CALIBRATING INITIAL SHOCKS IN BANK STRESS TEST SCENARIOS:

AN OUTLIER DETECTION BASED APPROACH

Olivier DARNE[1], Guy LEVY-RUEFF[2] and Adrian POP[3]

This version: *November 2012*

[Preliminary version]

Abstract: We propose a rigorous and flexible methodological framework to select and calibrate initial shocks to be used in bank stress test scenarios based on statistical techniques for detecting outliers in time series of risk factors. Our approach allows us to characterize not only the *magnitude*, but also the *persistence* of the initial shock. The stress testing exercises regularly conducted by supervisors distinguish between two types of shocks, *transitory* and *permanent*. One of the main advantages of our framework, particularly relevant as regards the calibration of transitory shocks, is that it allows considering various reverting patterns for the stressed variables and informs the choice of the appropriate stress horizon. We illustrate the proposed methodology by implementing outlier detection algorithms to several time series of (macro)economic and financial variables typically used in bank stress testing.

Keywords: Stress testing; Stress scenarios; Financial crises; Macroprudential regulation
JEL codes: G28, G32, G20, C15

---

[1] University of Nantes (LEMNA), Institute of Banking and Finance, Chemin de la Censive du Tertre, BP 52231, 44322 Nantes Cedex 3, France, Tel.: +33-2-40-14-17-33, Fax: +33-2-40-14-16-50, E-mail: olivier.darne@univ-nantes.fr

[2] Banque de France, Business Conditions and Macroeconomic Forecasting Division, 39 Rue Croix des Petits Champs, 75049 Paris Cedex 01, France, Tel.: +33-1-42-97-73-43, E-mail: guy.levy-rueff@banque-france.fr

[3] University of Nantes (LEMNA), Institute of Banking and Finance, Chemin de la Censive du Tertre, BP 52231, 44322 Nantes Cedex 3, France, Tel.: +33-2-40-14-16-54, Fax: +33-2-40-14-16-50, E-mail: adrian.pop@univ-nantes.fr. This research project was conducted while Adrian Pop was consultant to the French Prudential Supervision Authority. (*Corresponding author*)

1. Introduction and motivation

From a macroprudential perspective, the main objectives of a stress testing exercise are to properly identify the risk drivers and vulnerabilities that are most likely to generate financial instability and assess the resilience of the banking system to various macroeconomic or financial shocks. Particularly, they allow supervisors to identify the relevant transmission channels of extreme, but still plausible, events affecting the stability of the banking system.

Since the beginning of the nineties, stress testing exercises have been regularly conducted in central banks and supervisory authorities, as well as in major financial institutions. The Basel II Capital Accord that applies to large international banks since the beginning of 2007 formally ask bank managers to carry out regular stress tests under the Pillar 1 guidelines for internal model validation purposes (see notably BCBS, 2006, §434-435).[4] Moreover, Basel II also refers to stress tests under the supervisory approach in Pillar 2 by highlighting that banks should consider the results of such tests in their capital planning process (see BCBS, 2006, §726).

The topic of stress testing in banking has received renewed attention and has been vividly debated since the inception of the global financial crisis in the summer 2007. For instance, in the US, the Federal Reserve decided on May 2009 to disclose the results of stress tests conducted by the 19 major bank holding companies in the country. In the same vein, the European authorities followed the US in carrying out European Union-wide stress tests by September 2009. The main features of the European stress scenario, common to all participating banks, were calibrated by the European Central Bank, but the final scenario imposed to banks was adapted at the national level by each supervisory authority.

Some observers argued that the US and European stress scenarios were too mild to generate credible results. In the US for instance, at the end of the first quarter of 2009, actual data on the considered stressed variables, i.e. GDP growth, unemployment rate, and home prices,

---

[4] Interestingly, the market risk amendment to the Basel I capital accord proposed by the Basel Committee in 1996 already contains some references to stress testing (see BIS, 1996, pp. 46 *et passim*). Precisely, the supervisory approval to use internal models to set capital charges for market risks is granted only to banks that run regularly rigorous and comprehensive stress tests. The several examples of stress scenarios mentioned in the 1996 amendment of the first Basel capital accord are all based on past episodes of market turbulence like the October 1987 stock market crash, the Exchange Rate Mechanism crisis of 1992-1993 or the fall in bond markets at the beginning of 1994.

have been already *worse* than the adverse scenario proposed for 2010. In Europe, analysts expressed concerns that stress scenarios were underestimating the potential losses on banks' trading portfolios from depreciations of sovereign (particularly Spanish and Greek) bonds. By taking a more fundamental, macroprudential, perspective, Borio, Drehmann and Tsatsaronis (2012) critically review the state of the art in macro stress testing and conclude that stress tests failed at the very junctures when regulators needed them. In their opinion, (macro) stress tests are a useful tool for crisis management and resolution, but are unreliable as an early-warning device during tranquil periods of time, when risks are quietly building up. Indeed, recent empirical evidence suggests that financial crises tend to start developing far before the bust, at the peeks of medium-term financial cycles (see Drehmann, Borio and Tsatsaronis, 2012).

Clearly, the reliability and usefulness of the stress test results largely depend on the assumptions backing the stress scenario, as well as on the methodology used to link the risk factors to the financial strength indicators. A study on sound stress testing practices published by the Basel Committee in January 2009 sums up several deficiencies inherent in stress testing exercises conducted by the largest banks and their supervisory authorities before and during the current financial crisis (see BCBS, 2009). Besides a lack of integration of stress tests into the broad risk governance process, deficiencies resulted from disregarding specific risks, including securitization risks, and the failure to take into account feed-back effects, the study has also highlighted some basic flaws in the design of stress scenarios. Particularly, it seems that stress scenarios implemented before and during the subprime crisis only reflected mild and temporary shocks and assumed that those shocks were maintained over short stress periods. Most extreme scenarios were often considered as highly implausible by senior managers and hence have been routinely dismissed as improbable.[5]

---

[5] The FSA (2006) also criticizes the deficiencies in the calibration of stress scenarios by the vast majority of UK financial firms: "*We were struck by how mild the firm-wide stress events were at some of the firms we visited. On the evidence of our review, few firms were seeking out scenarios such as those that might […] result in shortfalls against capital requirements while still remaining plausible.*" Haldane (2009) attributes the modest severity of bank stress scenarios before the current financial crisis to three micro-economic frictions: (*i*) disaster myopia; (*ii*) network externalities; and especially (*iii*) misaligned incentives. In particular, risk managers simply didn't have enough incentive to conduct severe stress testing exercises and to report the results to senior managers.

In light of these weaknesses, the aim of the present paper is to contribute to the growing economic literature on stress testing in banking by focusing on a key methodological issue: the design and calibration of initial shocks to be used in stress scenarios. Although our analysis mainly concerns the "macro" stress testing, the basic approach presented in the paper is also applicable to sensitivity analyses and stress tests performed at a bank level.

We propose a rigorous and flexible methodological framework to select initial shocks based on statistical techniques conceived to detect outliers and structural breaks in financial and (macro)economic time series. Our statistical approach allows us to characterize not only the *magnitude* or *severity* of the initial shock, but also its *persistence* through time. This feature of our framework is relevant to the current debate around the appropriate methodology to generate "extreme, but plausible" shocks and stress scenarios. Indeed, the size of the shock assumed at the beginning of the stress period ("*how severe the stress should be?*") and the appropriate length of the stress period ("*over what time horizon the stress should be maintained?*") largely determine the quality of the outcome. Setting the magnitude of the initial shock too high or, more often, too low would undermine either the credibility or the meaning of the whole stress testing exercise.[6] In the same vein, choosing a too narrow or a too wide window for the stress period would either not be sufficient for risk factor to fully materialize or not be compatible with the usual assumptions that banks do not reallocate their portfolios during the whole stress period (see e.g. De Bandt and Oung, 2004). As pointed out by Drehmann (2008), there seems to be no golden rule for the optimal horizon to consider in stress testing. However, according to Isogai (2009) and Sorge (2004), although there is no generally accepted rule in setting the appropriate magnitude of the initial shock or the stress horizon, there is considerable room to apply "objective criteria" in specifying stress scenarios and to account for these criteria when interpreting the stress results.

The article is organized as follows. Section 2 focuses on the role of initial shocks in the design of stress scenarios and discusses the related literature containing conceptual, as well as methodological, contributions. Section 3 proposed a conceptual taxonomy of outliers based on the assumed dynamics of the stressed variables and describes the statistical procedures

---

[6] For more thorough discussions on the credibility of stress testing exercises, see Sorge (2004). Since there is a considerable level of discretion over the choice of the stress scenario, it is hardly difficult to set a clear standard of plausibility or to identify a threshold level of magnitude for the initial shock beyond which the plausibility of the stress scenario can be taken for granted.

for detecting outliers and structural breaks in risk factors or, more generally, in economic and financial time series. In section 4 we propose numerical examples that allow us to illustrate the relevance of the outlier detection methodology for the calibration of initial shocks and to highlight some practical implementation issues. Finally, Section 5 concludes.

2. Shocks and stress scenarios: related literature

There is a vast and growing literature analyzing the many facets of a stress testing exercise and highlighting a number of methodological issues and challenges. Given the aim of the present paper, we decided to focus our literature review on studies investigating the calibration of initial shocks and stress scenarios. To set the stage, we begin by briefly discussing some relevant conceptual papers and then we review several related methodological contributions to this literature.

Foglia (2009) proposes an extensive survey of the stress testing methodologies in use at major central banks and supervisory authorities. While the focus of her survey is on credit risk from a "macro" stress testing perspective, it already reveals a high level of methodological diversity in the approaches adopted by banking authorities. The same degree of heterogeneity in stress testing practices has also been noticed at a financial institution level (see the surveys conducted by the Committee on the Global Financial System, CGFS, 2000, 2001, 2005). It is however worth noting that whatever the adopted methodology is, the implementation of a stress testing exercise takes place in several common stages (see also Sorge, 2004; Bunn, Cunningham and Drehmann, 2005; Drehmann, 2009; Borio, Drehmann and Tsatsaronis, 2012): (*i*) the selection of the initial shock or combination of initial shocks; (*ii*) the quantification of the impact of the simulated shock on the macroeconomic environment; (*iii*) the assessment of the effect of the shock on the probability of default of borrowers and asset prices in general; and (*iv*) the measurement of the induced effect on the profitability and solvability of banks and the macro-financial stability.[7]

---

7 Some recent sophisticated models add a fifth stage to the stress testing architecture consisting of (*v*) the modeling of the so-called "feed-back" or "second round" effects between the financial (banking) sector and

As far as the first stage is concerned, Borio et al. (2012) correctly point out that the literature does not make a clear distinction between "shocks" and "scenarios." For some authors, the term "scenario" describes a set of exogenous shocks, while for others the same term designates both the set of exogenous shocks and their estimated, model-specific, impact on the macroeconomic environment. Generally speaking, in the first stage, the initial shocks are calibrated using either hypothetical scenarios based on experts' opinions and the economic expertise of staff or historical data.[8] Table 1 synthesizes the main advantages and drawbacks of each of the two approaches used to select stress scenarios, while Table 2 provides several examples of stress scenarios in use at major financial institutions. Hypothetical scenarios are particularly useful when the aim of the stress testing exercise is to assess the impact of shocks related to innovative risk factors and new financial products for which sufficiently long time-series are rarely, or simply not, available. The current financial crisis has revealed a worried lack of imagination among senior managers that led to an underestimation of the impact of low probability events due to cognitive biases and to a "false sense of security" (see Borio and Drehmann, 2009; BIS, 2009). Hence, it is not fortuitous that the Basel Committee strongly recommends that stress testing programs should include a large specter of hypothetical shocks and scenarios in order to scrutinize new emerging risk drivers related to innovative financial products that are impossible to replicate from previous crisis episodes. Hypothetical shocks are by their very nature forward-looking and help to address this concern. Another advantage of hypothetical, as opposed to historical-based, shocks and stress scenarios is that they may be tailored to the risk profile and specific composition of the bank asset portfolio. The most common examples of hypothetical scenarios in macro stress testing of system-wide credit risk are based on large movements in economic growth prospects, e.g. unexpected slowing of the global demand; rises in interest rates and oil prices; a widening of sovereign credit spreads; a severe decline in stock prices; geopolitical tensions or terrorist attacks (see Table 2 and CGFS, 2005 for a survey).

---

the real economy. The literature on macro feedbacks is promising, but still in early and incipient phases (see e.g. Jokivuolle, Virén and Vähämaa, 2010, and references therein).

[8] The taxonomy of stress scenarios is discussed at length in several related conceptual papers. See e.g. Blashke, Jones, Majnoni and Martinez-Peria (2001); Cihak (2004, 2007); CGFS (2005); Drehmann (2008, 2009); Isogai (2009); Sorge (2004). We only insist here on the main differences between historical-based and hypothetical scenarios and the need for rigorous quantitative methods in calibrating shocks and stress scenarios.

The most serious drawback of hypothetical scenarios is that they may lack credibility and are hence relatively easy to dismiss by senior managers as implausible. Indeed, as there is no objective rule or guidance in setting the magnitude or the persistence of shocks in hypothetical scenarios, which may sometimes include "unimaginable" events, they tend to be less credible and plausible than historical-based scenarios.

