

Primary submission: 09.04.2015 | Final acceptance: 05.08.2016

Scenario-based stress tests: are they painful enough?

Colin Ellis¹

ABSTRACT

Forecasts, models and stress tests are important tools for policymakers and business planners. Recent developments in these related spheres have seen greater emphasis placed on stress tests from a regulatory perspective, while at the same time forecasting performance has been criticized. Given the interlinkages between the two, similar limitations apply to stress tests as to forecasts and should be borne in mind by practitioners. In addition, the recent evolution of stress tests, and in particular the increasing popularity of scenario-based approaches, raises concerns about how well the shortcomings of the associated models are understood. This includes estimated stress cases relative to base cases – the degree of pain – that simple scenario modelling approaches engender. This paper illustrates this phenomenon using simulation techniques and demonstrates that more extreme stress scenarios need to be employed in order to match the inference from simple value-at-risk approaches. Alternatively, complex modelling approaches can address this concern, but are not widely used to date. Some policymakers seem to be aware of these issues, judging by the severity of some recent stress scenarios.

KEY WORDS: Stress test, VaR, Scenarios, Risk modelling

JEL Classification: G32, E37, G17

¹ University of Birmingham - Economics, United Kingdom of Great Britain and Northern Ireland

1. Introduction

Stress tests are an important tool in the arsenals of policymakers and risk managers. Since the advent of the global financial crisis in 2007, the function and nature of stress tests have significantly changed, with regulators taking more of a leading role and the tests themselves evolving towards more of a scenario-based approach that relies on some form of economic and/or financial model. At the same time, economic forecasters' poor performance during the crisis has led to

questions about the role and nature of these forecasts. Given the inherent linkages between models, forecasts and stress tests, and related issues associated with all three, this juxtaposition is odd. The aim of this paper is to explore these linkages, noting the recent evolution of stress tests, especially for banking, given regulatory developments such as Basel-related requirements, and the different stress testing approaches that can be employed. Importantly, following discussion of these different techniques, the paper then demonstrates that simple scenario-based stress tests that rely on estimated economic relationships can often result in stresses that are less pronounced than the implied stresses from simple historical value-at-risk based approaches, when both are conducted on a similar basis.

Correspondence concerning this article should be addressed to: **Colin Ellis**, University of Birmingham - Economics, Edgbaston, Birmingham B15 2TT, United Kingdom of Great Britain and Northern Ireland. E-mail: c.ellis@bham.ac.uk

As such, the degree of stress imposed by scenario-based approaches may be relatively less punitive – or painful – than simple historical approaches, when both are conducted on the same basis. In part, this may be one reason why regulatory authorities have focused on extreme scenarios to mitigate this issue.

The remainder of the paper is structured as follows. Section 2 provides some broad background on the linkages between models, forecasts and stress tests. Section 3 then outlines the evolving role of stress testing in the banking sector, including Basel-related regulatory requirements. Section 4 briefly details and compares alternative stress testing methods from previous research. Section 5 presents new simulations that compare and contrast scenario-based and value-at-risk stress testing approaches. Finally, Section 6 concludes.

2. Models, forecasts and stress tests: their uses and misuses

Economists make a living by analyzing prices, quantities or some other form of economic or financial variable. One type of tool often used for such analysis is a model. This can take many different forms: a model can often be a set of algebraic equations, for instance, describing how different factors such as demand and inflation interact. For econometricians, models have taken increasingly complex statistical forms, both to establish where they obviously fail – instances commonly referred to as errors or residuals – and exactly how to interpret those failures. But more generally, models can be diagrams, spreadsheets or complex interlinked calculations: there are no broad restrictions about the form that different models can take.

However, all economic and financial models – whether qualitative, statistical or algebraic – share one key feature. They are necessarily simplifications of how the real world works (Bank of England, 2003). The financial and economic linkages between companies, banks and governments, both within and across countries, are so complex that no model can completely (or arguably even adequately) capture them. Trying to do so typically results in a model that does not work in some fashion, for instance, because it is too complex to solve or because it fails to match perceptions of prior experience. More generally, simply including more data, or adding more complexity, is no guarantee of improving model or forecasting

performance, although in certain circumstances it can prove very powerful (Bernanke, Boivin, & Eliasziw, 2005; Mumtaz, Zabczyk, & Ellis 2009). Pagan (2003) provides a useful summary of the linkages between models and forecasting.

Instead, it is important to recognize the true role of models: they are essentially useful tools for consistently processing information. That consistency in turn makes statistically based models in particular very useful for large-scale simulations, where multiple Monte Carlo or bootstrap simulations are used to test and infer the properties of the model (Ellis, 2006). However, the real world is rarely completely consistent, so, even armed with these tools, analytical judgement will always be critical both in building and using models. Models are useful analytical tools but do not mechanically give ‘the answer’.

Forecasting is one area where analytical judgement is key. Economists, meteorologists and other forecasters cannot solely rely on mechanical processes but need to apply their own insights and judgement (Colletti, Hunt, Rose, & Tetlow, 1996). Because models are necessarily simplifications of the real world – and the related identification of shocks is inexact, particularly in real-time given underlying data uncertainty issues (Castle & Ellis, 2002; Croushore & Stark, 2001) – they will often miss specific issues or linkages, or potentially overstate them. Here again judgement is required before attempting to ‘fill the gaps’ in the model.

It can often seem like these interventions make things worse, rather than better. Economists’ forecasting record both during and since the financial crisis has been less than stellar. There remains considerable ‘clustering’, where point forecasts for data series from different forecasters tend to bunch together over time, and forecasters often assume that key macroeconomic variables such as growth and inflation revert to trend too quickly (Pain, Lewis, Dang, Jin, & Richardson, 2014), although views about trend may be more disparate now than prior to the crisis. Faced with this poor performance, it can be tempting to disregard forecasts altogether, and there is evidence that the general public’s expectations are often far removed from those of policymakers (Moessner, Zhu, & Ellis, 2011). However, in part, the lack of regard for forecasts may reflect a lack of understanding about the role that forecasts should play. By definition, a (modal) point forecast

should reflect the forecasters' best guess of what will happen over the coming months or years, but that forecast will almost certainly be precisely wrong, even if it is broadly right. The fundamental reason for this is that shocks hit the economy, or indeed a company's cash flow, all the time, and shocks – as the name suggests – are by definition technically unpredictable, being random in terms of both incidence and magnitude. Shocks are no more predictable than exactly which lottery numbers will emerge in each draw. The best a forecast can do is express how past shocks will affect a business or sector over time.

