Dealing with Systemic Risk When We Measure It Badly:

A Minority Report*

Jon Danielsson
London School of Economics
j.danielsson@lse.ac.uk

Kevin R. James
London School of Economics
k.james1@lse.ac.uk

Marcela Valenzuela
m.a.valenzuela@lse.ac.uk

Ilknur Zer
London School of Economics
i.zer@lse.ac.uk

First Draft: January, 2012

This Draft: August, 2012

*We thank the AXA Research Fund for financial support provided via the LSE Financial Market Group's program on risk management and financial regulation. We thank Bob DeYoung, Jean Helwege, Kostas Tsatsorinis and seminar participants at the Bank of England, the Bank of France, the Bank of Canada, the Financial Services Authority, the Reserve Bank of India, the Hong Kong Monetary Authority, the CCBS/LSE conference "Measuring Systemic Risk and Issues for Macroprudential Regulation" and the SUERF/Bundesbank Conference "The ESRB at 1" for helpful comments and discussions.

Dealing with Systemic Risk When We Measure It Badly:

A Minority Report

Abstract: While an omniscient regulator would base a bank's capital requirement upon its contribution to systemic risk, we show that a regulator who measures a bank's contribution to systemic risk badly will find it optimal to use a simple leverage ratio instead. We empirically analyze the performance of leading risk measurement methods and find that they are incapable of providing either precise estimates of an individual bank's contribution to systemic risk or reliable rankings of banks by the amount of systemic risk they create. We conclude that a simple leverage ratio dominates a policy of systemic risk based capital requirements. JEL Codes: D81, G28, G32, G38.

In the film *Minority Report* the PreCrime unit of the DC Police uses psychic powers to detect people who are about to commit crimes and then locks them up before they do so. Macro-prudential policy 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 pursuing high-risk strategies that may bring about such a crisis if they become stressed or fail, and the task of the PreCrime units of central banks/regulators is to detect these banks in advance and to require them to increase the proportion of costly capital in their balance sheets so as to enable them to survive the future shocks that would otherwise tip them over into the stressed state.[1] 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.[2] In the movie the government shuts down PreCrime as it turns out that the psychics' visions are not precise enough to justify locking people up on the basis of the information they provide. In this paper we examine the merits of a policy that gives central banks/regulators the authority to treat banks in a radically different manner on the basis of the systemic risk readings that their riskometers provide.

To keep our analysis focused on the merits of macroprudential policy in general rather than on the accuracy of any given set of riskometers in particular, we (optimistically) assume that each riskometer we examine implements a true measure of the amount of systemic risk a bank

---

[1] Our analysis here focuses upon the "cross-section" dimension of systemic risk (i.e., how risk varies across firms) rather than the "time-series" dimension (i.e., the evolution of risk over time for the financial system as a whole).

[2] We owe the term riskometer to Danielsson (2009).

creates (riskometer errors that arise from a flawed measure of systemic risk can be fixed by revising an existing measure or by devising a new one).[3] We instead focus on a common point of failure for all riskometers, viz., that any empirical implementation of even a systemic risk measure that truly captures the amount of systemic risk a bank creates will inevitably lead to risk readings that are (to at least some) degree imprecise.

Our analysis proceeds in two steps. First, we show theoretically that a policy of treating all banks the same by requiring that they abide by a simple leverage ratio dominates the macroprudential (MP) policy of basing bank capital requirements upon the regulator's endowed riskometer if that riskometer is sufficiently imprecise.[4] Second, we examine the empirical performance of three leading riskometers. We find that each of the riskometers we investigate provides an extremely imprecise reading of the systemic risk created by individual banks and a completely unreliable ranking of banks by the amount of systemic risk they create. We therefore conclude that, while systemic risk based capital requirements (or other regulatory interventions) may make sense for an omniscient being who can simply observe the risk fundamentals, they probably do not make sense for regulators who must base their decisions upon highly imprecise estimates of those fundamentals.[5]

---

[3] And indeed, there is now an enormous and still burgeoning literature on devising and improving systemic risk measures. See Bisias et. al.'s (2012) survey for an overview of the state of play.

[4] See D'Hulster (2009) for background on the leverage ratio.

[5] Hildebrand (2008) argues that flaws in risk models and/or gaming of the capital rules by banks may leave banks with too little capital and hence sees a role for the leverage ratio as a backstop for a MP policy.

We analyze the importance of riskometer imprecision in the context of a model in which a regulator acts to minimize the probability of a systemic crisis by setting each bank's capital requirement at either a high or a low level. The banking system consists of type H banks that always pursue a high risk strategy and type D banks that pursue a high risk (low risk) strategy if they operate with high levels (low levels) of capital. Banks pursuing a high risk strategy with low capital create the greatest risk of a systemic crisis, banks pursuing a high risk strategy with high levels of capital create a systemic crisis with a lower probability, and banks pursuing the low risk strategy never create a systemic crisis.

The regulator has three options: i) he can require all banks to operate with high capital (the HighCap policy); ii) he can require all banks to operate with low capital (the LowCap policy); or iii) he can use his imperfect but better than random riskometer to get a reading on a bank's type and base his capital decision upon that reading (the MP policy). All three policies lead to costly errors. Under the HighCap policy type H banks operate with the optimal amount of capital (lowering systemic risk), but the high capital requirement leads type D banks to pursue the high risk strategy (increasing systemic risk). Under the LowCap policy, type D banks can operate with low levels of capital and so pursue the low risk strategy (lowering systemic risk), but high risk banks can also operate with low levels of capital (increasing systemic risk). The regulator therefore eliminates the policy option that leads to the more expensive error (as a function of the model parameters). Under the MP policy, the regulator will make fewer errors as he uses his riskometer to get information on each bank's type, but he will make both types of error (allowing some H banks to operate with low capital and forcing some D banks into the high risk strategy). Hence, his errors will on average be more expensive. It follows that a regulator will find the MP policy optimal only if his riskometer is precise enough to compensate for the more costly errors he makes while using it–a riskometer that is merely "better than nothing" will not suffice.

With this result in mind, we empirically evaluate the performance of the Value-at-Risk (VaR) riskometer, the Marginal Expected Shortfall (MES) riskometer devised by Acharya et. al. (2010), and the CoVaR riskometer devised by Adrian and Brunnermeier (2008). We use Acharya et. al.'s (2010) sample of large US financial firms to do so, and we examine the pre-crisis (2003–2006) and crisis (2007–2010) periods separately. We assess riskometer performance along three dimensions: i) Absolute Imprecision; ii) Relative Imprecision; and iii) Rank Imprecision. A riskometer's Absolute Imprecision is low if it can pin down the level of a bank's risk reading to within a narrow numerical range. A riskometer's Relative Imprecision is low if that riskometer can pin down a bank's risk reading to within a narrow range of the distribution of risk readings for the sample banks. A riskometer's Rank Imprecision is low if it can reliably order a sample of banks by the level of systemic risk they create (as defined by that riskometer).