{Table 1}

By contrast, historical shocks tend to be based on more rigorous selection criteria and objective guidance grounded on past data. Consequently, they are harder to dismiss by senior managers as implausible and get more acceptance. They are also more credible and plausible than hypothetical shocks and are intuitively possible since the considered extreme movements in risk factors actually occurred at some point of time in the past. The main drawback of historical-based shocks and scenarios is that they are generally based on some parametric assumptions that may not be valid under stress conditions and especially on the assumption that future crises are somewhat similar to past crises. They may also lead to an underestimation of the impact in the particular case of innovative risk factors and new emerging financial products for which sufficiently long time-series are in fact not available.[9] The vast majority of historical stress scenarios focus on a number of major turbulence episodes observed in the past: the 9/11 terrorist attacks on the US; the "Black Monday" 1987 stock market crash; the 1998 market turmoil due to the LTCM collapse and the Russian debt default; the 1997 Asian crisis (see Table 2).

{Table 2}

Our brief description of the main advantages and drawbacks associated to historical and hypothetical shocks and scenarios suggests that important complementarities may (and should) exist between these two main approaches commonly used in the practice of stress testing.[10] Consequently, as both approaches are useful and help to address the same

---

[9] For instance, standard historical scenarios are not able to replicate extremely large and infrequent movements in risk drivers, such as those observed in the fast-growing market for credit derivatives (CDS, CDOs…) during the subprime crisis, simply because the series available before the crisis only covered a short-length expansionary period.

[10] Probably the best way to illustrate this kind of complementarities is to emphasize that, in practice, financial firms often embrace "hybrid" approaches, which consist in using extreme market movements observed in the

problem from different, but complementary, perspectives, the relevant question is not how to choose between them. However, as previously mentioned, the hypothetical approaches are by their very nature subject to arbitrariness and manipulation since no strict criteria apply for scenario selection. They are indeed more an art than a science. By contrast, rigorous statistical methods can and should be applied in order to calibrate historical-based, probabilistic, shocks.

The basic idea of the present paper is to emphasize that the initial shocks considered at the very early stage of any stress testing exercise are in fact "rare events" *per se*. Consequently, the principles of time series intervention analysis that we formally describe in the next section are particularly useful in this context. They allow us to detect outliers and structural breaks in (macro)economic and financial time series, to characterize the length of the stressed period, as well as the magnitude and the reverting dynamics of extreme (but still plausible) shocks to the main risk factors.

<div align="center">

*

*   *

</div>

There are few methodological papers that examine the design of stress scenarios using statistical approaches. As far as the macro-stress testing exercises are concerned, the "stressed" variables within a consistent scenario might be calibrated using structural econometric models or vector autoregressive / error-correction (VAR/VECM) models estimated by central banks for forecasting and monetary policy purposes.[11] An alternative statistical based approach to the calibration of stress scenarios is proposed by Boss, Breuer, Elsinger, Jandacka, Krenn, Lehar, Puhr and Summer (2006) within the Systemic Risk Monitor (SRM) tool developed at the Oesterreichische Nationalbank. Basically, stress scenarios are simulated by drawing randomly vectors of risk factor changes from a multivariate distribution constructed in two steps. First, marginal risk factor distributions are inferred

---

past -- not necessarily related to a specific crisis episode -- to form the basis for hypothetical scenarios. Another example is the use of exceptional events that occurred in the past (e.g. the 9/11 terrorist attacks; the 1972-1973 oil price shock; the 2003 Iraqi war), as well as the subsequent extreme movements in risk factors, to inform both hypothetical and historical scenarios.

[11] These macro-econometric models are extensively reviewed in Foglia (1999). Consequently, they will not be discussed here. Rather, in what follows, we focus our literature review on empirical studies that propose pure statistical approaches, more in line with the aim of the present paper.

based on uni-variate statistical tests that identify, for each relevant risk factor, an empirical model exhibiting the best out-of-sample performance. Second, the correlation structure of the risk factors is modeled by fitting a multivariate *t*-copula to the past data.[12]

The trade-off between plausibility and severity in calibrating stress scenarios is addressed explicitly by Breuer, Jandačka, Rheinberger and Summer (2009) within a one-period quantitative risk-management framework.[13] They propose a measure of plausibility based on the so-called *Mahalanobis* radius, i.e. the number of standard deviations of the multivariate move from the center of mass characterizing the distribution of the *systematic* risk factors to the stress test point. After retaining all scenarios above a minimum plausibility threshold level, the stress scenarios are identified by performing a systematic worst-case search over the given plausible admissibility domain. This framework has clear advantages over the traditional "hand-picked" approach to select stress scenarios: e.g. no harmful scenarios are missed; no implausible scenarios are retained; and stress scenarios are portfolio-specific. However, as noted by Isogai (2009) and Breuer and Csiszár (2012), the approach is only applicable to a family of elliptical (including the normal and *t*-Student) distributions of risk factors and thus may not be suitable for non-elliptically distributed risk factors.

The statistical approach of Breuer *et al*. (2009) was extended by Breuer, Jandačka, Mencía and Summer (2012) to a multi-period setting, in which stress scenarios are defined as paths (and not point-in-time values) of macroeconomic variables. Breuer *et al*. (2012) convincingly show the practical relevance of their approach. Compared with the stress testing results generated by traditional approaches, calibrated on real data prior to the subprime crisis, the worst-case scenario comes much nearer to the severity of the economic downturn observed since 2008.[14]

---

[12] While the statistical approach described in Boss *et al*. (2006) presents a certain number of advantages (e.g. explicit modeling of the joint behavior of the stressed variables and of the dependency structure), Foglia (2009) expressed some doubts that it is highly suitable for storytelling and communication purposes.

[13] This is a significant extension of a basic idea developed in an earlier work by Breuer and Krenn (2001).

[14] The traditional approaches considered in Breuer *et al*. (2012) are a synthetic one-step scenario, in which GDP growth decreases by three standard deviations in the first quarter and then reverts to its long term path, and a historical recession scenario (the 1992 recession in Spain). The worst-case scenario is calibrated by applying the original methodology proposed by the authors on Spanish data up to December 2006.

The length of the stress period plays a key role in a related analysis proposed recently by Varotto (2012). He estimates expected credit losses for individual exposures and representative bank portfolios under the most severe default scenario recorded in history: the Great Depression scenario. The worst case capital based on this extreme scenario is shown to be highly sensitive to the stress horizon. Specifically, by expanding the stress period from one year to three years, he reveals that the worst case capital increases more than three times. Consequently, over a longer (e.g. 3-year) stress horizon, banks having low quality portfolios would not be able to limit losses within their Basel II required minimum.

Finally, Dubecq and Gourieroux (2012) propose an interesting distinction between shocks on variables and shocks on distributions within a theoretical framework that allows examining the formal link between the two types of shocks. They illustrate the relevance of this distinction by applying a stress methodology to a portfolio of European sovereign bonds held by a financial institution. The methodology consists in first identifying a Eurozone systematic factor using a standard principal components analysis. Then, the distribution of the systematic factor before the financial crisis of 2007 is considered to set up the so-called "baseline" scenario, while the distribution over the crisis period is interpreted as the "contaminating" distribution (stress scenario). They focus their analysis on the effects of the stress scenario viewed as a "contamination" of the Eurozone systematic risk factor on both crystallized and optimally updated portfolios.

Compared with the other papers surveyed in this section, our paper focuses on the calibration of *initial* shocks at the very early stage of the stress testing exercise and addresses a number of somewhat different, albeit complementary, research questions. E.g., "*what should be the appropriate length of the stress period?*"; "*what are the most appropriate reverting dynamics for the initial shocks to risk factors if they are supposed transitory?*"; "*are these reverting patterns specific to the considered risk factors or to the stress horizon?*" It is worth noting that the objective of the present paper is *not* to design a whole coherent macro stress scenario, but to inform the calibration of initial shocks to be used at the very early stage of the stress test. Despite its appealing features, the method we describe in the next section, based on the detection of outliers and structural breaks in (macro)economic and financial data, has never been used, to our knowledge, in the stress testing literature.

Some recent contributions to the stress testing literature underline the need to calibrate *portfolio-specific* shocks, that is, worst-case shocks that reflect portfolio-specific dangers and suggest efficient risk reducing actions (see Breuer et al., 2009; 2012). The basic idea behind such a proposal is that a worst-case shock for a given portfolio may be harmless for another portfolio.[15] It is worth noting that the calibration of initial shocks proposed in the next section is by no means intended to be portfolio-specific. We see this feature as an advantage rather than a limitation of our method. Indeed, our analysis mainly applies to "macro" stress testing exercises (either top-down or bottom-up), for which the comparability of results across financial institutions is crucial. To give an extreme example, the Europe-wide banking stress tests are based on exogenous shocks and stress scenarios that are not (and should not be) portfolio-specific.

3. Outlier taxonomy and detection algorithms

The initial shocks considered at the very early stage of a stress testing exercise may be viewed as the materialization of rare or exceptional events that affect dramatically the dynamics of relevant risk factors. As such, they have their peers in time-series econometrics: the *outliers*. By their very nature, outliers are present in most economic time-series and are due to exceptionally rare events that cause sudden transitory or persistent shifts in the level of economic or financial variables. Time-series econometrics contains well-established statistical techniques to detect and characterize outliers, most of which are based on the so-called "intervention analysis" proposed in their seminal work by Box and Tiao (1975) and applied to linear models. Statistical procedures similar in scope were proposed by Tsay (1988) and improved by Balke and Fomby (1994).

This section describes a sophisticated iterative algorithm developed by Chen and Liu (1993) that can be implemented through the statistical package "TRAMO", originally developed by

---

[15] Breuer et al. (2009) cite the Basel Committee on Banking Supervision noting that a bank should "*develop its own stress tests which it identifies as most adverse based on characteristics of its portfolio*" (see BCBS, 2006, p.198, §718(LXXXIII)). However, besides the scenarios developed by the bank itself to capture specific characteristics of its portfolio (§718(LXXXIII) and §718(LXXXIV)), the Basel Committee also discuss supervisory scenarios (§718(LXXXI)) and other scenarios requiring a simulation by the bank (§718(LXXXII)): the 1987 equity crash; the ERM crisis of 1992 and 1933 or the bond market turbulence in the first quarter of 1994. These last scenarios are clearly not portfolio-specific, but they have the great advantage of preserving the comparability of results across institutions.

Gómez and Maravall (1997, 2001).[16] The algorithm was also applied in different contexts, e.g. by Darné and Diebolt (2004) to detect and correct outliers in macroeconomic time series.

Let y_t^* denoting the univariate time series of a risk factor relevant to the bank asset portfolio that can be described by the following $ARIMA(p, d, q)$ process:

$$\alpha(B)\phi(B)y_t^* = \theta(B)a_t \qquad [1]$$

where B is the lag operator; a_t represents a white noise; $\alpha(B)$, $\phi(B)$ and $\theta(B)$ are the lagged polynomials of degree p, d and q, respectively. Recall that the ARIMA process satisfies the usual conditions that all roots of $\phi(B)$ and $\theta(B)$ are outside the unit circle, while the polynomial $\alpha(B)$ has all roots on the unit circle.

The outliers may be expressed as regression polynomials in the following manner:

$$y_t = y_t^* + \sum_{j=1}^{m} \omega_j v_j(B) I_t(\tau_j) \qquad [2]$$

where y_t^* is the ARIMA process defined *infra* or intuitively the "uncontaminated" (but unobserved) series of the risk factor, while y_t describes the observed evolution of the risk factor; $v_j(B)$ is the polynomial characterizing the outlier occurring precisely at time $t = \tau_j$; ω_j measures the impact of the outlier on the series; $I_t(\tau_j)$ is an indicator function that takes the value of 1 if $t = \tau_j$ and 0 if $t \neq \tau_j$; m is the number of detected outliers. Note that the point in time $t = \tau_j$ at which the shock to the time series occurs may be unknown *ex-ante*.

According to the taxonomy proposed by Chen and Liu (1993), there are four main classes of outliers defined with respect to the expression assumed for the polynomial $v_j(B)$:

1. *Additive outliers* (AO), which have a "one-shot" effect on the observed series. They affect a one single observation at a given point in time, without influencing any other subsequent observation. In this particular case, the regression polynomial is naturally set to one: $v_{AO}(B) = 1$.
2. *Innovative outliers* (IO), which appear as atypical observations in the noise process and affect only temporarily the time series. The regression polynomial is expressed in

---

this case as: $\upsilon_{IO}(B) = \theta(B)/\alpha(B)\phi(B)$. Note however that IOs will produce *transitory* effects when the series are stationary and *permanent* level shifts when the series are non-stationary (see Chen and Liu, 1993, for further details).

3. *Level shifts* (LS), which increase or decrease all the observations from a certain point in time onward by some constant level. Consequently, LSs produce abrupt and persistent changes to the series. In this case, the regression polynomial becomes: $\upsilon_{LS}(B) = 1/(1 - B)$.

4. *Temporary or transitory changes* (TC), which produce an initial shock to the series and then the effect vanishes with passing time. The revert speed depend on the parameter δ, with $0 < \delta < 1$. In this final case: $\upsilon_{TC}(B) = 1/(1 - \delta B)$.