If shocks make point forecasts inaccurate, what is the point of forecasting? Forecasts are important because they help economists identify (future) shocks as they materialize; they also help to refine our understanding of economic and financial relationships and transmission mechanisms. If GDP growth is forecast to be 3% next year, and it comes in at 4%, the forecast was wrong. Either the forecaster will have misjudged the impact of past shocks, or new shocks have hit the economy. Distinguishing between these two outcomes is difficult, but starting from 3% at least quantifies the 'news' in the growth data, providing a gauge of how wrong the forecaster was. This has parallels with the time series modelling used by Blanchard & Fisher (1989) and implemented for instance by Flood & Lowe (1995) in an inventory modelling context to distinguish between expected and unexpected changes in demand.

This quantification is a critical step, in part, because it informs how future forecasts may need to be adjusted or refined. Without being clear about what was expected – the benchmark or counterfactual – we cannot begin to identify and quantify new or unexpected developments. Uncertainty will always be inherent in forecasts, but these same forecasts are still critical for any meaningful analysis of events.

Partly because of this inherent uncertainty, some forecasters deliberately choose to present their views as a range of different outcomes rather than simply presenting point forecasts for key data series. This can include distinct scenarios – for instance, poor, central or good scenarios – or potentially even illustrating the range of uncertainty around forecasts by plotting probability-based distributions of outturns. The Bank of England was a forerunner here with its famous 'fan

charts', as described in Britton, Fisher and Whitley (1998). These explicit distributions can be useful insofar as they tell us more about the assumed balance of risks, or forecasters' views about the uncertainty around their forecasts. Having a distribution of possible future outcomes – including the mode and the mean – is clearly more informative than simply providing point forecasts. However, ultimately even fan charts will also be wrong and fail to match the actual distribution of data over time. For instance, prior to the financial crisis, the Bank of England was criticized by Wallis (2004) for publishing fan charts that were too broad. However, the crisis exposed that the fan charts were actually not wide enough (see Bank of England, 2009).

One thing fan charts – and indeed models more generally – are often bad at is capturing the tails of the distribution of uncertainty. Most economic and econometric models still focus on central tendencies – the mean or mode of the distribution – and essentially try to ensure that the model lines up with the data on average. However, sometimes we may be more worried about what might happen if things turn out worse than expected and the economy experiences a so-called tail event. Growing interest in these outcomes, especially given the failure of banks' risk models during the recent financial crisis (Haldane, 2009), has led to a resurgence in stress testing.

If forecasting tries to predict what we expect to happen, then stress testing – by definition – tries to predict what we do not expect. There are good reasons for doing this; for instance, the costs of tail events may far exceed those implied by central forecasts, meaning that businesses must think about how to address them. Low probability but high hazard events need some kind of risk management framework. This is particularly true for policymakers, who in recent years have led a renewed focus on stress testing banks. By subjecting banks to unexpectedly bad economic conditions or scenarios, policymakers have sought to test their resilience to tail events and to provide the public with some reassurance that any future crisis will be less damaging than the one that emerged in 2007. The market leader here is probably the Federal Reserve in the US, which publishes its own annual assessment of banks' resilience (see Board of Governors of the Federal Reserve System, 2015).

However, if point forecasts – best guesses – are nearly always precisely wrong in terms of predicting what will happen, then there is no reason to think that stress tests – literally unexpected outturns – will be any better as a guide to tail events. This is again because of the unpredictable nature and transmission of shocks, as noted above, which may be even more pronounced for tail events. As such, an odd juxtaposition has arisen between the increased reliance on stress testing by regulators, and reduced confidence in central forecasts. In part, this may reflect the changing nature of many of the stress tests and models used, as discussed in the next section.

3. Stress tests: an evolving regulatory landscape for banks

Stress tests are used in a variety of industries and sectors, and not just from an economic or financial perspective. However, one prominent area where stress tests have been applied is in the financial sector, and it is instructive to consider the evolving use of stress tests in this field.

Prior to the 2007-8 financial crisis, stress tests were commonly used to estimate the risks on banks' balance sheets and hence inform how large capital buffers should be. The initial focus was on banks' balance sheets and, in particular, on the inherent riskiness of different asset types against which banks have to hold capital. Over time, regulatory stress tests for banks have evolved to consider market risks, operational risks and liquidity risks; the tests focus on the total balance sheet of the institution with an assessment of overall capital and liquidity adequacy, rather than just considering specific aspects of banks' behavior. Key international regulatory guidance on stress testing was provided by the Bank for International Settlements [BIS], (2009); the Board of Governors of the Federal Reserve System (2015) and the European Banking Authority (2014) describe different aspects and results of recent large-scale regulatory stress tests in the US and the European Union.

While these and other developments represent important changes in the regulatory framework for banking, they are beyond the scope of this paper. Instead, the focus of this paper is limited to comparing and contrasting aspects of two particular methods of stress testing that happen to be frequently applied in

the banking sector. In this context, it is most useful to focus and limit the initial discussion of stress testing to credit risk, as that aspect is one of the most directly applicable in terms of comparing and contrasting different risk assessment approaches. However, the differences discussed herein hold significant relevance for the broader stress testing sphere, rather than just being applicable to either to credit risks or indeed just to banks. As such, while the approach adopted herein is deliberately simplistic, it has potentially far-reaching implications for stress testing more generally. It is important to remember that stress tests are widely used in other sectors, including corporate cash flow analysis, social housing provision, and other non-economic realms.

The measurement and treatment of credit risk has evolved over time. With the advent of the Basel II accord, banks were allowed to calculate their own risk-weights for different types of lending under a so-called 'internal ratings based' (IRB) approach. This essentially meant that banks had to assess the credit risks posed by their own borrowing counterparties. Although these internal risk weights could be scrutinized by regulators and had to satisfy national supervisors' requirements, this change introduced more discretion into regulatory measures of capital adequacy such as the Tier 1 capital ratio.

To provide guidance to banks about capital requirements, particular risk weight functions were based on a specific model (Gordy, 2003) developed on behalf of the Basel Committee on Banking Supervision. A key feature of this model was portfolio invariance: the capital required for any given loan should depend only on the risk of that loan, and not on the portfolio that it is added to. In this way, the potential losses that banks face – which normally reflect specific factors such as the probability of default (PD), loss given default (LGD) and exposure at default (EAD) – are sufficient to determine the inherent risks of the loans, and hence capital charges.

