BANCA D'ITALIA
EUROSISTEMA

Unità di Informazione Finanziaria per l'Italia

Quaderni dell'antiriciclaggio

Analisi e studi

Looking at 'Crying wolf' from a different perspective: An attempt at detecting banks under- and over-reporting suspicious transactions

Mario Gara e Claudio Pauselli

novembre 2015

numero 4

BANCA D'ITALIA
EUROSISTEMA

Unità di Informazione Finanziaria per l'Italia

Quaderni dell'antiriciclaggio

Analisi e studi

Looking at 'Crying wolf' from a different perspective:
An attempt at detecting banks under- and over-reporting suspicious transactions

Mario Gara e Claudio Pauselli

Numero 4 novembre 2015

La serie Quaderni dell'antiriciclaggio *ha la finalità di presentare dati statistici, studi e documentazione su aspetti rilevanti per i compiti istituzionali della UIF — Unità d'Informazione Finanziaria per l'Italia, Banca d'Italia.*

La serie si articola in due collane: la collana Dati statistici *presenta, con periodicità semestrale, statistiche sulle segnalazioni ricevute e informazioni sintetiche sull'operatività dell'Unità; la collana* Analisi e studi *comprende contributi sulle tematiche e sui metodi in materia di contrasto al riciclaggio e al finanziamento del terrorismo.*

La collana Analisi e studi *comprende lavori realizzati all'interno della UIF, talvolta in collaborazione con altri settori della Banca d'Italia o con Istituzioni esterne. I lavori pubblicati riflettono esclusivamente le opinioni degli autori, senza impegnare la responsabilità delle Istituzioni di appartenenza.*

Unità di Informazione Finanziaria per l'Italia

Direttore responsabile
Claudio Clemente

Per la pubblicazione cartacea:
autorizzazione del Tribunale di Roma n. 1942013 del 30 luglio 2013
Per la pubblicazione telematica:
autorizzazione del Tribunale di Roma n. 1932013 del 30 luglio 2013

ISSN 2283-3498 (stampa)
ISSN 2283-6977 (online)

Looking at 'Crying wolf' from a different perspective: An attempt at detecting banks under- and over-reporting suspicious transactions

by Mario Gara[*] and Claudio Pauselli [*]

Abstract

By estimating an econometric model, this study aims to assess, from a quantitative point of view, the flow of suspicious transaction reports (STRs) filed by Italian banks from each of the provincial districts they operate in. Regressors include (i) indicators of banks' operational activities; (ii) measures of money laundering risk and (iii) proxies of economic activity, all of which at local level. The analysis presents some technical challenges which are addressed by adopting a Negative Binomial setting, commonly used to model count data variables. In addition, observations are split into two sub-samples, according to each bank's local scale of operation. Results show that the STR-filing strategies adopted by banks may be different from the 'crying wolf' approach, which is traditionally considered to be the most pressing threat to the effectiveness of anti-money laundering systems. Furthermore, at a more operational level, the model provides a useful tool that supervisory authorities can deploy when checking the compliance of individual intermediaries with anti-money laundering reporting regulations.

Sommario

Utilizzando un modello econometrico, lo studio si propone di fornire indicazioni sull'adeguatezza, in termini quantitativi, del flusso di segnalazioni di operazioni sospette trasmesse dalle banche italiane su base provinciale. Le variabili esplicative utilizzate dal modello sono di tre tipi: (i) indicatori relativi all'operatività delle banche, (ii) misure di rischio di riciclaggio e (iii) indicatori di attività economica, in tutti i casi misurati a livello locale. L'analisi ha presentato delle difficoltà tecniche che vengono affrontate utilizzando un modello Binomiale Negativo, comunemente applicato a variabili di conteggio. Inoltre, le osservazioni sono distinte in due sotto-campioni, in base alla scala di operatività espressa da ciascuna banca a livello locale. I risultati mostrano che il comportamento segnaletico delle banche non necessariamente si ispira all'approccio 'al lupo, al lupo' (sovra-segnalazione sistematica a scopo prudenziale), considerato tipicamente il maggior limite all'efficacia dei sistemi anti-riciclaggio. Per quanto riguarda le implicazioni operative dello studio, il modello rappresenta un utile strumento che le autorità antiriciclaggio possono impiegare nel monitorare la compliance *degli intermediari in materia di obblighi segnaletici.*

JEL Classification: C25, G21, G28, K23.

Keywords: Count Data, Negative Binomial, Anti-Money Laundering Regulation.

---

[*] UIF (Italian Financial Intelligence Unit), Bank of Italy. We thank for useful suggestions Domenico J. Marchetti, Marco Marinucci, Paolo Naticchioni and participants at the 2014 ISLE Conference in Rome and the 2015 UIF-Bocconi Workshop on "Quantitative Methods and the Fight to Economic Crime". All remaining mistakes are all the authors' fault.

Contents

1. Introduction

In his benchmark work on anti-money laundering and countering terrorist financing regimes, Takàts (2011) shows that, within the most common regulatory frameworks, such regimes risk being drowned in an overwhelming flow of suspicious transaction reports (or STRs).[1] Indeed, banks (and other intermediaries) mandated by law to file STRs have an incentive to over-report suspicious transactions rather than appropriately select them for authorities' consideration. This occurs since they face penalties if they fail to report money laundering cases — that is, transactions which are later prosecuted as money laundering or judged by authorities to be suspicious — whilst they substantially incur no sanctions for filing unfounded reports. As a result, anti-money laundering systems turn out to be affected by a 'crying wolf' syndrome, whereby their effectiveness is seriously undermined by an overload of useless information.

This study provides an empirical test of Takàts' conclusion, by making use of data on STRs filed by Italian banks in 2012 to *Unità di Informazione Finanziaria* (UIF), which is Italy's Financial Intelligence Unit (FIU henceforth), that is the national anti-money laundering authority. In particular, the number of high risk STRs filed by banks from the various Italian provinces is regressed against a wide set of explanatory variables, which include (i) measures of money laundering risk, (ii) indicators of banks' operational and financial activities and (iii) proxies of socio-economic conditions, all of which at local level.

Results show that banks' level of compliance with the reporting obligation varies significantly across intermediaries, and that potential under- and over-reporting banks can be detected. Moreover, the reporting policies of individual banks seem to be consistent across different provinces, partly as a consequence of the centralization of reporting decisions.

In addition to these theoretical results, this study aims at producing its most relevant outcomes at an operational level. In this perspective, the model being set up here can be effectively deployed as a tool for supervisory purposes by oversight authorities when checking the compliance of intermediaries with their reporting obligations. Indeed, the UIF used some of the results of earlier versions of the model to help define its on-site visit program in previous years.

The study is structured as follows. Section 2 states the research questions that we address. The model being estimated is described in Section 3, with a discussion of the explanatory variables and the rationale for their inclusion. Section 4 mentions some data modeling issues and describes how they have been tackled. Estimation results are presented and commented upon in Section 5, whilst Section 6 exposes the operational application of the results obtained for the purpose of assessing banks' level of compliance with reporting regulations. Some brief concluding remarks follow.

---

[1] According to most legislations and international standards, if a financial institution suspects that funds are the proceeds of a criminal activity, or are related to the financing of terrorism, is required to report promptly its suspicion to the national anti-money laundering authority.