We evaluate riskometer precision by using a stationary bootstrap (Politis and Romano (1991)) to construct 10,000 trial return histories for each sample period. For each trial we compute each riskometer's reading for each bank, and we set the plausible range of a given riskometer's reading for a given bank equal to the middle 95% of the that riskometer's trial readings for that bank. We measure a riskometer's Absolute Imprecision by the ratio of its highest plausible reading to its lowest plausible reading, and we find that this ratio averages between 1.5 (for VaR in the pre-crisis period) and 3 (CoVaR in the crisis period). We measure Relative Imprecision by comparing the plausible range of a given bank's risk reading to the span of the distribution of the point-estimate risk readings for the sample banks (in percentile space), and we find that that this range covers an average of between 40 (for VaR in the pre-crisis period) and 88 (for CoVaR in the crisis period) percentiles of that distribution.[6] That is, the imprecision of a riskometer reading for the typical bank is so large that one cannot generally rule out the possibility that a bank with a relatively low risk reading is in fact among the riskiest banks. We measure a riskometer's Rank Imprecision by computing the Spearman rank order correlation coefficient between banks ordered by riskometer reading for two trials selected at random, and we find that it is not statistically significantly different from 0. In short, we find that the riskometers we examine are not precise enough to support the demands that a macroprudential policy will place upon them even in the best-case scenario that we consider here.

While we assume that each riskometer implements a true measure of systemic risk and analyze the importance of riskometer precision, much of the research on this topic ignores the

[6] To illustrate, suppose for a given riskometer that the point-estimate readings of the sample were uniformly distribution on {10, 20} and that the plausible range of bank B's reading was {11, 19}. Then, our measure of B's Relative Imprecision would be 80 as its plausible range spans 80 percentiles of the point-estimate distribution.

role of precision and instead focuses almost exclusively upon devising the correct measure of systemic risk (Danielsson (2002, 2008) is an exception). One strand of this research seeks to derive a measure of the amount of systemic risk a bank creates from a micro-founded model of complex optimizing banks in a complex financial system (see, for example, Elsinger, Lehar, and Summer (2002), Goodhart, Sunirand, and Tsomocos (2004), and Danielsson, Shin, and Zigrand (2011)). However, while this strand of research does provide important qualitative insights into the nature of systemic risk, it has yet to yield a theoretically sound and empirically implementable measure of a bank's contribution to systemic risk.

In the absence of such a model, researchers devising practicable riskometers generally take the shock process (the initial shock plus any endogenous risk driven amplification/attenuation of that shock) as given and work with observable data on bank and financial system performance– with the idea being that performance data incorporate the impact of shocks that occur. While systemic risk measurement along these lines is a very active area of research (e.g., Segoviano and Goodhart (2009), Tarashev, Borio, and Tsatsaronis (2010), and Zhou (2010)), we focus here upon VaR, SES, and CoVaR, which are in our view the three most developed riskometers both theoretically and empirically.

The quest for the ideal systemic risk measure has led to both new risk measures and to constant tinkering with existing measures. Danielsson (2008) and Danielsson et. al. (2012) examine 6 methods of estimating VaR, Adrian and Brunnermeier (2008) discuss several implementations of their CoVaR measure, and Brownlees and Engle (2011) continue to develop and refine MES. So, if we were critiquing existing riskometers on the basis that they did not provide true measures of systemic risk, we would definitely be aiming at a moving target. But, that is not our critique. We assume that the quest of the true systemic measure is a success and are instead asking if it is possible to estimate that measure with the precision that macroprudential policy requires. For this purpose, the simplest version of each of the three riskometers we examine offers the best test-case for our analysis, and so we use that version of each riskometer in the analysis below.

Our analysis of riskometer imprecision builds upon Pástor and Stambaugh's (2012) analysis of the importance of incorporating estimation risk into an agent's optimization problem and Danielsson's (2002, 2008) critique of VaR. Pástor and Stambaugh show that even if stocks are fundamentally less volatile over the long run due to mean reversion, stocks are in fact more volatile over the long run from the perspective of investors because investors must imprecisely estimate rather than observe the fundamentals of the return process. Thus, they find that imprecision can have a substantial impact upon an investor's strategy. Danielsson (2002) explores the statistical assumptions behind the VaR measure, and Danielsson (2008) shows that the VaR riskometer performs poorly even in the relatively straightforward case of a $1000 investment in IBM. So, while few would disagree with the observation that riskometers will inevitably be imprecise to at least some extent, we show here that the imprecision of one's riskometer matters as it affects the optimal policy choice.

I. Optimal Policy When Using An Imprecise Riskometer

A. Set - Up and Assumptions

We examine the importance of riskometer precision by analyzing the case of a regulator who sets bank capital requirements so as to maximize the probability that the financial system remains stable. The banking system consists of two types of (large) banks, type *H* and type *D*, and the regulator can set a bank's capital requirement at either a high level or a low level.[7] An *H* banks always pursue a high risk strategy, and a *D* bank pursues the the low risk strategy unless the regulator requires it to operate with a high level of capital, in which case it pursues the high risk strategy.[8] A bank pursuing a high risk strategy with low capital sets in train a series of events that creates a crisis with a probability of χ_{HL}, and a bank pursuing a high risk strategy with high capital does so with probability χ_{HH}, with $\chi_{HL} > \chi_{HH}$. A bank pursuing the low risk strategy never creates a crisis. The banking system consists of Q_H (Q_D) banks of type *H* (*D*), and to simplify the exposition we further assume that the number of banks is large enough to be treated as a continuum.

---

[7] We assume that small banks do not create any systemic risk externalities and so do not include them in the analysis.

[8] We implicitly assume here that requiring a bank to operate with a high capital requirement increases its cost, and these higher costs make the low risk strategy unprofitable (as in, for example, Hellmann, Murdock, and Stiglitz (2000)).

We model this problem as a two period game. In the first period all banks operate with low capital, and so *D* (*H*) banks choose the low-risk (high-risk) strategy. In the second period the regulator sets sets each bank's capital requirement and then a crisis occurs with a probability that is a function of the capital requirements he sets. We impose the constraint (consistent with actual behavior) that the regulator cannot act in a capricious and arbitrary fashion. We operationalize this constraint by assuming that the regulator must: i) treat every bank that looks identical to him in an identical manner; and ii) require all banks that look identical to him to hold the same level of capital.

