The Pernicious Effects of Contaminated Data in Risk Management

Laurent Frésard*

Christophe Pérignon

Anders Wilhelmsson

Abstract: Banks hold capital to guard against unexpected surge in losses and long freezes in financial markets. The minimum level of capital is set by banking regulators as a function of the banks' own estimates of their risk exposures. As a result, a great challenge for both banks and regulators is to validate internal risk models. We show that a large fraction of US and international banks uses contaminated data when testing their models. In particular, most banks validate their market risk model using profit-and-loss (P/L) data that include fees and commissions and intraday trading revenues. This practice is inconsistent with the definition of the employed market risk measure. Using both bank data and simulations, we find that data contamination has dramatic implications for model validation and can lead to the acceptance of misspecified risk models. Our estimation reveals that the use of contaminated data reduces (market-risk induced) regulatory capital by around 17%.

Date: February 9, 2010

JEL Classification: G21, G28, G32

Keywords: Regulatory capital, proprietary trading, backtesting, value-at-risk, profit-and-loss

---

[*] Frésard and Pérignon are at HEC Paris, France; Wilhelmsson is at Lund University, Sweden. We are grateful to Thomas Gilbert, Uli Hege, Alexandre Jeanneret, Evren Ors, Jérome Taillard, Philippe Valta, and seminar participants at the Banque de France and at the 2009 International Meeting AFFI for their comments and suggestions. Frésard and Pérignon gratefully acknowledge the financial support of the Europlace Institute of Finance. This paper previously circulated under the title "Risk Model Validation with Contaminated Data". Emails: fresard@hec.fr, perignon@hec.fr, anders.vilhelmsson@nek.lu.se. Contact author: Laurent Frésard, HEC Paris, 1 Rue de la Libération, 78351 Jouy-en-Josas, France. Tel: (+33) 139 67 94 07, Fax: (+33) 139 67 70 85

1. Introduction

By gradually expanding their activities, modern banks have exposed themselves to a broader risk spectrum. In response, they have developed large-scale risk-management systems to monitor and aggregate risks within their banking and trading books. Over the past fifteen years, these internal risk models have been increasingly used by banking regulators to impose on banks minimum levels of capital. If inaccurate, the in-house risk assessments lead to inappropriate levels of regulatory capital. Hence, the validation process of internal risk models turns out to be of paramount importance to guarantee that banks have enough capital to cope with unexpected surge in losses and long freezes in financial markets. Nevertheless, the recent financial turmoil has cast serious doubt on current practices and calls for a more rigorous examination of banks' risk models. Following a series of risk management failures (Stulz, 2008, 2009), new proposals on capital regulation have flourished at an unprecedented pace (Basel Committee on Banking Supervision, 2009b). In this context of profound regulatory uncertainty, it has never been so imperative for banks to prove that their risk management systems are sound.

In this paper, we analyze the process by which banks appraise their risk models. Using a sample that includes the largest commercial banks in the world, our analysis reveals a key inconsistency in the way banks validate their models. We uncover that most banks use contaminated data when testing the validity of their models. In particular, we document that a large fraction of banks artificially boost the performance of their models by polluting their profit-and-loss (P/L) with extraneous profits such as intraday revenues, fees, commissions, net interest income, and revenues from market making or underwriting activities. We show that such a contamination has important implications for risk model validation, and hence materially impacts the level of banks' regulatory capital.

In order to understand the inconsistency identified in this paper, consider a simple bank that only trades one asset, say asset A. To measure its market risk and determine its regulatory capital, the bank typically computes its one-day ahead 99% Value-at-Risk (VaR), which is simply the VaR of asset A times the number of units owned at the end of a given day.[1] The "perimeter" of the VaR model includes all trading positions that are marked-to-market. Periodically, the banking regulator checks whether the VaR model is producing accurate figures. To do so, it compares the daily P/L of the trading portfolio to the daily VaR, a process known as backtesting. If the model is correctly specified, the bank should experience a VaR exception (i.e. P/L lower than VaR) one percent of the time, i.e., 2.5 days per year. To consistently validate its model, the bank faces two key requirements. First, as VaR is based on yesterday's positions, it is crucial that the P/L be also computed from yesterday's positions. Second, the P/L should only include items that lie in the VaR perimeter. As a result, the P/L should not include intraday trading revenues (due to changes in the number of assets owned) and revenues and fees from other activities. These requirements are clearly stated by the Bank for International Settlements (BIS) in the 1996 Amendment of the Basel Accord:

> *"The inclusion of fee income together with trading gains and losses resulting from changes in the composition of the portfolio should not be included in the definition of the trading outcome because they do not relate to the risk inherent in the static portfolio that was assumed in constructing the value-at-risk measure.* […] *To the extent that the backtesting program is viewed purely as a statistical test of the integrity of the calculation of the value-at-risk measure, it is clearly most appropriate to employ a definition of daily trading outcome that allows for an "uncontaminated" test.*"
>
> ***Basel Committee on Banking Supervision, BIS, January 1996***

---

[1] The one-day ahead 99% VaR indicates the amount of money a bank can loose on proprietary trading over the next day, using a 99% confidence interval. Banks compute firm-level VaR using parametric models (e.g. Monte Carlo) or non-parametric models (e.g. historical simulation).