2. Purpose of the research

The research question of this study is twofold and has relevant operational implications for anti-money laundering purposes.

Firstly, the study investigates to what extent the suspicious transaction reporting mechanism, which lies at the heart of all anti-money laundering regimes, is efficient. More notably, drawing from the literature analyzing the functioning of money laundering systems, we investigate whether it is possible to define an expected level of reporting flows by individual banks and which variables may be usefully taken into consideration for this purpose.

In this perspective, the study departs from the seminal works of Masciandaro (1999) and Masciandaro and Filotti (2001), which investigate the costs of anti-money laundering regulations for banks and how these may be reconciled with the socially valuable goal of detecting and disrupting money laundering and the financing of terrorism. Instead, the issues being dealt with here are more akin to those raised by Takàts (2011). In his formalization, the reporting regime is set up in a principal-agent framework with two players, i.e. the government and reporting banks. An excessively high level of fines established by the former to punish an inadequate amount of effort by the latter could result in a report overload, in turn diluting the informational significance of STRs. Dalla Pellegrina and Masciandaro (2009) extend this framework by introducing an additional player, a supervisory authority (which may well be identified in the FIU), whose role is twofold, i.e. that of assessing banks' reporting effort as well as the difficulty of detecting money laundering schemes. In this framework, banks may end up under-reporting useful STRs (and not only over-reporting them, as predicted by Takàts) since they put too little effort in identifying money laundering or money launderers are too sophisticated to detect. At the same time, though, the supervisor, with its insider knowledge, mitigates the asymmetric information distortions typically arising within the traditional principal-agent framework: banks are aware that the FIU can observe the actual state of money laundering technology and adequately assess their effort. As a result, the introduction of the FIU improves on alternative equilibria by inducing an adequate quality of banks' effort for low levels of fines.

Secondly, the study contributes to the more general literature on the use of quantitative methods, such as complex econometric models, for financial supervisory purposes, i.e. for assessing the effectiveness of financial regulations or the level of compliance of intermediaries. Econometric settings have been used typically to investigate which oversight approach tackles more efficiently the overall risk exposure of the national banking system of various countries, as in Barth, Caprio and Levine (2004) and Beck, Demirgüç-Kunt and Levine (2006). There are very few empirical studies (if none at all) on the functioning of money laundering regimes.[2]

[2] One exception is Masciandaro and Volpicella (2014), who, by setting up a political economy model, try to explain why after the 2001 Semptember 11th events anti-money laundering systems have increasingly featured law-enforcement FIUs, as opposed to financial ones.

3. The model

The model being estimated aims at determining the expected level of STRs filed by a given bank in a certain province in a given period of time. To this purpose, we use a set of explanatory variables which account for the size and type of the banks activity in that province, the area's socio-economic environment and some local measures of money laundering risk. The time span of the analysis is the whole of 2012 (if not stated otherwise) and the geographical dimension is Italy's provincial districts.

3.1. The dependent variable

The dependent variable is represented by the number of STRs that Italian banks (705 intermediaries) filed in 2012 from each province they operate in. More precisely, not all STRs are taken into account, but only those presenting a medium-to-high level of risk. The UIF's IT system for processing STRs embeds an automatic algorithm which assigns a risk score to every report as soon as it is received, measuring the degree to which it may be considered to be well-grounded and signaling its prospective relevance for anti-money laundering purposes. The initial, automatically-set score can be subsequently adjusted at the discretion of UIF financial analysts, according to what emerges from their investigations. The risk score may take five different values: 1 and 2 (respectively, low and medium-low risk), 3 (medium risk), 4 and 5 (medium-high and high risk). The dependent variable throughout the analysis is the overall number of STRs, per bank and province, scored at least 3 (denominated henceforth, for brevity, high risk STRs).

Since the aim of the paper is that of analyzing the actual level of compliance of intermediaries with the reporting obligation, reports that only add noise to the system (i.e. those with a very low risk) have been deliberately disregarded.

As for the regressors, the approach adopted here lies on the assumption that the number of STRs filed by each bank is affected by both the size of its overall activity and the degree of money laundering risk that it is exposed to.

3.2. The regressors: Financial measures of banks' local exposure to money laundering risk

A first set of regressors in the model draws on previous work carried out by the UIF, together with the Bank of Italy's Banking Supervision Department, aiming at setting up a system of indicators providing a measure of individual banks' exposure to money laundering risk. The experience acquired within the UIF from STRs analysis shows that financial conducts most often deployed to money laundering ends feature recurring operational traits. The indicators build on these recurring features.

Most underlying data come from the Aggregate Anti-Money Laundering Reports (SARA from the Italian acronym[3]), which banks and other intermediaries file every

[3] Full Italian name is *Segnalazioni Anti-Riciclaggio Aggregate*.

month to the FIU[4].

In particular, a first regressor used in our study refers to cash transactions. Indeed, illicit activities typically produce their proceeds in cash, which are normally those involved in the early stages of money laundering (so-called *placement*[5]). In addition to cash deposits and withdrawals, it is possible to extract from the SARA database the cash component of all other transactions, and thus compute the total amount of all transactions, recorded by each bank in each province, that is actually finalized in cash. This is the variable used in the model.

A second regressor refers to wire transfers with a group of countries that are considered high risk from a money laundering perspective because they either are designated as such by international organizations (so-called *black-listed* countries) or emerge as favourite destination or origin of ill-gotten funds from the analysis of STRs conducted at the UIF. Crucially, the SARA database contains information on the country where the intermediary of the counterpart of each wire transfer is located. A second regressor is thus obtained by summing the amount of all inward and outward transfers, respectively, from and to high risk countries, processed by each bank in every province on behalf of the respective clients.

A third regressor refers to another type of risky financial conduct, which takes place through over-the-counter transactions. Such transactions may typically occur in two cases: the customer of a bank is occasional (i.e., he does not hold an account at the bank) or, alternatively, he requires that the transaction is not recorded on his account. In the former case, the customer may try to exploit the fact that the bank does not have an ongoing relationship with him and thus fails to hold an in-depth knowledge of his financial profile. In the latter instance, the customer may be trying to render the transaction more difficult to trace, since it does not contribute to the turnover of an ongoing business relationship. In either case, the prejudice to financial transparency that both conducts are likely to cause is clear. Accordingly, the regressor is set equal to the total amount of over-the-counter transactions finalized by each bank in each provincial district.

An additional explanatory variable meant to provide a measure of money laundering risk is associated to the need for criminals to ease the transportation and transfer of high amounts of money, which in turn may require them to acquire high-denomination banknotes[6]. Alternatively, criminals may need to 'smurf' the proceeds of their activities to the end of minimizing the risk of detection, whence their demand for

---

[4] The Italian anti-money laundering law (Legislative Decree 231/2007) requires banks and other intermediaries to record all single transactions exceeding 15.000 euros in a specific archive (Single Electronic Archive). Each month intermediaries file these data to the UIF after aggregating individual records according to several criteria, which include the customer's place of residence and economic sector, the intermediary's branch where the transaction took place and the type of the transaction.