The regulator knows all of the parameters of the model, but he cannot observe a bank's type or its strategy. The only way that he can get information on a bank's type is to use his endowed riskometer to measure the risk of the strategy the bank pursued in period 1. The riskometer will give one reading of either "High Risk" or "Low Risk" per bank, and that reading will be correct (incorrect) with probability π ($1-\pi$), with $0.5 < \pi \leq 1$.[9] That is, we assume that the riskometer is always better than random.

B. Optimal Policy

Since the requirement that the regulator treat all banks that look identical to him in an identical manner rules out any policy that involves a mixed strategy, the regulator has the

[9] Intuitively, one could think that all banks try to appear to be low risk, and that the regulator can use his riskometer to (imperfectly) sort the actual high risk banks from the actual low risk banks. We have assumed for simplicity that the probability that the riskometer provides the correct reading is the same for both high risk and low risk banks, but allowing the probability to vary across bank types would not fundamentally alter the results below.

following policy options available to him. He can eschew his riskometer (in which case all banks do look the same to him) and simply require that all banks operate with a uniform level of capital. Capital can be set at either a low level (the LowCap policy) or at a high level (the HighCap policy). Alternatively, the regulator can choose to use his riskometer (the MP policy), in which case he must use it for all banks and he must require a bank with a "High Risk" ("Low Risk") reading to operate with a high level (low level) of capital in period 2. Consider the costs of each policy in turn.

If the regulator chooses the LowCap policy, then all banks operate with a low level of capital in period 2. It follows that *D* banks choose the low risk strategy and *H* banks the high risk. Hence, the probability of a stability is under this policy is θ_{LowCap}, with:

$$\theta_{LowCap} = \left(1 - \chi_{HL}\right)^{Q_H}. \tag{1}$$

If the regulator chooses the HighCap policy, then all banks operate with high levels of capital. Hence, both *D* and *H* banks choose the high-risk strategy, implying that the probability of banking system stability under this policy is $\theta_{HighCap}$, with:

$$\theta_{HighCap} = \left(1 - \chi_{HH}\right)^{Q_D + Q_H}. \tag{2}$$

If the regulator chooses the MP policy, then he treats each bank as if it is of the type indicated by his riskometer's reading. Consequently, the number of *H* banks operating with a high level of capital is equal to the number of *H* banks (Q_H) times the probability that the regulator's riskometer picks correctly identifies them (π), which is $\pi\, Q_H$. The number of H banks operating with a low level of capital is then $Q_H\,(1-\pi)$. The total number D banks that operate with low capital is equal to the number of D banks that the riskometer classifies correctly, which is $\pi\, Q_D$. The number of D banks that the regulator forces to operate with high capital (and so that pursue the high risk strategy) equals the number of D banks that the riskometer classifies incorrectly, which is $(1-\pi)\, Q_D$. Hence, the probability of banking stability under the MP policy is θ_{MP}, with:

$$\theta_{MP} = (1-\chi_{HL})^{Q_H(1-\pi)}(1-\chi_{HH})^{Q_H\pi+Q_D(1-\pi)} \,. \tag{3}$$

When $\pi = 0.5$,

$$\theta_{MP}\big|_{\pi=0.5} = (1-\chi_{HL})^{\frac{Q_H}{2}}(1-\chi_{HH})^{\frac{Q_H+Q_D}{2}} \,, \tag{4}$$

and when $\pi = 1$, we have

$$\theta_{MP}\big|_{\pi=1} = (1-\chi_{HH})^{Q_H} \,. \tag{5}$$

Obviously, $\theta_{MP}|_{\pi=1} > \theta_{MP}|_{\pi=1/2}$ and $\dfrac{\partial\theta_{MP}}{\partial\pi} > 0$.

Let us now consider the relationship between the probability of banking system stability across the regulator's policy options.

LEMMA 1. The regulator chooses the MP policy if his riskometer can perfectly identify a bank's type.

Proof. Under the MP policy with a perfect riskometer, all *D* banks operate with a low level of capital and so pursue the low risk strategy (and thereby create a crisis with probability 0), and all *H* banks operate with a high level of capital and so each cause a crisis with probability χ_{HH}. Under the LowCap policy all *D* banks operate with a low level of capital, but all *H* banks also operate with a low level of capital and so each cause a crisis with probability χ_{HL}, $\chi_{HL} > \chi_{HH}$. Obviously, the regulator prefers the MP policy in this case. Under the HighCap policy, all *H* banks operate with a high level of capital and so also each cause a crisis with probability χ_{HH} as under the MP policy. However, the *D* banks must also operate with a high level of capital and so also each cause a crisis with probability χ_{HH} instead of 0. Hence, the regulator prefers the MP policy with a perfect riskometer to the HighCap policy as well.

∎

LEMMA 2. The regulator chooses one of the leverage ratio policies if his riskometer works poorly.

Proof. Suppose that $\theta_{MP}|_{\pi = 1/2} > \theta_{HighCap}$. In then follows from equations 2 and 4 that

$$
\begin{aligned}
&(1-\chi_{HL})^{\frac{Q_H}{2}}(1-\chi_{HH})^{\frac{Q_H+Q_D}{2}} > (1-\chi_{HH})^{Q_D+Q_H} \\
&\Rightarrow (1-\chi_{HL})^{\frac{Q_H}{2}} > (1-\chi_{HH})^{\frac{Q_H+Q_D}{2}}
\end{aligned}, \tag{6}
$$

Now suppose also that $\theta_{MP}|_{\pi = 1/2} > \theta_{LowCap}$. In then follows from equations 1 and 4 that

$$
\begin{aligned}
&(1-\chi_{HL})^{\frac{Q_H}{2}}(1-\chi_{HH})^{\frac{Q_H+Q_D}{2}} > (1-\chi_{HL})^{Q_H} \\
&\Rightarrow (1-\chi_{HH})^{\frac{Q_H+Q_D}{2}} > (1-\chi_{HL})^{\frac{Q_H}{2}}
\end{aligned}, \tag{7}
$$

Equations 6 and 7 are inconsistent, and thus it is impossible for $\theta_{MP}|_{\pi = 1/2}$ to be greater than both θ_{LowCap} and $\theta_{HighCap}$ at the same time. The regulator therefore prefers one of the leverage ratio policies to the MP policy if his riskometer works poorly.

∎

Having established the key properties of the regulatory policy cost functions, we can now derive our main result.