Using this framework, banks could calculate expected and unexpected losses associated with each credit exposure. In particular, average PDs can be transformed into conditional PDs – or default rates arising from a more conservative assessment of risks – using a mapping based on a global regulatory approach (BIS, 2005). Essentially, the process posits and calibrates

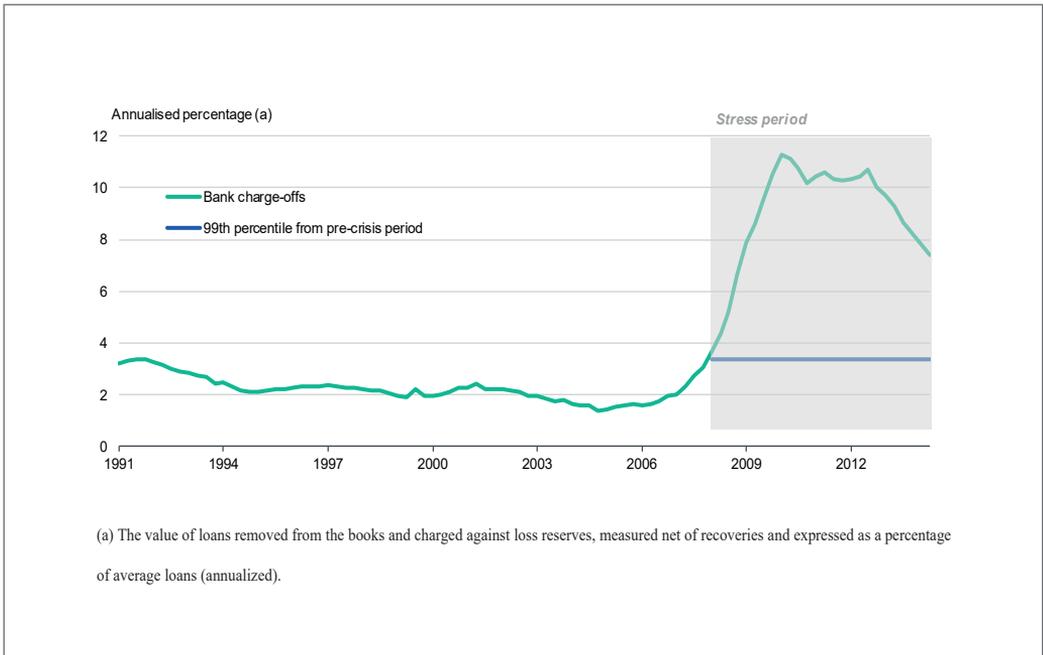


Figure 1. US banks' charge-offs on residential mortgages

Source: Adapted from “Stress testing the UK banking system: key elements of the 2014 stress test” by Bank of England (2014, April). Retrieved from <http://www.bankofengland.co.uk/financialstability/Documents/fpc/keyelements.pdf>

a distribution of potential losses and evaluates that distribution both at a central tendency – the expected loss – and at some conservative confidence level, which represents the Value-at-Risk (VaR). Under the IRB approach these then inform the capital requirement.

The regulatory mapping was derived from a version of the single asset model proposed by Merton (1974), which assumes that the change in value of a borrower’s assets is normally distributed. As such, the associated capital requirements depend on the properties of the normal distribution. (In principle, it would be possible to employ an alternative distributional assumption.) Importantly, some judgement is also required about the correlation between different borrowers; how the asset value of one borrower – which, in the Merton model, essentially determines the incidence of default – depends on the asset value of another borrower. In addition, the regulatory capital requirements also depended on the chosen degree of conservatism. Typically, under the original Basel II framework banks were required to hold capital to cover unexpected loss-

es associated with 99.9% VaR over a one-year horizon; in other words, capital buffers should be high enough to cover all but the most extreme of unexpected losses.

Because this approach specifies the entire distribution of losses, the stressed loss – literally the expected loss plus the further unexpected loss, at a given level of stress – can be expressed as a ‘multiple’ of the expected loss:

$$\text{Multiple} = \frac{\text{Stressed losses}}{\text{Expected losses}} \tag{1}$$

This idea is developed by Moody’s (2014) and serves as a useful gauge of how ‘stressful’ a given outcome is: the higher the multiple, the higher the stress. When EAD is the same in both instances, the multiple is simply the ratio of the stressed and expected loss rates.

VaR analysis offers some substantial advantages, such as its practical viability and conceptual attractiveness, as presented by Kupiec (1998) among others from a historical context, and the ability to consider and contrast multiple models and calibrations, for

instance, as demonstrated by Alexander and Sheedy (2008) in the context of currency pairings. Unfortunately, however, VaR analysis spectacularly failed to act as a useful guide to the losses that banks suffered during the financial crisis. In part, this likely reflected modelling and data limitations where risks were calculated on insufficient data or time series that did not incorporate tail events. Figure 1 presents an illustration; based on the previous time series for US banks' residential mortgage charge-offs prior to the financial crisis, simple distribution assumptions akin to a VaR-based approach would have suggested relatively low loss rates during a tail event. This is a deliberately simplistic example: pre-crisis stress tests were not all based on such simple historical observation, and many stress tests are hypothetical rather than historical. However, it is clear from this example that, with hindsight, these types of backward-looking stress tests were far from exacting and greatly underestimated the damage done to banks' balance sheets from the subprime crisis.

4. Developing alternative stress test approaches: a brief literature review

Following the widespread criticism of such VaR-based stresses, different approaches were (re)considered in both the private and official sectors. The sheer volume of academic and public policy work on this topic has been enormous, and this section highlights some themes and papers of interest.

One example of a different approach is so-called 'reverse' stress testing, where the focus is on testing a business plan to failure, i.e., calculating what level of loss would result in the failure (insolvency) of the firm. While reverse stress testing is a useful tool for risk managers, it implicitly relies on underlying models that are similar to those described earlier in order to calculate the probabilistic frequency of the assumed or calibrated loss (or the scenario that generates it) occurring. As such, reverse stress testing lives up to its name; it defines the loss or damage required for business models to fail and then works out the likelihood of that loss occurring. However, the literature on reverse stress tests remains relatively sparse: useful examples include Grundke and Pliszka (2013), who present a quantitative technique based on principal component analysis to reduce the number of risk factors, and McNeil and Smith (2012) who focus on

identifying plausible reverse stress tests, which they call the most likely ruin event.

However, most attention since the financial crisis has focused on a form of stress testing known as 'scenario analysis'. Essentially, this approach relies on some macroeconomic modelling (or assumptions) for key variables that are then 'stressed' based on downside economic scenarios: for instance, Jokivuolle, Virolainen and Vähämaa (2008) model banks' corporate credit risks using macroeconomic variables and then use that model to simulate the impact of a recession on banks' credit costs and capital positions. One key aspect of this approach is that it is explicitly designed to be 'forward looking'; it considers how conditions may evolve in the future. In addition, most scenario analyses focus on multiple periods, in part to try and reflect the way that stresses can evolve over time. It is important to note that neither of these points – either the forward-looking or multiple period aspects – necessarily differentiate scenario-based stresses from VaR-based stresses, as the latter can be both historical (based on past series) or hypothetical (based on posited, unobserved distributions), and conducted over multiple periods. Nonetheless, the popularity of these scenario-based approaches is such that some policymakers have argued that they – combined with other regulatory developments – should replace the IRB approach to risk-weighted capital (Tarullo, 2014).

