Comments on
MARKUS STAUB: The Term Structure of Interest Rates
and the Swiss Regional Bank Crisis – Empirical Evidence
and its Limits
GEORGE SHELDON, MARTIN MAURER: Interbank Lending and
Systemic Risk: An Empirical Analysis for Switzerland

EVA TERBERGER-STOY*

1. WHAT THE TWO PAPERS HAVE (NOT) IN COMMON

Going by their titles, the paper by STAUB and the paper by SHELDON and MAURER seem to form part of the same research program. They present the results of empirical work trying to assess the extent of systemic risk endangering the Swiss banking sector. STAUB’s analysis concentrates on changes in the term structure as a likely source of systemic risk. Using micro panel data from 1987 to 1994, it addresses the question if the inversion of term structure was a major cause of the Swiss regional bank crisis in that period. SHELDON and MAURER direct our attention to a further source of systemic risk. Using 1987 to 1995 accounting data for all banks operating in Switzerland and aggregate data on interbank lending, they try to assess the likelihood that the failure of one bank could trigger a chain reaction in the Swiss banking sector via domestic interbank lending relationships. Considering their closely related research aim, it made sense to give a joint comment on the two papers.

Looking at their main findings, however, both papers seem to be of quite different nature. While SHELDON and MAURER present an answer to their initial question, STAUB does not. «This paper is essentially destructive» is STAUB’s opening statement, pointing to the fact that his analysis turned out to be less about systemic risk and the Swiss regional bank crisis, but more about the limits of econometrics as a means to answer the posed question.

* University of Heidelberg
I would like to thank HANS GERSBACH for helpful discussion.
2. Staub's RESULTS: ENOUGH REASON TO DISMISS ECONOMETRIC ANALYSIS AS A TOOL TO ANALYSE THE INFLUENCE OF TERM STRUCTURE INVERSION ON SWISS REGIONAL BANKS?

Why did Staub start out to do an empirical analysis about the influence of the term structure on the Swiss regional bank crisis to end up with writing a paper about econometric methodology?

Staub's research was motivated by the observation that parallel to the inversion of interest rates starting in 1988 the number of Swiss regional banks began to decline, leading to the disappearance of approximately 40 percent of the Swiss regional banks from the Swiss official banking statistic until the beginning of the nineties. «Can the coincidence in time and the apparent correlation be statistically sustained?» (p. 664) Staub wanted to provide an answer based on «rigorous econometric evidence» (p. 664). Rigorous econometric evidence, however, he could not find.

Following the line of recent empirical research on banking crises, he tested his hypothesis in numerous different model specifications. Besides the term structure as an explanatory variable for the disappearance of regional banks, he included as right hand variables all those factors as well which were classified as having a significant influence on bank failures in former research on banking crises. Therefore, nobody would be able to blame him for predetermining an explanatory power of the term structure in his model design by leaving out other important factors of influence like liquidity risk, the risk in the banks' credit portfolio etc. Furthermore, Staub controlled for problems of multicollinearity—a common problem with empirical tests based on accounting data. Staub even defined the variable representing term structure influence in three alternative ways to check for any inadequate measurement of his most important influence factor. The results were always the same, or rather—just the opposite—they lacked in similarity. Depending on the model specification, the term structure was either of high significance on the disappearance of Swiss regional banks or of no significance at all. The testing results of almost all other right hand variables showed the same lack of robustness. Without being able to identify a pattern which was responsible for these unstable results the explanatory variables changed from significant to non-significant, depending on the model specification. Furthermore, most of the regression coefficients had implausible signs and their signs even switched with a change in the model specification.

What could Staub do with these most frustrating results? He chose to turn the vice into a virtue by redefining his research aim. As the only result, which «...paradoxically, proves to be relatively robust, is the lack of robustness» (p. 684), Staub decided to put this result into the center of his arguments: He wrote a paper on why «it seems practically impossible to rigorously prove term structure influence in a panel framework of micro balance sheet and income statement data.» (p. 684)

No wonder that reading his paper I was reminded of the times when I was into reading literature on economic methodology. Staub's paper seems to be the perfect reply de-
fending the econometric profession against the harsh criticism of leading economic methodologists which is summed up by the following citations:

«...researchers feel free to find inferences implied by many different «models»...From this set of estimated models the researcher or a computer chooses the model that yields the most congenial results. In my language this is a «specification search». These searches have no foundation in any clearly articulated theory of inference and the meaning that attaches to the resulting inferences is consequently unclear and open to discussion.» (LEAMER 1985, p. 298.)