[5] See, for instance, '*The Money-Laundering Cycle*', United Nations Office on Drugs and Crime (UNODC) (http://www.unodc.org/unodc/en/money-laundering/laundrycycle.html).

[6] That such conducts are widely spread among criminals and launderers is consistent with the finding of an on-going work by the Financial Action Task Force (FATF), the international anti-money laundering standard setter, on the cross-border physical transportation of cash.

low-denomination banknotes. In either case, all transactions in which a customer requires the bank to change the denomination of the banknotes that he holds may be considered as featuring some money laundering risk. Accordingly, one of the regressors of the model is the value of all transactions of the kind extracted from the SARA database, again at bank-province level.

A further regressor is a proxy for usury. Usury is a wide-spread and most alarming criminal activity in Italy, often linked to powerful criminal organizations using it as a tool to invest the proceeds of their illicit activities, as well as to strengthen their grip on a given region and to permeate its legal economy. Hence, the amount of unpaid cheques is added into the model as an attempt to measure the possible level of usury taking place in a province.

A final control variable being considered is the total number of transactions recorded by a bank in a province; by providing a measure of the local scale of the activity by each intermediary, this variable helps to capture its overall money laundering risk, which can be held to increase as the turnover of a bank steps up. In fact, such variable is also used to aptly partitioning the data sample so as to accommodate for peculiar features of our dependent variable, as will be explained in detail in Section 4 below.

3.3. The regressors: Local socio-economic and criminal indicators

The magnitude of a bank's reporting activity has to be gauged also against the socio-economic context in which it operates. A first regressor that we considered, in this perspective, is the local per capita income level[7].

Alongside the latter, we also used a synthetic *indicator of financial and economic vulnerability* aiming to provide a measure of the local socio-economic conditions as a whole. Such indicator has been originally developed within the UIF; by summing up several variables, its aim is that of capturing the different features of a given area which, by making such area weaker from a financial or socio-economic point of view, may render it more 'vulnerable' to organized crime penetration, mainly through the illegal provision of funding.[8]

Additional explanatory variables added to the model account for criminal activity at provincial level, since STRs may certainly be affected by local crime rates. Hence we included the number of reports to law enforcement (per 100,000 people) for each of the

[7] The relevant data refer to 2010, the most recent year for which the information was available when the analysis was carried out.

[8] The vulnerability indicator factors in several variables. A first set refers to features of the credit market and to measures of potential bottlenecks in the local supply of banking and financial services; in particular, the variables included in the indicator are the total amount and number of lines of credit granted by banks and other financial intermediaries, and the total amount and number of non-performing loans. Additional components account for local economic conditions (i.e., the businesses birth-to-death ratio), households consumption patterns (i.e., the number of big distribution outlets, new vehicles registrations), and the extent of the underground economy (i.e., per-capita energy consumption) . Provinces are ranked against each and every variable according to a growing scale of risk. The average rank for each province provides the values of the final vulnerability indicator.

following crimes: (i) money laundering, (ii) corruption, (iii) criminal conspiracy (both plain and mafia-style) and (iv) fraud.

The complete set of variables included in the model are summed up in Table 1, which also provide the shorthand notation that will be used henceforth and some descriptive statistics.

Table 1
List of variables

Variable	***Shorthand notation***	***Descriptive statistics***	
		Mean	*Standard deviation*
1. *Dependent variable*			
Number of high risk STRs	*High risk STRs*	5.3	22.7
2.*Regressors: Financial measures of local exposure to money laundering risk*			
Value of cash transactions (log)	*Log cash*	55.2 mln.	210.3 mln.
Value of wire transfers to and from high risk countries (log)	*Log high risk transfers*	67.5 mln.	962.8 mln.
Value of transactions unrecorded on on-going business relationships (log)	*Log over-the-counter transactions*	208.9 mln.	8.3 bln.
Value of banknote denomination exchanges (log)	*Log denomination exchange*	7,685.0	57,421.0
Value of unpaid cheques (log)	*Log unpaid cheques*	11.8 mln.	73.8 mln.
Overall number of transactions (log)	*Log total transactions*	55,209.8	221,062.1
3. *Regressors: Local socio-economic indicators and measures of criminal activity*			
Taxable *per capita* income (log)	*Log income pc*	22,537.6	2,433.4
UIF's financial vulnerability index	*Vulnerability index*	53.0	8.7
Number of reports to law enforcement agencies for conspiracy and fraud (per 100,000 people)	*Conspiracy, mafia and fraud*	53.3	22.9
Number of reports to law enforcement agencies for money laundering (per 100,000 people)	*Money laundering*	4.0	4.9
Number of reports to law enforcement agencies for corruption (per 100,000 people)	*Corruption*	6.8	12.2

Note: the number of obs. is 4,991; all variables are measured as of 2012, except for taxable per capita income, whose data refer to 2010; log variables are natural logs.
Data sources: UIF, Bank of Italy and Law enforcement database (SDI).

4. Data modelling issues

The independent variable of our model is the total number of high risk STRs that each Italian bank filed in 2012 from each province where it has at least one branch. Figure 1 provides a diagrammatic representation of how the variable is distributed across its range of values.

It is straightforward to notice that the distribution of STRs displays all the typical features of count data variables for rare events, such as distributions describing ship accidents (McCullagh and Nalder, 1983) or defects of manufactured items (Lambert, 1992). Figure 1 shows that our independent variable is a non-negative discrete variable with a clear preponderance of observations with zero or low values and a limited number of very high values, with consequently a very large variance. This latter feature may signal a case of over-dispersion, which in turn may be caused by:

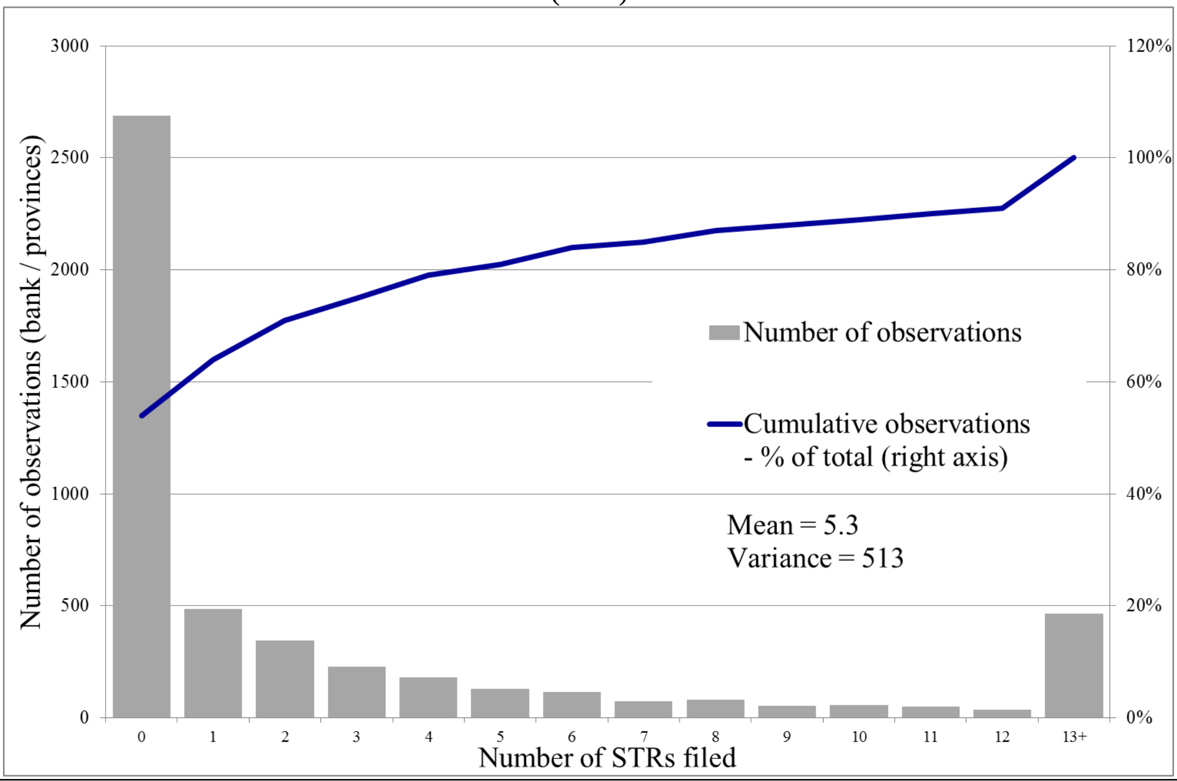
· unobserved inter-individual heterogeneity (e.g., different processes originating the variable);