As noted in the previous section, any forward-looking scenario analysis implies some reliance on a (backwardly estimated or implied) model in order to generate forward scenarios. In essence, the backward-looking aspects of the exercise is often hidden by the modelling apparatus. Previous work such as Berkowitz (1999) had looked at means of incorporating such scenarios into VaR-based approaches. Typically, these applications look to combine distributional assumptions from VaR and scenario-based approaches. However, this work gained little initial traction in part due to the high sensitivity of outcomes to assumptions about how to combine (the implied) distributions (see Aragones, Blanco, & Dowd, 2001).

In a typical scenario-based framework, some worse-than-expected ('downside') outcome is normally asserted or assumed for the aggregate economy. This downside scenario is often summarized using key economic indicators such as GDP, unemploy-

ment and inflation; alternatively, macroeconomic factors (and shocks) can be generated from a large number of different data series, for instance using the type of modelling framework described in Mumtaz *et al* (2009). The scenarios should generally be internally consistent: typically, this might include a country falling into recession and unemployment rising. However, if inflation soared in a recessionary stress scenario, then in the absence of relative price shocks (which by definition are unpredictable) the scenario would probably imply that monetary policy had failed, as it had not managed to anchor the nominal side of the economy.

A key advantage of the scenario-based approach is that it offers a compelling ‘narrative hook’ for non-technicians. In a banking context, the supervisor or the banks themselves can then try to work out what this ‘scenario’ would imply for credit losses, net interest income, and other relevant determinants of bank solvency. This point was first noted by Lopez (2005) in terms of linking potential losses to a ‘specific and concrete’ set of events. However, the specificity and solidity of said events is far from assured given the forward-looking and ultimately hypothetical nature of scenario-based stress tests; to date, no scenario posited in regulatory stress tests has materialized with the effects implied in the stress test. (As noted earlier, this is not surprising given that the exercise seeks to predict the unexpected.) At the same time, the narrative hook can give non-technicians unwarranted comfort about the stress testing approach if the transmission mechanism between the scenario and stressed outcomes – i.e., the implicit modelling approach – is ignored. In fact, scenario-based stress tests are every bit as subject to the modelling uncertainty that is inherent in forecasts, as described in the previous section of this paper.

There have been other criticisms of stress tests based on macroeconomic scenarios; Borio, Drehmann and Tsatsaronis (2012) provide a useful summary, noting that they are not suitable early warning devices and would benefit from complementary information. Concerns have also been raised about the appropriate modelling framework. In many instances, the models that are used to construct the scenario, or relate it for instance to banks’ credit losses, will often have been built around the center of the distribution (Bunn, Cun-

ningham, & Drehmann, 2005; Hoggarth, Sorensen, & Zicchino, 2005). Typically, we might estimate the relationship between default rates and macroeconomic variables over a number of years and then plug a downside scenario into that model to generate losses. However, financial and economic relationships, as in other spheres, can be very different in the tails of the distribution than in the middle of it. One particular concern is the role that using an empirical model to map scenarios onto loss rates, or some other ‘stress’ variable, will unduly influence the range of outcomes. Put simply, if the model is geared towards fitting the central tendency of some dependent variable or data series, it may be poor at capturing the tails of the distribution, which is what stress tests try to explore. For this reason, some recent research has focused on more flexible models that allow these relationships to change as we move into the tail of the distribution (see Covas, Rump, & Zakrajsek, 2013). These quantile models, as introduced by Koenker & Hallock (2001), may offer a better guide to stressed outcomes, but they are not yet widely employed.

However, even these models may still fall short. Critically, the fit of the model will still play an important role in determining stressed outcomes in such conditional analysis: if the overall fit of the model is poor – even for central tendencies – then we should not expect it to perform well when predicting tail outcomes.

5. Simulation analysis: where scenarios fall short

To investigate this phenomenon, a variety of simulations were constructed to compare and contrast the performance of simple proxies for scenario-based and VaR-based stress tests. The following simulation results are not meant as sophisticated statistical advancements of either VaR or scenario-based stress-testing approaches; instead, they are deliberately simple in order to clearly illustrate a key issue when gauging differences between the two approaches. The focus in this section is on comparing simple stress tests based on past events relative to out-of-sample scenario-based approaches.

The methodological framework for the simulation testing is deliberately simple in order to illustrate the differences between VaR and scenario-based ap-

proaches in an accessible manner. The starting point is to define an underlying data-generating process (DGP) for a hypothetical reference series X_t , which can be thought of as a simulated series for bank charge-off rates. A variety of different DGPs were assumed for the reference series: for instance, a simple autoregressive process (equation (2)); and a more complex autoregressive GARCH process (see Bollerslev, 1986; equation (3)).

Simple autoregressive DGP:

$$X_t = \alpha + \varnothing X_{t-1} + \delta_t \quad (2)$$

$$\delta_t \sim N(0, \sigma_\delta^2), \alpha > 0$$

Autoregressive GARCH DGP:

$$X_t = \tau + \beta X_{t-1} + \vartheta_t \quad (3)$$

$$\sigma_{t,\vartheta}^2 = \omega + \gamma \vartheta_{t-1}^2 + \rho \sigma_{t-1}^2, 1 > \beta, \gamma, \rho > 0.$$

Alongside this series, an 'indicator' variable, Y_t , was generated for the reference series, where the error between the indicator and reference series was constrained to be gaussian, and the signal-noise ratio was randomly selected:

Indicator series:

$$Y_t = X_t + \varepsilon_t \quad (4)$$

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

This linkage between the indicator and reference series is meant as a simple representation and illustration of the broader economic modelling approach used in stress tests; it does not imply that all macroeconomic scenario models use a limited number of indicators. However, the distance between the reference series and all explanatory (macro) factors in any model can always be summarized in a single time series, namely the model residual: the simulation approach herein implicitly exploits this fact.

The relationship between the generated reference and indicator series was then estimated using simple regressions, reflecting the associated nature of the two

data series. Both series were generated for 125 periods, and the model was estimated using ordinary least squares (OLS) over this sample:

$$X_t = A + BY_t \quad (5)$$

Illustrative outturns for the reference and indicator series, based on the simple DGP in (2), are shown in Figure 2.

Having constructed in-sample relationships, the simulations then constructed different out-of-sample stresses. As a proxy for a scenario-based stress test, past observations from the tail of the indicator series were selected based on a given level of stress such as a '1 in 100' event, corresponding to the 99th percentile of the distribution. Importantly, varying this stress level had no impact on the broad results presented below. This 'stressed indicator' was then used to construct stressed values for the reference series over the subsequent 25 periods, using the regression model estimates from (5).