«That only significant results get published has long been a scandal among statistical purists: they fear with some reason that at a five percent level of significance something like five percent of the computer runs are successful.» (McCloskey 1985, p. 139)

With his paper, STAUB proves himself to belong to the group of the statistical purists. He did not feel free to choose and present the model that yields the most congenial results, although he could have easily done so. In some of his model specifications he found exactly what he was looking for: A highly significant influence of the inversion of term structure on the failure of regional banks. He chose to present not only these confirming results, but all of the other computer runs as well whose results contradicted his hypothesis—even if this statistical purism prevented him from writing a nice plausible paper on the subject he wanted to write about. Therefore, the work of STAUB deserves the highest respect. Its scientific standard is beyond all methodological doubt. The paper stands for the scientific honesty of econometricians.

Having STAUB’s paper in mind and thereby realizing how easily one can be misled by a naive trust in the presented results of an econometric analysis, it seems difficult to return to everyday business and discuss the paper of SHELDON and MAURER as if nothing has happened. It rather seems more appropriate to first try to restore the trust in the explanatory power of econometric analysis before going on to the next paper. Therefore, I want to close the discussion of STAUB’s result by asking if the question which he tried to answer in his analysis really gave econometrics a fair chance of delivering a satisfactory answer. So, let us turn back to his original story once more.

STAUB started out with the hypothesis that the inversion of term structure was a major cause of the Swiss regional bank crisis. Without doubt, inversion is an extreme constellation of term structure and, for sure, it can cause insolvency problems in the banking industry which is specialized in transforming short-term deposits into long-term credits. What is not plausible, however, is the implicit assumption of STAUB that regional banks were hit more severely by the inversion of interest rates than other Swiss banks. Exactly because interest rate risk is part of systemic risk it should be expected that the whole banking sector has to suffer from unhedged term structure risk. Why should regional banks suffer more than the average bank and get into a crisis which makes about forty percent of these institutions disappear? Could the true reasons for their disappearance rather be that regional banks are smaller than the average Swiss bank or that their equity ratio is lower than average? STAUB did not compare regional banks with other Swiss
banks which had to handle the same term structure scenario. If he had done so he might have found out more about the roots of their problems.

Staub himself mentions that the Swiss regional bank crisis might have had other causes than the inversion of term structure, without taking these other causes into account, however, in his analysis later on. In his introduction he hints at the massive structural change which the banking industry in Switzerland went through. Smaller banks seemed to have comparative disadvantages in the growing competition, leading to the necessity of regional banks to cooperate with each other or with larger partners. Staub points out as well that the disappearance of regional banks from the official banking statistics, although treated in his analysis as an indicator for failure, does not mean that these banks were closed down. Rather, disappearance «normally indicates a takeover by another bank. We therefore implicitly assume that takeovers (at least in the regional bank sector) are a good proxy for (potential) failures or at least serious problems, abstracting from the possibility that a regional bank can join another bank although it is in a strong financial position.» (p. 668/669) Perhaps it would have been wise to question this implicit assumption, because Staub's own empirical results give strong indication that this assumption might have been wrong. Besides the robust result of a general lack in robustness which Staub puts into the center of his argument there is one more robust result which he ignores in his comments: All of his model specifications show that the disappearance of banks is significantly influenced by their rentability. And, contrary to the plausible expectation, the regression coefficient of the explanatory variable rentability does not have a negative, but has a positive sign. The probability of «failure», which is really a probability for being taken over, seems to rise with rentability. A most implausible result, unless a take over is rather a good proxy for relative success than a good proxy for problems and possible failure.

In my opinion, this highly implausible and, at the same time, robust result gives enough reason to rethink the initial hypothesis and test for the explanatory power of an alternative story. This alternative story speculates on structural changes in the banking industry as the major cause of the regional bank crisis: Structural changes lead to diminishing margins in the whole banking industry putting pressure on all banks to look for economies of scale. The pressure is hardest for the smallest market participants, the regional banks. They need to cooperate with bigger partners or sell their business if they want to survive, even if they lose their independence. The most attractive take over targets are those regional banks which have not yet run into financial problems, which, although they are small, still show a satisfactory rentability. With this alternative story as a starting hypothesis, econometric analysis might have had a better chance of producing robust results than with the story Staub started out with.

