The reputational risk impact of internal frauds on bank customers: a case study on UniCredit Group

Daniele Ruspantini, Head of Reputational Risk Measurement team, UniCredit SpA
Alessandra Sordi, Reputational Risk Measurement, UniCredit SpA

Abstract

This paper assesses the reputational risk impact arising from the customers’ negative reaction to 20 internal fraud cases occurred in some UniCredit Group Italian retail branches in the period 2008-2010. The reputational risk impact is measured applying the event study methodology: the clients’ assets under custody and management have been used to explain the customers’ reaction to the event. It is demonstrated that these frauds have a statistically significant reputational risk impact that is relevant both in terms of strength and length and that internal frauds detected due to customers’ complaints are more damaging from a reputational risk point of view than the ones detected due to internal controls.

Keywords: reputational risk measurement, Event Study methodology, internal frauds, customers’ behavior

Acknowledgements

The article authors are very grateful to Gabriele Maucci and Maria Gaia Soana for their valuable contributions.

1. Introduction

In the Basel Committee on Banking Supervision (BCBS) document “Enhancement to the Basel II Framework” (July ‘09), the need of introducing reputational risk measurement techniques has been strongly underlined by bank regulators for the first time. It is in fact repeated many times that banks should be able to evaluate the reputational risks associated with their off-balance sheets exposures and that reputational risk scenarios should be inserted in regular stress-test performing but, above all, it is stated that “in order to avoid reputational damages and to maintain market confidence, a bank should develop methodologies to measure as precisely as possible the effect of reputational risk in terms of other risk types (e.g. credit, liquidity, market or operational risk) to which it may be exposed”.

In this context, UniCredit Group has approached the reputational risk measurement issue by starting developing a comprehensive methodology allowing to measure the reputational risk knock-on impact of four types of primary risks (credit, market, liquidity
and operational risk) on four stakeholder categories (customers, investors, employees and institutional counterparties).

The analysis which is shown in this paper is inserted in this context: the focus is on the customer stakeholders’ category and the aim is to evaluate the strength and length of the reaction that internal frauds perpetrated at bank retail branches provoke on clients. Such reaction might in fact result in the bank lack of capability in maintaining the existing or establishing new business opportunities: the reputational risk impact of the event.

The article is structured as follows: section two provides a brief overview on the literature context which the article belongs to; section three details the methodology applied; section four describes the sample that is used in the performed empirical analyses; section five reports the analyses results and the related comments; section six briefly summarizes the article most important considerations.

2. Context

This paragraph provides some brief considerations on the reference context for the performed analyses: readers who are interested in deepening this topic may refer to the articles which are quoted in the bibliography.

Before the change in regulators’ perspective which has been mentioned in the article introduction, operational risk was deemed the most important source for the reputational one. For this reason contributions from financial literature have focused on the reputational losses arising from the operational losses announcements for listed companies. The reputational losses were evaluated on the companies market capitalization meaning that the stakeholder category investors was taken into account. Looking very quickly to the findings of some of the latest analyses performed in this context, Cummins, Lewis and Wei (2006) studying the impact on the market value of American banks and American insurance companies found out that reputational losses are more marked for insurances than for banks; De Fontnouvelle and Perry (2005) detected that reputational losses are particularly relevant for banks characterized by strong shareholders’ right; Schwizer, Soana, Maucci, Ruspantini (2010) analyzing for the first time five out of the seven regulatory operational risk event types highlighted that the most relevant reputational losses are suffered with external frauds and more over that small operational losses can lead to reputational damages equal to and in some cases slightly greater than, those of the higher operational losses (the threshold between the two classes is set to 10 millions USD $)

The aim of this article is to start overcoming these limits: operational risk is still considered but the reputational risk impact is evaluated keeping into consideration the customers’ stakeholders category. The article authors hope this can be the starting point for a number of analyses by financial practitioners and academics taking into account all the different aspects of reputational risk (e.g. different primary risks and/or different stakeholder categories).
3. Methodology

As already mentioned this paper aims to evaluate the reputational risk impact arising from the customers’ negative reaction to 20 internal fraud cases occurred in some UniCredit Group Italian retail branches in the period 2008-2010. This impact is evaluated in terms of the strength of the reaction that customers show to the event and in terms of the length of this reaction: in the UniCredit Reputational Risk Measurement framework these are referred to as **Intensity of the customers’ reaction** and **Time to Reputation Recovery**.

