Journal of Banking & Finance 36 (2012) 332–340

Contents lists available at SciVerse ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

A systematic approach to multi-period stress testing of portfolio credit risk

Thomas Breuer [a], Martin Jandačka [a], Javier Mencía [b], Martin Summer [c,*]

[a] Research Centre PPE, Fachhochschule Vorarlberg, Hochschulstr. 1, A-6850 Dornbirn, Austria
[b] Financial Stability Department, Banco de España, Alcalá, 48, E-28014 Madrid, Spain
[c] Oesterreichische Nationalbank, Otto Wagner Platz 3, A-1090 Vienna, Austria

ARTICLE INFO

Article history:
Received 21 January 2011
Accepted 21 July 2011
Available online 31 August 2011

JEL classification:
G28
G32
G20
C15

Keywords:
Stress testing
Credit risk
Worst case search
Maximum loss

ABSTRACT

We propose a new method for analysing multi-period stress scenarios for portfolio credit risk more systematically than in current macro stress tests. The plausibility of a scenario is quantified by its distance from an average scenario. For a given level of plausibility, we search systematically for the most adverse scenario. This ensures that no plausible scenario will be missed. We show how this method can be applied to some models already in use by practitioners. While worst case search requires numerical optimisation we show that we can work with reasonably good linear approximations to the portfolio loss function. This makes systematic multi-period stress testing computationally efficient and easy to implement. Applying our approach to data from the Spanish loan register we show that, compared to standard stress test procedures, our method identifies more harmful scenarios that are equally plausible.

1. Introduction

When supervisory authorities or financial institutions are asked to produce stress tests for particular aggregate or individual portfolios they face a basic challenge: While scenarios should be extreme they should at the same time also be plausible. Clearly the more extreme the scenarios which are considered the less plausible they become: There is a trade off between severity and plausibility of stress scenarios. Once this trade off has been made and some measure and level of plausibility have been determined we still have to make sure not to ignore any severe scenario for a given plausibility. Unfortunately, current stress test procedures generally use a few hand-picked scenarios. This makes stress testing subject to considerable arbitrariness. If, for which reason whatsoever, an institution does not pick the worst scenarios but just *some* scenarios involving large risk factor moves it might get or give an *illusion* of safety. For instance, this has been the case of the Irish banking system despite the hopeful results of its 2010 stress test. There is, therefore, a real danger that 'stress testing results will continue to lull users into a false sense of security' (Borio and Drehmann, 2009).

In this paper we offer a systematic approach to stress testing. We specify a measure of plausibility based on the Mahalanobis distance of the changes in a set of factors. In this way, we can control how far we go into the tails of the risk factor distribution in our search for stress scenarios. All scenarios above the minimal level of plausibility are considered on an equal footing, all scenarios under the minimal plausibility level are left out of consideration. This makes transparent the trade off between plausibility and severity one always has to make. Then, we determine the worst case by maximising a loss function among all scenarios above the minimal level of plausibility. If an institution can take this worst case loss it can be sure to survive any scenario at this level of plausibility. At the same time our method allows for an analysis of appropriate risk reducing actions.

The idea that it might be useful to conduct stress testing analysis in different and complementary ways to traditional stress tests has been widely discussed among practitioners for some time. Ultimately a stress test asks which scenarios lead to big losses. Starting from a big loss and working backward to identify how such a loss could occur, is commonly referred to among risk management professionals as "reverse stress testing". The Counterparty Risk Management Policy Group (2008, IV, 5, p. 84) in its section on reverse stress testing states that "the conduct of such a reverse stress test, would be a very challenging exercise". One way to look at our contribution is that we propose a fairly straightforward,

* Corresponding author. Tel.: +43 1 40420 7200; fax: +43 1 40420 7299.
E-mail addresses: thomas.breuer@fhv.at (T. Breuer), martin.jandacka@fhv.at (M. Jandačka), javier.mencia@bde.es (J. Mencía), martin.summer@oenb.at (M. Summer).

0378-4266/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbankfin.2011.07.009

computationally efficient way of implementing this idea by relying on credit risk models already widely in use in risk management. Worst case search over plausible domains is one specific way to make reverse stress testing ideas operational.

While the theory and the concepts of worst case search have already been developed in a one-period framework by Breuer et al. (2009), the current paper introduces two major novel contributions. First, we extend the method to a multi-period setting, in which scenarios are paths of macroeconomic variable values rather than values at one point in time. Second, we show the practical usefulness of this approach by means of a meaningful and relevant empirical application, which makes use of standard data and models already in use by stress testing practitioners. We use loan register data from the central loan register of the Banco de España and a credit risk stress testing model developed by Jiménez and Mencía (2009). Similar credit risk portfolio models have been used by practitioners in many other central banks. In fact, portfolio credit risk models with similar characteristics have been estimated using data from Canada (Misina et al., 2006), Finland (Virolainen, 2004), Hong Kong (Wong et al., 2008), Italy (Fiori et al., 2007) or the UK (Drehmann, 2005; Drehmann et al., 2006). An international model can also be found in Pesaran et al. (2006). These models follow Wilson (1997a,b) in trying to model the default frequencies of loans to different economic sectors as functions of macroeconomic conditions. Compared to the previous literature, Jiménez and Mencía (2009) contain two important novel features. First, they not only consider corporate loans but also loans to households such as mortgages and consumption loans. Second, they introduce latent factors in order to model the correlations between different types of loans. However, since the methodology is otherwise consistent with the previous literature, the stress testing approach that we use can be directly applied to other countries. In this sense, we demonstrate that the application to real world stress testing problems is straightforward and does not raise complicated implementation issues.