After resuming the underlying DGP for the reference series for another 25 observations, stress multiples (equation 1) were constructed for these scenario-based stresses. This is, in essence, a very simple scenario-based stress test.

At the same time, to provide a useful comparison within the same methodological framework, an equivalent backward-looking VaR-based stress was also simulated over the out-of-sample period. In these instances, the stressed reference series were constructed purely from the distribution of the observed reference series over the preceding 125 periods. Importantly, the level of stress was aligned with that for the scenario-based stress; thus, if the stress scenario corresponded to a '1 in 100' event, so too did the simple VaR-based stresses. As before, resulting stress multiples were calculated using equation (1).

These approaches are deliberately simplistic, but they still present a powerful means of comparing the two stress-testing approaches. More complex models and relationships did not have a significant impact on the key results reported below, as confirmed by the use of different DGPs. By running the simulations multiple times as a Monte Carlo exercise, and comparing the multiples from the scenario-based and VaR-based approaches, it was then possible to gauge which approach

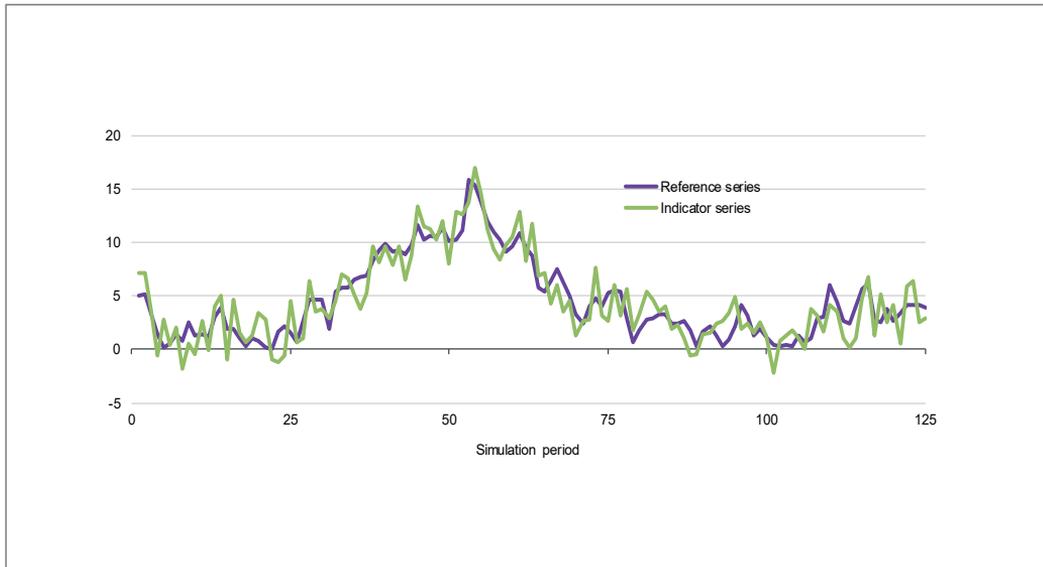


Figure 2. Illustrative reference and indicator series from simulations

was typically more ‘stressful’ in terms of generating higher stressed multiples.

Results for the simple autoregressive DGP are presented below where the autoregressive term in (2) was a random draw between zero and one. Figure 3 presents stress multiples for the scenario-based approach on the left-hand side and multiples for the VaR-based approach on the right-hand side. The results are based on 10,000 simulations, stressing to a ‘1 in 100’ event. The horizontal axis in both charts corresponds to the fit of the indicator model used in the scenario-based stresses. Importantly, both vertical axes have been truncated so that the pattern of main results is not compressed by outliers. Consequently, Figure 4 also summarizes the distributions of the simulated stressed outcomes.

Two key results emerge from the simulations. First, model fit does appear to influence the degree of stress in scenario-based approaches; in the left-hand panel of Figure 2, the results cluster closer to 1 (indicating no gap between stressed and expected outcomes) when model fit is poor and are higher and more dispersed when model fit is better. However, the impact of model fit is relatively small overall: even models that captured the reference series (i.e., fitted the DGP) relatively well

could still result in multiples that are not very different from badly-fitting models. Importantly, the higher level and dispersion of stress multiples for better-fitting models in the scenario-based approach was not consistently evident across different DGPs for the reference series.

Second, and more importantly, the degree of ‘stress’ from the scenario-based approach is typically significantly less than that from the VaR-based approach. The full distribution results presented in Figure 4 confirm this, and formal statistical tests also indicate that the VaR-based approach was more stressful than the scenario-based one. Importantly, these broad results were unaffected even when the underlying DGP for the reference series was changed (for instance, from equation (2) to equation (3)).

This result reflects the ‘error’ between the indicator and reference series as noted in equation (4). In the presence of any volatility in this error term, which in this context can be taken as a proxy for the inability of the scenario model to completely explain the variance of the reference series, the distribution of fitted outcomes from the scenario model will be smaller than the distribution of the historical reference series by definition. As long as the variance of

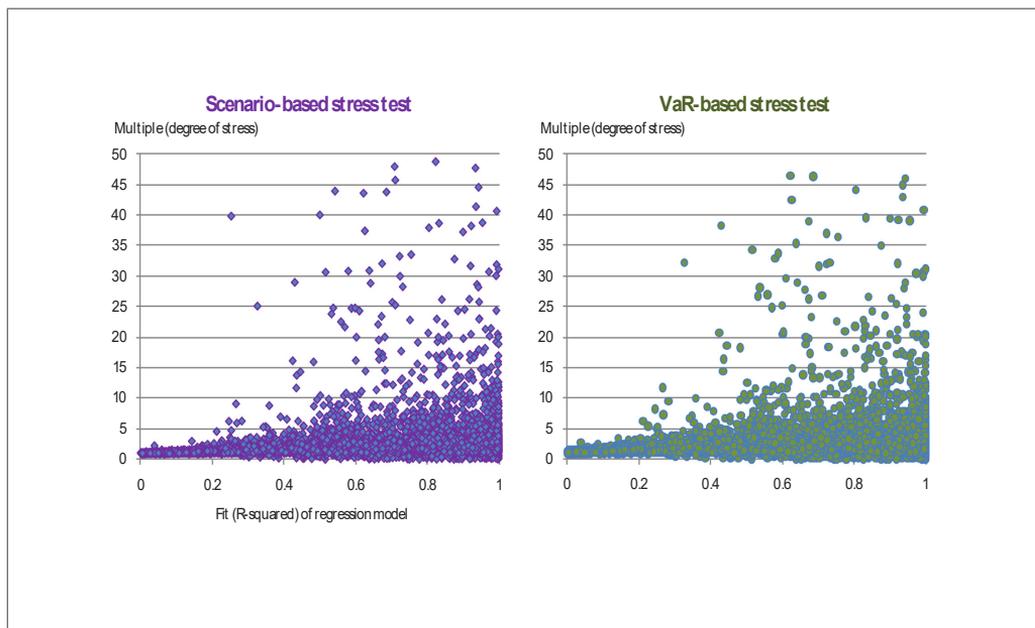


Figure 3. Simulated multiples from different stress-testing approaches (truncated)

Percentage of simulations	Scenario-based stress	VaR-based stress
Multiple > 10	2.30	2.53
5 < Multiple < 10	3.64	4.04
2.5 < Multiple < 5	15.01	18.04
1.25 < Multiple < 2.5	44.29	50.39
Multiple < 1.25	34.76	25.00

Figure 4. Simulated multiples from different stress-testing approaches (summary)

the noise series, σ_ϵ^2 , is non-zero, then there will be a gap between the historical reference series and the implied model-fitted distribution.