PROPOSITION 1. The regulator prefers the leverage ratio to the MP policy if his riskometer is sufficiently imprecise.

Proof. We know from Lemmas 1 and 2 that

$$\theta_{MP}\Big|_{\pi=0.5} < \text{Max}\left[\left\{\theta_{HighCap}, \theta_{LowCap}\right\}\right] < \theta_{MP}\Big|_{\pi=1}, \tag{8}$$

and from equation 3 that $\dfrac{\partial \theta_{MP}}{\partial \pi} > 0$. It follows that there exists a π^* such that

$\theta_{MP}\Big|_{\pi=\pi^*} = \text{Max}\left[\left\{\theta_{LowCap}, \theta_{HighCap}\right\}\right]$. The regulator therefore prefers either the HighCap or the LowCap policy if $\pi < \pi^*$ and the MP policy if $\pi > \pi^*$.

∎

Our analysis demonstrates that the regulator's optimal policy choice depends upon the quality of his riskometer. If the regulator finds himself in possession of a perfect riskometer, then a policy of setting capital requirements on the basis of that riskometer's readings is indeed optimal. However, as the regulator's riskometer becomes less precise, there comes a point where (poorly measured) systemic risk based capital requirements become dominated by other policies. So, what riskometers do we have and just how precise are they?

II. Riskometer Design: VaR, MES, and CoVaR

A riskometer $\Psi\gamma$ is an algorithm that implements an approach γ for measuring a bank's contribution to systemic risk. Provided with the data concerning a given bank *B* that the riskometer requires, it provides a quantitative reading of *B*'s contribution to systemic risk, denoted by $\Psi\gamma_{,B}$.

A. The VaR Riskometer

If one assumes that the marginal increase in the probability of a financial crisis due to a bank entering into a stressed state is independent of the state of the financial system, then a sensible measure of the systemic risk that a given bank creates is just the probability that the bank enters into a stressed state. The probability that a bank enters into a stressed state increases with the magnitude of a negative shock to its value (all else equal). Hence, banks that experience larger negative shocks create more systemic risk. If a bank does experience larger negative shocks, then one might expect its return distribution to have a longer left tail. A conceptually simple (if manipulatable) measure of the length of the tail is the position of a given extreme quantile of the return distribution. This line of thought leads to the VaR riskometer, where a bank's systemic risk measure is set equal to the value of a given extreme quantile of its return distribution.

To empirically implement the VaR riskometer, one must choose a return quantile and a method of estimating the location of that quantile. To begin with the quantile decision, the choice of exactly which quantile to use is of course arbitrary. The trade-off when selecting the quantile is that while a quantile further out in the tail provides a better estimate of the return consequences of a plausible worst case event, the further one goes out into the tail of the distribution, the fewer observations one has with which to estimate the location of the quantile for which one is aiming. The standard solution to this trade-off is to set the VaR quantile at either the 1th percentile or the 5th percentile of the return distribution. We use the 1th percentile here.[10]

There exist a wide variety of methods that one can use to estimate the location of the VaR quantile one selects. To avoid needless complexity, we simply take as our estimate of the 1th percentile of the return distribution the 1th percentile value of the return series we examine (that is, we use the Historical Simulation method of estimating VaR).[11]

We denote this quantile plus method specification of VaR as Ψ_{VaR}. So, Ψ_{VaR} for a bank B with a return series $\{R_{B,1}, R_{B,2} \ldots R_{B,T}\}$ is $\Psi_{VaR,B}$, with

$$\Psi_{VaR,B} = \text{Quantile}\left[\left\{R_{B,1}, R_{B,1} \ldots R_{B,T}\right\}, 0.01\right]. \tag{9}$$

We implement this measure using daily market returns for sample banks.

B. The MES Riskometer

The MES riskometer takes its inspiration from the idea that, as Andrew Crocket (2000) put it, "the financial system is a system". That is, since the real economy requires the services that the financial system provides in order to function properly, a bank creates a risk to the overall economy (i.e., systemic risk) to the extent that it contributes to stress in the financial system as a whole. Building upon this idea, Acharya et. al. (2010) construct a theory yielding the result that

[10] We obtain the same results when using the 5th percentile.

[11] Danielsson et. al. (2012) show that all standard methods of estimating VaR are highly imprecise. Since there is no general agreement on which method is best, adding in the imprecision that arises from the fact that one can estimate VaRs in different ways would strengthen our empirical results below.

a bank creates systemic risk if it performs poorly at the same time as the economy as a whole performs poorly. This conclusion in turn leads to the MES riskometer, which is defined as a bank's expected return conditional upon the market performing poorly.[12] Acharya et. al. (2010) operationalize MES as follows: a bank *B*'s MES is equal to its average return on days when the market return R_{Market} is at or below its 5th percentile value for the sample period ($VaR^{5\%}_{Market}$), determining the 5th percentile value of the market return series using the Historical Simulation method as above. Denoting this specification of MES as Ψ_{MES}, a bank *B*'s Ψ_{MES} is then

$$\Psi_{MES,B} = \text{Mean}\left[\left\{R_{B,z_1}, R_{B,z_2} ... R_{B,z_W}\right\}\right] , \tag{10}$$

where $\{z_1, z_2 ... z_W\}$ is the set of days such that $R_{Market,z} < VaR^{5\%}_{Market}$. Following Acharya et. al. (2010), we compute this measure using daily returns and we use the return on the value-weighted CRSP index as our measure of R_{Market}.[13]

C. The CoVaR Riskometer

CoVaR shares with MES the idea that a bank creates systemic stress when stress at the bank coincides with stress in the financial system as a whole, and it shares with VaR the idea that the

---

[12] MES is the key building block of Acharya et. al.'s (2010) full systemic risk measure SES, which also includes a leverage measure. Given that SES builds upon MES, any imprecision in MES will necessarily affect SES. Thus, we concentrate our analysis here upon the MES part of SES.

[13] One aspect of MES that is worth noting is that, as Acharya et. al. (2010) acknowledge, the 5th percentile threshold is not really extreme enough to properly serve as a defining bound for "market stress". However, Acharya et. al. (2010) argue that this bound is far enough out into the tail of the market return distribution to lead to MES readings that are proportional to what MES readings would be if there was enough data to estimate them with a more realistically extreme bound to define episodes of market stress.

location of the 1th percentile of a return distribution provides a good measure of risk. Combining these two ideas, Adrian and Brunnermeier (2008) posit that a bank B creates systemic risk if risk in the financial system as a whole increases as stress at the bank increases. Their measure of risk in the financial system is $VaR_{FinSys}^{1\%}$, which is the 1% VaR of the portfolio of financial firms.