3.1. The considered variables

The articles, which have been already mentioned in the previous paragraph, evaluate the “reputational losses” that listed companies suffer on the market as a consequence of the announcement of relevant operational losses. They investigate in fact the stock return changes resulting from the announcement of operational losses: the registration of any returns significantly differing from the predicted ones is interpreted as an anomaly to be attributed to the new information made available to the market. It can thus be concluded that they measure the reputational risk impact of these events taking into consideration only the stakeholder category investors.

The proposed analysis wants instead to investigate the reaction of the stakeholder category customer. A different variable is thus needed, allowing to describe the behaviour of customers over time and sufficiently “sensible” to catch any reaction that customers might have as a consequence of an internal fraud event. The assets under custody and management (hereafter AUC&M) have been chosen: for each considered case: the total AUC&M of the branch where the fraud has been perpetrated is chosen as explanatory variable while the total AUC&M of the geographical area the branch belongs to (i.e. the sum of the total AUC&M of all the branches of the geographical area where the defrauded branch is located) is considered as benchmark. It has been chosen to consider as explanatory variable the AUC&M of all the customers of the impacted branch and not only the AUC&M of the defrauded clients as it is deemed that reputational risk is particularly damaging when a sort of contagion effect makes the number of stakeholders (within a chosen stakeholder category) reacting to the primary event higher than the number of the impacted ones.

The data used in the analysis were available on a monthly basis. This is a huge difference with the analyses investigating the reactions of the market where daily data are available. A couple of consideration shall thus be made. From a practical point of view, the adoption of monthly data is considered a proper choice to describe the customers’

---

1 It shall be underlined these represent only two out of the five factors which UniCredit Reputational Risk Measurement and Control uses to evaluate the reputational risk impact of events. The five factors are: the degree of the exposure to the primary risk, the probability of the trigger event, the intensity of the stakeholder reaction, the event amplifier factor and the time to reputation recovery.

2 More in details, “markets” are only one of the components of the stakeholder category investors (e.g. shareholders).
reaction: customers in fact need time to be aware of an event, get a sense of it and coherently react, thus eventual changes in their behaviours cannot be appreciated with short period variables (e.g. daily variables). From the statistical point of view instead, the usage of monthly data makes more questionable the robustness of the achieved results: a lower number of observations is in fact available for the regression analysis as it is not possible to go too far back in time without the risk of including in the analysis some distortion elements. By the way, as a practitioner approach wants to be followed, the positive considerations on the capability and reasonableness in describing the customers’ behaviours have prevailed on the concerns about the statistical robustness.

3.2. Methodology description

Accordingly with the studies already quoted, Event Study Methodology has been applied: by comparing the values of AUC&M effectively experienced in the branch with the ones which would have been experienced if the event had not happened, it allows to detect if customers show an abnormal reaction when they are hit by the fraud events. The existence of an abnormal trend for their AUC&M is then interpreted as the reputational risk impact of the event.

For each considered event $i$, the “normal” behavior of the branch AUC&M variable has been estimated applying the market model:

$$R_{i,t} = \alpha_i + \beta_i R_{mk_i,t} + \varepsilon_{i,t}$$

where
- $R_{i,t}$ is the logarithmic return of the explanatory variable (AUC&M of the branch)
- $R_{mk_i,t}$ is the logarithmic return of the benchmark (AUC&M of the geographical area)
- $\alpha_i$ and $\beta_i$ are the linear model coefficients estimated with the Ordinary Least Squares regression
- $\varepsilon_{i,t}$ is the residual of the model, that is the part of $R_{i,t}$ which could not be explained by the benchmark; for hypothesis $i$, $iid \sim N(0, \sigma^2_i)$.

