}=\text{Median}[R_{B_q}]}. \quad (12)$$

Following Adrian and Brunnermeier (2008, 2011), we estimate the relationship between VaR_F^{01} and R_{B_q} with a quantile regression:

$$\text{VaR}_F^{01} |_{R_{B_q}} = \alpha + \beta R_{B_q}. \quad (13)$$

We use equation (13) to calculate $\omega_{q,\Delta\text{CoVaR}}$ in equation (12).

We measure bank returns by (easily observable) daily common stock returns rather than by (difficult to infer) changes in the market value of bank assets. Adrian and Brunnermeier (2008) report that both methods lead to similar results in practice, and we chose daily stock return measure so as to avoid very difficult errors in variable problem.

4.2 Data and Cases

We estimate riskometer reliability using daily return data for large US financial firms for the period 1974–2012.¹² We use the CRSP Value-Weighted Index as our measure of the market rate of return (used to calculate MES) and the Fama-French Finance Industry

¹²We obtain our return data from CRSP, and we count a firm as a financial firm if its 4 digit SIC code is in the set (6020, 6021, 6022, 6023, 6025, 6026, 6030, 6035, 6036, 6052, 6060, 6110, 6111, 6112, 6120, 6122, 6123, 6130, 6140, 6141, 6145, 6146, 6150, 6153, 6159, 6160, 6162, 6163, 6199, 6210, 6211, 6310, 6311, 6330, 6331, 6411, 6500, 6510, 6512, 6513, 6514, 6515, 6519, 6531, 6552, 6553, 6710, 6711, 6712, 6719).

portfolio return as our measure of the financial sector return (used to calculate CoVaR).

In our base case we: i) use a sample consisting of the largest 200 financial firms at the end of each year; ii) assume that 10% of these firms are systemically risky (denoted by $\psi_{10\%}$); and iii) estimate systemic risk scores by using rolling 5 year estimation windows. To identify the largest financial firms, we first drop firms with missing returns over the year Y to $Y - 4$ window and calculate the end-of-year Y market capitalization for the remaining firms. Our sample then consists of the 200 firms with the highest market cap. For a given riskometer, we calculate guilt probabilities (that is, the probability that a sample firm is in the upper decile of the risk score distribution) for each firm and year. We then calculate the reliability for each riskometer for each year.

To assess the sensitivity of our results to the parameters of our base case, we create a series of alternative cases in which we vary the sample size (200 or 300), our assumption on the proportion of banks that create systemic risk (5%, 10%, or 25%), and the length of the estimation window (2 year or 5 year). In particular, we calculate guilt probabilities and riskometer reliability for the following cases:

1. 200 Sample Firms/ $\psi_{10\%}$ /5 Year Estimation Windows;
2. 300 Sample Firms/ $\psi_{10\%}$ /5 Year Estimation Windows;
3. 200 Sample Firms/ $\psi_{25\%}$ /5 Year Estimation Windows;
4. 200 Sample Firms/ $\psi_{25\%}$ /2 Year Estimation Windows;
5. 200 Sample Firms/ $\psi_{5\%}$ /5 Year Estimation Windows.

We find that our results do not vary materially across the cases we examine, so we did not explore all possible sample size/ ψ /estimation window combinations.

4.3 Guilt Probabilities

For each bank and riskometer in a given year, we use the bootstrap method described in Section 3 to calculate guilt probabilities for the cases listed above. In Figure 1, we plot guilt probabilities for each riskometer for the 200 Sample Firms/ $\psi_{10\%}$ /5 Year Estimation Windows case and the 200 Sample Firms/ $\psi_{25\%}$ /5 Year Estimation Windows case. The results for the other cases are very similar. To construct the plot, we sort banks from the highest to the lowest on the basis of a bank's median risk score across the bootstrap trials for a riskometer/year/case combination. In the first case we assume that 20 banks create systemic risk (i.e., $\psi = 10\%$), so we plot the guilt probabilities for the banks with the highest, 10th highest, and 20th highest median risk score. In the second case we assume that 50 banks create systemic risk (i.e., $\psi = 25\%$), so we plot the guilt probabilities of the banks with the highest, 25th highest, and 50th highest median risk score.