We find that over the period 2005-2008, less than 6% of the largest 200 commercial banks in the world evaluate their risk models in a way that is consistent with the above quote. This proportion has remained pretty constant over the sample period and in particular has not increased during the recent financial crisis. Only 28.2% of the sample banks screen out intraday revenues, and 7.1% of the sample banks do remove fees and commissions from their P/L.[2] We also show that the use of clean data is more popular among the largest banks and also more common in Europe.

We show that data contamination has a substantial economic impact. First, we document that data contamination has a major effect on backtesting outcomes. In particular, we find that banks using contaminated data have much fewer days with trading losses and much fewer VaR exceptions than banks that rely on uncontaminated data. While the average number of VaR exceptions is 3.18 for the entire sample, it is equal to 6.12 for banks that use uncontaminated data. We also show, in a multivariate regression setting, that the most critical variable to explain the annual number of exceptions is the P/L contamination, and not the bank's level of risk or VaR methodology, nor the regulatory environment and market conditions. Another direct impact of inflating P/L with fees and intraday trading revenues is to lower the rejection rate of standard validation techniques used by banking regulators. In our sample, 23.5% of the risk models are rejected when tested with uncontaminated P/L, whereas only 10.8% of the risk models are rejected when tested with P/L that include both fees and intraday trading revenues.

Second, we show that data contamination has a material effect on the level of the regulatory capital of banks. Under the current regulatory framework, banking regulators increase capital requirements for banks experiencing an excessive number of VaR exceptions. As contamination tends to lower the number of exceptions, it mechanically reduces the

---

[2] In the following, the term "fees and commissions" refers to fees, commissions, net interest income, reserves, revenues from market-making, and revenues from underwriting activities.

penalty imposed by banking regulators. We show that, for an average sample bank, data contamination leads to a 17% reduction in market-risk induced capital.

Third, we further characterize the importance of data contamination by analyzing four years of clean and contaminated daily P/L for a leading European bank. This is the only bank in our sample to publicly disclose backtesting results based on both clean and contaminated P/L. We show that the contamination component accounts for a very large fraction of the total P/L. Extraneous revenues inflate both the mean and the volatility of the P/L by a surprisingly large magnitude. For instance, the average daily P/L shrinks from around 200,000 euros to close to zero when fees and intraday trading revenues are excluded. In addition, we show that the pernicious effect of data contamination has strengthened during the recent financial crisis.

Finally, we investigate how P/L contamination affects the performance of standard statistical tests used to backtest VaR models. To do so, we conduct a Monte Carlo experiment to measure the sensitivity of each test to mean and/or volatility-inflated P/L. Using a battery of backtesting methods, we show that data contamination severely distorts conclusions about model accuracy. In particular, the average rejection rates of common VaR evaluation tests vary from being six times too high to being eight times too low depending on the nature of the data contamination. The main take-away from this simulation exercise is that all available backtesting methods are highly sensitive to data contamination.

This paper contributes to the vast literature on bank capital requirements, and more specifically on internal-model based regulatory capital. Allen and Saunders (2004) investigate the procyclicality of risk-sensitive capital requirements for credit, market, and operational risks. Other research focuses on the effects of internal-model based capital requirements on the risk-taking behavior of banks (Basak and Shapiro, 2001; Dangl and Lehar, 2004; Repullo and Suarez, 2004; Leippold, Trojani and Vanini, 2006; Daníelsson, Shin and Zigrand, 2009), on truthful revelation of risk (Lucas, 2001; Morrison and White, 2005; Cuoco and Liu, 2006), and on the quality of the risk management systems (Daníelsson, Jorgensen, and de Vries, 2002). To the best of our knowledge, this paper is the first to study the role of data contamination in internal risk model validation and capital determination. Although some authors mention the presence of intraday revenues and/or fees and commissions in the P/L of banks (Hendricks and Hirtle, 1997; Berkowitz and O'Brien, 2002, 2007; Hirtle, 2003; Jorion, 2006; Pérignon, Deng and Wang, 2008; Berkowitz, Christoffersen and Pelletier, 2009; Christoffersen, 2009b; Pérignon and Smith, 2010b), none of them examine the extent and economic implication of this concern. Our first contribution is to document that data contamination is a widespread phenomenon among US and international banks.

Furthermore, while the financial risk-management literature primarily focuses on designing risk models (Christoffersen, 2003, 2009a, and Alexander, 2008) and developing backtesting methodologies (Christoffersen, 2009b), it largely assumes that the appropriate data are used. Our second contribution is to show that data contamination strongly distorts backtesting results and materially impacts capital requirements. Overall, this study reveals that the quality of the data is of first-order importance in financial risk management.