In practice, the best that the scenario-based approach can achieve, where the scenario is based on

past values of the indicator series, is to match the performance of the backward-looking VaR approach: this would correspond to the noise series having zero variance, i.e., being a constant that is subsumed by the estimated OLS coefficients. As long as any noise in the

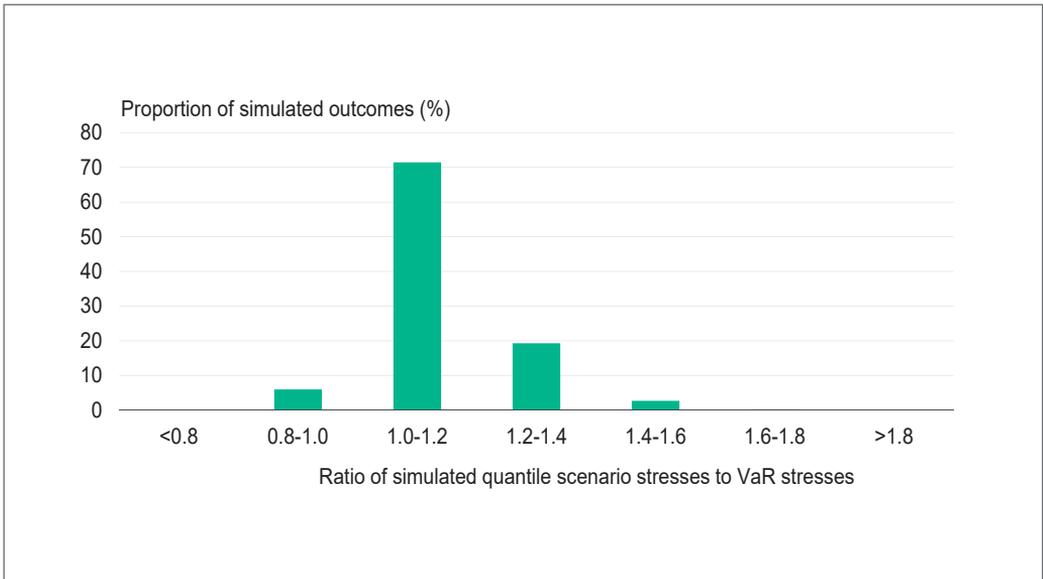


Figure 5. Relative performance of VaR stresses compared with a quantile-model scenario approach

indicator series is orthogonal to the reference series – as implied by OLS models – then the VaR-based approach will tend to generate more stressful outcomes, on average, than simple scenario-based models.

To obtain more extreme stressed outcomes from the simple indicator model, more extreme scenarios – that is, those that are outside the historical experience captured by the indicator series – must be applied. The analogous approach, in a VaR-based stress, would be to calibrate the stress based on some non-observed shift in the series. Such a calibration would be subject to considerable uncertainty, as indeed would the choice of an ‘extreme’ economic scenario. However, in principle, one approach would be to base it on past shocks seen in other spheres or instances, such as that shown in Figure 1. A key finding from this analysis is that to match the implied degree of stress from VaR-based approaches, simple scenario-based stress tests need to employ more extreme scenarios – in this instance, worse than ‘1 in 100’ – than those used in those VaR-based approaches.

Importantly, this gap between VaR-based and scenario-based stresses was only overturned when the functional form of the scenario model changed, in particular when a quantile model was estimated in place of equation (5), with the estimation quantile specified

to a similar percentile as the stresses. However, the gap between the two approaches was not especially pronounced, as illustrated by Figure 5: the mean (median) extra stress from the quantile scenario approach was approximately 13% (10%). Quantile modelling therefore buys stress testers more challenging outcomes than using a simple VaR-based approach, but at the cost of increased complexity.

This analysis indicates that the recent focus on scenario-based approaches does not automatically deliver more stressful outcomes, relative to a VaR-based approach; when both are conducted on a similar basis using historical models, the stress is likely to be lower in the former approach than in the latter.

Perhaps accordingly, some recent research has focused on reviving interest in combining VaR approaches with scenario-driven stress tests, building on Berkowitz (1999). Abdymonunov, Blei and Ergashev (2015) is a good example here, extending and applying the ‘worst in a certain period’ framework proposed by Ergashev (2012) to credit and market risk, while Chen and Skoglund (2013) demonstrate how to incorporate stress scenarios – defined as ‘plausible’ events not necessarily seen in historical data – into model-based risk analysis using Markov-switching simulations. In prin-