Assume that B is under stress when $R_B = \Psi_{VaR,B}$ and not under stress when R_B = Median[$\boldsymbol{R}_B$], where $\boldsymbol{R}_B$ is the return series for B. We then get the CoVaR riskometer's measure of systemic risk at bank B $\Psi_{CoVaR,B}$, with

$$\Psi_{CoVaR,B} = VaR_{FinSys}^{1\%}\Big|_{R_B=\Psi_{VaR,B}} - VaR_{FinSys}^{1\%}\Big|_{R_B=\text{Median}[\boldsymbol{R}_B]} \ . \tag{11}$$

Following Adrian and Brunnermeier (2008), we estimate $VaR_{FinSys}^{1\%}\Big|_{R_B}$ with

$$VaR_{FinSys}^{1\%}\Big|_{R_B} = \alpha + \beta R_B + \varepsilon \ . \tag{12}$$

using a quantile regression.[14] We then calculate $\Psi_{CoVaR,B}$ with the estimated parameters from equation 12.

There are obviously many ways of implementing this general CoVaR idea, and Adrian and Brunnermeier (2008) discuss several specifications in their paper. In particular, one must decide whether to measure returns by common stock returns or by changes in the market value of bank

[14] We thank Johannes Ludsteck for sharing and updating the code he developed to estimate quantile regressions in Mathematica.

assets. Given that Adrian and Brunnermeier (2008) report that both methods lead to similar results in practice, we choose to measure returns with (easily observable) stock returns rather than by (difficult to infer) changes in the market value of bank assets so as to avoid a very significant errors in variable problem.[15] While Adrian and Brunnermeier estimate Ψ_{CoVaR} with weekly returns, we use daily returns in our analysis below to be consistent with our analysis of Ψ_{VaR} and Ψ_{MES}.[16] We set R_{FinSys} equal to the return on the Fama-French Finance Industry portfolio.

IV. Riskometer Performance

A. Gauging Performance: Criteria and Method

Since the goal of the MP policy is to make a bank's capital requirement a function of the amount of systemic risk that the bank creates, a regulator pursuing a MP policy needs a riskometer that enables him to precisely measure the amount of systemic risk that each individual bank creates. We gauge riskometer precision along three dimensions, Absolute, Relative, and Rank. A riskometer is precise in the Absolute dimension if it can pin down the numerical measure of the amount of systemic risk a bank creates (as defined by that riskometer) to within a

---

[15] The errors in variable problem can arise through several channels. Among these channels are the following. First, banks report asset values on a quarterly basis. One is therefore left with either very few return observations or one must use some ad hoc interpolation rule to trace the path of asset values over the quarter. Second, a considerable proportion of assets are reported at book rather than market value, so one must use some ad hoc method to transform book values to market values. Third, reported asset values are highly vulnerable to accounting manipulation. Fourth, publicly reported asset values do not include (at times substantial) off-balance sheet items and so may present a very incomplete picture of true assets.

[16] We repeated the analysis below using weekly returns and found that doing so did not alter our results.

narrow range. A riskometer is precise in the Relative dimension if it can pin down the amount of risk that a given bank creates to within a narrow range relative to the amount of systemic risk that other bank creates. A riskometer provides Rank precision if it can reliably order a sample of banks by the amount of systemic risk they create.

As there is no simple analytical formula available to compute the plausible ranges of and the correlations between the bank riskometer readings that we need to measure riskometer precision, we examine riskometer imprecision by using a bootstrap. For each sample period, we generate 10,000 return histories ("trials"). Since our return data incorporates both cross-sectional and time series relationships, our trials must capture these aspects of the data. We therefore generate the trials as follows. To capture the cross-sectional relationships between R_{Market}, R_{FinSys}, and individual bank returns, we construct our trials by drawing sample days. So, if bank *B*'s return has a cross-sectional relationship with R_{Market}, that relationship will appear in the trials as we will have R_{Market} and R_B observations for the same days. To capture the time series relationships in the data, we select the days we include in a trial using a stationary bootstrap (Politis and Romano (1991)). In this technique one initially draws an observation day at random from the sample and then includes a random number of following days. The number of additional days that one includes is given by a draw from a geometric distribution with parameter g, with $g = T^{-1/3}$, where T is the number of days in the sample period. One then repeats this process until the number of days in the trial equals T. Each trial τ produces one reading for each riskometer γ and each bank

B, denoted by $\Psi_{\gamma,B,\tau}$. The list of trial risk readings for a given bank/riskometer pair is then $\Psi_{\gamma,B,Boot}$, with

$$\Psi_{\gamma,B,Boot} = \left\{\Psi_{\gamma,B,1}, \ldots \Psi_{\gamma,B,10000}\right\} \tag{13}$$

We measure a riskometer's Absolute Imprecision for a bank B by the ratio of its highest plausible risk reading to its lowest plausible risk reading, and we set that plausible range equal to the middle 95% of the distribution of riskometer readings generated by the bootstrap. Labeling this ratio for riskometer γ and bank B as $\mathrm{AbP}_{\gamma,B}$, it follows that

$$\mathrm{AbP}_{\gamma,B} = \frac{\mathrm{Quantile}[\Psi_{\gamma,B,Boot}, 0.975]}{\mathrm{Quantile}[\Psi_{\gamma,B,Boot}, 0.025]} \tag{14}$$

We measure a riskometer's Relative Imprecision for a bank B by measuring the plausible range of its riskometer reading relative to the distribution of the point-estimate risk readings for the sample banks. We label this measure as $\mathrm{RelP}_{\gamma,B}$. We calculate this measure as follows. Denote the distribution of the point-estimate risk readings produced by riskometer γ by Ω_γ. Then,

$$\mathrm{RelP}_{\gamma,B} = 100 * (\mathrm{CDF}\left[\Omega_\gamma, \mathrm{Quantile}[\Psi_{\gamma,B,Boot}, 0.975]\right] - \mathrm{CDF}\left[\Omega_\gamma, \mathrm{Quantile}[\Psi_{\gamma,B,Boot}, 0.025]\right]) \tag{15}$$

To illustrate, suppose that the risk readings generated by riskometer γ applied to a sample of banks were uniformly distributed on the interval $\{10, 20\}$, and that the confidence interval for bank B's reading was {11, 19}. Then, $\text{RelP}_{\gamma,B}$ equals 100 * (CDF[Ω_γ, 19] – CDF[Ω_γ, 11]) = 100 * (0.9 – 0.1) = 80.