This market model is performed for each event $i$ on a pre-event window including a number of observations varying from 12 to 24 before the chosen event window. The event window is the interval ranging between $-\tau_1$ months before and $\tau_2$ months after the event (the time $t=0$ representing the event discovering date is also comprised) in which the event reputational consequences are evaluated.

It shall be noted that the pre-event windows of the different cases have different lengths, depending on data availability: the lower and upper limits for the range (respectively 12 and 24 months) are set considering that less than 12 observations are deemed not sufficient to obtain statistical relevant considerations, and that including more than two years can add some distortion factors in the model estimation. The statistical significance of the market model has been tested considering the $R^2$ values and verifying the stationarity of the logarithmic returns time series of both the explanatory variable and benchmark: positive results have been obtained.
More over, it wants to be underlined that three types of event windows have been considered: the ones preceding the event, the ones foregoing the event and the ones including months both before and after the event. The event windows of the first type have been taken into account as they allow to study if reputational consequences are suffered even before the bank officially recognizes the fraud.

The evidences of the market model are then used in a predictive way to determine the Abnormal Returns $AR(t)$ at each time $t$ in the event window $[-\tau_1, \tau_2]$. The Abnormal Returns are defined as the differences between the really experienced AUC&M values for every month $t$, $R_t^*$, and the values calculating using the coefficients $\alpha_i$ and $\beta_i$ estimated with the market model:

$$AR_i(t) = R_t^* - \alpha_i - \beta_i * R_{mk,i,t}^*$$

where the symbol $*$ means that only data in the event window are considered.

For each event $i$, the Cumulative Abnormal Returns are then evaluated as the sum of consecutive Abnormal Returns in the chosen interval $[-\tau_1, \tau_2]$:

$$CAR_i(t) = \sum_{s=1}^{t} AR_i(s)$$

The average $\overline{CAR}$ time series is instead obviously defined as

$$\overline{CAR}(\tau_1, \tau_2) = \frac{1}{n} \sum_{i=1}^{n} CAR_i(\tau_1, \tau_2)$$

where $n$ is the total number of cases under consideration.

The statistical significance of the achieved results has been studied by considering the one tail test which has been originally presented by Campbell, Lo and Mackinley (1997) and which verifies the null hypothesis whereby the event has no influence on the cumulated returns. The considered statistic is thus

$$T = \frac{\overline{CAR}(\tau_1, \tau_2)}{\hat{\sigma}^2(\tau_1, \tau_2)} \approx N(0,1)$$

where $\hat{\sigma}^2(\tau_1, \tau_2)$ is the estimated variance of the $\overline{CAR}$ time series.

This test is based on the assumption of the normal distribution for both $AR$ and $CAR$ time series: the Jarque Bera test has been performed to investigate this hypothesis. As it will be shown in paragraph 5, the normal distribution condition is always respected.

4. Sample description

The analysis has been performed on a sample of 20 internal frauds which have been extracted from the official UniCredit database for operational risk according to the following criteria:
1. the internal fraud has occurred in a retail branch;
2. the event has happened in the Italian perimeter and its observation date (the date in which the event has been officially discovered by the company) is comprised between the last quarter of 2008 and the first quarter of 2010;
3. more than a customer has been affected by the fraud event;
4. the operational loss has an amount higher than € 50.000.

The first two criteria aim to define and limit the application scope of the analysis in terms of the type of stakeholder investigated (mass market customers) and in terms of the geographical region the stakeholder belongs to (Italy) and the time where his/her behavior is investigated.