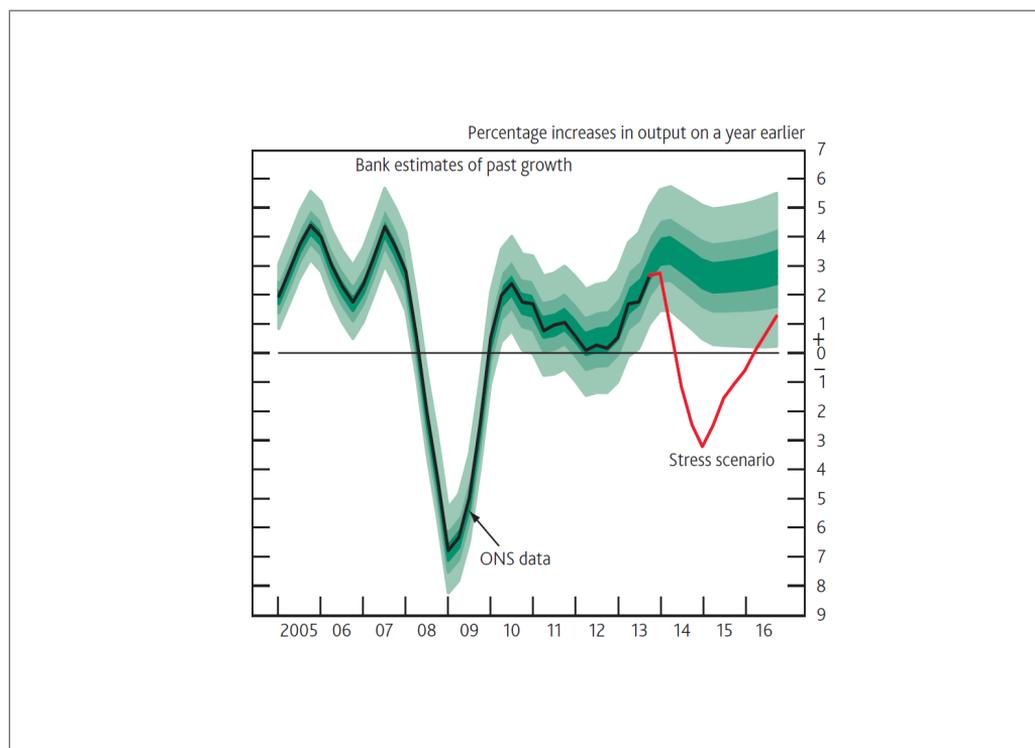


Figure 6. Bank of England forecast and stress test in 2014

Source: Adapted from "Stress testing the UK banking system: key elements of the 2014 stress test" by Bank of England (2014, April). Retrieved from <http://www.bankofengland.co.uk/financialstability/Documents/fpc/keyelements.pdf>

principle, such combined approaches could address concerns about scenario-based stresses being too reliant on in-sample model fit, thereby increasing stresses accordingly. However, these types of scenario-enhanced VaR approaches are not yet widely applied either by regulators or private sector participants.

Furthermore, there are signs that policymakers may already be aware of the relatively less stressful nature of scenario-based stress tests. In its 2014 stress tests, the Bank of England proposed paths for GDP growth, inflation and interest rates that were different from its central forecasts, and in particular represented 'downside' macroeconomic outcomes. Given that the Bank explicitly publishes its fan charts in the form of two-part normal distributions (see Britton *et al*, 1998) and also publishes the parameters behind these fan charts, it is possible to (re)construct the entire forecast distribution and focus on different percentiles as desired.

Figure 6 presents the Bank of England's 2014 scenarios. By examining the parameters underpinning the green fan chart, it is possible to calculate exactly where in the Bank's probability distribution each point in the stress scenario (the red line) falls. The trough in GDP growth at the end of 2014 is below -3%. According to the Bank's own published parameters at the time of its stress test, that observation is weaker than the 0.2nd percentile implied by the fan chart – more extreme than a '1 in 500' event.

It is unlikely that policymakers genuinely want to hold banks to such an exacting standard. But, if they are aware that scenario analysis can potentially understate stresses relative to other approaches, policymakers may be incentivized to create highly stressful scenarios. Unfortunately, one longer-term risk is consequently that these scenarios are deemed to be too incredible and lose public and political support over

time, especially if complex modelling approaches are required. As such, one natural mitigant would be to ensure that alternative approaches such as VaR-based stresses are also employed.

A broader underlying concern is that models based on historical data may not be appropriate when considering stress scenarios that may not have happened in the past. However, this critique applies equally to both VaR and scenario-based stresses. If policymakers (or other stress testers) are concerned that past data may not be a good basis for VaR stresses, they can employ assumptions that are outside the observed historical distribution, in the same way that posited macro scenarios can be outside historical experience. In particular, the implicit shock could be calibrated against previous episodes, for instance, by scaling downside risks for credit cards or corporate loans on the basis of the jump in US banks' mortgage charge-offs during the crisis (Figure 1); in principle, this is no different to choosing an extreme downside (and as yet unseen) macroeconomic scenario for scenario-based tests, as discussed earlier. However, given the inherent uncertainties associated with economic and financial models, as noted above, and the relative performance of different stress-testing approaches, it is clear that multiple approaches are likely to prove more robust than individual ones. Indeed, formal guidance on stress tests for large US banks recommends that the banks consider using a variety of stress testing approaches (see Board of Governors of the Federal Reserve System, 2012). As such, VaR analysis should still have a role to play in stress testing frameworks.

6. Conclusion

Forecasts and models – and increasingly stress tests – are important tools for policymakers and business planners. However, there is a risk that they are not properly understood. Forecasts, for instance, are best thought of as guides to evolving events that let analysts gauge future 'news' in events and data, rather than as completely accurate guides to that future. Similar limitations apply to stress tests, which by definition try to describe what analysts *do not expect* to happen.

Furthermore, the recent shift in focus from VaR-based stress tests to scenario-based stress tests raises

particular concerns. First, the appealing narrative hook of a scenario-based approach could distract non-technicians from the inherent complexities and shortcomings of the underlying modelling approach, which will necessarily be flawed and incomplete. Second, some facets that may be thought of as distinguishing scenario-based tests from VaR-based tests, such as being forward-looking or covering multiple periods, do not in fact distinguish the two approaches; VaR-based tests can be both forward-looking and cover multiple periods. Third, as the analysis in this paper has demonstrated, the stresses arising from simple scenario-based approaches will often be less onerous than those from a similarly-calibrated yet simple VaR-based approach when they are both conducted on the same basis. To address this concern, more complex modelling approaches are required, which in turn may not be widely understood. Happily, there are indications that some policymakers are already aware of this issue, judging from the extreme stress scenarios applied in some recent regulatory stress tests.

References

- Abdymomunov, A., Blei, S., & Ergashev, B. (2015). Integrating stress scenarios into risk quantification models. *Journal of Financial Services Research*, 47(1), 57-79.
- Alexander, C., & Sheedy, E. (2008). Developing a stress testing framework based on market risk models. *Journal of Banking & Finance*, 32(10), 2220-2236.
- Aragones, J., Blanco, C., & Dowd, D. (2001). Incorporating stress tests into market risk modeling. *Derivatives Quarterly*, 7(3), 44-49.
- Bank for International Settlements. (2005, July). An explanatory note on the Basel II IRB risk weight functions. Retrieved from <http://www.bis.org/bcbs/irbriskweight.pdf>
- Bank for International Settlements. (2009, May). Principles for sound stress testing practices and supervision. Retrieved from <http://www.bis.org/publ/bcbs155.htm>
- Bank of England. (2003). Bank's response to the Pagan Report. *Bank of England Quarterly Bulletin*, 43(1), 89-91. Retrieved from <http://www.bankofengland.co.uk/archive/Documents/historicpubs/qb/2003/qb030107.pdf>