To begin with MES, we find that the firm with the highest median systemic risk score in each case is indeed guilty beyond a reasonable doubt of creating systemic risk (Figure 1, Panels A and B). However, the guilt probability of the 10th firm in the $\psi_{10\%}$ case and the 25th firm in the $\psi_{25\%}$ is often less than our 90% threshold for guilt beyond a reasonable doubt. The guilt probability of the 20th firm in the $\psi_{10\%}$ case and the 50th firm in the $\psi_{25\%}$ is always far less than our 90% threshold for guilt beyond a reasonable doubt.

Turning to CoVaR, the firm with the highest median systemic risk score is again generally guilty beyond a reasonable doubt of creating systemic risk. However, for this riskometer it is almost always the case that the guilt probabilities of even the firm with the 10th highest median systemic risk score in the $\psi_{10\%}$ case and the firm with the 25th highest median systemic risk score in the $\psi_{25\%}$ are too low to establish guilt beyond a reasonable doubt.

[Figure 1 approximately here]

To sum up, our results reveal that while the bank with the highest estimated systemic

risk score is generally guilty beyond a reasonable doubt of creating systemic risk, guilt probabilities decline rapidly as a bank's position in the list of banks sorted by systemic risk score falls. In other words, it is impossible to reliably identify the banks in the upper decile (or quartile) of banks ranked by their systemic risk scores.

4.4 Riskometer Reliability

A riskometer's reliability is equal to the number of firms that are guilty beyond a reasonable doubt of creating systemic risk divided by the number of systemically risky firms. As the guilt probabilities that we discuss above would lead one to expect, both MES and CoVaR are unreliable riskometers.

We plot riskometer reliability by year in Figure 2 and provide summary statistics on reliability in Table 2.¹³ The average reliability for the MES riskometer is 38.63% in our base case (standard deviation: 15.82%). The average reliability of the CoVaR riskometer is even lower, equalling only 6.44% in our base case (standard deviation: 6.21%). Reliability for both riskometers is highest for the 200 Sample Firms/ $\psi_{25\%}$ /5 Year Windows case, with MES's reliability increasing to 53.69% and CoVaR's to 20.96%.

[Figure 2 approximately here]

[Table 2 approximately here]

In Section 2 we argue that a regulator finds the targeted macroprudential policy the optimal choice only if the reliability of their riskometer exceeds 75% (or, if one assumes that the cost of bank capital is 50% higher than is reported in BCBS10, 62%). Our results here show that the riskometers we examine do not meet this requirement in any year, let alone on average (we can reject the hypothesis that reliability = 75% and the hypothesis that reliability = 62% at the 1% level for both riskometers). In other words, we find

¹³While examining the dynamics of riskometer performance is beyond the scope of this paper, we note that Danielsson et al. (2015) explore this topic.

that even if one supposes that the riskometers we examine here are technically perfect, estimation error alone renders them incapable of providing an adequate foundation for *TMP*.

5 Conclusion

Andrew Crockett’s (2000) vision for targeted macroprudential policy is a compelling one: since a financial firm only generates a negative externality for the economy if it creates systemic risk, a regulator should target a bank for intervention only if it is guilty of creating systemic risk. However, by classifying a firm as “systemically risky”, a regulator places that firm at a disadvantage relative to its non-systemic commercial rivals. Common sense suggests (and recent evidence confirms) that firms will therefore fiercely oppose being designated as systemically risky. Consequently, in a post-crisis environment, we think that a regulator will need to be able to show that a bank is guilty beyond a reasonable doubt of creating systemic risk before subjecting it to the capital punishment penalty that the macroprudential code sets forth.