Note that, by analogy, "level shifts" correspond to "permanent shocks," while "temporary changes" correspond to "transitory shocks" in the stress testing terminology. The outlier detection algorithm consists in fitting an ARIMA model to y_t^* in equation [1] and then obtaining the estimated residuals, denoted by $\hat{a}_t$:

$$\hat{a}_t = \pi(B)y_t \qquad [3]$$

where

$$\pi(B) = \frac{\alpha(B)\phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 - \cdots \qquad [4]$$

Depending on the type of outlier modeled by the regression polynomials described by the equation [2], the estimated residual are expressed as follows:

$$\begin{cases} \text{AO: } \hat{a}_t = a_t + \omega_{AO}\pi(B)I_t(\tau) \\ \text{IO: } \hat{a}_t = a_t + \omega_{IO}I_t(\tau) \\ \text{LS: } \hat{a}_t = a_t + \omega_{LS}\ [\pi(B)/(1 - B)]I_t(\tau) \\ \text{TC: } \hat{a}_t = a_t + \omega_{TC}\ [\pi(B)/(1 - \delta B)]I_t(\tau) \end{cases}$$

These expressions can be viewed as regression models where the dependent variable is the estimated residual $\hat{a}_t$, that is:

$$\hat{a}_t = \omega_i\, x_{i,t} + a_t, \text{ for } i \in \{AO, IO, LS, TC\} \qquad [5]$$

where

$$x_{i,t} = \begin{cases} 0 \text{ for all } i \text{ and } t < \tau \\ 1 \text{ for all } i \text{ and } t = \tau \end{cases}$$

Finally, for $t > \tau$ and $k \geq 1$,

$$\begin{cases} \text{AO: } x_{AO,t+k} = -\pi_k \\ \text{IO: } x_{IO,t+k} = 0 \\ \text{LS: } x_{LS,t+k} = 1 - \sum_{j=1}^{k} \pi_j \\ \text{TC: } x_{TC,t+k} = \delta^k - \sum_{j=1}^{k-1} \delta^{k-j}\pi_j - \pi_k \end{cases}$$

The detection of the various types of outliers is based on likelihood ratio statistics computed as follows:

$$\begin{cases} \text{AO: } \hat{\tau}_{AO}(\tau) = [\hat{\omega}_{AO}(\tau)\hat{\sigma}_a] / \left(\sum_{t=\tau}^{n} x_{AO,t}^2\right)^{1/2} \\[2ex] \text{IO: } \hat{\tau}_{IO}(\tau) = \dfrac{\hat{\omega}_{IO}(\tau)}{\hat{\sigma}_a} \\[2ex] \text{LS: } \hat{\tau}_{LS}(\tau) = [\hat{\omega}_{LS}(\tau)/\hat{\sigma}_a] / \left(\sum_{t=\tau}^{n} x_{LS,t}^2\right)^{1/2} \\[2ex] \text{TC: } \hat{\tau}_{TC}(\tau) = [\hat{\omega}_{TC}(\tau)/\hat{\sigma}_a] / \left(\sum_{t=\tau}^{n} x_{TC,t}^2\right)^{1/2} \end{cases}$$

where $\hat{\omega}_i(t) = \sum_{t=\tau}^{n} \hat{a}_t x_{i,t} / \sum_{t=\tau}^{n} x_{i,t}^2$, for $i \in \{AO, LS, TC\}$ and $\hat{\omega}_{IO}(t) = \hat{a}_t$.

$\hat{\omega}_i(\tau)$, for $i \in \{AO, IO, LS, TC\}$ denotes the estimation of the effect of the shock at time $t = \tau$, and $\hat{\sigma}_a$ is an estimate of the standard deviation of the residual process (see Chang *et al.*, 1988, for details). When the risk factor is a financial variable (e.g. stock index), a potential source of misspecification is the (conditional) heteroscedasticity, which may inflate artificially the standard errors of estimators. Therefore, in this case we use heteroscedasticity-consistent covariance matrices proposed by Newey and West (1987, 1994).

Outliers are identified through running a sequential procedure, consisting of outer and inner iterations. In an outer iteration, assuming that there are no outliers in the series, an $\text{ARIMA}(p,d,q)$ model is estimated and the resulted residuals are retained. The results from the outer iteration are then used in the inner iteration to detect outliers. For that purpose, likelihood ratio test statistics are computed separately, for each observation and for each of the four types of outliers. The largest absolute value of these test statistics,

$$\hat{\tau}_{max} = \max\{| \hat{\tau}_i(\tau) |\} \tag{6}$$

14

is then compared to a pre-specified critical value based on simulation experiments. If the test statistic is larger than the critical value, an outlier is found at time $t = \tau$.

When an outlier is detected, the effect of the outlier is removed from the data using the following "cleaning" procedure: the observation y_t is adjusted at time $t = \tau$ to obtain the corrected value y_t^* using equation [2] and the estimation $\widehat{\omega}_i$:

$$y_t^* = y_t - \widehat{\omega}_i v_i I_t(\tau) \tag{7}$$

The outlier search algorithm consists in repeating this procedure until no more outliers are found. Next, we return to the outer iteration and re-run the ARIMA model on the "corrected" data, y_t^*, and start the inner iteration again. Once again, we repeat indefinitely the procedure described *supra*, until no outlier is found. Finally, we run a multiple regression on the detected outliers in order to identify the so-called "spurious" outliers (see Tolvi, 2001, for technical details).

4. Illustration and numerical applications

This section discusses some numerical applications aiming at illustrating the statistical procedure for selecting initial shocks to risk factors based on the outlier detection methodology described in the previous section.

4.1. Basic intuitions

The basic intuition behind our approach can be sketched as follows. Suppose the time evolution of a macroeconomic or bank-specific risk factor likely to be at the origin of an extreme but plausible shock depicted as a solid line in Figure 1. The risk factor may be a stock index, an interest rate or the GDP growth, for instance. The red dashed line in the same figure describes several examples of the most common outlier dynamics depending on the nature of the event affecting the series: (*i*) an instantaneous shock; (*ii*, *iii*) two transitory shocks; and (*iv*) a permanent shock. The statistical technique described in the previous section allows us to characterize not only the *magnitude* of the shock, but also the *persistence* of the initial shock across the time dimension.

{Figure 1}

This feature of our framework is quite important because the macro stress testing exercises conducted by the supervisory authorities distinguish between two types of shocks (see e.g. Martin and Tiesset, 2009):

- *transitory shocks*, which are implemented gradually over the considered stressed period. The idea here is that after some time (e.g. several quarters, depending on the presumed persistence of the initial shock) the stressed variable follows a reverting process toward the long term trend that is reached at the end of the stress testing horizon.[17] The main advantage of our methodology is that it allows us to consider different mean reverting patterns for the stressed variables depending on the nature of the initial shock and the considered risk factor.

- *permanent shocks*, which are maintained at the same stressed level throughout the entire considered stress period. Market or monetary policy scenarios (e.g. exchange rate shocks; yield curve shocks) are the most common examples of this kind of shocks. The advantage of our methodology is that the nature of the shock may be inferred *endogenously* instead of supposing it as given on an *ad-hoc* basis, by characterizing the dynamics of the identified outlier in the time series of risk drivers.

Our aim is to apply the proposed methodology to infer the most appropriate dynamics of the identified outliers from the historical time series of the variables of interest. Figure 1 summarizes various possible shocks and their dynamics according to the taxonomy proposed by Chen and Liu (1993) and briefly described in the previous section. The basic idea in this section is to calibrate these shocks based on the identified patterns characterizing the outliers detected in real-world historical time series of risk drivers.

---

[17] According to Drehmann (2009), the first generation macro stress tests supposed a one-year horizon for the stress period, although some observers argued at the time that the credit risk losses may take much more to fully materialize. Consequently, central bankers often consider nowadays a three-year stress period. For instance, Drehmann et al. (2008) show that credit risk losses take about three years to impact on UK banks' balance sheets (see also Boss et al., 2006, for a similar stress horizon for Austrian banks). In the same vein, Laviola et al. (2009) point out that a time horizon of two years is the minimum needed if the objective of the stress test is to fully capture the impact of the Italian business cycle on the credit cycle. Finally, according to De Bandt and Oung (2004), a two-year horizon corresponds to the estimated average maturity of French banks' portfolios and is reasonably well-matched with the hypothesis of no portfolio reallocation by French banks.

4.2. Shock severity and stress horizon

As we have already noted, in the particular case of transitory shocks, the proposed methodology allows us to jointly characterize the severity of the shock and the stress horizon. Figure 2 suggests a simple and intuitive definition of the *severity of the shock* as the maximum spread between the *observed* and the *corrected* outlier-free risk factor series around the detection date (segment *c*, Figure 2):

$$\Delta y = y_t - y_t^* = \hat{\omega}_{TC} v_{TC}(B) \tag{8}$$

Note that there is a striking analogy between our definition of the shock severity and the notion of "abnormal return" used in conventional event studies. Indeed, in the standard financial economics literature the abnormal return for security *i* on event date *t*, AR_{it}, is defined as the difference between actual returns R_{it} and the returns predicted by some factor market model, $\mathrm{E}[R_{it}|\Phi_t]$, where Φ_t may contain the return to the market index; the yield on the ten-year government bond to account for the interest rate sensitivity of equity returns; or the so-called Fama-French-Carhart factors:

$$AR_{it} = R_{it} - \mathrm{E}[R_{it}|\Phi_t] \tag{9}$$

where $\mathrm{E}[R_{it}|\Phi_t]$ is the return expected in the absence of the event. *Mutatis mutandis*, the *shock severity* can be viewed as the difference between the actual observed value for the risk factor y_t and the value computed under the baseline scenario, y_t^*, i.e. in the absence of the shock.

{Figure 2}

According to our approach, the *stress horizon* is intuitively defined as the time frame between the first materialization of the shock -- i.e. the first point in time when the observed and outlier-free series begin to diverge -- and the date of the reversal to the long-term trend (segment *d*, Figure 2). This definition allows us to infer a useful segmentation of the stress horizon into two regions, which are particular relevant when calibrating transitory shocks, implemented gradually over the stressed period:

- the time period during which the vulnerabilities need to crystallize to reach the maximum severity point (segment *a*, Figure 2)

- the time frame over which the shock is absorbed and the risk factor reverts to the long-term trend (segment *d*, Figure 2)

4.3. Real-world implementation of detection algorithms and implications for stress testing

The aim of this section is to implement outlier detection algorithms using real data on macroeconomic and financial time series. In order to do this, we first collect series of risk factors from various data sources (Bloomberg, Datastream Thomson Financial, Reuters and Banque de France). In selecting key risk factors commonly used in (macro) stress testing exercises, we follow the recommendations and guidelines provided in the technical notes on the macro-economic scenarios published by banking authorities (e.g. Committee of European Banking Supervisors, 2010; European Banking Authority, 2011). Of course, the statistical procedures described and proposed in the present paper may be applied to a large specter of risk drivers and should not be limited to those presented in this section for illustration purposes.

The most representative stressed variables in the macro-stress testing exercises may be classified in six broad categories:

- Macroeconomic, e.g. Gross Domestic Product (GDP); Consumer Price Index (CPI); Unemployment; Commercial Property Prices; Residential Property Prices
- Equity, e.g. Eurostoxx 50; S&P 500; Nikkei; Emerging Market Stock Indices
- Commodities, e.g. brent; brent volatility; gold; other commodities
- Credit, e.g. Itraxx, CDX, ABX, CMBS, CMBX, RMBS Indices
- Interest rates, various currencies and maturities, and slopes of the yield curves
- Foreign exchanges, e.g. EUR/USD; JPY/USD; GBP/USD

To illustrate the relevance of the outlier detection method for the calibration of historical shocks, we select from the above categories a few series that exhibit interesting and interpretable reverting patterns ("transitory changes") and structural breaks ("level shifts"). Tables 3, 4 and 5 summarize the main results of applying the detection algorithms to various classes of macroeconomic and financial series. For each series, we report the type of the identified outlier (AO, IO, LS or TC); the detection date (τ); the measure of the impact of the outlier on the series ($\widehat{\omega}_i(\tau)$, for $i \in \{AO, IO, LS, TC\}$); the likelihood ratio statistics (*LR-stat.*);

and a brief description of the historical event explaining the identified outlier. All of the outliers reported in Tables 3, 4 and 5 are identified using a significance level $\alpha = 0.05$, implying that the null hypothesis of no shocks in the data is rejected 5% of the time when it is in fact true. In our view, the choice of a significance level of 0.05 is a good compromise between the frequency of observation of the shocks and their severity. In practice, however, the significance level may be set to reflect a different trade-off between identifying a large number of less extreme shocks *vs.* detecting a few more extreme shocks.

{Table 3}

Tables 3, 4 and 5 emphasize that most series for which data are available for long periods of time exhibit several episodes of stress related to various historical events (geopolitical; economic; regulatory changes etc.). The most common shocks are either level shifts (LS) or transitory changes (TC). AO are interpreted as short-term "one-shot" temporary shocks having a less severe impact of the stressed variable than LS or TC. We are not able to detect any IO in the considered series of risk factors. Consequently, we focus our discussion on the LS and particularly TC, which are the most relevant shocks from a stress testing perspective.

In some cases, we observe series of consecutive shocks that seem to have a cumulative effect on the stressed variables if the stress horizon is extended to cover a longer period of time. This is particularly the case of the level shifts in Brent prices at the beginning of the seventies (Table 5, panel A); the transitory changes in the same series at the beginning of the nineties (Table 5, panel A); the clustering of shocks in the Japanese CPI series in the inflationary environment of the early seventies (Table 3, panel C) or in the interest rate and yield series (2-year and 10-year Treasury yields; the slope of the Treasury yield curve) during the early eighties recession in the United States (see Table 4).