To measure Rank Imprecision for a given riskometer we calculate the Spearman rank order correlation coefficient ρ between trials selected at random. If banks tend to appear in the same order when ranked on the basis of riskometer readings from trial to trial (even if the value of those readings varies a great deal), then ρ will be positive. We obtain the distribution of ρ by computing it for 10,000 pairs of trials, with each trail in the pair selected at random. We sort the set of ρ's we obtain by value and take the 99% confidence interval to be equal to the bounds on the middle 99% of that distribution.

B. Data

Systemic risk concerns arise in connection with large financial firms. We therefore construct our sample by beginning with the large financial firm sample used in Acharya et. al. (2010). This sample consists of the 101 financial firms with a market cap in excess of $5 billion as of June 2007. We examine a pre-crisis sample period consisting of the years 2003–2006 and a crisis sample period consisting of the years 2007–2010. We include a bank in a sample period if we have a return observation for that bank for at least 75% of sample days. Our 2003–2006 sample consists of 92 firms, and our 2007–2010 sample consists of 77 firms. We obtain our firm returns from CRSP.

C. Absolute Imprecision

We report the distribution of the AbP for the sample banks for each riskometer and each sample period in Table I. The best performing riskometer is VaR, which produces an average AbP of 1.5 for the 2003/2006 sample period and of 1.8 for the 2007/2010 sample period. If the amount of capital the regulator requires a bank to hold is a linear function of its risk reading, then the imprecision in the VaR readings could lead to the amount of capital a bank is required to hold to vary by 50% to 80%. So, even the best performing riskometer does not perform very well.

The VaR riskometer does not derive from a model of how systemic risk arises. The MES and CoVaR riskometers, on the other hand, do rest upon elegant and theoretically coherent foundations. However, a theoretically sound riskometer is probably a more complicated riskometer, and a more complicated riskometer will necessarily place greater demands upon the data. Hence, we were not surprised to see that Absolute Imprecision for both MES and CoVaR is much higher than it is for VaR.

In the case of MES, we find that the average AbP ratio is about 2 for both sample periods, and in the case of CoVaR the average AbP ratio is about 3 for both sample periods. That is, if a bank's capital requirement is a linear function of its risk reading, then the capital requirement a regulator using the MES or CoVaR riskometer imposes upon a bank could vary by a factor of 2 to 3 purely due to the difficulty the regulator has in estimating the parameters of his risk model.

D. Relative Imprecision

We report the distribution of RelP for the sample banks for each riskometer and each sample period in Table II. The best performing riskometer is, once again, VaR (average RelP of about 40), followed by MES (average RelP of about 50) and then CoVaR (average RelP of 70 to 80).

In the case of VaR and MES, then, the imprecision of the riskometer readings means that the regulator cannot generally be confident that a bank with a riskometer reading that places it well into the lower half of the risk reading distribution for sample banks does not in fact belong in the upper half of that distribution. In the case of CoVaR, the confidence interval for the typical bank's risk reading spans fully 70 to 80 percentiles of the risk-reading distribution. Here the regulator could not be confident that a bank with a risk reading placing it among the least risky banks in the sample was not in fact one of the riskiest banks in the sample.

E. Rank Precision

Our analysis of Absolute and Relative Imprecision demonstrates that individual bank riskometer readings vary tremendously across trials. But, it could be the case that individual bank riskometer readings covary from trial to trial as well. That is, some trials may produce high riskometer readings for all banks, and some trials may produce low riskometer readings for all banks. So, our results on Absolute and Relative Imprecision results above do not rule out the possibility that a regulator can still rank banks by the amount of risk they create. If this is the case, then a regulator would at least find it possible to require high risk banks to hold more capital than low risk banks.

We analyze this possibility by computing the Spearman rank order correlation coefficient ρ for banks ordered by risk-reading for pairs of trials selected at random. We report the results of this analysis in Table III. We find for each riskometer and each sample period that the ranking of banks by riskometer reading are essentially uncorrelated from trial to trial, and hence that ρ is not statistically significantly different from 0. It follows that a regulator can not use the riskometers we examine here to order banks by the level of systemic risk they create.

V. Conclusion

Basing a bank's capital requirement (or other regulatory intervention) upon the level of systemic risk that the bank creates is undoubtedly far superior to basing its capital requirement upon a cruder policy such as a simple leverage ratio...at least in theory. Yet, the macroprudential policy will only be superior in practice if the regulator possesses a riskometer that enables him to precisely measure the amount of systemic risk that individual banks create. We examine the three leading riskometers, viz., VaR, MES, and CoVaR, and find that they are incapable meeting the demands that macroprudential policy places upon them.

We note that our critique of the riskometers we examine is not that they are theoretically unsound or that they are mis-specified. If this were the case, then a different riskometer or a different specification for an existing riskometer might solve the problem. But this is not the case: MES and CoVaR in particular have strong theoretical foundations, and these foundations lead to plausible empirical risk measures. Instead, we show that the three analytically distinct riskometers that we do examine are all empirically unreliable. Since the empirical implementation of any systemic risk measure–no matter how theoretically attractive–will have to deal with similar problems to those that caused the VaR, MES, and CoVaR riskometers to produce such imprecise risk readings, we think it likely that riskometers in general will struggle to provide the precise risk readings that macroprudential policy requires. We therefore conclude that, in practice, less informationally intensive policies such as the leverage ratio may offer a more sensible approach to dealing with systemic risk than a more theoretically attractive macroprudential policy.

We draw two implications from our analysis. First, on a methodological note, our analysis highlights the importance of taking measurement uncertainty seriously. A policy option that may make sense for an omniscient being that can observe all of the relevant parameters directly may not make sense for regulators who must operate with imperfect estimates of those parameters. Second, while we do think that the leverage ratio may dominate macroprudential policy in practice, this is not because the leverage ratio is itself ideal–it is just that we don't know enough to get macroprudential policy to work well. It follows that capital will inevitably be an inefficient tool with which to address systemic risk. And while the inefficiency of capital by no means implies that a better tool exists, it does suggest that it would be worth looking hard to find one (or more).

Our analysis of systemic risk based capital requirements brings to mind the story of Phaeton, the long lost son of the god Apollo. One day Phaeton appeared on Mount Olympus and Apollo was so delighted to see him that he offered to grant him anything that he desired. Phaeton asked to drive the chariot that pulled the Sun across the heavens. Apollo argued that this was a terrible idea as the chariot was built for him–a god–and not for a mortal. But Phaeton stood firm and Apollo had given his word, so the next morning Phaeton set off. Of course, he was completely incapable of controlling Apollo's chariot and was thus in the process of destroying the world when Zeus stuck him down with a thunderbolt. The moral of the story is that one should be wary of using tools that require divine powers to operate effectively if one doesn’t have them. This is wisdom that we forget at our peril.