In practice, the problem that we document can be addressed in two ways. First, banking regulators can precisely state what data must be used in backtesting. Second, banks could have the option of defining their backtesting P/L as they want. This choice would be based on the most appropriate measurement given the nature of the bank's business. In both cases, the perimeter of the risk model must perfectly match the one of the P/L calculation. For instance, if a bank decides to include market-making income in its backtesting P/L, it must reflect the risk inherent in this activity in its VaR model. There is conceptually nothing wrong about intraday profits being part of the P/L since the riskiness of the bank does include the intraday trading performance. In that case, however, the dynamics of intraday profits must be included upfront within the risk model.

The rest of the paper is organized as follows. We provide in Section 2 some background information about risk model validation. In Section 3, we investigate the use of contaminated P/L among US and international commercial banks. In Section 4, we assess the economic impact of data contamination on backtesting and regulatory capital using both actual bank data and simulations. Section 5 summarizes and concludes our study.

2. Background

For the past 15 years, VaR has been the standard metric to measure and manage aggregate market risk and determine capital requirements (Basel Committee on Banking Supervision, 1996, 2009a). VaR is defined as prob($VaR_{t+1|t} > PL_{t+1}$) $= p$ where $VaR_{t+1|t}$ is the worst expected loss within a given confidence level $1\text{-}p$ over day $t+1$ given the bank's positions at the end of day t, and PL_{t+1} is the P/L on day $t+1$. Risk managers typically compute the bank VaR at the end of each trading day from the current positions w_t of all securities included in the bank's portfolio.

To assess and validate their risk management approach, banks need to backtest their VaR models. This backtesting procedure requires contrasting one-day ahead VaR on day t to the P/L on day $t+1$. In this context, it is essential that the ex-post recorded P/L arises directly from the portfolio used to make the ex-ante VaR computation (w_t). In practice, the actual P/L may contain some extraneous cash-flows such as fees and commissions, as well as cash-flows from intraday trading. To account for these additional revenues, we define four types of P/L:

[1] **Clean Hypothetical P/L** (PL_t^{CH}) measures the change in the value of the portfolio that would arise from previous-day positions. This is the uncontaminated P/L.

[2] **Clean Actual P/L** measures the change in the portfolio that would arise from previous-day positions, as well as intraday revenues (π_t): $PL_t^{CA} = PL_t^{CH} + \pi_t$

[3] **Dirty Hypothetical P/L** measures the change in the value of the portfolio that would arise from previous-day positions, as well as fees and commissions (ϕ_t): $PL_t^{DH} = PL_t^{CH} + \phi_t$

[4] **Dirty Actual P/L** measures the change in the portfolio that would arise from previous-day positions, as well as intraday revenues and fees and commissions. The variable ε_t is the contamination term: $PL_t^{DA} = PL_t^{CH} + \pi_t + \phi_t = PL_t^{CH} + \varepsilon_t$

Based on this classification, any observed difference between uncontaminated and contaminated P/L must originate in the characteristics of the contamination term ε. Hence, the theoretical effect of contamination on backtesting results depends on how ε distorts the P/L distribution. To better understand the theoretical impact of data contamination on backtesting outcomes, we consider a simple numerical example. For simplicity, we assume that the "clean" P/L is *iid* standard normal distributed. Similarly, we posit that the contamination term ε is also *iid* normally distributed with mean μ and variance γ^2. Under this set of distributional assumptions, the probability of getting a VaR exception (at the 1% level) in the presence of contaminated data is given by:[3]

$$p\left(VaR_{t+1|t} > PL_{t+1}\right) = \int_{-\infty}^{-2.33} f\left(PL \middle| \mu, (1+\gamma^2)\right) dPL \qquad (1)$$

---

[3] We will consider more realistic dynamics for the P/L and the contamination term in the simulation study presented in Section 4.4.

where f is the probability density function of the normal distribution. Using this simple framework, we can gauge the effect of data contamination on the number of VaR exceptions.

< Insert Figure 1 >

Figure 1 displays the expected number of exceptions when the clean P/L follows a standard normal distribution and the contaminated data has mean ranging from 0 to 0.5 and variance ranging from 1 to 1.3. We consider 1,000 observations (4 years) and a VaR defined at the 1% level so that we expect 10 exceptions in the absence of any contamination. Clearly, we note that the number of VaR exceptions decreases when the mean of the contamination term increases. As the inclusion of fees and commissions shifts the P/L distribution to the right, we expect banks using dirty P/L to experience fewer VaR exceptions. In contrast, Figure 1 shows that a larger dispersion of the contamination term magnifies the P/L variability, thereby increases the number of exceptions. As a result, the inclusion of volatile intra-day revenues in the P/L definition may actually boost the number of exceptions. Overall, this simple numerical example highlights that the net effect of contamination on risk model validation very much depends on the nature of the contamination. As to which effect dominates is an empirical question. In the following we start by investigating the type of P/L used by banks to backtest their models, and then measure the actual impact of contamination on risk model validation and regulatory capital.