We believe that our approach is able to address some of the more recent criticism raised against stress testing. One such concern is the plausibility of stress scenarios. Several authors try to increase the plausibility of stress scenarios by taking into account not just the size of risk factor changes but also the correlations. For example, Bonti et al. (2006) apply conditional scenarios which specify for a few key risk factors a range of adverse values and leave the other risk factor values unspecified. Thereby they truncate the risk factor distribution and conserve the correlations between the risk factors. Their choice of key risk factors reflects portfolio characteristics. While this approach increases the plausibility of scenarios and takes portfolio characteristics into account, it provides no guarantee that there are no more harmful scenarios of the same or even higher plausibility.

Alfaro and Drehmann (2009) criticise the realism of using mainly domestic macroeconomic factors in the description of stress scenarios. They question the ability to construct plausible yet severe scenarios and they diagnose a lack of robustness of models during a crisis. While our method is agnostic with respect to which are the relevant risk factors, our model shows a way how plausible and severe scenarios can be constructed. In our example we see that our method of worst case search applied during the boom year 2006 is clearly superior to historical scenario stress testing in gauging the subsequent crisis. While we use the historical data to delineate the plausibility of the stress scenario the worst case looks clearly different from the historical recession scenario and comes much nearer to the actual downturn that materialised in 2008. Hence, we show that worst case search could have generated scenarios as severe as the current crisis based only on information prior to 2007. Notice, though, that we do not mean to imply that our model would have predicted the crisis, but rather that its use could have forewarned practitioners that such an unprecedented scenario could occur within reasonable levels of plausibility.

Alfaro and Drehmann (2009) conclude that the unsatisfactory state of stress testing models calls for a stronger role of judgement in stress testing. Our approach is open to this requirement. While Alfaro and Drehmann (2009) want to more strongly weight judgement with respect to the output of a stress test, our approach would argue to the contrary that the judgement should enter more strongly on the side of the input. We propose to pin down ex ante in a transparent way the plausibility one is willing to consider and then accept the result of the stress test. The advantage we see in this approach is that the current practice and the practice augmented with stronger judgement as suggested in Alfaro and Drehmann (2009) makes stress testing an exercise extremely prone to all kinds of manipulation. We believe that our approach is a step towards manipulation proofness of stress testing. As such our paper contributes not only to the methodology of stress testing but has wider implications for the political economy of stress testing.

Finally a word of caution on model risk. Traditional risk measurement models have also been criticised for assuming naively that the risk factor distribution in times of crisis is the same as usual (see Berkowitz, 2000). Traditional stress tests partially overcome this problem because the design of stressed scenarios is mainly based on judgement, and not on a probability distribution. However, not even these tests are fully immune to model risk, because a model typically calibrated with historical data is employed to assess the impact of the stressed scenarios on the variable of interest. This criticism also applies to our framework. We measure plausibility of scenarios by its Mahalanobis distance from the expectation. This measure of distance involves a fixed distribution of risk factors, which is estimated from historical data. With respect to a different distribution, which might prevail in a crisis period, the plausibility is different. In the end, there is some arbitrariness in whether we interpret a crisis as a period of adverse risk factor realisations, as we do in this paper, or as a regime of an adverse risk factor distribution from which not so extreme realisations are drawn.[1]

The paper is organised as follows. Section 2 shows how the worst case search can be applied to a set up where risk factor movements are forecasted over multiple time steps. Section 3 describes our data of the Spanish loan market. Section 4 describes and discusses our main results and Section 5 concludes.

2. Worst case search and credit risk

2.1. To which credit risk models can worst case search be applied?

We want to apply the method of worst case search over risk factor domains of certain plausibility to the analysis of portfolio credit risk. In order to do so we need to be able at each point in time t to measure the expected loss of position i in sector k, for $k = 1,\ldots,K$. We need a model structure that can express the expected loss as a function of risk factors. Typically these risk factors are innovations of macroeconomic variables at future times, as well as idiosyncratic events. We write these categories of risk factors separately as two vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}$, with $\boldsymbol{v}$ describing the innovations of macroeconomic variables at different future times.

At a given time t the loss of a position i is given by $L_{i,t}(\boldsymbol{u}_i, \boldsymbol{v})$. In typical stress tests we are interested in assessing what happens

[1] New results by Breuer and Csiszár (2010) generalize our approach to address this question. In this more general framework questions of parameter stability and model risk can be directly analysed. Following Berkowitz (2000) and Bonti et al. (2006), that approach uses a more general concept of scenarios. It takes a scenario not to be a realisation but an alternative risk factor distribution.