REFERENCES

Acharya, Viral, Lasse Pederson, Thomas Philippon, and Matthew Richardson (2010), Measuring Systemic Risk, NYU Working Paper

Adrian, Tobias and Marcus Brunnermeier (2008), CoVaR, Federal Reserve Bank of New York Working Paper no. 348

Bisias, Dimitrios, Mark Flood, Andrew Lo, and Stavros Valavanis (2012), A Survey of Systemic Risk Analytics, Office of Financial Research Working Paper #1

Brownlees, Christian and Robert Engles (2010), Volatility, Correlation, and Tails For Systemic Risk Measurement", NYU Working Paper

Brunnermeier, Markus, Andrew Crockett, Charles Goodhart, Avinash Persaud, and Hyun Shin (2009), The Fundamental Principles of Financial Regulation, *Geneva Reports on the World Economy*

Crockett, Andrew (2000), Marrying the Micro - and the Macro - Prudential Dimensions of Financial Stability, BIS Speech

D' Hulster, Katia (2009), The Leverage Ratio, *World Bank Crisis Response,* Note 11

Danielsson, Jon (2002), "The Emperor Has No Clothes : Limits to Risk Modelling", *Journal of Banking and Finance 26*, 1273–1296

______ (2008), "Blame the Models", *Journal of Financial Stability* 4, 321–328

______ (2009), "The myth of the riskometer", VoxEU.org, http://voxeu.org/index.php?q =node/ 2753

Danielsson, Jon, Hyun Shin, and J.P. Zigrand (2011), "Endogenous and Systemic Risk", http://www.riskresearch.org, 2011

Danielsson, Jon, Kevin R. James, Marcela Valenzuela, and Ilknur Zer (2012), Model Risk of Systemic Risk Models, *www.riskresearch.org*

Efron, Bradley and R.J.Tibshirani (1994), *An Introduction to the Bootstrap*, 1st ed. (New York, NY: Chapman and Hall, Monographs on Statistics and Applied Probability 57)

Elsinger, Helmut, Alfred Lehar, and Martin Summer (2002), Risk Assessment for Banking Systems, Oesterreichsiche NationalBank Working Paper No.79

Goodhart, Charles, Pojanart Sunirand and Demitrios Tsocomos (2004), A Model to Analyse Financial Fragility : Applications, *Journal of Financial Stability* 1, 1 - 30

Hellmann, Thomas, Kevin Murdock, and Joseph Stiglitz (2000), Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are capital requirements enough?, *American Economic Review* 90, 147 – 165

Hildebrand, Philipp (2008), Is Basel II Enough?The Benefits of the Leverage Ratio, http://www.bis.org/review/r081216d.pdf

Pástor, Lubos and Robert F. Stambaugh (2012), Are Stocks Really Less Volatile in the Long Run?", *Journal of Finance* 67, 431 - 477

Politis, Dimitris N. and Joseph P. Romano (1991), The Stationary Bootstrap, Technical Report #91–03, Department of Statistics, Purdue University

Segoviano, Miguel and Charles Goodhart (2009), Banking Stability Measures, *IMF Working Paper* WP/09/04, 2009

Tarashev, Nikola, Claudio Borio and Kostas Tsatsaronis (2010), Attributing Systemic Risk to Individual Institutions, BIS Working Paper 308

Zhou, C. (2010), Are banks too big to fail? Measuring systemic importance of financial institutions, *International Journal of Central Banking* 6, 205–250

TABLE I

Absolute Imprecision

We measure a riskometer's Absolute Imprecision AbP for a given bank by the ratio of its maximum plausible value to its minimum plausible value. In this table we plot the distribution of AbP for sample banks for three riskometers (VaR, MES, and CoVaR) and two sample periods (2003/2006 and 2007/2010). The rectangle in the middle of each plot gives the interquartile range, the bars at the end mark the middle 95% of the distribution (with numerical values), and the vertical line in the middle indicates the average AbP value (with its numerical value underneath). We compute AbP by: i) using a stationary bootstrap to construct 10,000 trial return histories for each sample; ii) calculating riskometer readings for each bank for each trial; and iii) setting the plausible range of a bank's riskometer reading equal to the middle 95% of its trial riskometer readings. AbP for bank *B* is thus equal to the ratio of the 97.5th percentile value of its trail riskometer readings to the 2.5th percentile value of those readings. Our base sample consists of the 101 financial firms with a market cap in excess of $5 billion as of June 2007. We include a firm in a sample period if we have a return observation for that firm for at least 75% of sample days. Our 2003/2006 sample consists of 92 firms, and our 2007/2010 sample consists of 77 firms.