Having somehow restored confidence in econometric analyses' ability to answer questions about the influence of systemic risk on the viability of the Swiss banking sector, let us turn to the results of Sheldon and Maurer.
3. THE RESULTS OF SHELDON AND MAURER: ENOUGH REASON NOT TO WORRY ABOUT INTERBANK LENDING RISK IN SWITZERLAND?

SHELDON and MAURER manage to provide an answer to their question on how severe is the potential threat stemming from domestic interbank lending relationships in Switzerland. Their answer is rather comforting to all those who might have been worried by interbank lending as a source of systemic risk: «...although the likelihood of a bank insolvency in any given year is quite high, the chances of a bank failure propagating itself through the banking system via the network of interbank loans are quite low.» (p. 710)

Of course, SHELDON and MAURER did not forget the appropriate warnings to prevent any reader from being too comforted by their result. First of all, their data base restricted them to studying the risk from domestic interbank loans. These loans, however, are outsized by cross-border interbank loans by more than double. Secondly, their analysis concentrates on the likelihood of chain reactions triggered off by a single idiosyncratic shock. «A series of simultaneous default shocks would undoubtedly place a greater strain on the network of interbank loans» (p. 710). Thirdly, shortcomings of the data at hand lead to the necessity to fill the gaps in the empirical information with assumptions, and the chosen assumptions rule out clumped risks. Finally, «one should bear in mind that a low risk of contagion needs to be weighed against the costs of its consequences before a final judgment is felled» (p. 710). In the light of such careful assessment of their own work any further comments seem superfluous.

But, as it is the task of a discussant to give some additional insights, one has to start looking for the most promising starting point to fulfill this task. A point which seems to deserve further questioning is SHELDON and MAURER’s third remark about the shortcomings of their analysis. If lack of data prevent SHELDON and MAURER from calculating the realistic exposure to systemic risk due to domestic interbank lending, how can their estimation of risk which partly rests on assumptions be interpreted? Do they choose assumptions which allow them to calculate the lower bound of risk, or do they aim at calculating the upper bound of risk? Ruling out clumped risks by assumption points to the former. But how could a banking supervisor be comforted by an analysis which points out that there is little danger of contagion in a best case scenario? Most likely they try to come close to a realistic risk scenario, which lies somewhere in-between the upper and the lower bound. But if this is the case, it seems worthwhile to take a closer look at their assumptions to find out how realistic these assumptions really are.

SHELDON’S and MAURER’S analysis starts off with a set of empirical data on the marginal distributions of domestic interbank lending among the 12 bank groups in Switzerland consisting of cantonal banks, commercial banks, regional banks etc. On the aggregate level of the different bank groups this scenario is therefore realistic. The research, however, aims at calculating default probabilities due to interbank lending on the level of the individual bank. Consequently, SHELDON and MAURER have to break the aggregate empirical data down to individual banks’ level, and at this point their first assumption comes in. Using the method of entropy maximization, they fill the missing cells in the in-
terbank lending matrix based on the assumption that within each bank group borrowed funds are distributed evenly among all group members. The authors justify this assumption by it being the one with the least additional information input. Assuming even distribution among group members, however, means assuming a high degree of diversification in interbank lending as well. Although this assumption does not "minimize the amount of idiosyncratic risk in the system" (p. 702) as the authors claim—the risk minimizing assumption would match the most stable banks with the biggest interbank loan exposure—the scenario of Sheldon and Maurer comes very close to a best case scenario.