If a regulator has to be able to show beyond a reasonable doubt that a firm does create systemic risk, then the regulator will need to use a formal model—a riskometer—to do so.¹⁴ It follows that a regulator will only be able to effectively pursue the *TMP* option if his riskometer is reliable enough to show that a sufficiently high proportion of truly systemically risky firms are guilty beyond a reasonable doubt of creating systemic risk. Riskometer reliability is therefore of central importance to macroprudential policy, and we are, to the best of our knowledge, the first to analyze this topic.

Given current estimates of the benefits of avoiding a financial crisis and the social cost of

¹⁴In antitrust policy, for example, a regulator wishing to block a merger on anti-competitive grounds cannot just claim that a given merger will adversely affect the market based upon regulatory judgement. Instead, the regulator must be able to show using formal analysis and evidence that the merger will adversely affect the market. We think that macroprudential policy will move in this direction as opposition from firms will force regulators to justify their actions.

bank capital, we find that the *TMP* option is optimal only if the regulator can show that at least 75% of systemically risky banks are guilty of creating systemic risk. To see if it is possible for riskometers to achieve this level of reliability, we estimate the reliability of two leading systemic risk models, viz., MES and CoVaR, in a best case scenario. In this scenario we assume that each riskometer is technically perfect and concentrate upon assessing the impact of estimation risk alone (which will affect even the ideal riskometer). We find that even in this best case scenario the riskometers we examine are far less reliable than the *TMP* option demands.

Contemplating what a truly sound riskometer would need to capture to accurately measure a bank's contribution to systemic risk (e.g., the non-linearities in bank behavior or endogenous risk), it is difficult to see how any systemic risk measure that is of sufficient technical quality to serve as a proper foundation for *TMP* will be *easier* to estimate than the versions of MES or CoVaR that we examine here. Thus, while the substantial research effort now underway that aims to develop better systemic risk measures will certainly make progress on the resolving the theoretical and implementation problems that we have assumed away, it seems to us that this progress may well come at the price of making the estimation risk that we do examine here even worse. So, while targeted macroprudential policy would clearly be optimal for an omniscient regulator who could directly observe a bank's impact upon systemic risk, it is not at all clear that it makes sense for actual regulators given the riskometers that we now have or are likely to have.

References

- ACHARYA, V. V., R. ENGLE AND M. RICHARDSON, “Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risk,” *American Economic Review* 102 (2012), 59–64.
- ACHARYA, V. V., L. H. PEDERSEN, T. PHILIPPON AND M. RICHARDSON, “Measuring Systemic Risk,” (May 2010), Working Paper.
- ADMATI, A. AND M. HELLWIG, *The Bankers’ New Clothes: What’s Wrong with Banking and What to Do about It* (Princeton University Press, 2013).
- ADRIAN, T. AND M. K. BRUNNERMEIER, “CoVaR,” (2008), Federal Reserve Bank of New York Working Paper no. 348.
- , “CoVaR,” (2011), Working Paper, NBER–17454.
- BASEL COMMITTEE ON BANKING SUPERVISION, “An assessment of the long-term economic impact of stronger capital and liquidity requirements,” Technical Report, Basel Committee on Banking Supervision, 2010.
- BENOIT, S., G. COLLETAZ, C. HURLIN AND C. PERIGNON, “A Theoretical and Empirical Comparison of Systemic Risk Measures,” (June 2013), HEC Paris Research Paper No. FIN-2014-1030.
- BERKOWITZ, J. AND J. O’BRIEN, “How Accurate Are Value-at-Risk Models at Commercial Banks?,” *Journal of Finance* 57 (2002), 977–987.
- BISIAS, D., M. FLOOD, A. LO AND S. VALAVANIS, “A Survey of Systemic Risk Analytics,” Technical Report, U.S. Department of Treasury, January 2012.
- BROWNLEES, C. T. AND R. ENGLE, “SRISK: A Conditional Capital Shortfall Measure of Systemic Risk,” (October 2015), Working Paper, NYU Stern School of Business.