3. Frequency of Data Contamination

To identify the nature of the P/L used to validate risk models, we collect specific information on risk management practices from the annual reports of the largest 200 US and international commercial banks (based on total assets in USD, as of fiscal year end 2006). The sample spans the 2005-2008 period. To empirically distinguish between the four types of P/L and to accurately classify banks as using contaminated or uncontaminated data, we use the exact nature of the P/L used for backtesting purposes as described in each annual report.[4] Based on this information, we categorize banks as using *clean* P/L when a bank specifically states that it excludes fees and commission when generating its P/L. On a similar ground, we classify a bank as using *hypothetical* P/L when it explicitly states in its report that its P/L does not include intraday trading revenues. With this strict classification, only banks that use clean and hypothetical P/L are considered as using uncontaminated data. According to our typology, this corresponds to Type [1] banks. On the contrary, we classify all the other banks as using contaminated data, but distinguish between the types of contamination. Hence, Type [2] comprises banks whose P/L includes intraday revenues but exclude fees, Type [3] contains banks that add fees but exclude intraday revenues, while Type [4] comprises banks that include both fees and intraday revenues in their P/L.

Because the disclosure of backtesting information is made on a voluntary basis, one problem may arise if some banks that use clean or hypothetical P/L do not mention it in their annual report. Although we cannot completely rule out this possibility, we note that (1) banks have strong incentives to explicitly mention that they have removed some revenues from their disclosed trading revenues, and (2) such a misclassification would bias us against finding any difference between banks using contaminated data and banks using uncontaminated data.

< **Insert Table 1** >

---

[4] Annual reports are obtained from Bankscope and banks websites. We only consider annual reports written in English. We start by performing case-insensitive text search for the words *var* and *value at risk* (with and without hyphen). We then manually validate every returned item and exclude any item unrelated to VaR (e.g. *var* may stand for variation). We eliminate bank-year with no reference to VaR in their annual report. After this first screen, we search for the word *backtest* (also in two words, with and without hyphen) to locate the section of the annual report dealing with market risk management. As with VaR, we discard items unrelated to market risk model validation. Importantly, we only consider occurrences when the bank discloses some quantitative information (typically the number of exceptions) about VaR backtesting.

Table 1 presents the main results of our survey procedure. The total number of valid annual reports is 714, corresponding to a sample of 189 different banks.[5] First, we observe that on average 88% of the largest US and international commercial banks disclose some information about their VaR models. Notably, this proportion has significantly increased over the sample period and turns out to be higher for the largest banks. Also, we observe that banks from Europe and the Pacific region appear to provide investors more systematically with some information on their VaR than their North American and Asian peers. Similar patterns emerge concerning the disclosure of specific information on VaR backtesting. In particular, we find that almost 44% of the sample banks release some quantitative information about the backtesting of their VaR models. Again, larger banks appear to unveil more backtesting information.

Turning to the type of P/L, Table 1 reveals that only a very small fraction of commercial banks use uncontaminated P/L to backtest their VaR models. Less than 6% of the sample banks report using uncontaminated P/L (Type [1]). Moreover, we find that 1.5% of banks exclude fees but include intraday trading revenues when computing their P/L (Type [2]), while 22.5% actually incorporate fees but not intraday trading revenues in their P/L (Type [3]). Remarkably, the proportion of banks working with uncontaminated data has remained remarkably low over time: from 5.3% in 2005 to 6.2% in 2008. Furthermore, we observe a substantial heterogeneity among banks and geographic areas. As a matter of fact, while almost 15% of the largest 50 banks use uncontaminated data, this proportion falls to zero for smaller banks. Also, we note that the use of contaminated P/L appears to be slightly less common among European banks. In all, the results in this table provides clear-cut evidence that among the largest US and international commercial banks, the vast majority of

---

[5] Note that the lower number of available annual reports for the year 2008 is mainly due to the fact that Japanese banks ends their fiscal year in March 2009 and some of them have not released (the English-version of) their annual report yet. Other reasons include crisis-triggered mergers, acquisitions, and nationalizations.

them use P/L that are polluted by fees and/or intraday trading revenues when validating their risk management models.

One particularly puzzling result is the difference between US and European banks in their use of contaminated data. Part of this may come from differences in banking regulations. In Europe, the directive 2006/49/EC on the Capital Adequacy of Investment Firms and Credit Institutions is precise about the way backtesting should be conducted in practice.