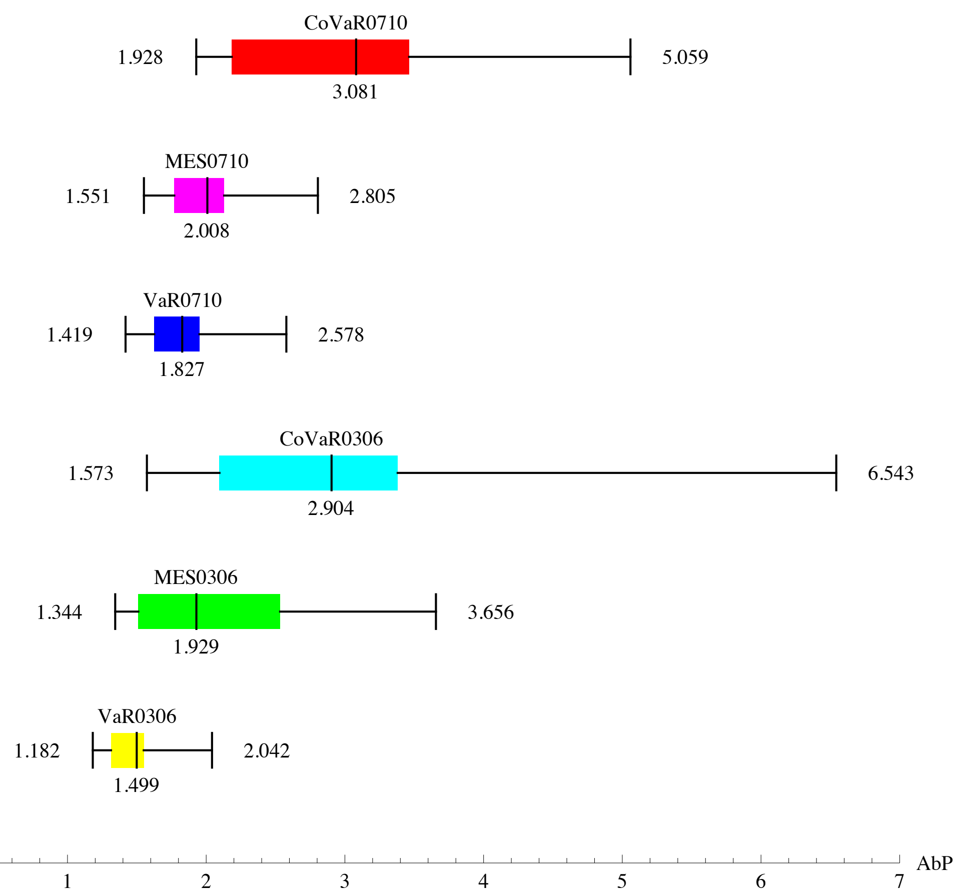

TABLE II

Relative Imprecision

We measure a riskometer's Relative Imprecision RelP for a given bank by the number of percentiles of the sample point estimate risk reading distribution covered by the plausible range of the riskometer reading for that bank. To illustrate, if the sample point estimate risk readings for a given riskometer are uniformly distributed on the interval {10, 20} and if the plausible range of the riskometer reading for a given bank B is {11, 19}, then RelP for that bank is 80 as the plausible range for that bank's riskometer reading covers 80 percentiles of the sample point estimate risk reading distribution. In this table we plot the distribution of RelP for sample banks for three riskometers (VaR, MES, and CoVaR) and two sample periods (2003/2006 and 2007/2010). The rectangle in the middle of each plot gives the interquartile range, the bars at the end mark the middle 95% of the distribution (with numerical values), and the vertical line in the middle indicates the average RelP value (with its numerical value underneath). We compute RelP by: i) using a stationary bootstrap to construct 10,000 trial return histories for each sample; ii) calculating riskometer readings for each bank for each trial; iii) setting the plausible range of a bank's riskometer reading equal to the middle 95% of its trial riskometer readings; and iv) comparing that plausible range to the point-estimate risk reading distribution. See Table I for a description of the sample.

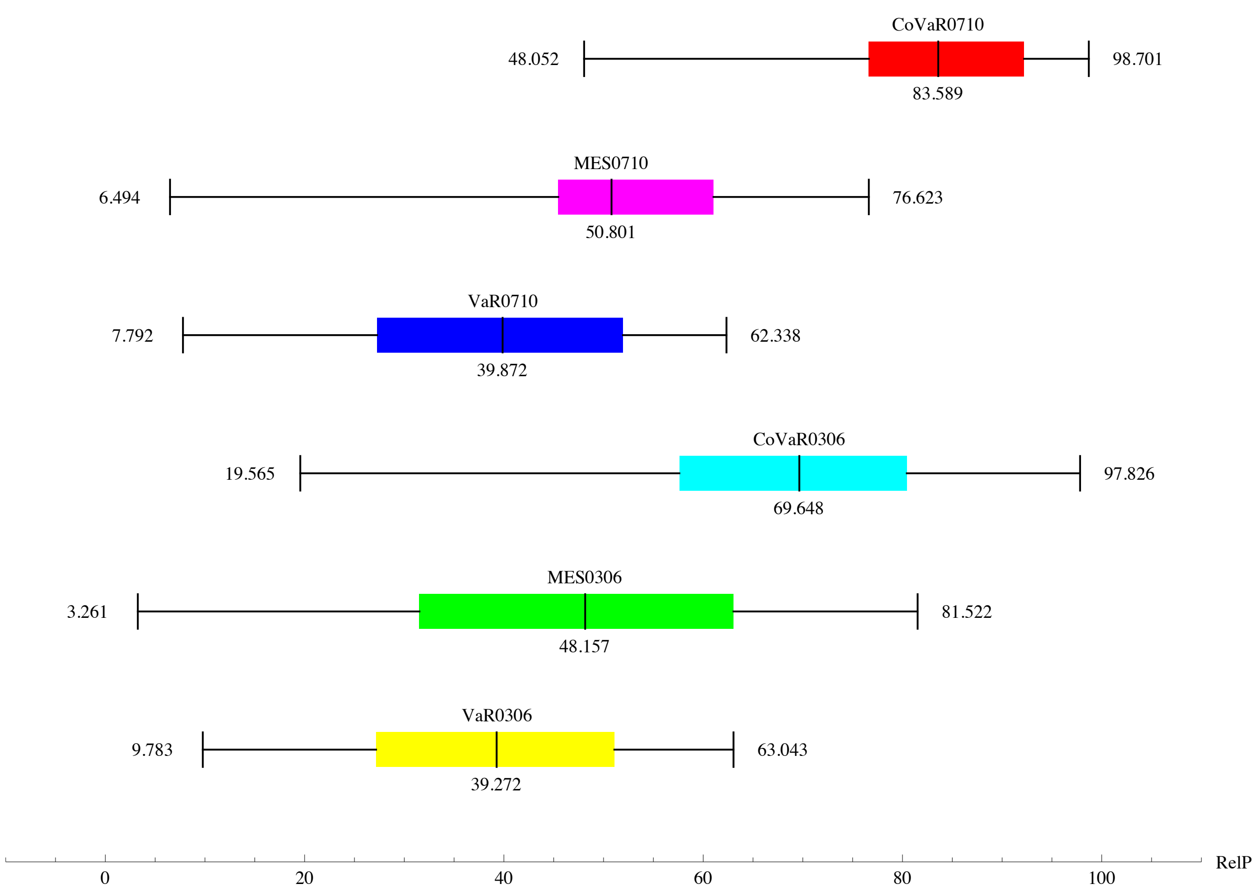

TABLE III

Rank Imprecision

In this table we report summary statistics for the Spearman rank order correlation coefficient for each riskometer for each sample period. We compute these statistics by: i) using a stationary bootstrap to construct 10,000 trial return histories for each sample; ii) ordering the sample firms by riskometer readings for each trial; and iii) calculating the Spearman coefficient for 10,000 pairs of trials selected at random. See Table I for a description of the sample.

Sample	Riskometer	Median	99% Confidence Interval
2003/2006	VaR	0.05	{-0.22, 0.32}
	MES	0.01	{-0.26, 0.28}
	CoVaR	0.03	{-0.24, 0.29}
2007/2010	VaR	0.12	{-0.19, 0.41}
	MES	0.09	{-0.21, 0.39}
	CoVaR	0.08	{-0.2, 0.36}