- CROCKETT, A., “Marrying the micro- and macro-prudential dimensions of financial stability,” (2000), the General Manager of the Bank for International Settlements; <http://www.bis.org/review/rr000921b.pdf>.
- DANIELSSON, J., “The Emperor has no Clothes: Limits to Risk Modelling,” *Journal of Banking and Finance* 26 (2002), 1273–1296.
- , “Blame the Models,” *Journal of Financial Stability* 4 (2008), 321–328.
- DANIELSSON, J., K. JAMES, M. VALENZUELA AND I. ZER, “Model Risk of Risk Models,” Technical Report, Federal Reserve Board, 2015, Finance and Economics Discussion Series, 2014–34.
- DE RAMON, S., Z. ISCENKO, M. OSBORNE, M. STRAUGHAN AND P. ANDREWS, “Measuring the Impact of Prudential Standards On the Macroeconomy,” (2012), FSA Occasional Paper no. 42.
- EUROPEAN CENTRAL BANK, *Report on the First Two Years of the Macro-prudential Research Network*, ECB (October 2012), www.ecb.europa.eu.
- HENDRICKS, D., “Evaluation of Value-at-Risk Models Using Historical Data,” Technical Report, FRBNY Economic Policy Review, April 1996.
- IDIER, J., G. LAME AND J. MESONNIER, “How Useful is the Marginal Expected Shortfall for the Measurement of Systemic Exposure?: A Practical Experiment,” ECB Working Paper Series 1546, European Central Bank, 2011.
- KRAGH-SORENSEN, K., “Optimal capital adequacy ratios for Norwegian banks,” Staff memo, Norges Bank, 2012.
- LOPEZ-ESPINOSA, G., A. MORENO, A. RUBIA AND I. VALDERRAMA, “Short-term Wholesale Funding and Systemic Risk: A Global CoVaR Approach,” *Journal of Banking and Finance* 36 (2012), 3152–3162.

- MCCAULIFF, C., “Burdens of proof: degrees of belief, quantification of evidence, or constitutional guarantees?,” *Vanderbilt Law Review* 35 (1982), 1293–1335.
- POLITIS, N., D. AND P. ROMANO, J., “The Stationary Bootstrap,” Technical Report 91–3, Department of Statistics, Purdue University, 1991.
- ROSE, P. AND J. WALKER, C., “The Importance of Cost-Benefit Analysis in Financial Regulation,” (2013), report for U.S. Chamber of Commerce; Law and Capital Markets, Ohio State Public Law Working Paper No. 208.
- WEINSTEIN, J. AND I. DEWSBURY, “Comment on the Meaning of Proof Beyond a Reasonable Doubt,” *Law, Probability, and Risk* 5 (2007), 167–173.
- ZHANG, Q., F. VALLASCAS, K. KEASEY AND C. CAIF, “Are Market-Based Rankings of Global Systemically Important Financial Institutions Useful for Regulators?,” *Journal Money Credit and Banking* Forthcoming (2015).

Figure 1: Guilt Probabilities

In this figure we plot the probability that selected banks are guilty of creating systemic risk. We assume that a bank is guilty of creating systemic risk if it is in the upper decile (panels (a) and (c)) or upper quartile (panels (b) and (d)) of banks sorted by their systemic risk scores. In Panels (a) and (b), we measure the risk scores by the MES riskometer, whereas, we use ΔCoVaR riskometer in Panels (c) and (d). To calculate bank guilt probabilities given a sample S , we use a block bootstrap with 1,000 trials. In each trial we draw a bootstrap sample and calculate each bank's MES and ΔCoVaR systemic risk score and note each bank's position in the list of banks sorted (from highest to lowest) by their riskometer scores. The probability that a bank creates systemic risk is equal to the proportion of trials in which that bank is in the upper decile or quartile of banks. We use 5 year estimation windows, and our sample for a year Y consists of the 200 financial firms (classified on the basis of 4-digit SIC codes) with the highest market capitalization on the last trading day of year Y for which we have return observations for each trading day in years Y to $Y - 4$. Our sample period consists of the years 1974 to 2012. In panels (a) and (c) we assume that 20 banks create systemic risk, so we plot the guilt probabilities of the banks with the highest, 10th highest, and 20th highest median risk scores across the bootstrap trials. In panels (b) and (d) we assume 50 banks create systemic risk so we plot the guilt probabilities of the banks with the highest, 25th highest, and 50th highest median risk scores across the bootstrap trials.