· occurrence dependence between events;

· relevant outliers.

In accordance with the literature, all these factors suggest that the common regression models are unsuitable and that the least square estimator can be improved on by adopting a model that helps account for all these features.

Figure 1
Distribution of high-risk STRs
(2012)



4.1. Basic parametric models

Guo and Trivedi (2002) suggest the adoption of some semi-parametric and non-parametric models, which they recognize as significantly demanding in terms of computational implications. Hence, in line with Cameron and Trivedi (1998), we decide to resort to more traditional non-linear distributions which are typically referred to in these cases, such as the Poisson regression model.

The model assumes that the values y_i of the dependent variable Y are drawn from a Poisson distribution with parameter μ. Consequently, Y is distributed according to the following function:

$$\mathrm{Prob}[\mathrm{Y} = \mathrm{y_i} | X_i] = \frac{\mathrm{e}^{-\mu_i} \mu_i^{y_i}}{\mathrm{y_i}!}, \qquad \mathrm{y_i} = 0,1,2, \ldots \ldots \mathrm{n} \qquad [1]$$

where μ_i is related to the regressors X_i according to the following

$$\ln \mu_i = \beta' X_i .$$

Hence,

$$\mathrm{E}[\mathrm{Y}_i | \mathrm{X}_i] = \mathrm{Var}\,[\mathrm{Y}_i | \mathrm{X}_i] = \mu_i = \mathrm{e}^{\beta' \mathrm{X_i}}.$$

Thus, μ_i represents both the mean and the variance of Y, which, for this reason, is held to be equi-dispersed. Real data do not typically verify the equi-dispersion condition; in fact, they are usually over-dispersed. Hence the Poisson model is generally replaced by the Negative Binomial (NB) regression model, which is distributed according to the following function:

$$\mathrm{Prob}[\mathrm{Y} = \mathrm{y_i} |\, \mu_i, \alpha] = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i}\right)^{\mathrm{y_i}} \quad \alpha \geq 0$$

$$\mathrm{y}_i = 0,1,2, \ldots$$

where we have

$$\mathrm{E}[\mathrm{Y}_i | X_i] = \mu_i \qquad \text{and} \qquad \mathrm{Var}\,[\mathrm{Y}_i | \mathrm{X}_i] = \mu_i + \alpha \mu_i^2 .$$

Hence, $E[Y|X_i] \leq Var[Y|X_i]$. If $\alpha = 0$, then we revert to the Poisson.

At the estimation stage, the parameters α and μ are commonly estimated for both the Poisson and the NB model using maximum likelihood techniques.

The Poisson distribution function belongs to the Linear Exponential Family (LEF), whilst the NB distribution function belongs to the Linear Exponential Family with Nuisance parameter (LEFN). In order to have consistent estimates for the βs, LEF functions require that the function of μ is correctly specified, whilst LEFN functions require that this condition be met for both functions of α and μ. Empirically, the quadratic specification for Var[Y] of NB models ensures that the data over-dispersion is adequately accounted for.

4.2. Zero-inflated models

Following Lambert (1992), Poisson (and hence NB) models can be adjusted so as to accommodate distributions featuring large number of zeroes, as the one being analysed here.

In order to do this, it is assumed that the distribution of Y, the observed variable, is some distribution of zeroes with probability p and a Poisson with parameter μ with probability *(1 - p)*. Formally we have:

$$\mathrm{Prob}[\mathrm{Y} = 0] = \mathrm{p} + (1 - \mathrm{p})e^{-\mu},$$

whilst

$\text{Prob}[Y = y_i]$ is as in [1] for $y_i \neq 0$.

Normally it is assumed that μ and p are related to the regressors X_i and hence they are linked to each other according to some functional form. A typical specification of the relation between μ and p is the following:

$\ln(\mu) = X_i\beta$ and $\text{logit}(p) = -\tau X_i\beta$,

which implies that

$p = (1 + \mu^{\tau})^{-1}$.

As in the non zero-inflated Poisson, the parameters, and hence the βs, are estimated using a maximum likelihood approach, granting the same properties that apply to non zero-inflated Poisson and NB.