{Table 4}

It is worth noting that all the detected shocks reported in Tables 3, 4 and 5 are obtained by applying the iterative algorithm described in the methodological section of the paper to identify *simultaneously* the four classes of outliers ("*joint detection*"). In order to draw cleaner inferences about the nature of the shocks, we also considered an alternative way to answer our main research question by proceeding with a *separate* identification of outliers,

by class ("*separate detection*"). This alternative detection procedure is particularly useful when we discuss and interpret the results in a graphical and descriptive statistical manner.

{Table 5}

Not surprisingly, we detect in the vast majority of the considered series significant level shifts and/or transitory changes related to the global financial crisis that broke out in the summer 2007 and exacerbated over the last four years. This is not indicative of an end-point problem inherent to the outlier detection methodology, i.e. higher incidence of trend shifts and breaks systematically observed at the end of sample. Indeed, when we remove the recent tumultuous period related to the 2007 financial crisis from our sample, we don't detect any outliers at the end of the remaining tranquil period.

To gain further insights from the analysis, we also examine the nature of the identified shocks in a graphical and descriptive statistical manner. We select only a few representative series for each class of risk drivers that enable us to clearly disentangle the impact of various types of outliers. The discussion is structured by broad categories of macroeconomic and financial variables, with a particular focus on transitory changes. The final part of the section brings together the key themes to compare the severity and stress horizons across series and discusses some implications for stress testing.

4.3.1. Macro series: Gross Domestic Product (GDP)

One of the most common stressed variables used in (macro) stress tests is the Gross Domestic Product (GDP). For illustration purposes, Figure 3 depicts the evolution of the GDP growth rates (chain linked, % quarter-on-quarter) for Germany, one of the most resilient countries during the recent global financial crisis. A significant transitory change is detected in the four quarter of 2008, just after the global financial market panic triggered by the collapse of Lehman Brothers on September 14th, 2008 and the ambiguous policy responses by public authorities in the early stages of the crisis (e.g. the announcement of the first version of the Troubled Asset Relief Program, TARP).

{Figure 3}

20

The implied stress horizon, as defined supra §4.2, i.e. the time frame between the first materialization of the shock and the date of the reversal to the long-term trend, is equal to four quarters (for a useful analogy, see also Figure 2, segment *d*). This value for the length of the stress horizon is somewhat narrower than suggested in recent macro stress tests. In practice, central bankers often consider nowadays a larger stress horizon (a two- or three-year stress period). For comparison purposes, De Bandt and Oung (2004), Boss et al. (2006) and Laviola et al. (2009) suggest an eight-quarter stress horizon for French, Austrian and Italian banks, respectively, while Drehmann et al. (2008) propose a twelve-quarter stress period for UK banks. However, the first generation macro stress tests, as well as the upcoming 2013 adverse supervisory scenario in the US, use a one-year horizon for the stressed GDP variable, which is more in line with the order of magnitude of the transitory change depicted in Figure 3 (see e.g. Drehmann, 2009, and references therein, and Board of Governors of the Federal System, 2012).[18] It should also be noted that the resilience of the German economy to the subprime shock could partially explain the shortness of the stress horizon discernible in Figure 3.

As far as the severity of the shock is concerned, Figure 3 suggests a maximum spread of –4.5% between the stressed and the outlier-free series around the detection date (for a useful analogy, see also Figure 2, segment *c*). The order of magnitude for the shock is in line with that used in the EU macro stress tests, where GDP growth is supposed to be about 4% lower than in the benchmark scenario (see EBA, 2011). The 2012 US stress test was somewhat more stringent, with the GDP growth dropping about 6% under the worst case scenario. The same is true for the upcoming 2013 US "severely adverse" scenario, which supposes a nearly 5% decline in GDP between the third quarter of 2012 and the end of 2013, i.e. over a four-quarter stress horizon (see Board of Governors of the Federal Reserve System, 2012).

---

18 To be more precise, the US adverse scenario features a moderate recession that begins in the fourth quarter of 2012 and lasts until early 2014. However, all scenarios start in the fourth quarter of 2012 (2012:Q4) and extend through the fourth quarter of 2015 (2015:Q4).

4.3.2. Real estate prices

Another common stressed variables used in (macro) stress tests is the real estate (commercial and residential property) prices. Figure 4 presents the evolution of a broad US property index (quarterly time series). We detect a short extreme movement in the property index at the beginning of the nineties and o more severe shock to the series in the first quarter of 2008 due to the worsening mortgage crisis in the US.

{Figure 4a}

The reverting pattern (TC) depicted in Figure 4a implies a length of the stress horizon of two-years (or eight quarters) and a magnitude of the shock of about −10%. These figures are in line with the proposed values used to calibrate macro stress testing exercises (see also our discussion *supra*, §4.3.1).

A similar pattern is detected in the evolution of the China real estate index depicted in Figure 4b, as shocks to property prices around the world also spread to Asia. However, the transitory change observed at the end of 2008 seems to be shorter (four quarters) and much less severe than in the US.

{Figure 4b}

Finally, Figure 4c shows the time evolution of land prices in Japan (nationwide, biggest cities, % year-on-year) and two transitory changes. The first one, at the beginning of the eighties, is rather mild. By contrast, the second one is longer (five years) and much more severe; it starts at the beginning of 1987, ends in 1993 and corresponds to the economic bubble in Japan, in which real estate and stock prices were greatly overvalued. The bubble subsequently collapsed, but lasted for more than a decade. Note that the subprime shock in 2008 is not detected in Figure 4c because it corresponds to a significant level shift (not represented in the same figure) starting in the nineties, just after the real estate bubble burst.

{Figure 4c}

4.3.3. Inflation rates

The Consumer Price Index (CPI) is another macroeconomic variable that is often mentioned in stress testing methodological documentations, although the retained stressed values reflect mild inflationary scenarios: e.g. an increase by 2.2% and 2.4% over a two-year period for France and Germany, respectively (EBA, 2011); and an increase by 4% over a one-year stress horizon for the US (Board of Governors of the Federal Reserve System, 2012). Figure 5 illustrates the application of the outlier detection algorithm to an extreme inflationary environment: the Great Inflation experience of the early seventies in Japan. The literature on inflation mentions several plausible causes of the high-inflation episode in Japan (see e.g. Nelson, 2007, and references therein): political factors (e.g. politicians' willingness to sacrifice price stability in favor of full employment); policy errors due to inaccuracies in the output gap measurement; and the extensive use of nonmonetary devices, such as wage and price controls, to fight inflation (the so-called "monetary-policy-neglect hypothesis").

{Figure 5}

Figure 5 exhibits an extreme transitory change starting in September 1973 and ending 18 month later, in February 1975. The reverting pattern suggests a common stress horizon of 1.5 years, but an unusual magnitude for the inflationary shock. We attribute the mild severity of inflationary scenarios used nowadays in stress testing to the effectiveness of modern central banks in successfully fighting against inflation.

4.3.4. Interest rates and yields

Stress scenarios involving a flattening of the yield curve, a quick increase in short-term interest rates reflecting rising inflation, as well as a slight decline of the long-term government yields, are quite common in the practice of stress testing. Figures 6a and 6b illustrate the application of the outlier detection procedure to short-term interest rates and yields (3-month Libor rate and 2-year US Treasury yield), while Figures 6c and 6d presents the evolution of long-term US Treasury yields during the early eighties recession and the recent financial turmoil, respectively.

{Figures 6a & 6b}

The behavior of short-term interest rates and 10-year Treasury yields during the recent global financial crisis exhibit transitory changes starting in January and December 2008, respectively. The shocks are characterized by a reverting pattern implying a relatively short stress horizon, ranging between nine months and one year, and a moderate severity (about −100 bps, Figures 6a and 6d).

It is worth noting that a *decrease* in interest rates and Treasury yields is uncommon in the practice of stress testing in banking. However, scenarios of falling interest rates may have significant negative effects in the insurance industry, as the duration of insurers' liabilities is general longer than the duration of their assets. The interest risk in the insurance industry is also known as the asset/liability mismatching risk. The 2011 EU-wide stress test in the insurance sector supposes *negative* shocks on short-term (long-term) interest rates of −125 bps (−62.5 bps). For more details see European Insurance and Occupational Pension Authority (2011).

{Figures 6c & 6d}

Figures 6b and 6c show the behavior of short-term and long-term Treasury yields during the early eighties, when interest rate series exhibit positive shocks (i.e. sharp increases), which are more relevant for the banking industry. In 1980 the US economy experienced a severe recession that, at the time, was the worst since the Great Depression. One of the causes of the early eighties recession was P. Volker's restrictive monetary policy, which led to a slow economic growth. Both shocks depicted in Figures 6b and 6c started in February 1980 and were short-lived (stress period of about six months). The shock is more severe on the short-term (+4%) than the long-term yield (+2%). The order of magnitude is slightly larger than what has been considered in the EU-wide stress tests (see EBA, 2011), e.g. a deviation of the long-term interest rates from the baseline scenario of +0.5% (in France), +1.4% (in Italy) or +1.7% (in Spain). The 2013 supervisory scenario in the US considers a sharp increase (+2%) in short-term interest rates over a one-year stress period. The yield on the long-term Treasury notes increases by less (+1.1%). Both movements imply a shock on the yield curve, which becomes both higher and flatter (see Board of Governors of the Federal Reserve System, 2012).

The shocks on the slope of the yield curve, depicted in Figure 7, are short-lived (stress period less than six months), moderate, but clustered over a relatively short and tumultuous period (1980—1983).

{Figure 7}

4.3.5. Commodity prices

Compared with the other macroeconomic and financial variables considered in the previous sections, commodities (e.g. oil, gold) are relatively less common in the practice of stress testing. Moreover, the impact of shocks on commodity prices on banks' balance-sheets is found to be week. For instance, De Bandt and Oung (2004) report that a macroeconomic stress scenario involving an increase of nearly 50% in the price of oil for two years, with or without a monetary policy reaction, have a modest impact on the solvency ratios of French banks (–0.8% under Basel I and –0.11% under Basel II calculations).

{Figures 8a and 8b}

Figures 8a and 8b presents the detected transitory changes in Brent crude oil spot prices (Bloomberg Oil Index), over different periods, while Figure 8c depicts the evolution of gold spot prices at the beginning of the eighties. The first shock detected in August 1990 occurred in response to the Iraqi invasion of Kuwait. It lasted only 18 months and was less extreme than the recent shock on the oil prices (2007—2008). Yet, the rise in oil prices at the beginning of the nineties is widely believed to have been a significant recessionary factor. The oil shock of 2007—2008 implies a similar length of the stress horizon (18 months), but it has a much larger effect on the economy (i.e. on consumption spending and purchases of domestic automobiles in particular). Hamilton (2009) compares the oil shock of 2007—2008 with other shocks observed in the past (including the August 1990 shock) and concludes that absent the dramatic price moves in 2007 and 2008, it is unlikely that the period 2007Q4–2008Q3 would have been characterized as one of recession for the US.

{Figure 8c}

In January 1980, gold fully played its role as strategic reserve asset in a highly uncertain economic environment characterized by inflation fears, high oil prices, the Soviet invasion in Afghanistan and the Iranian revolution. As a consequence, gold prices rose dramatically in late December 1979 and January 1980. Figure 8c reveals a transitory change in gold spot prices in January 1980, caused by geopolitical reasons (Soviet invasion and Iran hostage crisis) and high inflation. The length of the stress period was less than one year, but the sharp increase in gold prices was severe (+40% in less than three months).

*

* *

Summarizing the results of the outlier detection algorithms reported in this section, the dynamics of shocks identified in the selected macroeconomic and financial variables, as well as the length of the stress horizons or the shock severity, are sensitive to the type of initial shock and the nature of the risk factor. For instance, the inferred stress period is longer (one to two years) for macroeconomic variables like GDP, CPI, real estate and oil prices, than for interest rate variables and the slope of the yield curve (six months). To conclude this section, we would like to mention several general issues related to the interpretation of the shocks and the implementation of the outlier detection algorithms that deserve further investigation:

- What kind of initial shocks are more dangerous, long lasting and severe but *isolated* shocks or short-lived and moderate shocks, which are *clustered* over a relatively short period?
- Is there any effect of the *data frequency* (daily, monthly, quarterly etc.) on the outlier detection procedure?
- Does the way the stressed variables are measured (levels, growth rates etc.) have any effect on the detection algorithms? For instance, short-lived shocks (AO or TC) detected in variables expressed as a *percentage growth rate* become persistent changes (LS) in the same variables measured in *levels*. This is not surprising if we take into account the basic definition of and the link between the various types of outliers.

5. Conclusion

The bank stress testing exercises have gain new momentum since the inception of the global financial crisis in the summer 2007. The crisis has revealed several important deficiencies inherent in the stress tests conducted by the largest banks and their supervisory authorities: lack of integration of stress tests into the broad risk governance process; disregard for specific risks, such as those related to securitization; failure to take into account feed-back (second round) effects; other flaws in the design of stress scenarios. Particularly, it seems that stress scenarios implemented before and during the subprime crisis only reflected mild and temporary shocks and assumed that those shocks were maintained over short-length periods.