The last two criteria are instead introduced to describe the relation (“trigger”) between the primary risk event (in this case an operational one) and the secondary reputational risk impact. These two criteria have been chosen by the articles’ authors for their reasonableness. It is in fact easy to understand that
1. customers may be more sensible to primary events directly hurting them rather than to primary events not impacting on them (i.e. an internal fraud where money is embezzled from clients is considered more damaging from a reputational risk point of view than an internal fraud where a branch employee steals cash from the deposit vault)
2. the higher is the amount of the primary loss, the higher is the coverage which might be potentially granted to the event by the media (even at local level), this potentially increasing the set of the stakeholder that being aware of the event might react to it (e.g. if an internal fraud event is reported by media, also the branch customers who have not been affected are aware of it and thus can react to it).

In Table 1, some statistics about the considered operational losses are reported (values are expressed in Euro):

**Table 1: operational losses statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Min</strong></td>
<td>53,500</td>
</tr>
<tr>
<td><strong>Max</strong></td>
<td>3,568,000</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>756.606</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>358.900</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>1,020.211</td>
</tr>
</tbody>
</table>

As already stated in paragraph 3.1, for each one of the detected cases the monthly time series of the total AUC&M of the branch where the event has occurred and of the total AUC&M of the branches of the geographical area to which the defrauded one belongs have been considered. Please note again that the benchmark is described by the sum of this second set of monthly total AUC&M.
5. Results

The analysis of the considered data, has thus allowed to evaluate the reputational risk impact arising from the customers reaction to the internal fraud events which have been selected. It is considered that a reputational risk impact exists when a statistically significant negative value for the CAR is obtained.

The analysis has been performed at two different level:
- the cases have been investigated altogether
- the cases have been divided in two samples, the internal frauds discovered due to customers complaints and the ones discovered due to internal control. This has been done to detect if a statistical evidence can be found for the hypothesis that the internal fraud cases which are detected by customers are potentially more damaging for the bank from a reputational risk point of view than the ones which are detected due to internal controls.

The results of the first exercise (data investigated altogether) are reported in Table 2: the columns respectively represent the investigated event window, the number of considered cases, the CAR, the value of the statistic T and the p-value of the Jarque Bera test.

<table>
<thead>
<tr>
<th>Event window</th>
<th>N. cases</th>
<th>CAR</th>
<th>T</th>
<th>Jarque Bera test (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-3; 0]</td>
<td>20</td>
<td>-0.031</td>
<td>-1.906**</td>
<td>0.848</td>
</tr>
<tr>
<td>[-3; 3]</td>
<td>20</td>
<td>-0.063</td>
<td>-2.678***</td>
<td>0.711</td>
</tr>
<tr>
<td>[-3; 5]</td>
<td>20</td>
<td>-0.077</td>
<td>-2.709***</td>
<td>0.665</td>
</tr>
<tr>
<td>[-3; 7]</td>
<td>20</td>
<td>-0.082</td>
<td>-2.504***</td>
<td>0.587</td>
</tr>
<tr>
<td>[0; 1]</td>
<td>20</td>
<td>-0.024</td>
<td>-2.241**</td>
<td>0.846</td>
</tr>
<tr>
<td>[0; 3]</td>
<td>20</td>
<td>-0.041</td>
<td>-2.525***</td>
<td>0.835</td>
</tr>
<tr>
<td>[0; 5]</td>
<td>20</td>
<td>-0.055</td>
<td>-2.563***</td>
<td>0.824</td>
</tr>
<tr>
<td>[0; 7]</td>
<td>20</td>
<td>-0.061</td>
<td>-2.311**</td>
<td>0.707</td>
</tr>
</tbody>
</table>

* Statistically significant at 10% (one-tailed test)
** Statistically significant at 5% (one-tailed test)
*** Statistically significant at 1% (one-tailed test)