4.3. Partitioning the data sample

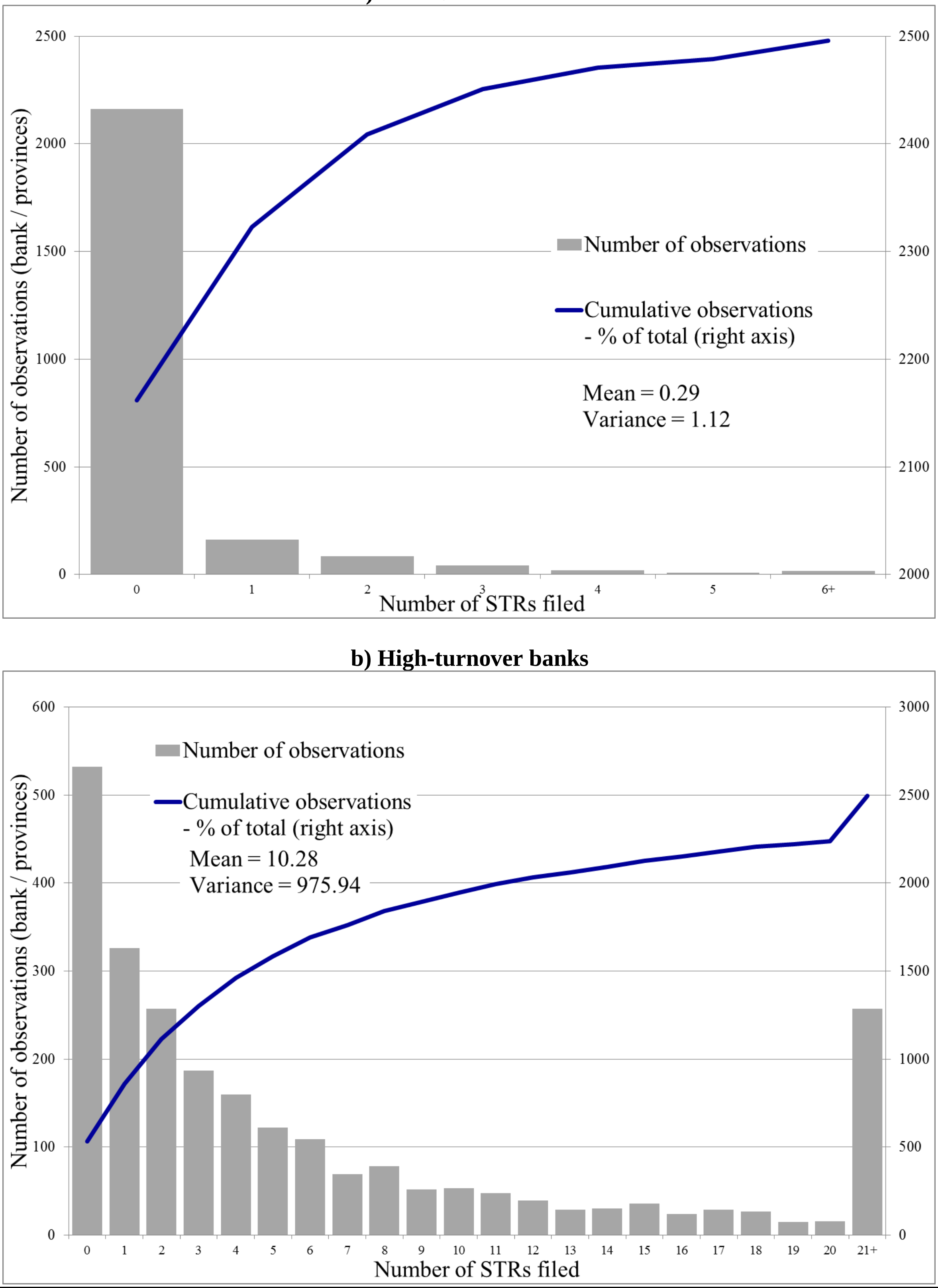
Besides adopting traditional and zero-inflated parametric non-linear models, an additional measure has been taken so as to accommodate one of the possible causes of the over-dispersion of our dependent variable, i.e. heterogeneity. The latter is relevant in our context since different scales of operation by banks may induce significant divergences in reporting behavior, stemming, for example, from differences in the organizational approach to anti-money laundering or from returns-to-scale in monitoring and reporting activity etc. In econometric terms, this would imply that the stochastic process whereby STRs are distributed in low-turnover banks is different from the one underlying high-turnover banks' STRs distribution. Hence, there is a rationale for mitigating the heterogeneity of the data by splitting the sample.

As a useful example, we can use the analogy with car accidents; the latter variable can be properly modeled by making use of count data models. All other things equal, a driver is more likely to have an accident the more miles he drives. Furthermore, and more at a 'structural' level, the driving behavior of occasional, short-mileage drivers is found to be quite different from that of professional, high-mileage drivers (such as truck drivers and commuters). Accordingly, the determinants of the respective accident rate may differ, and the use of different models may be suitable to capture such divergence. Similarly, banks with different 'mileage' (size of business) may be deemed to be driven by different factors as far as reporting is concerned.

In our case, the sample of all observations has been split into two subsets against one of the regressors, namely the total number of transactions performed by each bank in a province. The split value has been set equal to the median, so as to obtain two sub-sample of equal size and thus maximize the number of observations of both subsets. We thus have a subset of (locally) low-turnover banks and one of (locally) high-turnover banks. The same bank can belong to one subset for its activity in one province and to the other subset with reference to its activity in another province, since its turnover may significantly differ in size depending on how deeply rooted its presence is in a given area and how large is the number of its customers there.

Figure 2 shows to what extent the distributions of the two sub-samples differ.

Figure 2
Distribution of high-risk STRs
(2012)



In the first subset (low-turnover banks; fig. 2a) almost all banks file very few STRs or, more often, none at all. Consequently, the range of the observed values of the dependent variable is extremely narrow. Conversely, in the second subset (high-turnover banks; fig. 2b), though observations are highly concentrated on the left-hand side of the distribution, zeroes are way much rarer and the right-hand tail is extremely long, resulting in a significantly high variance.

Both groups remain over-dispersed — as shown by tests based on an auxiliary regression (see Cameron and Trivedi, 2005) — with the first one being possibly zero-inflated.[9]

For all the reasons mentioned above, we thus used several estimators and two sub-samples. More precisely, we used the two benchmark models for count data (i.e., Poisson and negative binomial) for the sub-sample of high-turnover banks; in the case of low-turnover banks, we added the two corresponding zero-inflated variants.

5. Estimation results

This section illustrates the outcome of the estimation process. In line with the relevant literature using count data models, estimates have been obtained by maximum likelihood techniques.

In order to appreciate the level of fitness, reference will be made to commonly applied diagnostics used in non-linear regressions. In addition to BIC and AIC, several other indicators were taken in consideration:

· squared correlation between actual STRs and estimated ones;
· Pearson statistic[10];
· pseudo R-squared measures[11];
· count fitting, which compares the estimated frequency corresponding to each modality of STRs with the actual frequency[12].

The estimation results for, respectively, low- and high-turnover banks are reported in Tables 2 and 3. For reasons of confidentiality, values shown are not the parameter estimates, but rather the corresponding measures of statistical significance (so-called Zs).

5.1. Estimates for low-turnover banks

For low-turnover banks, the two benchmark models seem preferable to the two corresponding zero-inflated variants as they ensure that a larger number of regressors is statistically significant.

---

[9] In particular, null hypothesis of equidispersion is rejected for both samples with a p-value < 0.0001.

[10] Sum of squared Pearson's residuals (Cameron and Trivedi, 2013).

[11] The so called 'likelihood ratio index', $R^2=1-(L_{fit}/L_0)$, where L_{fit} is the likelihood of the fitted model and L_0 is the likelihood of the no parameters model (Cameron and Trivedi, 2013).

[12] The count fitting is equal to $\sum\frac{\left(n\bar{p}_j-n\hat{p}_j\right)^2}{n\hat{p}_j}$, where $\hat{p}_j$ is the expected STRs (as a share of the total) of the j^{th} observation and $\overline{p_j}$ is the corresponding observed value.