In this paper, we propose a rigorous and flexible methodological framework to select initial shocks to be used in stress scenarios based on statistical techniques for detection of outliers in time series of risk factors. The advantage of our framework is twofold. First, it allows us to characterize not only the magnitude, but also the persistence of the initial shock. Second, it allows considering various reverting patterns for the stressed variables and informs the choice of the appropriate time horizon. This is important because extreme but plausible stresses that have the most harmful impact on the banking sector are of the transitory but long lasting type; they do not necessarily imply structural changes, which are hard to make plausible before the bust, but keep having effects during a sufficiently long period, so that they cannot be dissimulated by accounting techniques or regulatory arbitrage.

We illustrate the proposed methodology by implementing outlier detection algorithms to a few representative time series of economic and financial variables typically used in bank stress testing. Summarizing the main results reported in the paper, the dynamics of shocks identified in the selected macroeconomic and financial variables, as well as the length of the stress horizons or the magnitude of the shock, are sensitive to the type of initial shock and the nature of the risk factor. For instance, the inferred stress period is longer (one to two years) for macroeconomic variables like GDP, CPI, real estate and oil prices, than for interest rate variables and the slope of the yield curve (six months).

A natural extension of our work would be to implement the outlier detection algorithms described in the present paper within a multivariate framework. The passage from a

univariate to a multivariate framework would improve the interpretation of the shocks by taking explicitly into account the *correlation structure* between simultaneous risk factors. This idea is left for future research.

Acknowledgements

References

Balke N. and Fomby T. (1994). Shifting Trends, Segmented Trends, and Infrequent Permanent Shocks. Journal of Monetary Economics, 28, pp. 61–85.

Basel Committee on Banking Supervision (BCBS) (2006). International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Bank for International Settlements.

Basel Committee on Banking Supervision (BCBS) (2009). Principles for Sound Stress Testing Practices and Supervision. Consultative Document, Bank for International Settlements.

Blaschke W., Jones M., Majnoni G. and Martinez Peria S. (2001). Stress Testing of Financial Systems: An Overview of Issues, Methodologies, and FSAP Experiences. IMF Working Paper 01/88.

Board of Governors of the Federal Reserve System (2012). 2013 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule. Released on November 15th.

Borio C. and M. Drehmann (2009). Towards an Operational Framework for Financial Stability: 'Fuzzy' Measurement and its Consequences. BIS Working Paper #284.

Borio C., Drehmann M. and Tsatsaronis K. (2012). Stress-testing Macro Stress Testing: Does It Live Up to Expectations? BIS Working Paper #369.

Boss M., Breuer T., Elsinger H., Jandacka M., Krenn G., Lehar A., Puhr C. and Summer M. (2006). Systemic Risk Monitor: A Model for Systemic Risk Analysis and Stress Testing of Banking Systems. Technical Report, Oesterreichische Nationalbank.

Box G. and Tiao G. (1975). Intervention Analysis with Applications to Economic and Environmental Problems. Journal of the American Statistical Association, 70, pp. 70–79.

Breuer T. and Csiszár I. (2012). Systematic Stress Tests with Entropic Plausibility Constraints. Journal of Banking and Finance, forthcoming.

Breuer T., Jandačka M., Mencía J. and Summer M. (2012). A Systematic Approach to Multi-period Stress Testing of Portfolio Credit Risk. Journal of Banking and Finance, 36(2), pp. 332-340.

Breuer T., Jandačka M., Rheinberger K. and Summer M. (2009). How to Find Plausible, Severe, and Useful Stress Scenarios. International Journal of Central Banking, 5(3), pp. 205-224.

Bunn P., Cunningham A. and Drehmann M. (2005). Stress Testing as a Tool for Assessing Systemic Risks. Bank of England, Financial Stability Review.

Chang I., Tiao G. and Chen C. (1988). Estimation of Time Series Parameters in the Presence of Outliers. Technometrics, 30, pp. 193–204.

Chen C. and Liu L. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. Journal of the American Statistical Association, 88, pp. 284–297.

Cihak M. (2004). Stress Testing: A Review of Key Concepts. Czech National Bank, Working Paper.

Cihak M. (2007). Introduction to Applied Stress Testing. IMF Working Paper #59.

Committee of European Banking Supervisors (2010). Aggregate Outcome of the 2010 EU Wide Stress Test Exercise.

Committee on the Global Financial System (2000). Stress Testing by Large Financial Institutions: Current Practice and Aggregation Issues. CGFS Publications #14.

Committee on the Global Financial System (2001). A Survey of Stress Tests and Current Practice at Major Financial Institutions. CGFS Publications #18.

Committee on the Global Financial System (2005). Stress Testing at Major Financial Institutions: Survey Results and Practices. CGFS Publications #24.

Darné O. and Diebolt C. (2004). Unit Roots and Infrequent Large Shocks: New International Evidence on Output. Journal of Monetary Economics, 51, pp. 1449-1465.

De Bandt O. and Oung V. (2004). Assessment of Stress Tests Conducted on the French Banking System. Banque de France, Financial Stability Review, 5, pp. 55-72.

Drehmann M. (2008). Stress Tests: Objectives, Challenges, and Modelling Choices. Riksbank Economic Review, 2, pp. 60-92.

Drehmann M. (2009). Macroeconomic Stress-testing Banks: A Survey of Methodologies. In: Stress-testing the Banking System: Methodologies and Applications, M. Quagliariello (ed.), Cambridge University Press.

Drehmann M., Borio C. and Tsatsaronis K. (2012). Characterising the Financial Cycle: Don't Lose Sight of the Medium Term. BIS Working Paper #380.

Drehmann M., Sorensen S. and Stringa M. (2008). The Integrated Impact of Credit and Interest Rate Risk on Banks: An Economic Value and Capital Adequacy Perspective, Working Paper No 339, Bank of England.

Dubecq S. and Gourieroux C. (2012). Shock on Variable or Shock on Distribution with Application to Stress-Tests. Banque de France Working Paper #368.

European Banking Authority (2011). EU-Wide Stress Test Aggregate Report.

European Insurance and Occupational Pension Authority (2011). Specifications for the 2011 EU-wide Stress Test in the Insurance Sector. March.

Financial Services Authority (2005). Stress Testing. Discussion Paper #2.

Financial Services Authority (2006). Stress Testing Thematic Review. Letter to Chief Executives, October.

Financial Services Authority (2008). Stress and Scenario Testing. Consultation Paper #08/24.

Foglia A. (2009). Stress Testing Credit Risk: A Survey of Authorities' Approaches. International Journal of Central Banking, 5(3), pp. 9-45.

Gómez V. and Maravall A. (1997). Programs TRAMO and SEATS: Instructions for the User (Beta Version: June 1997), Working Paper No. 97001, Dirección General de Análisis y Programación Presupuestaria, Ministerio de Economía y Hacienda, Madrid.

Gómez V. and Maravall A. (2001). Automatic Modeling Methods for Univariate Series. In: Peña, D., Tiao, G.C., Tsay, R.S. (eds.), A Course in Time Series Analysis. Wiley, New York, pp. 171–201.

Hamilton, J. (2009). Causes and Consequences of the Oil Shock of 2007–08. Brookings Papers on Economic Activity, Spring, pp. 215-261.

Isogai T. (2009). Scenario Design and Calibration. In: Stress-testing the Banking System: Methodologies and Applications, M. Quagliariello (ed.), Cambridge University Press.

Jokivuolle E., Virén M. and Vähämaa O. (2010). Transmission of Macro Shocks to Loan Losses in a Deep Crisis: The Case of Finland. In: Model Risk: Identification, Measurement and Management, D. Rösch and H. Scheule (eds.), Risk Books, pp. 183-208.

Maravall A. (2005). Brief Description of the Programs TRAMO, SEATS and TERROR. Banco de España, Statistics and Econometrics Softwares.

Martin C. and Tiesset M. (2009). From the Macro to the Micro: The French Experience on Credit Risk Stress Testing. In: Stress-testing the Banking System: Methodologies and Applications, M. Quagliariello (ed.), Cambridge University Press.

Moretti M., Stolz S., and Swinburne M. (2008). Stress Testing at the IMF. IMF Working Paper #08/206.

Nelson, E. (2007). The Great Inflation and Early Disinflation in Japan and Germany. International Journal of Central Banking, 3(4), pp. 23-76.

Newey W. and West K. (1987). A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55, pp. 703–708.

Newey W. and West K. (1994). Automatic Lag Selection in Covariance Matrix Estimation. Review of Economic Studies, 61, pp. 631–653.

Sorge M. (2004). Stress-testing Financial Systems: An Overview of Current Methodologies. Working paper No 165, Bank for International Settlements.

Sorge M. and Virolainen K. (2006). A Comparative Analysis of Macro Stress-testing Methodologies with Application to Finland. Journal of Financial Stability, 2, 113-151.

Tolvi J. (2001). Outliers in Eleven Finnish Macroeconomic Time Series. Finnish Economic Papers, 14, pp. 14–32.

Tsay R. (1988). Outliers, Level Shifts, and Variance Changes in Time Series. Journal of Forecasting, 7, pp. 1–20.

Varotto, S. (2012). Stress Testing Credit Risk: The Great Depression Scenario. Journal of Banking and Finance, 36(12), pp. 3133–3149.

Table 1: Historical-based *vs*. hypothetical stress scenarios

Historical-based scenarios	Hypothetical scenarios
Rigorous and objective, less discretionary, transparent, based on clear selection criteria	*Ad hoc*, arbitrary, subjective, discretionary, opaque, subject to manipulation since no strict criteria apply for scenario selection
Objective guidance based on past data	No rule or guidance in setting the magnitude or the persistence of the initial shock
Backward-looking by definition	Forward-looking by nature
Hardly to dismiss by managers as implausible, increased acceptance	Easy to dismiss by senior managers as implausible
Credible, plausible, intuitively possible since the extreme movements actually occurred in the past	Less credible, less plausible, may include unimaginable events
Underestimation of the severity of the shock in the case of innovative risk factors and new products for which sufficiently long time-series are rarely available	Based on experts' opinions and the economic expertise of staff, which may underestimate the impact of low probability events (cognitive bias)
Underestimate the possibility that statistical patterns may break down differently in the future	A "failure of imagination" may lead to a false sense of security
Not necessarily worst-case scenarios	May include worst-case scenarios
The composition and risk profile of the bank portfolio is only taken into account when selecting the risk factors	The composition and risk profile of the bank portfolio may be considered
Computational-intensive, but likely to be fully automated	Labor-intensive, more judgment involved
Based on parametric assumptions and on the assumption that future crises resemble to past crises	
Do not include the possibility of inexperienced events	

Table 2: Examples of historical-based and hypothetical stress scenarios in use at major banks

Historical-based scenarios	Hypothetical scenarios
Interest rates	
1994 bond market sell-off	US economic outlook
1997 Asian financial crisis	Global economic outlook
1998 LTCM, Russia	Increase in inflation expectations
2001 09/11 terrorist attacks	
Equities	
1987 Black Monday	Geopolitical unrest
1997 Asian financial crisis	Terrorist attacks
2000 bursting of IT bubble	Global economic outlook
2001 09/11 terrorist attacks	
Foreign exchange	
1992 ERM crisis	Collapse of currency pegs
1997 Asian financial crisis	
1998 Russia	
Commodities	
	Oil price scenario
	Geopolitical unrest in the Middle East
Credit	
1997 Asian financial crisis	Emerging markets economic outlook
1998 Russia	Euro area economic outlook
2001 09/11 terrorist attacks	Global economic outlook
	Natural disasters
	US GSEs
	Terrorist attack
Property	
	Fewer than three per hypothetical episode
Other	
	Bank funding
	Global economy

Note: This table is based on the last (2005) survey on stress testing practices conducted by the Committee on the Global Financial Stability (CGFS, 2005), covering 64 banking organizations and other global financial players headquartered in 16 different countries.