Table 2 shows that a statistically significant reputational risk impact (negative CAR) is detected for each considered event window: more in detail the CAR in the event windows [-3;0], [0;1], [0;7] are statistically significant at 5% while the average CAR in the event windows [-3;3], [-3;5], [-3;7], [0;3], [0;5] are statistically significant at 1%. Quite surprisingly, a reputational risk impact is even detected in the event window [-3;0], this meaning that customers starts reacting before the event is officially recognized by the bank.
It shall also be noted that both the event windows of the type \([-3,x]\) and the ones of the type \([0,x]\) track an increasing trend for the $\overline{\text{CAR}}$ (see figure 1 for the second type): this means that the customers’ reaction becomes more severe with the passing of time. Moreover, considering that for both the event window types, the increasing trend is not ended with the last available observation, it has to be concluded that, on average, seven months after the event detection the bank has not yet been able to stop the negative consequences. Thus, considering that the effect seems starting three months before the event detection (the analysis was not repeated for previous months as it seemed that otherwise too few data would have been used in the market model) it can be concluded that the reputational risk impact lasts at least for 11 months.

**Figure 1: $\overline{\text{CAR}}$ trend**

![Internal Frauds analysis](image)

Considering now the second part of the analysis, Table 3 and 4 show the results respectively for the cases the fraud is detected due to internal control and the fraud is detected due to customers complaints.
Table 3: fraud detected due to internal control

<table>
<thead>
<tr>
<th>Event window</th>
<th>N. cases</th>
<th>CAR</th>
<th>T</th>
<th>Jarque Bera test (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-3; 0]</td>
<td>10</td>
<td>-0.011</td>
<td>-1.146</td>
<td>0.768</td>
</tr>
<tr>
<td>[-3; 3]</td>
<td>10</td>
<td>-0.021</td>
<td>-1.497*</td>
<td>0.684</td>
</tr>
<tr>
<td>[-3; 5]</td>
<td>10</td>
<td>-0.034</td>
<td>-2.018***</td>
<td>0.605</td>
</tr>
<tr>
<td>[-3; 7]</td>
<td>10</td>
<td>-0.027</td>
<td>-1.428*</td>
<td>0.615</td>
</tr>
<tr>
<td>[0; 1]</td>
<td>10</td>
<td>-0.005</td>
<td>-0.756</td>
<td>0.846</td>
</tr>
<tr>
<td>[0; 3]</td>
<td>10</td>
<td>-0.008</td>
<td>-0.874</td>
<td>0.75</td>
</tr>
<tr>
<td>[0; 5]</td>
<td>10</td>
<td>-0.021</td>
<td>-1.682**</td>
<td>0.871</td>
</tr>
<tr>
<td>[0; 7]</td>
<td>10</td>
<td>-0.015</td>
<td>-0.988</td>
<td>0.791</td>
</tr>
</tbody>
</table>

* Statistically significant at 10% (one-tailed test) ** Statistically significant at 5% (one-tailed test)
*** Statistically significant at 1% (one-tailed test)

The two tables show that while in the case the fraud is detected due to customers’ complaints a statistically significant reputational risk is observed for each considered event window, in the case the fraud is detected due to internal control only the event window [-3;3], [-3;5], [-3;7] and [0;5] are statistically significant (the first and the third one at the 10%, the other ones at the 5%).

The obtained $\overline{CAR}$ are different not only for their statistical significance but also for their values. While in the internally detected case the four statistical significant values range between -0.021 and -0.034 for the customers’ complaints detected case the average $\overline{CAR}$ values range between -0.044 and -0.138.

Thus the hypothesis to be tested seems confirmed: as both the values and the statistical significance of $\overline{CAR}$ are higher in the case the fraud is detected due to customers’ complaints, it seems possible to conclude that this kind of internal frauds is more damaging from a reputational risk point of view than the internal detected ones.