Table 2
Estimation results for low turnover banks
Measures of statistical significance (so-called Zs)

Dependent variable: High risk STRs	**Poisson**	**Zero-inflated Poisson (ZIP)**	**Negative binomial (NB)**	**Zero-inflated negative binomial (ZINB)**
Financial measures of banks' local exposure to money laundering risk				
Log cash	2.11**	0.76	2.14**	1.59
Log high risk transfers	3.25***	0.24	3.26***	1.22
Log over-the-counter transactions	4.12***	1.01	4.26***	1.86*
Log denomination exchange	0.83	1.61	1.17	1.07
Log unpaid cheques	1.26	1.11	1.49	0.06
Log total transactions	5.64***	1.70*	6.21***	2.16**
Local socio-economic and criminal indicators				
Log income pc	0.20	0.07	1.03	-0.65
Vulnerability index	-0.84	-0.99	-0.70	-1.35
Conspiracy, mafia and fraud	-0.41	0.44	0.01	-0.46
Money laundering	1.62	1.10	1.31	1.61
Corruption	0.84	0.89	0.03	1.28
Constant	-0.93	-0.18	-1.74*	0.29
Alfa			2.43	2.04
Diagnostics				
Obs.	2,496	2,496	2,496	2,496
GDL	11	11	11	11
Pseudo R^2	0.32	0.02	0.20	0.01
Chi^2	273.23	20.13	317.21	25.59
P (X> Chi^2)	0.00	0.04	0.00	0.01
AIC	2,868	2,490	2,398	2,372
BIC	2,938	2,630	2,474	2,517
$Correlation^2$	0.15	0.15	0.15	0.15
Pearson P	5,611	4,388	3,902	3,459
Count fitting	4.9e+18	3.9e+13	102.28	442.64

Note: Robust standard errors., **, *** indicate statistical significance at, respectively, the 10%, 5% and 1% level.*

Indeed, in the zero-inflated scenarios the model would amount to very little, since only a very limited number of independent variables would be retained: in particular, the Poisson component of the ZIP model has only one significant regressor (see Table 2) and the zero-inflated component two additional ones[13]; the zero-inflated negative binomial model (ZINB) has just two significant explanatory variables as a whole (with the zero-inflated component having none).

The overall set of diagnostics does not provide clear-cut conclusions on which model, Poisson or negative binomial, provides a better fit. AIC and BIC show that the negative binomial specifications take account of the over-dispersion of the data with a better adaptation compared to the Poisson models. The squared correlations are equivalent for all models and in line with what can be expected from cross sectional samples.

On the whole, the benchmark negative binomial specification (NB) seems preferable, since it addresses over-dispersion better than the Poisson models and in a more parsimonious way compared to its zero-inflated variant (ZINB). It also grants a satisfactory pseudo R^2 and by far the best goodness of fit as measured by the count fitting.

Since the values of the coefficients (not shown here for confidentiality purposes) in the benchmark Poisson and negative binomial models do not differ much, that may provide an indirect confirmation that estimates in the NB specification are reliable (since Poisson estimates are consistent).

Focusing on the specific estimates for the benchmark negative binomial specification, most financial measures of money laundering risk turn out to be highly statistically significant (the exceptions being unpaid cheques and denomination exchanges); on the other hand, quite strikingly, none of the local socio-economic and criminal indicators is significant. While this latter result is not easy to explain, a possible interpretation is offered at the end of the next paragraph, after taking into consideration the results of high-turnover banks.

5.2. Estimates for high-turnover banks

With regard to the sub-sample of high-turnover banks, since this subset does not feature zero-inflation symptoms, only the two benchmark models have been estimated.

Within this sample, the diagnostics tilt the balance decidedly in favour of the benchmark negative binomial (NB) specification, based on the results of the AIC and BIC criteria, the extremely good measure of goodness of fit and the high Chi square. The NB model also ensures that a wider range of regressors is significant, thus making the model quite comprehensive. Most financial measures of money laundering risk remain highly significant (with the noticeable exception of wire transfers to and from high risk countries); in addition, most socio-economic and criminal regressors turn out

[13] The estimated parameters for the logit part of the model (that explaining zeroes inflation) are described in the Appendix.

to be statistically significant (the only exception being corruption reports).

Table 3
Estimation results for high turnover banks
Measures of statistical significance (so-called Zs)

Dependent variable: High risk STRs	**Poisson**	**Negative binomial**
Financial measures of banks' local exposure to money laundering risk		
Log cash	3.07***	2.76***
Log high risk transfers	-0.26	1.01
Log over-the-counter transactions	3.81***	7.94***
Log denomination exchange	1.53	2.88***
Log unpaid cheques	-0.20	-0.72
Log total transactions	2.94***	7.40***
Local socio-economic and criminal indicators		
Log income pc	3.40***	4.18***
Vulnerability index	0.45	4.18***
Conspiracy, mafia and fraud	2.17**	1.95*
Money laundering	1.93*	1.95*
Corruption	-0.87	-0.43
Constant	-4.84***	-6.13***
Alfa		1.04
Diagnostics		
Obs.	2,495	2,495
Gdl	11	11
Pseudo R^2	0.53	0.12
Chi^2	713.04	1,353.2
P(X> Chi^2)	0.00	0.00
AIC	34,399	13,949
BIC	34,469	14,024
$Correlation^2$	0.37	0.33
Pearson P	59,690	4,661
Goodness of fit	720	100

*Note: Robust standard errors are estimated. *, **, *** indicate statistical significance at, respectively, the 10%, 5% and 1% level.*

5.3. *Comparing the estimates across sub-samples*

Since for both sub-samples the benchmark negative binomial (NB) specification turns out to the preferable model, we can focus on the corresponding results in order to compare the two sets of estimates and attempt to improve our understanding of the evidence obtained.

The most conspicuous difference in the results across the two subsamples regards wire transfers involving high risk countries, which are highly statistically significant for low-turnover banks and not significant for the other banks. A possible explanation relies on the differences in the frequency and use of this type of financial instrument across the two sub-samples of intermediaries: in those fewer, relatively rare cases in which smaller intermediaries process transfers *vis-à-vis* high risk countries, they might be more likely to generate an STR, compared to bigger banks, which process many more transfers of this type, because of the much larger size — and potentially higher sophistication — of the financial flows that they manage.

The other striking difference in the results across the two subsets regards the significance of socio-economic and criminal variables, none of whose coefficient ends up being statistically different from zero in the case of low-turnover banks. While again there is no obvious interpretation for this finding, it seems that, for banks whose business is relatively tiny, the reporting behaviour tends to be affected primarily (or exclusively) by the own activity's intrinsic features rather than by external factors. Once again the analogy with car accidents may help providing a possible explanation: should a driver use his car very seldom and for a negligible amount of miles, the expected number of accidents he may have is likely to reflect mainly his style of guide and the relative weaknesses or strengths. On the other hand, for long-haul drivers, the risk of being involved in an accident is likely to reflect also, to a larger extent, external factors such as traffic congestion and the quality of road maintenance.

6. Using the results for supervision purposes and compliance checks

Besides improving our understanding of banks' reporting behavior, our estimation results seem potentially useful for more operational ends, most notably as a tool for supervisory purposes.

Indeed, one can compare the actual bank's behavior, i.e. the flow of STRs that the intermediary filed from a given province, with the results of the model, i.e. the expected flow. Although of course no automatism at all has to be used in interpreting the results, nevertheless the size of the deviation can be used as providing some statistical indications (akin to a 'red flag' indicator) on the level of compliance of individual intermediaries.