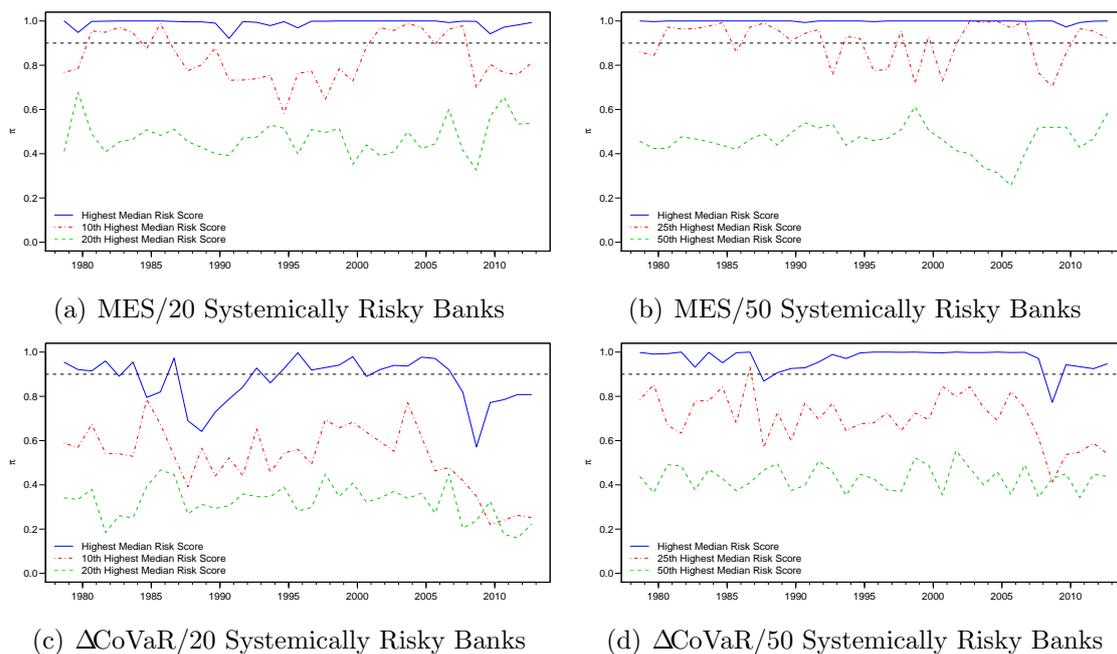
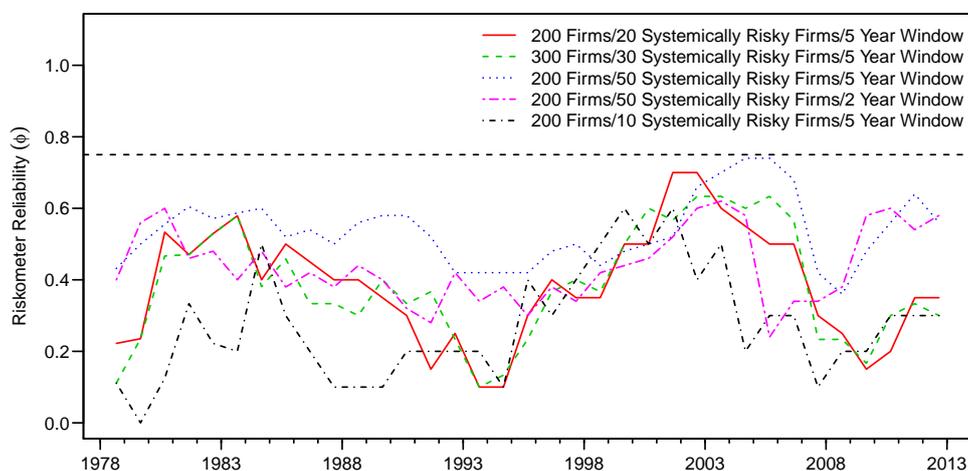
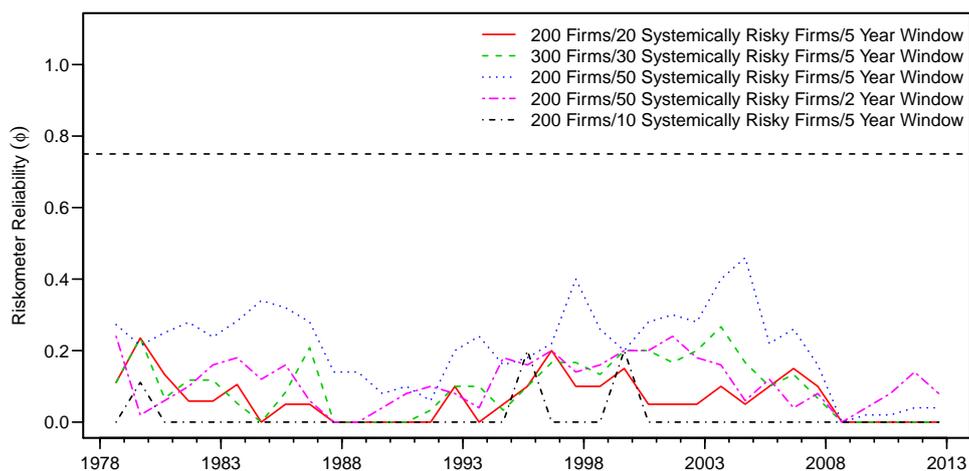