Table 4: fraud detected due to customers’ complaints

<table>
<thead>
<tr>
<th>Event window</th>
<th>N. cases</th>
<th>CAR</th>
<th>T</th>
<th>Jarque Bera test (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[-3; 0]</td>
<td>10</td>
<td>-0.051</td>
<td>-1.641**</td>
<td>0.875</td>
</tr>
<tr>
<td>[-3; 3]</td>
<td>10</td>
<td>-0.105</td>
<td>-2.340***</td>
<td>0.738</td>
</tr>
<tr>
<td>[-3; 5]</td>
<td>10</td>
<td>-0.120</td>
<td>-2.214**</td>
<td>0.653</td>
</tr>
<tr>
<td>[-3; 7]</td>
<td>10</td>
<td>-0.138</td>
<td>-2.183**</td>
<td>0.605</td>
</tr>
<tr>
<td>[0; 1]</td>
<td>10</td>
<td>-0.044</td>
<td>-2.112**</td>
<td>0.846</td>
</tr>
<tr>
<td>[0; 3]</td>
<td>10</td>
<td>-0.074</td>
<td>-2.373***</td>
<td>0.85</td>
</tr>
<tr>
<td>[0; 5]</td>
<td>10</td>
<td>-0.089</td>
<td>-2.163**</td>
<td>0.766</td>
</tr>
<tr>
<td>[0; 7]</td>
<td>10</td>
<td>-0.107</td>
<td>-2.115**</td>
<td>0.745</td>
</tr>
</tbody>
</table>

* Statistically significant at 10% (one-tailed test) ** Statistically significant at 5% (one-tailed test)
*** Statistically significant at 1% (one-tailed test)
A last consideration regards the bank possibility to recover its reputation: looking at Tables 3 and 4 and at Figure 2 and 3 it is in fact possible to understand that while in the customers’ complaints case the negative increasing CAR trend is not finished at the end of the seventh month after the event detection (the last data available), in the internal detected case the negative increasing CAR trend stops at the end of the fifth month after the event detection. This seems to underline that in the case the fraud is internally discovered it is possible for the bank to soften the reputational risk impact not only in terms of strength of the customer reaction but also in terms of the time which is needed to recover its pre-event reputation level. On the other side instead it seems that when the internal fraud is discovered due to customers’ complaints it is not anymore possible for the bank to recover the pre-event reputation level (at least on a year time horizon) meaning that the relationship of trust with customers is structurally compromised.

**Figure 2: CAR trend for the internal detected frauds**

[Graph showing CAR trend for internal frauds]
6. Conclusion

The article wants to evaluate the reputational risk impact arising from the customers’ negative reaction to 20 internal fraud cases occurred in some UniCredit Group Italian retail branches in the period 2008-2010. Cases have been selected as they respect some reputational risk “trigger criteria” (see paragraph 4). The reputational risk impact (to be expressed in terms of business capability) has been evaluated in terms of strength and length of the customers’ reaction to the event: the assets under custody and management have been chosen to describe them.

The performed analysis has first of all allowed to detect that these internal frauds have produced a relevant reputational risk impact. More in detail, the reputational risk impact starts three months before the event is officially recognized by the bank and it is still increasing at the end of the seventh month after it (the last data available for the analysis). It can thus be concluded that the bank is not able to recover its pre-event reputation level (with regards to customers) in a one year time horizon.

The second part of the analysis aims to compare the frauds detected due to customers’ complaints with the ones discovered due to internal controls as it is considered that the
first ones are more damaging from a reputational risk perspective. The performed analysis confirms this hypothesis: while for the first group obtained results are higher than in the case where all the frauds have been considered, for the second group results are statistically significant only in few cases and with a very low value.

A final consideration that needs to be underlined regards the bank capability of restoring the pre-event level of reputation in terms of outstanding level of business. While in the case the fraud is internally discovered a recovery of the pre-event level of reputation seems tracked, in the other case it seems not possible for the bank to recover the pre-event relationship of trust with customers at least on a one year time horizon.

In the end, just a couple of brief considerations on the fields of possible development for this kind of analysis. The authors think in fact that useful enhancements might be achieved by considering other kinds of variables to describe the customers reaction and/or by refining the reputational risk “trigger criteria” which are used to select the events to be investigated.
References