Instead of referring to the actual difference between the observed and the expected (fitted) number of STRs, residual analysis for the purpose of detecting potential banks' deviant behavior has been carried out referring to the estimated model's Anscombe residuals, which are a particular type of standardized residuals. We use this type of residuals rather than the raw residuals, since using the latter would amount to overly relying on the model's capacity of precisely pinpointing the number of STRs expected by each bank/province pair. Hence, we prefer to select potentially deviant banks against a more reliable statistical yardstick which is widely applied in this type of analysis for count data models, in the context of Poisson and negative binomial specifications. An additional advantage granted by this type of residuals is that it implicitly accounts for the relative size of the raw residuals (since they are normalized by a function of the variance of each estimated value). Thus deviant banks are identified also considering by how much their behavior diverges from what is predicted by the model relative to the overall observed number of STRs they actually file.

The banks for which the deviation between actual and expected behavior is statistically more anomalous have thus been identified as those belonging to the first and the last percentile of the distribution of the Anscombe residuals. In other words, we used this criterion to single out the pairs of bank/province for which the model features the worst fit. The observations belonging to the first percentile refer to *potential* cases of banks under-reporting suspicious financial conducts, whilst the pairs located above the 99th percentile correspond to banks filing a higher number of STRs than that predicted by the model (potential over-reporting).

The aggregate statistics on the observations corresponding to, respectively, the 1^{st} and 99^{th} percentile of the distribution of the Anscombe residuals are reported in Table 4 a) and b), for each sub-sample of banks as well as for the total (details on individual banks are of course omitted for confidentiality reasons).

Let us first examine the cases of *potential* under-reporting (Table 4, upper panel). As to the interpretation of these findings, it is worth reminding that our dependent variable does not include all STRs filed to the Italian FIU, but only those featuring a medium to high risk score. Hence, the bank/province observations for which the observed number of STRs is lower than that fitted by the model are associated to banks which *potentially* fail to detect financial conducts which are likely to be of some relevance for money laundering prevention. Overall, the 50 bank/province observations of this type refer to 33 different banks and 32 provinces. As it can be seen, cases of *potential* under-reporting are fairly distributed among all categories of banks and regions of the country. When a category features seemingly high or low figures, such figures have to be gauged taking into account the overall number of bank/province observations for that category (e.g., foreign banks feature very low figures, but this is roughly proportional to their presence in the territory).

The evidence for potential over-reporting is to some extent similar (Table 4, lower panel). The 50 bank/province observations of this type refer to 27 different banks and 28 provinces. Again, cases of apparent over-reporting are widespread among different categories of banks and regions of the country.

As to the interpretation of these cases, over-reporting banks may be seen as particularly effective in spotting anomalous financial behaviors possibly associated to money laundering, beyond the average bank's detecting ability captured by our estimated coefficients. This might occur, presumably, because they have a more effective internal monitoring regime for preventing money laundering, due for example to a better organization or a higher quality of their soft information on the respective customers. For several of these intermediaries, however, an alternative, less favorable interpretation emerges from a further, parallel analysis that was carried out by focusing on the bank/province flow of *low risk* STRs (i.e., STRs scored 1 or 2 in terms of risk; see below). Interestingly, many banks appearing to over-report high risk STRs emerge as over-reporting also when the model is applied to the data of low risk STRs (regressions not shown here for the sake of brevity). Taken all together, this evidence seems to be consistent, at least for some banks, with the 'crying wolf' theory: there seem to be a group of banks which show a larger-than-average propensity to report suspicious operations, regardless of the level of risk. In other words, such banks appear to over-report *both* high risk and low risk STRs.

Table 4
Summary statistics of deviant banks

a) **Under-reporting banks**			
	Low-turnover banks	**High-turnover banks**	**Total***
Number of bank / province observations	25	25	50
Banks involved	20	15	33
· large	2	7	9
· medium and small	7	5	10
· minor	10	2	12
· foreign banks	1	1	2
Provinces involved	20	17	32
· North-West	6	7	11
· North-East	9	1	10
· Centre	4	3	5
· South and Islands	1	6	6

b) **Over-reporting banks**			
	Low-turnover banks	**High-turnover banks**	**Total ***
Number of bank / province observations	25	25	50
Banks involved	18	14	27
· large	1	4	5
· medium and small	4	4	7
· minor	11	5	13
· foreign banks	2	1	2
Provinces involved	19	15	28
· North-West	6	4	8
· North-East	3	0	3
· Centre	4	6	9
· South and Islands	6	5	8

* Some totals may differ from the sum of the corresponding figures for each subset, since the same bank or province can appear in both the low- and high-turnover subsets.

Another interesting result is that, with very few exceptions, the sets of *potentially* under- and over-reporting intermediaries do not overlap. More precisely, by comparing the lists of the 33 and 27 individual banks which correspond to the cases summarized in, respectively, Tables 4 a) and b), it emerges that only 3 banks belong to both groups. In other words, those banks which file less-than-expected high-risk STRs in one province do not come out as filing more-than-expected STRs in another: intermediaries seem to be implementing a *consistent reporting policy* across the country.[14]

A further noticeable result is that the banks which appear in Table 4 in *both* the high- and the low-turnover sub-samples — there are 7 of them in the 57-strong group of potentially deviant banks — show the same conduct regardless of the local size of their business, i.e. they result to under- or over-report both in provinces where they are classified as low-turnover as well as in those when they are high-turnover banks.[15] This is another indication of the consistency of individual banks' behavior across provinces.

Most interestingly, also for supervisory purposes, it is possible to have an overall assessment of the reporting performance of each bank, comparing the occurrence of anomalies (if any) with the dimensions of the bank's presence over the territory. To this end, each bank has been rated according to the share of provinces (over the total number of provinces where it operates) for which there is model-based evidence of under-reporting (i.e., the bank/province pairs correspond to the first percentile of the distribution of the Anscombe residuals), if any.

We tested formally whether such share is statistically larger than the respective 'physiological' value (it should be reminded that anomalies were defined as the observations corresponding to the bottom 1% residuals); the banks for which this is the case are reported in Table 5, left column (banks' denomination is duly omitted because of confidentiality).

As it can be seen, the banks which appear to be chronically under-reporting according to this criterion are fairly distributed among all size categories. Analogous rating has been computed for the share of provinces with evidence of over-reporting (again, if any). The results are reported in column 2; once again, all dimensional groups are represented in the subset of seemingly over-reporting banks.

The list of potentially under-reporting banks can provide useful information and be used, eventually in conjunction with other information available, to direct the on-site inspection program of anti-money laundering authorities; this is the case, for example, of the Italian FIU, whose functions include overseeing banks' compliance with the reporting obligation. Moreover, if this estimating approach were applied routinely (e.g. annually) to updated sets of data, the overall evidence collected would allow to analyze the reporting performance of individual banks over time, thus obtaining more robust evidence on under-reporting (this would be particularly useful with regard to low-turnover banks, for which the difference between the observed and the expected number of STRs is tiny).

[14] This is presumably due, to a large extent, to the significant centralization of the STR selection process which takes place in Italian banks.

[15] There is only one exception, with one bank appearing to under-report in four provinces and over-report in three provinces.