Figure 2: Riskometer Reliability

In this figure we plot the reliability of the MES and ΔCoVaR riskmeters for each sample year. A riskometer's reliability is equal to the proportion of systemically risky banks that can be shown to be guilty beyond a reasonable doubt of creating systemic risk on the basis of that riskometer's readings. We assume that a bank is guilty of creating systemic risk if it is the upper 5%, 10%, or 25% (depending upon the case) of banks sorted by their systemic risk scores as measured by a given riskometer. The probability that a bank is guilty of creating systemic risk given a sample S is equal to the probability that the bank is in the upper 5%, 10%, or 25% of banks sorted by their riskometer scores. We describe how we calculate these probabilities in Figure 1. A bank is guilty beyond a reasonable doubt of creating systemic risk if that probability is greater than or equal to 90%. Our sample for a year Y consists of the 200 or 300 financial firms (classified on the basis of 4-digit SIC codes) with the highest market capitalization on the last trading day of year Y for which we have return observations for each trading day in years Y to $Y - 4$. Our sample period consists of the years 1974 to 2012. We plot riskometer reliability for 5 cases: i) 200 Sample Firms/20 Systemically Risky Firms/5 Year Estimation Windows; ii) 300 Sample Firms/30 Systemically Risky Firms/5 Year Estimation Windows; iii) 200 Sample Firms/50 Systemically Risky Firms/5 Year Estimation Windows; iv) 200 Sample Firms/50 Systemically Risky Firms/2 Year Estimation Windows; and v) 200 Sample Firms/10 Systemically Risky Firms/5 Year Estimation Windows.