Table 3: Outlier detection in macro series – some illustrations

		Panel A: Gross Domestic Product (GDP)			
Germany GDP	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	TC	2008:Q4	-4.553	-6.89	
		Panel B: Real estate prices			
US Property Index	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	AO	1991:Q4	-5.019	-5.99	
	LS	2010:Q1	3.338	3.56	
	TC	2008:Q4	-9.076	-9.65	
China House Prices	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	LS	Apr. 2010	-0.914	-4.05	
	TC	Dec. 2008	-3.000	-22.13	
Japan Land Prices	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	LS	1992	-10.389	-6.21	
	LS	2009	-6.176	-4.42	
	TC	1980	4.327	3.66	
	TC	1987	10.448	5.28	
		Panel C: Consumer Price Index (CPI)			
Japan CPI Index	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	LS	Sept. 1973	2.285	4.91	
	LS	Jan. 1974	3.589	7.72	
	LS	Jan. 1975	-3.933	-8.37	
	LS	Apr. 1998	-1.806	-3.88	
	TC	Sept. 1972	-2.271	-5.31	
	TC	Sept. 1973	2.116	4.78	

Table 4: Outlier detection in interest rates and yield series – some illustrations

		Panel A: Short-term interest rates and yields			
Libor 3m	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	AO	Aug. 1986	-0.737	-6.19	
	AO	Sept. 1987	0.852	7.16	
	AO	July 1989	-0.504	-4.24	
	LS	Jan. 2008	-1.420	-7.17	
	TC	Feb. 1985	0.738	4.17	
	TC	May 1985	-0.837	-4.77	
	TC	Feb. 1989	0.923	5.29	
	TC	Nov. 1990	0.730	4.18	
	TC	Jan. 2008	-1.563	-8.95	
US Treasury 2y	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	AO	Jan. 1981	-0.933	-5.59	
	AO	May 1981	1.106	6.67	
	LS	Nov. 1980	1.140	4.10	
	LS	Oct. 1982	-1.452	-5.26	
	TC	Feb. 1980	1.963	7.08	
	TC	Nov. 1981	-2.092	-8.59	
	TC	June 1982	1.021	4.19	
		Panel B: Long-term yields			
US Treasury 10y	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	AO	June 1981	-0.658	-4.69	
	LS	Oct. 1982	-1.096	-4.73	
	TC	Feb. 1980	1.466	6.67	
	TC	Nov. 1981	-1.480	-7.23	
	TC	June 1982	0.862	4.21	
	TC	Dec. 2008	-0.838	-4.09	
		Panel C: Yield curve slopes			
US yield curve slope	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	AO	Oct. 1973	0.772	5.35	
	AO	Feb. 1974	0.618	4.29	
	LS	May 1980	2.937	10.75	
	TC	Aug. 1974	-1.626	-7.56	
	TC	March 1980	-2.158	-9.90	
	TC	Dec. 1980	-1.224	-5.69	
	TC	May 1981	-2.255	-10.48	
	TC	Feb. 1982	-1.165	-5.42	
	TC	Aug. 1982	1.146	5.33	
Swap Curve Slope	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	LS	Oct. 2008	0.785	5.62	

Table 5: Outlier detection in commodity prices – some illustrations

		Panel A: Brent crude oil spot prices			
Brent Index	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	AO	March 1986	-0.244	-5.29	
	AO	July 1986	-0.287	-6.22	
	AO	July 2000	-0.182	-3.95	
	LS	Oct. 1973	0.506	7.64	
	LS	Jan. 1974	0.833	12.59	
	LS	Nov. 1979	0.283	4.27	
	LS	Jan. 1986	-0.295	-4.46	
	LS	Feb. 1986	-0.401	-6.07	
	LS	March 1999	0.280	4.21	
	LS	Oct. 2008	-0.390	-5.89	
	LS	May 2009	0.252	3.81	
	TC	Aug. 1990	0.269	4.37	
	TC	June 2007	0.803	4.59	
		Panel B: Gold spot prices			
Gold spot prices	*Outlier type*	*Event date*	*ω-value*	*LR-stat*	*Event explanation*
	AO	March 1982	-0.095	-3.83	
	AO	May 2006	0.108	4.33	
	LS	March 1983	-0.168	-4.16	
	TC	Jan. 1980	0.295	8.25	
	TC	Sept. 1982	0.173	4.28	
	TC	Oct. 1999	0.153	4.27	

Figure 1: Shocks as outliers -- a conceptual taxonomy and various patterns

Figure 2: Shock topography – severity, stress horizon and reverting pattern

'Corrected' series   Shock path
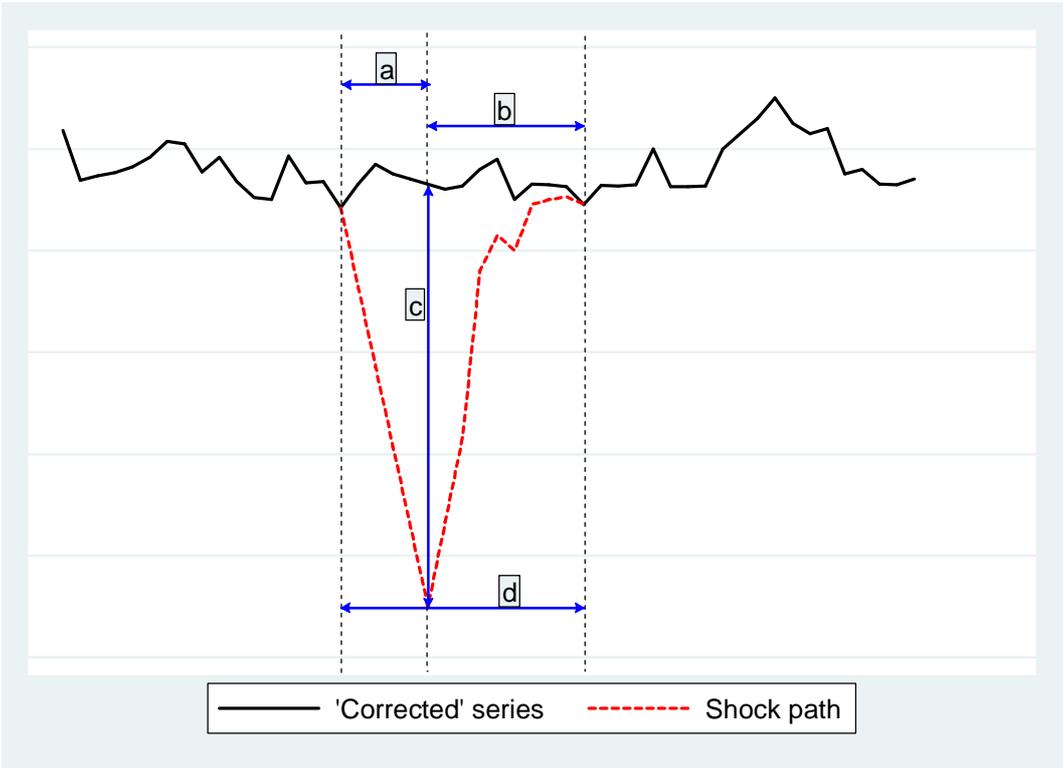
Figure 3: Outliers detection in GDP series
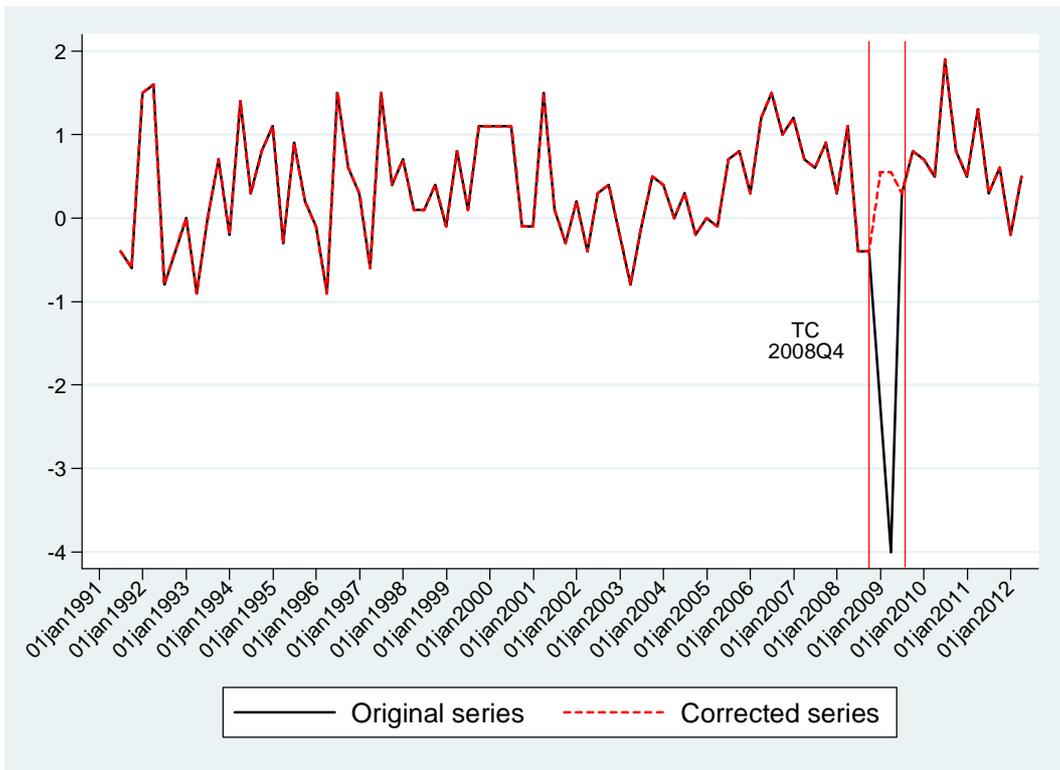
a. Germany

Figure 4: Outliers detection in Property Prices
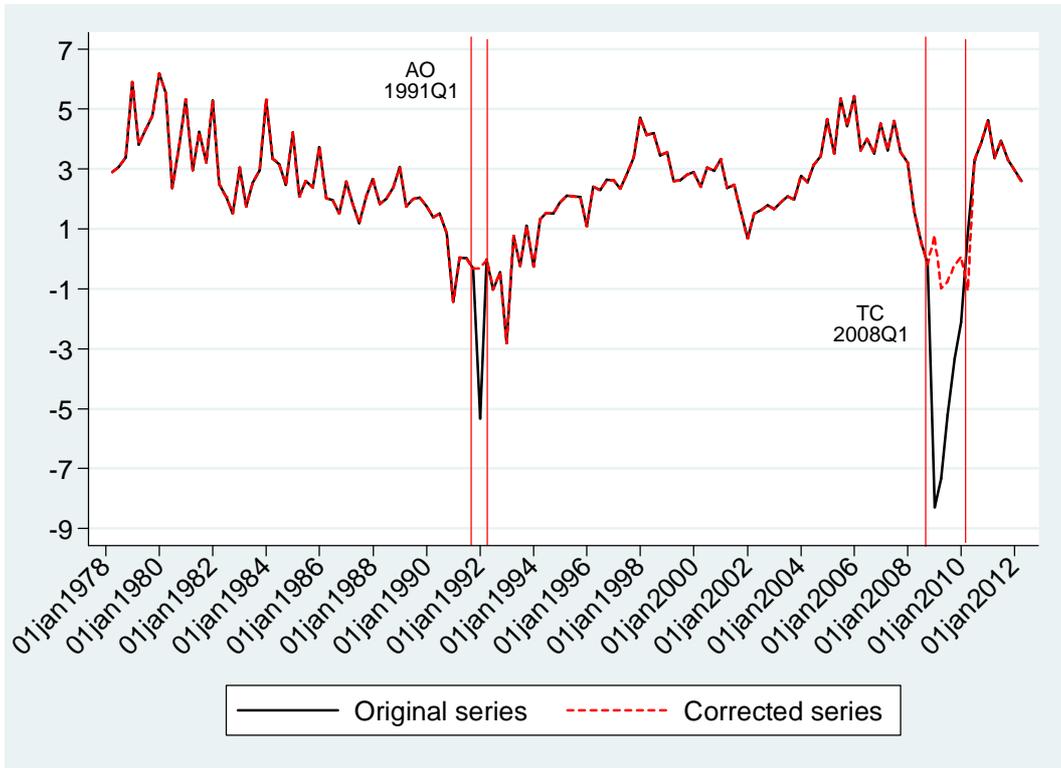
a.   United States
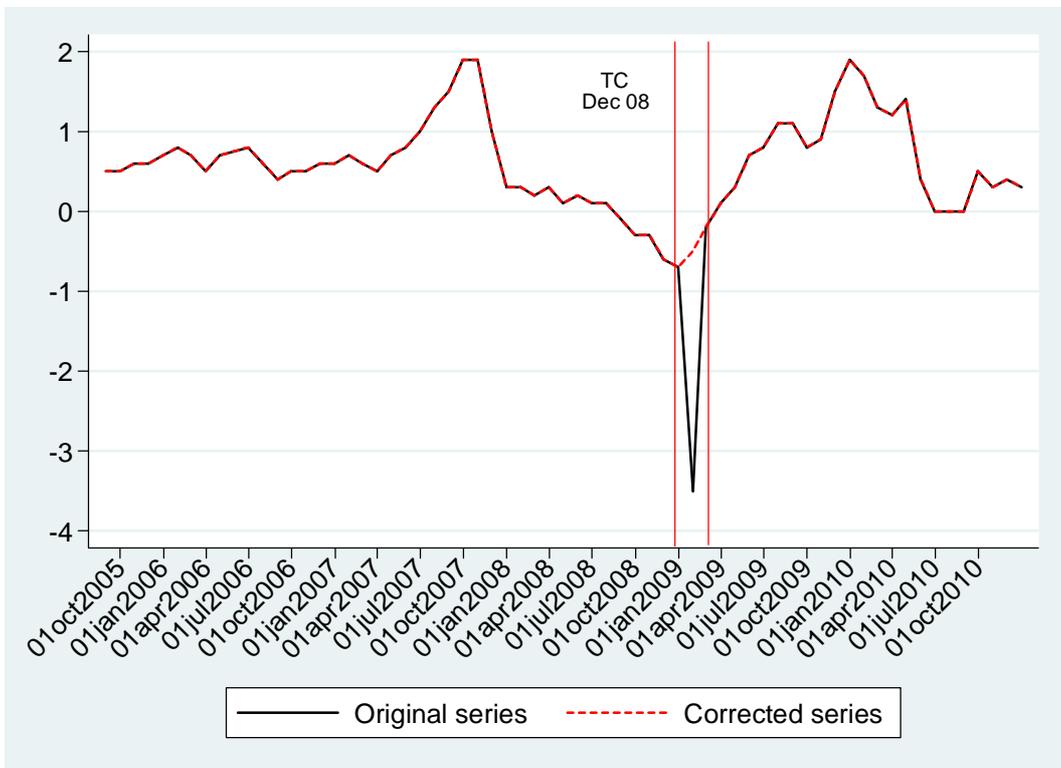
b.   China