Table 5
Potential cases of deviant banks

a) **Under-reporting banks**				b) **Over-reporting banks**			
Bank	**Dimensional group**	**Share of anomalous provinces (%)**	**Statistical significance**	**Bank**	**Dimensional group**	**Share of anomalous provinces (%)**	**Statistical significance**
1	Large	60.0	***	1	Medium & small	7.0	***
2	Large	25.0	***	2	Medium & small	15.4	***
3	Minor	28.6	***	3	Medium & small	30.0	***
4	Large	6.4	***	4	Medium & small	5.0	***
5	Large	100.0	**	5	Minor	37.5	***
6	Large	3.7	**	6	Large	9.7	***
7	Medium & small	20.0	**	7	Medium & small	100.0	**
8	Minor	100.0	**	8	Medium & small	100.0	**
9	Minor	100.0	**	9	Minor	20.0	**
10	Medium & small	50.0	**	10	Minor	100.0	**
11	Medium & small	25.0	**	11	Minor	100.0	**
12	Medium & small	9.5	**	12	Foreign	50.0	**
13	Minor	25.0	**	13	Minor	100.0	**
14	Medium & small	33.3	**	14	Minor	28.6	**
15	Minor	100.0	**	15	Minor	100.0	**
16	Minor	100.0	**	16	Minor	100.0	**
17	Minor	100.0	**	17	Minor	100.0	**
18	Minor	100.0	**	18	Minor	33.3	**
19	Minor	100.0	**	19	Minor	100.0	**
20	Minor	100.0	**	20	Minor	33.3	**
21	Minor	100.0	**	21	Large	2.8	*
22	Foreign	50.0	**	22	Minor	16.7	*
23	Large	14.3	*	23	Foreign	10.8	*
24	Medium & small	4.7	*				
25	Minor	10.0	*				
26	Medium & small	14.3	*				
27	Medium & small	12.5	*				
28	Medium & small	11.1	*				

7. Concluding remarks

As formally explored in the three-player principal-agent setting of Dalla Pellegrina and Masciandaro (2009), the money laundering supervisory authorities are entrusted with the dual task of (i) verifying the overall level of money laundering risk and (ii) measuring banks' effort in detecting money laundering. The model presented in this study aims at being deployed as a tool for discharging that dual function by matching a measure of banks' effort (high risk STRs) against a set of indicators gauging the level of money laundering risk in the areas where banks operate, together with proxies of the local socio-economic conditions.

The analysis presented some modelling issues originating from the nature of the dependent variable (high risk STRs per bank/province). The resulting estimation problems have been treated in accordance with the framework commonly adopted for data count variables, although in our case additional complexity originated from the high number of zeroes and the consequent tendency towards over-dispersion. Both issues have been addressed by splitting the data into two sub-samples, according to each bank's local business size. As a result, some robust evidence has been attained.

The results obtained show that banks may pursue reporting strategies other than solely aiming at providing a deceitful measure of their efforts, as in Takàts (2011). In fact, we do find evidence pointing to substantial under-reporting of high-risk STRs by selected banks. While the estimation results are not, *per se*, conclusive evidence of under-reporting, they can significantly help anti-money laundering authorities to target their monitoring interventions.

We thus come to the main implications of this study, which admittedly lie at an operational level. By concentrating on under-reporting intermediaries, the model, in its earlier versions, has been deployed in recent years to help design Italy's FIU on-site inspection program. Preliminary results of such program are encouraging, but comprehensive evidence is yet to be collected. Nonetheless, the approach appears to provide a promisingly robust data-based tool for identifying critical intermediaries and areas from an anti-money laundering perspective. Data from further inspections will be crucial for a refinement and fine-tuning of the model.

8. Appendix

Table A1
Estimation results for low turnover banks - Parameters of the inflated model (logit)
Measures of statistical significance (so-called Zs)

Dependent variable: High risk STRs	**Zero-inflated Poisson (ZIP) Z**	**Zero-inflated negative binomial (ZINB) Z**
Measures of money laundering risk		
Log cash	-1.09	0.14
Log high risk transfers	-3.74***	-1.37
Log over-the-counter transactions	-3.43***	-1.39
Log denomination exchange	0.1	-1.32
Log unpaid cheques	-0.51	-1.12
Log total transactions	-2.75***	-1.21
Socio-economic indicators		
Log taxable income	-0.48	-1.46
Vulnerability index	-0.58	-1.12
Conspiracy, mafia & fraud	0.95	-0.71
Money laundering	-0.43	1.25
Corruption	0.14	1.42
Constant	0.89	1.52
*Note: Robust standard errors are estimated. *, **, *** indicate statistical significance at the 10%, 5% and 1% level.*		

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6): 716-23. Institute of Electrical and Electronics Engineers, December.

Barth, J.R., Caprio, G. and Levine, R. (2004). Bank regulation and supervision: what works best? *Journal of Financial Intermediation*, 13 (2): 205–248. Elsevier, April.

Beck, T., Demirgüç-Kunt, A. and Levine, R. (2006). Bank supervision and corruption in lending. *Journal of Monetary Economics*, 53 (8): 2131-2163. Elsevier, November.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data*. Cambridge University Press (2nd edition).

Cameron, A.C., and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications.* Cambridge University Press.

Costa S., (2008). Implementing the New Anti – Money Laundering Directive in Europe: Legal and Enforcement Issues. The Italian Case. *Paolo Baffi Centre Research Paper* No. 2008-13. Bocconi University, May.

Dalla Pellegrina, L. and Masciandaro, D. (2009). The Risk-Based Approach in the New European Anti-Money Laundering Legislation. A Law and Economics View. *Review of Law and Economics*, 5(2): 931–952. De Gruyter, December.

Donato, L., Saporito, A. e Scognamiglio, A. (2012). Aziende sequestrate alla criminalità organizzata: le relazioni con il sistema bancario. *Questioni di Economia e Finanza*, n. 202. Banca d'Italia, September.

Guo, J.Q. and Trivedi, P.K. (2002). Flexible Parametric Models for Long-tailed Patent Count Distributions. *Oxford Bulletin of Economics and Statistics*, 64 (1): 63-82. Wiley, February.

Lambert, D. (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, 34 (1): 1-14. American Statistical Association, February.

Levitt, S.D. (2012). Identifying Terrorists using Banking Data. *The B.E. Journal of Economic Analysis & Policy*, 12 (3): 1935-1682. De Gruyter, December.

Masciandaro, D. (1999). Money Laundering: the Economics of Regulation. *European Journal of Law and Economics*, 7 (3):225–240. Kluwer Academic Publishers, May.

Masciandaro, D. and Filotti, U. (2001). Money Laundering Regulation and Bank Compliance Costs: What do your Customers Know? Economics and the Italian Experience. *Journal of Money Laundering Control*, 5 (2): 133-145. Emerald, October.

Masciandaro, D. and Volpicella, A. (2014). Economics and Politics in Designing Supervision: the Case of the FIUs against Money Laundering. *Paolo Baffi Centre Research Paper* No. 2014-156. Bocconi University, July.

McCullagh, P. and Nalder, J. (1983). *Generalized Linear Models*. Chapman & Hall.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6 (2): 461-464. The Institute of Mathematical Statistics, March.

Takàts, E. (2011). A Theory of "Crying Wolf": The Economics of Money Laundering Enforcement. *Journal of Law, Economics, and Organization*, 27 (1): 32-78. Oxford University Press, April.

Unger, B. (2009). Money Laundering – A Newly Emerging Topic on the International Agenda. *Review of Law and Economics*, 5 (2): 807–819. De Gruyter, December.