The assumption of ruling out clumped risks is not the only assumption in their analysis. The authors move on in their argument by calculating the probability of one bank failing due to idiosyncratic risk, failing being defined as all states of nature in which an individual bank's losses exceed the bank's equity. Calculation of default probability is based on individual banks' accounting data and on two different assumptions about the distribution of revenues, one of them representing a good case (normal distribution) and one of them representing a bad case (distribution relying on Chebychev inequality). Their result that one bank failing in Switzerland within one year is almost certain seems to rest on solid, realistic grounds. What is less realistic, however, is the assumption that only one bank fails. Even if systemic risk is ignored at this stage and only uncorrelated idiosyncratic risks are taken into account, the probability of more than one bank failing within one period is higher than the probability of just one bank failing. Using the good case assumption of normally distributed revenues, the average default probability for each bank in Switzerland according to Sheldon's and Maurer's calculations is 0.8 percent. Using this default probability and the information that there are almost 500 banks in Switzerland, the probability of at least one bank failing is close to 1, but the probability that only one bank fails is much lower than the probability that idiosyncratic risk causes four banks to fail at the same time. As the following arguments of Sheldon and Maurer build on only one bank failing, there seems to be another best or at least good case assumption involved, which drives their result towards the lower risk bound. The analysis could have been improved if the authors had assumed the most likely case for each bank group concerning the number of banks defaulting—an information which could have been drawn out of the existing data.

If default happens, the authors assume default to be complete. The lending banks lose the entire book value of their interbank loans to the defaulting bank. The assumption of complete loss is not quite consistent with the underlying definition of default, because losses consuming the entire equity do not necessarily mean an asset value of zero. On the other hand, complete loss is an assumption which does not underestimate, but overestimates risk and, therefore, pushes the results of the analysis a bit towards the upper risk bound. A further underestimation of risk, however, goes along with the next assumption concerning the distribution of the defaulted loans among the groups of lending banks. Here it is assumed that the origin of the defaulting bank's interbank funds corresponds to the empirical data on its group average. Again, clumped risks are ruled out by assumption. The group's average is the aggregate result of numerous special interbank relation-
ships. Probably none of the banks within one group borrows exactly according to the group average. The average picture, therefore, is most likely far away from the realistic picture of a single group member.

Finally, we come to the last step of Sheldon’s and Maurer’s analysis. Based on additional assumptions about the distribution of defaulted loans among the single members of the lending groups, the authors calculate the probability of the initial default due to idiosyncratic risk triggering off a chain reaction by causing a lending bank to fail. Sheldon and Maurer try to present a worst case scenario and a realistic case scenario. Their worst case scenario assumes that the default shock has to be absorbed by one single group member because the defaulted loans were concentrated in the loan portfolio of one bank within each lending group. «As to be expected, almost no average size bank would be able to withstand the full force of a loan default from a large bank. ... If we assume instead and more realistically that the borrowings of a defaulting bank from a specific group are uniformly distributed within a bank group so that the shock hits all banks in the group equally hard, then no lending bank should expect to fail.» (p. 708) Without wanting to question if the assumption of uniform distribution really is more realistic or if it should rather be classified as optimistic, there is one other flaw in their argument. Both of these results are based on the assumption that the banks which are hit by the default shock would have earned the expected value of returns if the default shock had not occurred. Taking expected returns as the benchmark from which losses caused by interbank lending are deducted ignores any risk due to the idiosyncratic volatility of the lending banks’ revenues. If this individual risk had been taken into account—and the authors could have done so without the need of additional data—the probability of chain reactions would have been much higher, even in Sheldon’s and Maurer’s good case scenario.

To sum up, Sheldon’s and Maurer’s comforting result seems to be biased towards the lower bound of risk stemming from domestic interbank lending. Therefore, it seems wise not to stop worrying about even one single component of systemic risk in Switzerland: the risk due to domestic interbank lending. Nevertheless, Sheldon and Maurer present an innovative technic to calculate that risk. With more appropriate micro level data on interbank lending they no longer will have to base part of their analysis on assumptions. Therefore, their work is a decisive step towards developing an unbiased picture which comes close to reality.

4. CONCLUSION: EMPIRICAL WORK IS HARD WORK

Although both papers turned out to be very different in nature, there is at least one common lesson which is taught by both of them: Empirical work is hard work, and empirical work on systemic risk is extremely hard. The authors’ attempt to take a closer look at one single aspect of systemic risk proved to be very difficult. How many more difficulties would arise if one exchanged these modest research aims against the more ambitious one of empirical research on systemic risk in general? The main feature of systemic risk had
to be ignored by both papers to have a fair chance of being able to handle their problem. The characteristic feature of systemic risk is the interlinkage of different types of risk, and this is exactly what makes systemic risk so dangerous. The question if systemic risk is uncalculable risk will continue to be a challenge for future empirical research.

LITERATURE