**Figure 4:** Outliers detection in Property Prices (cont.)
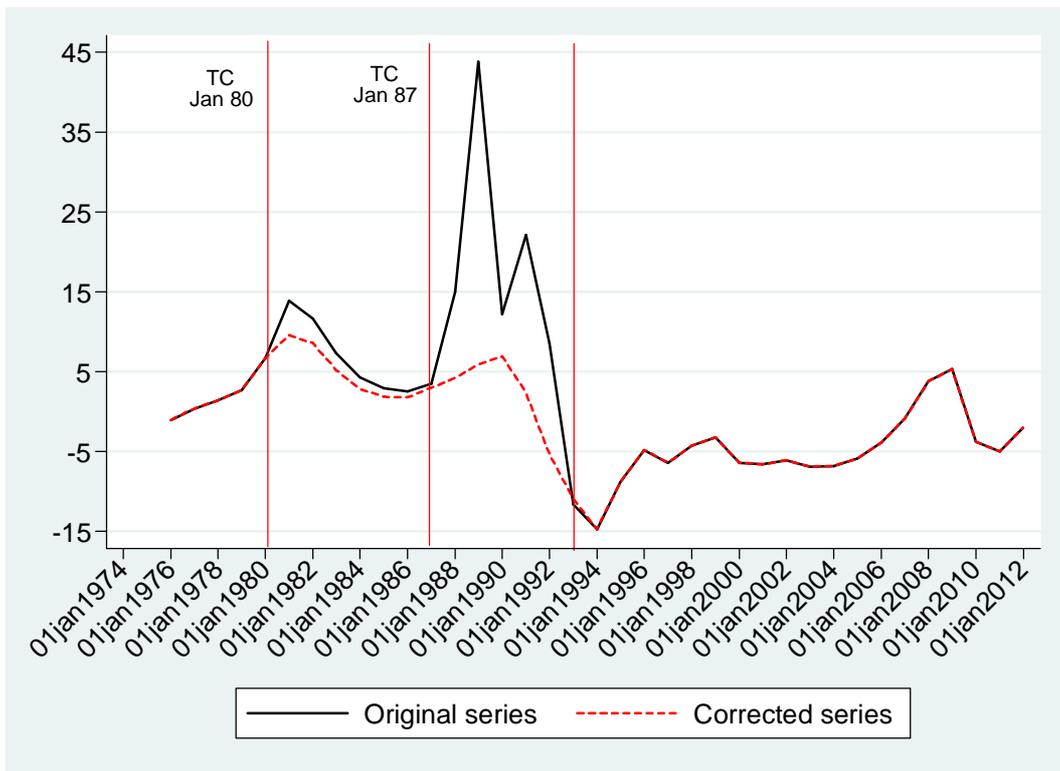

c. Japan

Figure 5: Outliers detection in inflation rates
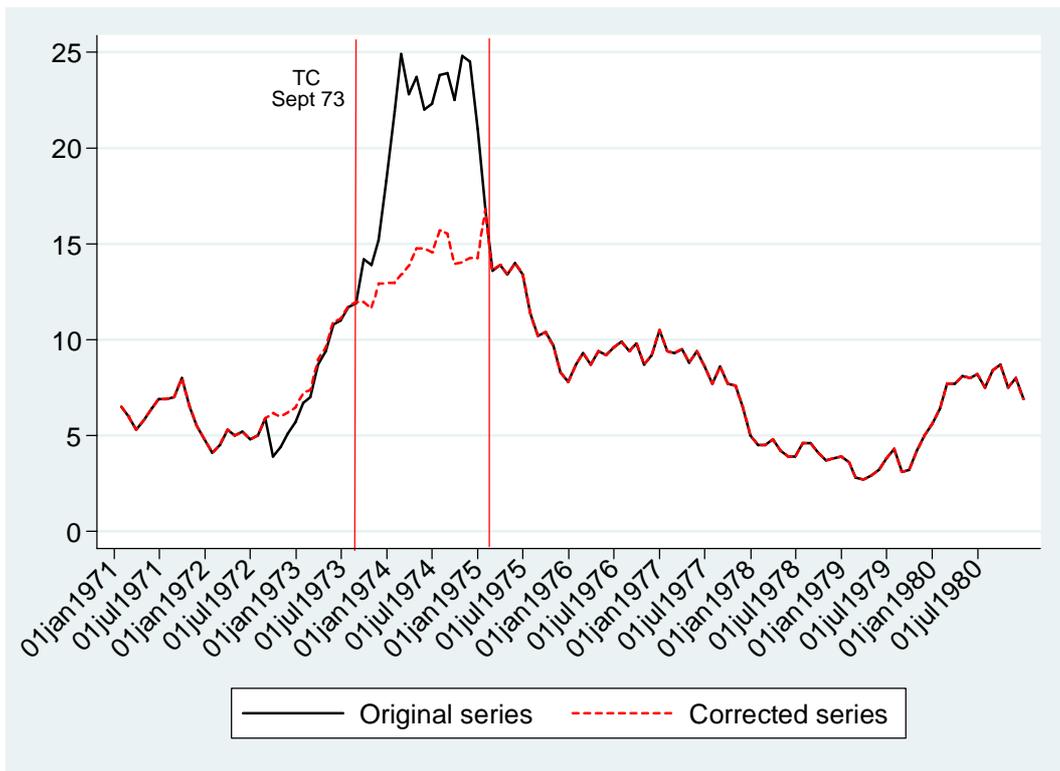
a.   Japan, 1970--1980

Figure 6: Outliers detection in interest rates
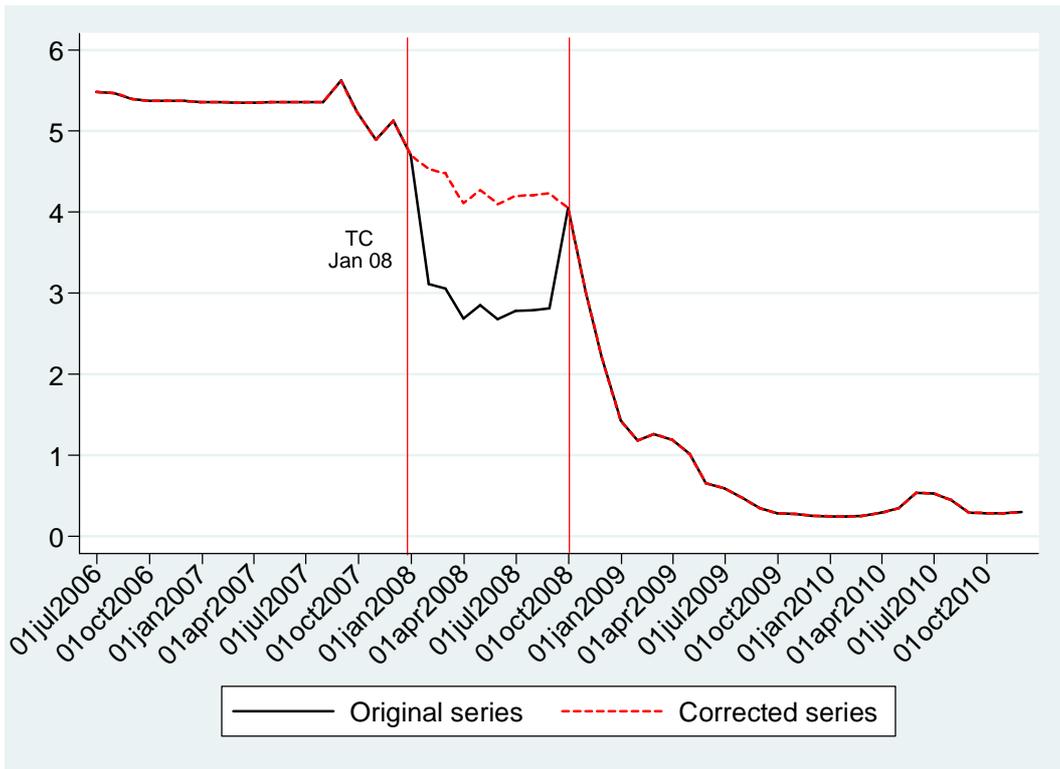
a.   3-month Libor (2006--2010)
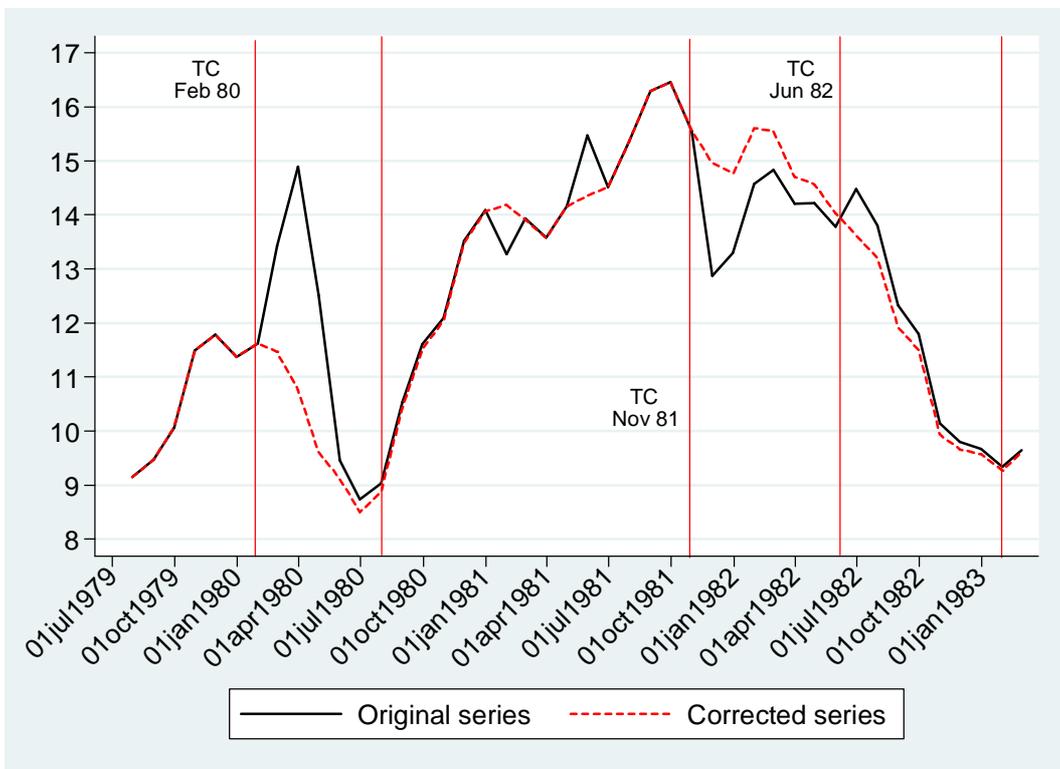
b.   US Generic Govt. 2Y yield (1979--1983)

Figure 6: Outliers detection in interest rates (cont.)
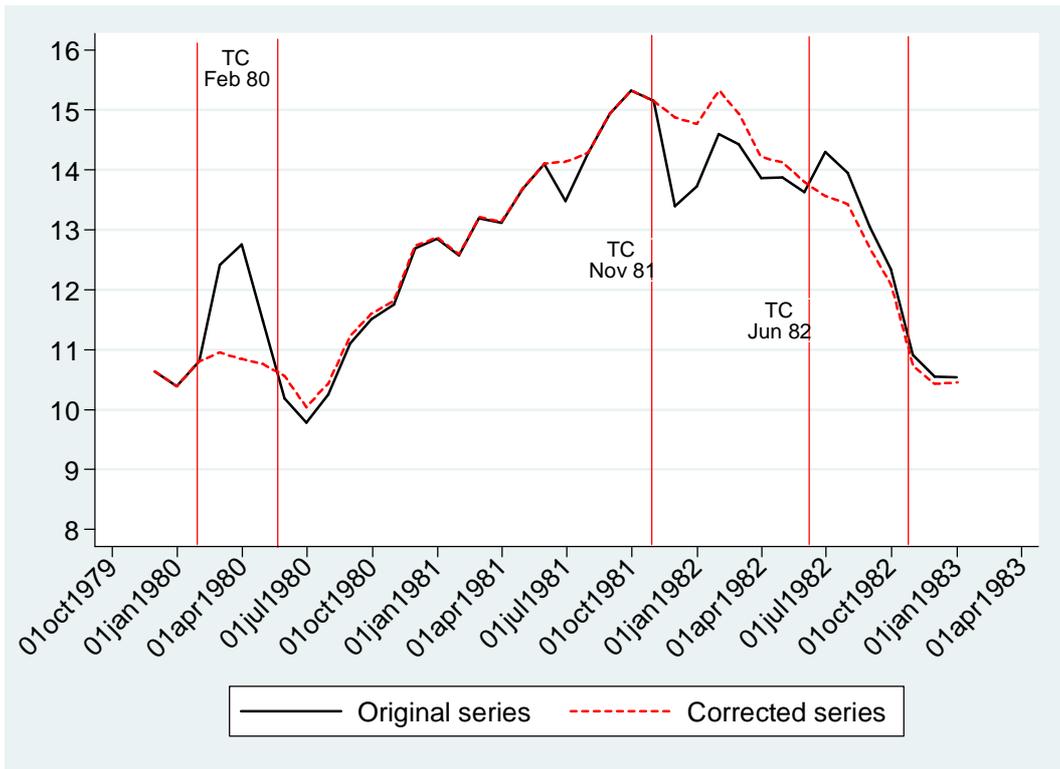
c. US Generic Govt. 10Y yield (1979--1983)
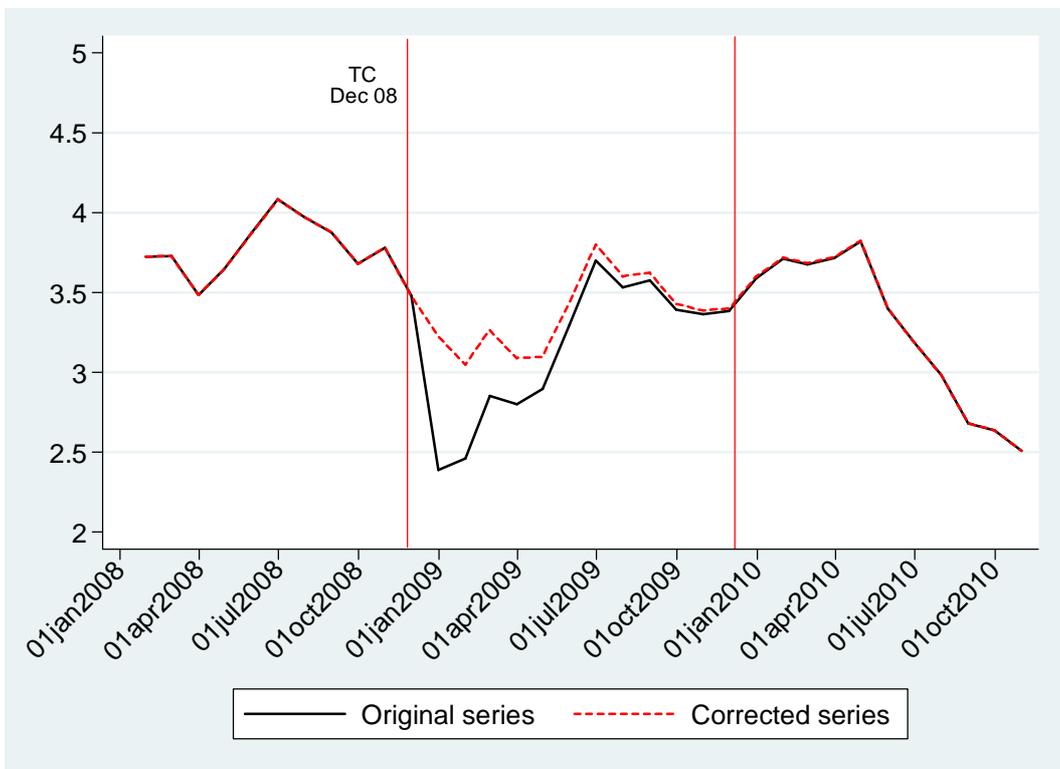
d. US Generic Govt. 10Y yield (2008--2010)