> *"Competent authorities shall examine the institution's capability to perform back testing on both actual and hypothetical changes in the portfolio's value. [...] Competent authorities may require institutions to perform back-testing on either hypothetical (using changes in portfolio value that would occur were end-of-day positions to remain unchanged), or actual trading (excluding fees, commissions, and net interest income) outcomes, or both"*
>
> ***Directive 2006/49/EC of the European Parliament and of the Council of 14 June 2006***

The transposition of the Directive 2006/49/EC in national law has been completed for all the major EU member states.[6] For instance, in the UK, the Financial Services Authority states that:

> *"A backtesting exception is deemed to have occurred for any business day if the clean profit and loss figure for that business day shows a loss, which in absolute magnitude, exceeds the one-day VaR measure for that business day [...] A firm must also perform backtesting against a clean hypothetical profit and loss figure"*
>
> ***FSA Handbook, 7.10.103R and 7.10.111R, October 2009***

In the US, the regulation in place provides extensive flexibility to construct backtests. In 2006, a joint Notice of Proposed Rulemaking entitled "Risk-Based Capital Standards: Market Risk" originated from the Board of Governors of the Federal Reserve System, the Federal Deposit Insurance Corporation, the Office of the Comptroller of the Currency, and the

---

[6] For a list of national transpositions of the Directive 2006/49/EC within EU member states, see http://www.c-ebs.org/documents/Supervisory-Disclosure/spreadsheets/rules/Rules_directive2006-49.aspx.

Office of Thrift Supervision. Among the changes proposed by the agencies was the exclusion of fees, commissions, reserves, and net interest income for the trading P/L used for regulatory backtesting. The second modification suggested by the agencies was to base regulatory backtesting on hypothetical P/L. While most of the leading US banks agreed that model validation based on clean hypothetical P/L would make more sense than the current practice, they claim that it would prove extremely burdensome in practice.[7] For instance, net interest income was said to often be inextricably embedded in actual P/L and removing intraday trading revenues would be a real operational challenge for some lines of business. In their public comments on the Risk-Based Capital Standards proposal, Wells Fargo (2007) states that "separating those profit components is very difficult and *the results would rarely be significantly different*". In the next section, we check whether the latter part of the quote holds true for a large sample of US and international banks.

4. Economic Impact of Data Contamination

We have seen in Section 3 that a large fraction of commercial banks use contaminated P/L when validating their risk models. The obvious next step is to investigate whether this phenomenon has a material impact on backtesting results and regulatory capital. In this section, we address this question in four different ways.

4.1. Effect of contamination on backtesting results

As a first step, we examine whether and how the use of contaminated P/L is related to banks' backtesting performance. To do so, we collect additional information about the actual performance of risk management models from the banks' annual reports. In particular, for the

---

[7] Public comments from US banks on the 2006 Risk-Based Capital Standards proposal can be found on http://www.federalreserve.gov/generalinfo/foia/index.cfm?doc_id=R%2D1266&doc_ver=1.

sub-sample of banks with available information on backtesting, we search specifically for the number of days over which the bank has experienced a trading loss (negative P/L) as well as the number of days for which the realized loss is greater than the VaR (VaR exceptions).

Table 2 presents a univariate analysis on both the number of days with negative P/L and the number of VaR exceptions. Over the 714 available annual reports, 27% contains information about the actual number of days with negative P/L. On average, the representative sample bank makes losses on 84 days per year. Noticeably, the number of days with losses is considerably different for banks that use uncontaminated data compared to banks that use contaminated data. Indeed, Table 2 indicates that banks that use uncontaminated (Type [1]) data realize losses 122 days per year on average. In sharp contrast, banks relying on P/L that are polluted with fees and intraday revenues (Type [4]) experience 50% fewer days with losses (65 days). This clear difference is largely confirmed when we consider medians or min-max ranges. Unambiguously, this pattern suggests that the inclusion of fees and intraday revenues in the calculation of P/L shifts the P/L distribution to the right, thereby artificially decreasing the number of days with actual losses.

< **Insert Table 2** >

Table 2 further reveals that these differences in the P/L distribution have direct implications for the backtesting results of banks' VaR models. Specifically, we notice substantial disparities in the number of VaR exceptions across the four types of P/L. While uncontaminated bank reports an average of 6.12 exceptions per year, contaminated banks display on average number of exceptions ranging from 4.75 (Type [2]) to 2.14 (Type [4]). These descriptive figures confirm that banks that employ contaminated P/L experience much fewer exceptions. This result, combined with the fact that contaminated data are very popular among banks, provide an explanation for the puzzling fact that banks tend to have too few (often zero) VaR exceptions in period of normal market conditions (Berkowitz and O'Brien, 2002, Pérignon, Deng and Wang, 2008, and Pérignon and Smith, 2010a).

To shed additional light on the incidence of data contamination on the validation of risk models, we apply the backtesting "traffic light" approach developed by the Basel Committee. The Basel rules for backtesting are derived directly from a failure rate test and aim at classifying the number of VaR exceptions into a "green light" zone, a "yellow light" zone, and a "red light" zone. To avoid a penalty on capital requirement, banks must stay in the green light zone, which the Basel Committee has decided to cap at four annual exceptions.[8] Following this regulatory approach, Table 2 reports the fraction of green light zone rejection for each type of P/L contamination. For the whole sample, we observe that around 15% of the banks fall outside the green light zone, whereas this fraction rises up to 23.5% for banks that use uncontaminated P/L. Yet, the fraction of green light zone rejection is only 10.8% for contaminated banks (Type [4]). Again, the different outcomes that obtain from using contaminated or uncontaminated P/L are striking. On average, the use of uncontaminated P/L appears to push VaR models outside the no-rejection zone and mechanically leads to penalties and heightened supervisory intervention.