(a) Panel A: Riskometer Reliability– MES



(b) Panel B: Riskometer Reliability– ΔCoVaR

Table 1: Optimal Policy And Riskometer Reliability

In this table we calculate the regulator’s optimal choice of Targeted Macroprudential Policy (*TMP*) or Blanket Macroprudential Policy (*BMP*) as a function of the social cost of a 15% bank capital requirement (rows), the social cost of a financial crisis (columns), and riskometer reliability calculated as in (4) (cell values). The regulator chooses *TMP* if the reliability of its riskometer is greater than or equal to the cell value. We derive our estimates for the social cost of bank capital and the social cost of a financial crisis from Basel Committee on Banking Supervision (2010) report (BCBS10). We consider three cases for the social cost of bank capital. In the first, we set it equal to that assumed in BCBS10 (0.72% of output/year), and for robustness we also consider cases in which we increase and decrease that cost by 50%. We also consider three scenarios for the social cost of a crisis. The key factor that determines the impact of a crisis is the extent to which a crisis has persistent effects. BCBS10 considers three scenarios: moderate persistent effects, large persistent effects, and minimal persistent effects. In the moderate case, the expected social cost of a crisis is 2.64% of output/year, in the Large case the cost is 6.62% of output/year, and in the minimal case that cost is 0.8% of output/year. A riskometer’s reliability is equal to the proportion of systemically risky banks that the regulator can prove guilty beyond a reasonable doubt of creating systemic risk when using that riskometer, and we assume that 10% of banks create systemic risk.

Cost of Capital	Crisis Persistency Effects		
	Moderate	Large	Minimal
BCBS10	75%	90%	11%
50% Higher	62%	85%	0%
50% Lower	88%	95%	58%

Table 2: Riskometer Reliability

In this table we report the mean and standard deviation of the reliability of the MES and ΔCoVaR riskometers. A riskometer's reliability is equal to the proportion of systemically risky banks that can be shown to be guilty beyond a reasonable doubt of creating systemic risk on the basis of that riskometer's readings. We assume that a bank is guilty of creating systemic risk if it is the upper 5%, 10%, or 25% (depending upon the case) of banks sorted by their systemic risk score as measured by a given riskometer in the limiting case of an infinitely large sample. The probability that a bank is guilty of creating systemic risk given a sample S is equal to the probability that the bank is in the upper 5%, 10%, or 25% of banks sorted by their riskometer scores. We describe how we calculate these probabilities in Figure 1. A bank is guilty beyond a reasonable doubt of creating systemic risk if that probability is greater than or equal to 90%. Our sample for a year Y consists of the 200 or 300 financial firms (classified on the basis of 4-digit SIC codes) with the highest market capitalization on the last trading day of year Y for which we have return observations for each trading day in years Y to $Y - 4$. Our sample period consists of the years 1974 to 2012. We report riskometer reliability statistics for 5 cases: i) 200 Sample Firms/20 Systemically Risky Firms/5 Year Estimation Windows; ii) 300 Sample Firms/30 Systemically Risky Firms/5 Year Estimation Windows; iii) 200 Sample Firms/50 Systemically Risky Firms/5 Year Estimation Windows; iv) 200 Sample Firms/50 Systemically Risky Firms/2 Year Estimation Windows; and v) 200 Sample Firms/10 Systemically Risky Firms/5 Year Estimation Windows. The last two columns indicate whether one-sided mean comparison tests reject the null of $\phi = 75\%$ and $\phi = 62\%$ at the 1% level.

Specification	Riskometer	Mean	St. Dev.	Reject $\phi = 75\%$?	Reject $\phi = 62\%$?
200 Sample Firms/ 20 Systemically Risky Firms/ 5 Yr Estimation Windows	MES	38.63%	15.82%	Yes	Yes
	ΔCoVaR	6.44%	6.21%	Yes	Yes
300 Sample Firms/ 30 Systemically Risky Firms/ 5 Yr Estimation Windows	MES	38.38%	15.81%	Yes	Yes
	ΔCoVaR	9.50%	8.07%	Yes	Yes
200 Sample Firms/ 50 Systemically Risky Firms/ 5 Yr Estimation Windows	MES	53.69%	9.54%	Yes	Yes
	ΔCoVaR	20.96%	11.51%	Yes	Yes
200 Sample Firms/ 50 Systemically Risky Firms/ 2 Yr Estimation Windows	MES	43.89%	10.03%	Yes	Yes
	ΔCoVaR	11.53%	7.03%	Yes	Yes
200 Sample Firms/ 10 Systemically Risky Firms/ 5 Yr Estimation Windows	MES	27.40%	15.34%	Yes	Yes
	ΔCoVaR	1.46%	5.00%	Yes	Yes