**Figure 7:** Outliers detection in the slope of the yield curve (US, 1972--1983)
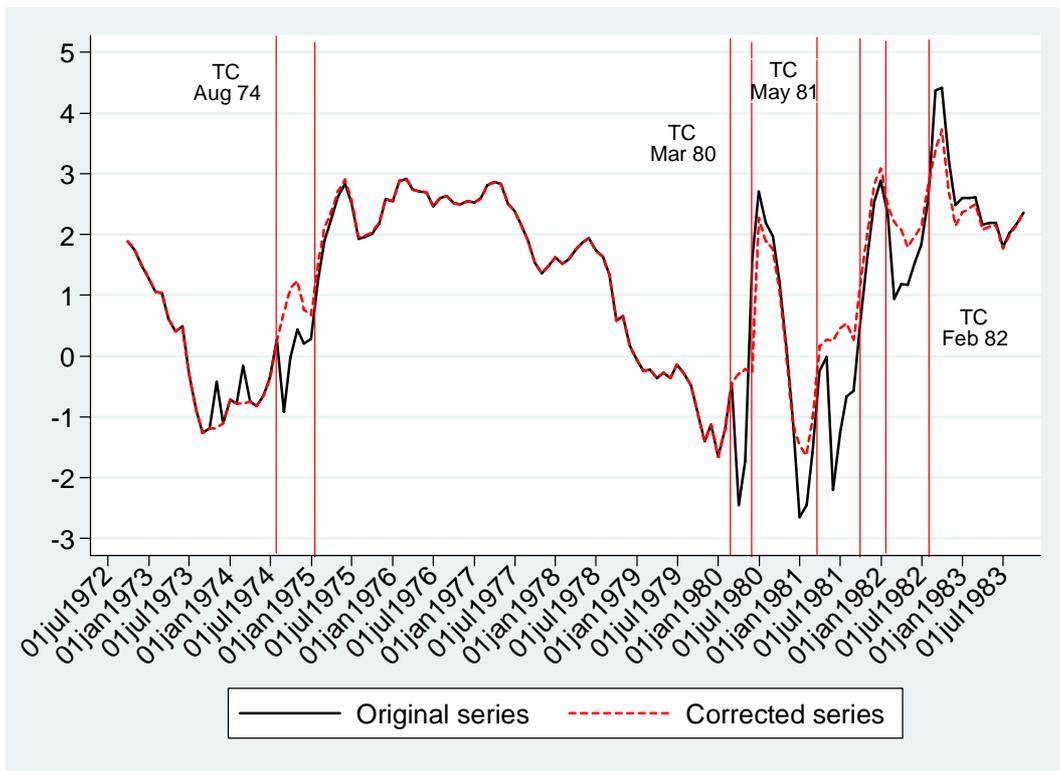
Figure 8: Outliers detection in commodity prices
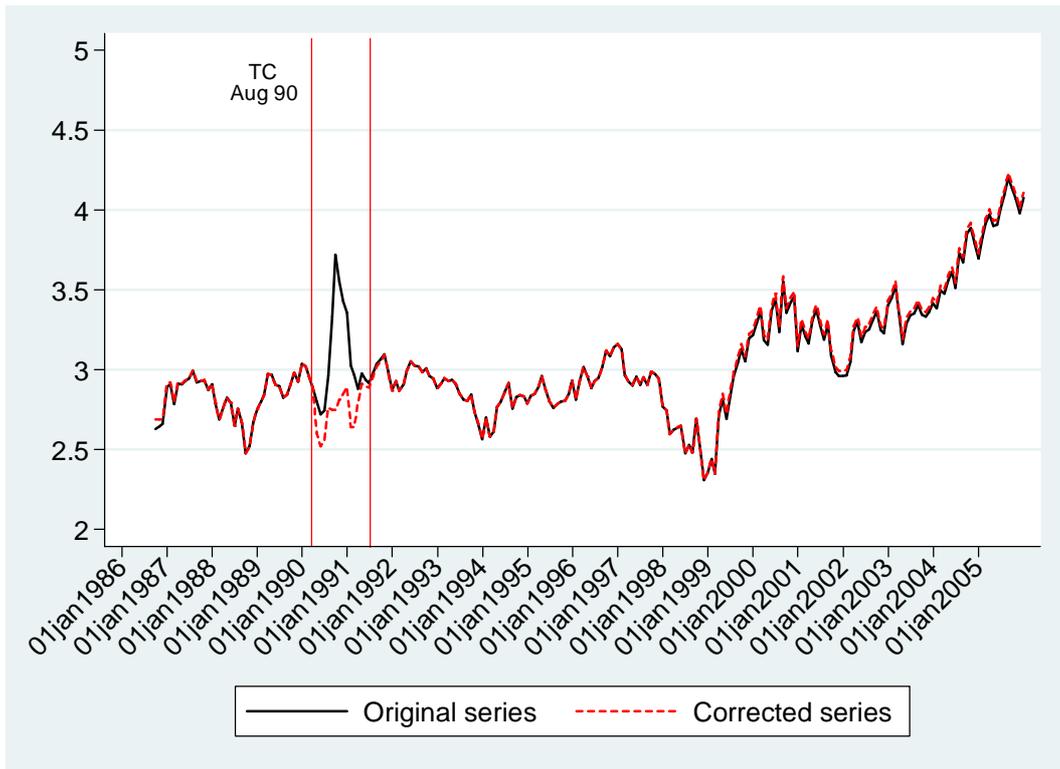
a.   Bloomberg Oil Price Index, 1986—2005 (logarithmic scale)
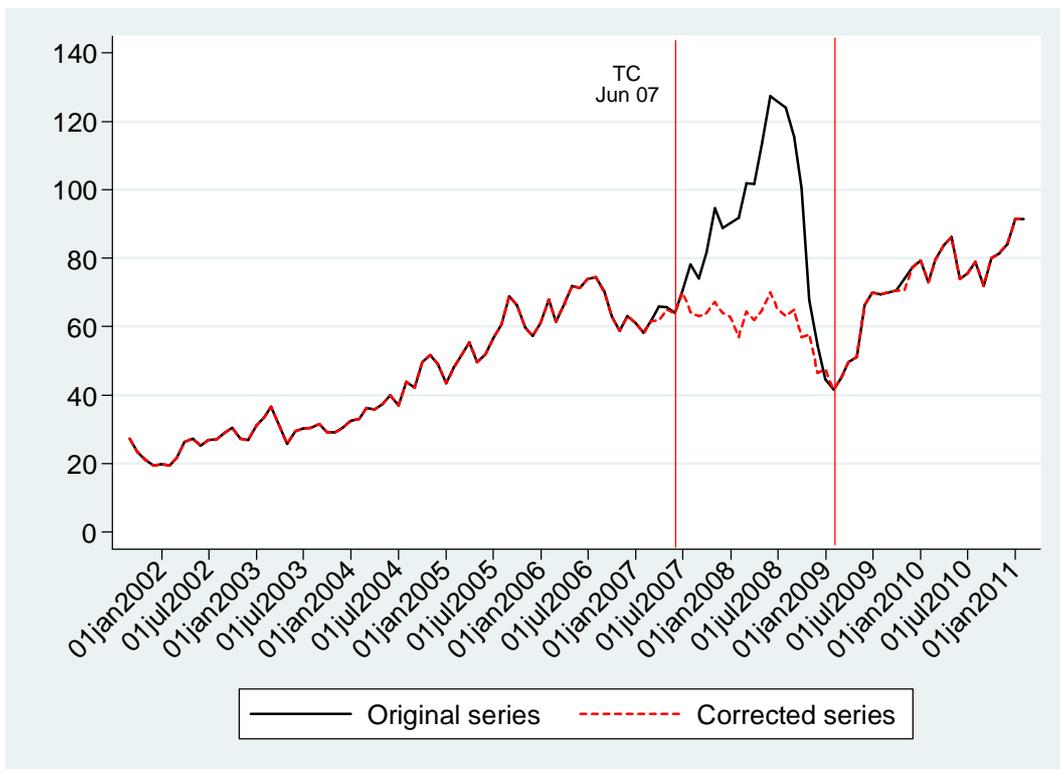
b.   Bloomberg Oil Price Index, 2002—2011 (linear scale)

Figure 8: Outliers detection in commodity prices (cont.)
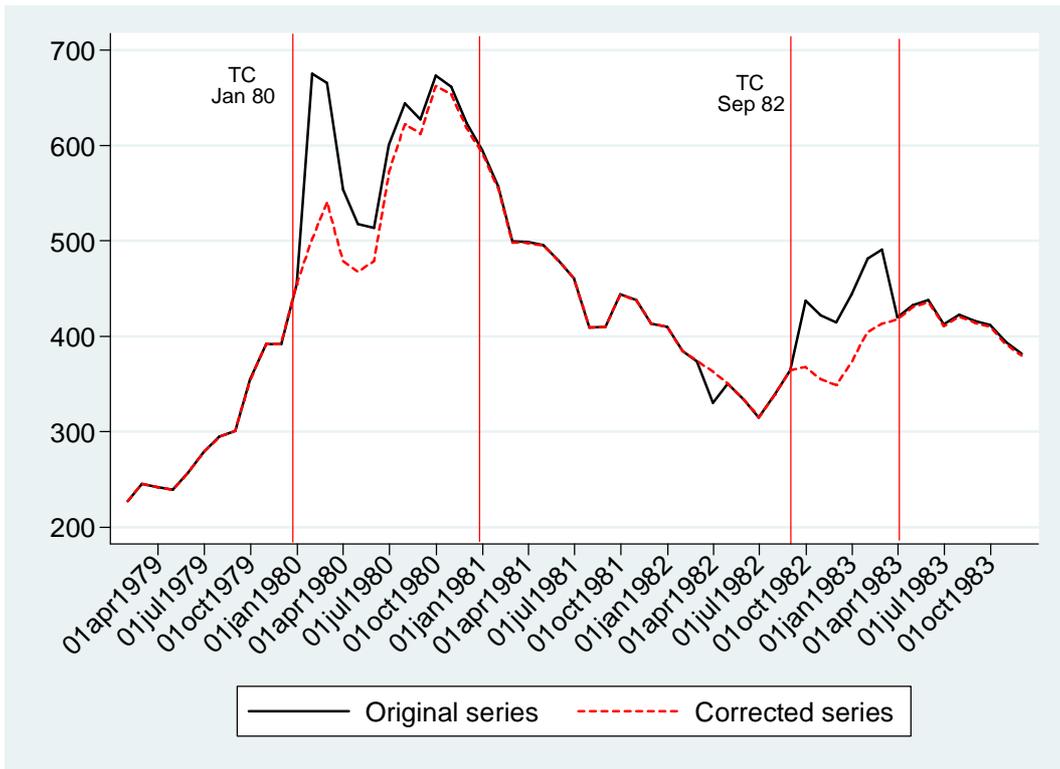
c. Gold Spot Prices, 1979--1983