To further validate these results, Table 3 presents a multivariate analysis of the number of VaR exceptions. In this table, we regress the annual number of exceptions on dummy variables reflecting the type of data contamination and a set of control variables. Specifically, we control for banks' characteristics, market conditions, and the regulatory environment. Because the number of exceptions is a count variable, we use a Poisson regression approach (Cameron and Trivedi, 2005). Bank specific financial information is gathered from Bankscope, the annual S&P 500 volatility is computed using CRSP data, and the Capital Requirement Index, which measures regulatory oversight of bank capital, is from Barth,

---

[8] See Jorion (2006) for a more detailed explanation about the Basel traffic light rules.

Caprio and Levine (2006). To further capture systematic differences across regulatory settings and time periods, we include country and year fixed effects.

< **Insert Table 3** >

In all specifications, the coefficient on the Clean Hypothetical (Type [1]) variable is positive and highly statistically significant. All else equal, banks that use uncontaminated P/L experience more VaR exceptions. Also, we continue to observe a pecking order in the magnitude of the effect of contamination. Indeed, the coefficients on Clean Actual (Type [2]) range from 0.82 to 1.19, and those on Dirty Hypothetical (Type [3]) are comprised between 0.58 and 0.65. All are statistically significant. Overall, the control variables display the expected sign. In particular, publicly traded banks and banks with larger trading positions (Securities/Assets) have more exceptions. Furthermore, while column (3) indicates that the use of historical simulation has no particular effect on backtesting results, column (4) shows that stock market conditions materially impact the likelihood of model failures. Finally, column (5) suggests that a strong regulatory oversight of banks' capital tends to limit the number of exceptions.

In summary, the results in this subsection clearly show that data contamination materially impacts risk model validation. We find that banks using contaminated P/L have much fewer days with trading losses, much fewer VaR exceptions, and a higher model rejection rate than other banks. These findings suggest that the mean effect (fees increase the mean of P/L) dominates the volatility effect (intraday trading increase P/L volatility).

4.2. Effect of contamination on regulatory capital

Bank capital serves as a cushion against unexpected losses. Minimum regulatory capital is commonly defined as a percentage of the risk-weighted bank assets. Under Basel II for instance, it is equal to 8% of the sum of the credit, market, and operational risk-weighted assets (RWA). The RWA component due to market risk or market risk charge depends on two main variables: the market risk VaR of the bank and a scaling factor k (for simplicity, we neglect the specific risk charge and the averaging of VaR):

$$\text{Market RWA} = (1 / 8\%) \times k \times \text{VaR} = 12.5 \times k \times \text{VaR}. \tag{2}$$

The role of the first 12.5 coefficient is to permit the aggregation between the market risk charge and the credit risk charge, whereas the scaling factor aims at accounting for model and estimation risk and generating a sufficiently conservative market risk charge (Jorion, 2006). The value of the scaling factor depends on the number of annual VaR exceptions. Since the 1996 Amendment of the Basel Accord, the scaling factor is set to three as long as the annual number of VaR exceptions remains strictly below five. A penalty component is added to k if there are more than four exceptions. For instance, with five (respectively 10) exceptions, the scaling factor increases to 3.4 (respectively to 4).

We show in Table 2 that the average sample bank experiences 3.18 exceptions per year, which corresponds to a scaling factor of 3. The number of exceptions is twice higher for banks using (Type [4]) uncontaminated data, and then the applicable k is 3.5. A back of the envelope calculation implies that, on average, the use of uncontaminated data increases the Market RWA by:

$$\%\Delta \text{ Market RWA } = \frac{12.5 \times 0.5 \times VaR}{12.5 \times 3 \times VaR} = 0.167 \;\; or \; 16.7\% \,. \tag{3}$$

This simple calculation highlights that P/L contamination has a material impact on the market risk charge. Moreover, if we define m, $0 \leq m \leq 1$, as the fraction of the risk-weighted assets due to market risk, then the overall regulatory capital increases by:

$$\%\Delta \text{ Regulatory Capital } = \frac{0.08 \times 12.5 \times 0.5 \times VaR}{0.08 \times RWA} = \frac{m}{6} \;\; or \;\; \frac{m}{6} \times 100\% . \qquad (4)$$

We see that the impact on regulatory capital directly depends on the relative importance of market risk compared to other sources of risk.