- Bank of England. (2009, August). Inflation Report: August. Retrieved from <http://www.bankofengland.co.uk/publications/Documents/inflationreport/ir09aug.pdf>
- Bank of England. (2014, April). Stress testing the UK banking system: key elements of the 2014 stress test. Retrieved from <http://www.bankofengland.co.uk/financialstability/Documents/fpc/keyelements.pdf>
- Berkowitz, J. (1999). A coherent framework for stress testing. *Journal of Risk*, 2(2), 1-11.
- Bernanke, B., Boivin, J., & Elias, P. (2005). Measuring the effect of monetary policy: a factor augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1), 387-422.
- Blanchard, O., & Fischer, S. (1989). *Lectures in Macroeconomics (6th ed.)*. Cambridge, MA: MIT Press.
- Board of Governors of the Federal Reserve System. (2012, May 14). Supervisory guidance on stress testing for banking organizations with more than \$10 billion in total consolidated assets. Retrieved from <http://www.federalreserve.gov/bankinforeg/srletters/sr1207.htm>
- Board of Governors of the Federal Reserve System. (2015, March). Comprehensive Capital Analysis and Review 2015: Assessment Framework and Results. Retrieved from <http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20150311a1.pdf>
- Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.
- Borio, C., Drehmann, M., & Tsatsaronis, K. (2012, January) *Stress-testing macro stress testing: does it live up to expectations?* (Working Paper No. 369). Bank for International Settlements. Retrieved from <http://www.bis.org/publ/work369.pdf>
- Britton, E., Fisher, P., & Whitley, J. (1998). The inflation report projections: Understanding the fan chart. *Bank of England Quarterly Bulletin*, 38(1), 30-37.
- Bunn, P., Cunningham, A., & Drehmann, M. (2005). Stress testing as a tool for assessing systemic risk. *Bank of England Financial Stability Review*, June, 116-126.
- Castle, J., & Ellis, C. (2002). Building a real-time database for GDP(E). *Bank of England Quarterly Bulletin*, 42(1), 42-49.
- Chen, W., & Skoglund, J. (2013). *An integrated stress testing framework via Markov switching simulation*. *Journal of Risk Model Valuation*, 7(2), 3-27.
- Coletti, D., Hunt, B., Rose, D., & Tetlow, R. (1996). *The Bank of Canada's new quarterly Projection Model. Part 3, The dynamic model: QPM* (Technical Report No. 75). Bank of Canada. Retrieved from <http://www.bankofcanada.ca/1996/05/technical-report-no75/>
- Covas, F., Rump, B., & Zakrajsek, E. (2013). Stress-testing U.S. bank holding companies: A dynamic panel quantile regression approach. *International Journal of Forecasting*, 30(3), 691-713.
- Croushore, D., & Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, 105(1), 111-130.
- Ellis, C. (2006, July). *Elasticities, Markups and Technical Progress: Evidence from a State-Space Approach* (Working Paper No. 300). Bank of England. Retrieved from <http://www.bankofengland.co.uk/research/Documents/workingpapers/2006/WP300.pdf>
- Ergashev, B. (2012). A theoretical framework for incorporating scenario analysis into operational risk modelling. *Journal of Financial Services Research*, 41(3), 145-161.
- European Banking Authority. (2014, October 26). Results of 2014 EU-wide stress test; Aggregate results. Retrieved from <https://www.eba.europa.eu/documents/10180/669262/2014+EU-wide+ST-aggregate+results.pdf>
- Flood, D., & Lowe, P. (1995). Inventories and the business cycle. *Economic Record*, 71(1), 27-39.
- Gordy, M. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, 12(3), 199-232.
- Grundke, P., & Pliszka, K. (2013). A macroeconomic reverse stress test (Discussion Paper No. 30/2015). Deutsche Bundesbank. Retrieved from <https://www.econstor.eu/bitstream/10419/119318/1/835092100.pdf>
- Haldane, A. (2009, February 13). Why banks failed the stress test. Retrieved from <http://www.bankofengland.co.uk/archive/Documents/historicpubs/speeches/2009/speech374.pdf>
- Hoggarth, G., Sorensen, S., & Zicchino, L. (2005, November). *Stress tests of UK banks using a VAR*

- approach* (Working Paper No. 282). Bank of England. Retrieved from <http://www.bankofengland.co.uk/archive/Documents/historicpubs/workingpapers/2005/wp282.pdf>
- Jokivuolle, E., Virolainen, K., & Vähämaa, O. (2008). *Macro-model-based stress testing of Basel II requirements* (Research Discussion Paper No. 17). Bank of Finland.
- Koenker, R., & Hallock, K. (2001). Quantile Regression. *Journal of Economic Perspectives*, 15(4), 143-156.
- Kupiec, P. H. (1998). Stress testing in a value at risk framework. *The Journal of Derivatives*, 6(1), 7-24.
- Lopez, J. (2005). Stress tests: Useful complements to financial risk models. *Federal Reserve Board of San Francisco Economic Letter*, 2005-14. Retrieved from <http://www.frbsf.org/economic-research/files/el2005-14.pdf>
- Merton, C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29(2), 449-470.
- McNeil, A. J., & Smith, A. D. (2012). Multivariate stress scenarios and solvency. *Insurance: Mathematics and Economics*, 50(3), 299-308.
- Moessner, R., Zhu, F., & Ellis, C. (2011, February). *Measuring Disagreement in UK Consumer and Central Bank Inflation Forecasts* (Working Paper No. 339). Bank for International Settlements. Retrieved from <http://www.bis.org/publ/work339.pdf>
- Moody's (2014). Expected and unexpected bank losses: revisiting the Basel approach. Available from https://www.moody.com/researchdocumentcontentpage.aspx?docid=PBC_167213
- Mumtaz, H., Zabczyk, P., & Ellis, C. (2009, March). *What lies beneath: what can disaggregated data tell us about the behaviour of prices?* (Working Paper No. 364). Bank of England. Retrieved from <http://www.bankofengland.co.uk/research/Documents/workingpapers/2009/wp364.pdf>
- Pagan, A. (2003). Report on modelling and forecasting at the Bank of England. *Bank of England Quarterly Bulletin*, 43(1), 60-88.
- Pain, N., Lewis, C., Dang, T. T., Jin, Y., & Richardson, P. (2014). OECD Forecasts During and After the Financial Crisis: A Post Mortem (Working Papers No. 1107). OECD Economics Department. Retrieved from <http://www.oecd-ilibrary.org/docserver/download/5jz7311qw1s1-en.pdf?expires=1495806565&id=id&accname=guest&checksum=EA2FEFBBF460EE6BFDEBAECE5E4224FA4>
- Tarullo, D. (2014, May 8). Rethinking the Aims of Prudential Regulation. Retrieved from <http://www.federalreserve.gov/newsevents/speech/tarullo20140508a.htm>
- Wallis, K. (2004). An assessment of Bank of England and National Institute inflation forecast uncertainties. *National Institute Economic Review*, 189(1), 64-71.