4.3. A case study: La Caixa

We complement the survey of large banks using more detailed information for one large European bank, namely La Caixa. This is the third largest Spanish bank and it ranks 68th in the world based on total assets. A unique feature of this bank is that it publicly discloses backtesting results separately for both clean hypothetical (hereafter "clean") and dirty actual (hereafter "dirty") P/L. This information allows us to get a sense of the relative importance of contamination in the P/L of a typical sample bank. We extract daily data on VaR and P/L from the firm's annual reports between 2004 and 2008. We use the data extraction numerical procedure of Pérignon and Smith (2010b) and convert the original graph from the annual report into a machine-readable format. We obtain three time series of 1,017 daily observations (VaR, clean P/L, and dirty P/L).[9]

Panel A of Figure 2 displays the daily VaR and the clean P/L of La Caixa between January 1, 2005 and December 31, 2008. We observe that this bank experienced 13 VaR exceptions, which is practically equal to the average number of exceptions in the sample (i.e., 4 × 3.18 = 12.72). We note that all the exceptions have occurred during the second half of the period (2007-2008). This pattern is typical in the sample since most banks experienced very

---

[9] The extracted series are noisy estimates of the true VaR and P/L series. However the extensive simulations conducted in Pérignon and Smith (2010b) suggest that the estimation error remains typically below 3%. A user guide for this data extraction method can be found in the appendix of the working paper version of Pérignon and Smith (2010b) available at http://ssrn.com/abstract=952595.

few exceptions before the financial crisis.[10] Panel B of Figure 2 shows the evolution of the contamination term, which is obtained by subtracting the clean P/L from the dirty P/L of the bank. The most striking result in this figure is the importance of the contamination term. Indeed, fees and intraday trading revenues account for a large fraction of the bank's P/L.

< **Insert Figure 2** >

We present in Table 4 some descriptive statistics about the VaR, clean and dirty P/L, as well as the contamination term for La Caixa. The main features of the data are as follows. First, the VaR increased enormously over the sample period. The 300% increase in VaR is comparable with those of other leading banks (Bank for International Settlements, 2009, page 42). Second, we see that subtracting fees and intraday trading revenues strongly affects the P/L: the average daily P/L drops from 194,000 euros (dirty) to 3,000 euros (clean). Third, contamination also substantially contributes to the volatility of the dirty P/L. Fourth, the level and the variability of the contamination term has increased over time and has spiked during the financial crisis. This is a crucial result as model validation is particularly important when financial markets are under stress. Finally, consistent with the results in Table 2, we find that both the number of trading days with a negative P/L and the number of VaR exceptions are higher with clean data. The effect of contamination on the number of exceptions is particularly interesting. Indeed, we see that in 2007, the VaR is exceeded eight times by clean P/L and only three times by dirty P/L. This is mainly due to the high mean of the contamination term. Differently in 2008, the number of exceptions was higher with dirty data because the volatility effect of the contamination dominates the mean effect.

< **Insert Table 4** >

---

[10] In the entire sample, the average number of VaR exceptions is 0.75 in 2005, 0.77 in 2006, 3.72 in 2007 and 7.48 in 2008.

4.4. Effect of contamination on VaR evaluation tests

To complete the analysis, we appraise the effect of data contamination on the performance of popular tests used to backtest VaR models. To do so, we adapt the Monte Carlo framework used by Berkowitz, Christoffersen and Pelletier (2009). Our procedure can be summarized as follows. First, we artificially generate clean P/L series and add different sources of contamination. Then, we measure whether and how the inclusion of data contamination alters the statistical performance of model validation tests.

Following Berkowitz, Christoffersen and Pelletier (2009), we use the Kupiec test (LR_{UC}), the Independence test (LR_{IND}), the Markov test (LR_{CC}), the Ljung-Box test (LB(1)), and the Caviar test. The Kupiec test checks whether the actual number of exceptions is significantly different from the expected number of exceptions. The Ljung-Box test, Independence test, and Caviar test build on the insight that the probability of a VaR exception should be independent of all information that was available when the VaR forecast was made. If it is not the case, the information can be used to improve the VaR forecast. The Ljung-Box and Independence tests investigate if the probability of having an exception depends on prior exceptions. The Caviar test uses a richer alternative hypothesis by also including prior VaR forecasts in the information set and also tests whether the number of exceptions is correct. Finally the Markov tests is a combination of the Independence test and Kupiec test that jointly tests for a correct number of exceptions that are evenly distributed over time. Appendix A provides technical details for each test.

To produce realistic P/L, we follow Berkowitz and O'Brien (2002) and assume a standard GARCH structure. Specifically, clean P/L (PL^{CH}) and VaR series are generated by:

$$PL_t^{CH} = \sigma_t z_t \tag{5}$$

$$z \sim iid\, N(0,1) \quad \text{with} \quad \sigma_{t+1}^2 = \omega + \delta_1 (\sigma_t z_t)^2 + \delta_2 \sigma_t^2 \tag{6}$$

$$VaR_{t+1|t} = \phi^{-1}(p) \cdot \sigma_{t+1} \qquad (7)$$

We use the parameter values $\omega = 0.05$, $\delta_1 = 0.15$, and $\delta_2 = 0.80$, and $\phi^{-1}(p)$ is the inverse of the standard normal density function evaluated at the VaR confidence level (1-*p*). This set of parameters gives an unconditional variance of 1 and a persistence (memory of a variance shock) of 0.95, which is in accordance with the parameter estimates on actual P/L reported by Berkowitz, Christoffersen and Pelletier (2009). To account for the different types of data contamination, we define the contamination term as being normally distributed with mean values of 0, 0.25 and 0.5 and variance values of 0, 0.1 and 0.2. This simple parameterization enables us to investigate the sensitivity of the main backtest methodologies to mean and/or volatility-inflated P/L.

Following Berkowitz, Christoffersen and Pelletier (2009), we conduct all simulations for a 99% VaR and we focus on test significance at the five percent level. Also, we use 250 observations because backtesting is typically conducted once a year using daily data (Basel Committee on Banking Supervision, 1996, 2009a).[11] For each set of parameters, we simulate 100,000 P/L series. Then, for each simulated sample, we calculate the different tests and count how often the tests reject the VaR model. Since we have constructed all tests to have correct size for uncontaminated P/L (see the Appendix), this procedure allows us to assess directly their performance in the presence of different kinds of contamination.

< **Insert Table 5** >

Table 5 presents the results of the Monte Carlo study. Panel A first shows the impact of data contamination on the most popular backtest method, namely the Basel traffic light

---

[11] We have also conducted similar simulations with a larger sample of 1,000 observations, different significance levels, and more combinations of mean and variance contamination. We have also used contamination that is skewed, leptokurtic and heteroscedastic and further introduced dependence in the random shocks between the P/L and contamination, as well as different volatility regimes. In all cases we obtain qualitatively similar results (see Appendix B).

categorization. When the data are uncontaminated (both mean and variance of the contamination term equal zero) the status of a correct model is green 89.22 % of the time, yellow 10.76% and red 0.02% of the time. However, we observe that contamination has a dramatic impact on the Basel classification. An increase in the mean of the contamination term, which may correspond to the inclusion of fees for instance, artificially increases the percentage of green light from 89.22% to more than 99%. Hence, a slight shift in the P/L distribution is sufficient to classify almost all models as green. Alternatively, contamination that only increases the variance of the P/L has the exact opposite effect. For example if the P/L variance is increased by 20% the model will now be classified as green in 44.8% of the time, as yellow 52.8% and as red 1.9% of the time. When both the mean and the variance of the contamination term are large (0.5 and 0.2, respectively) the Basel traffic light is green 98.9% of the time and yellow 1.1% of the time.

Panel B displays the Monte Carlo results for the five other tests. Across the different tests, we notice a substantial effect of contamination on the rejection rates. Indeed, in the absence of contamination, the rejection rate is 5% by construction. Strikingly, when we inflate the P/L distribution by increasing the mean of the contamination term, the rejection rate for the Kupiec test jumps from 5% to 16% (with mean equal to 0.25) or even 30.4% (with mean equal to 0.50). This over-rejection comes from the fact that a positive shift in the P/L often leads to zero exceptions.[12] Similarly, when we increase the variance of the P/L distribution, the rejection rate goes from 5% to 21.8%. Interestingly, we observe that the biasing effect of mean and volatility contamination tends to cancel each other out. For instance, when the contamination term has a mean 0.25 and a variance of 0.2, the rejection rate is almost correctly specified at 4.9%. For the other tests, a unified picture emerges. The tests tend to

---

[12] When the mean of the contamination term is 0.50 and the variance is zero we get zero exceptions about 60% of the time. However for zero exceptions the simulated critical value and the test statistic are exactly equal at the 5% level. When this happens, it is standard procedure to reject in 50% of the cases and accept in 50% of the cases. This is the reason why the rejection rate is in that case around 30%.

reject the VaR models too infrequently when the P/L distribution is shifted to the right by the contamination ($\mu > 0$). In contrast, the tests tend to reject the VaR models too often when the P/L distribution is widened ($\gamma > 0$). The net effect on the rejection rate depends on the relative magnitude of the mean and variance contamination.

Overall, the simulation results reveal that all popular backtesting methods are extremely sensitive to the presence of data contamination. In turn, inappropriate data lead to a severe misperception of the quality of the risk model.

5. Conclusion

The latest financial crisis has demonstrated that miscalculating risk exposures can be lethal for financial institutions. Inaccurate risk assessments can lead to both excessive risk exposures and capital charges that are not sufficient to absorb losses. This concern is at the center of the current debate on the regulation of financial institutions. As a result, it has never been so urgent for banks to convince the general public and politicians that the risk management systems in place are sound and efficient.

In this study, we identify a major inconsistency in the way banks validate their risk models. We find that most banks use contaminated data when assessing the quality of their models. This practice significantly alters backtesting results and leads to inadequate regulatory capital. Overall, we show that the quality of the data used in risk-management can be as important as the risk model in place.

There are two ways of addressing the problem we document in this paper. One way is for the banking regulators to clearly state what needs to be included in the P/L, and what needs to be stripped. Alternatively, regulators can let each bank choose the P/L definition that best fits their business lines. The former approach has the advantage of standardizing risk disclosure and easing comparison across banks. The benefit of the latter approach is that it offers the necessary flexibility to the banks to choose the risk management practices that fit their needs. In both cases, risk managers and regulators must check that the data used to validate a given risk model only include items that are modeled in this risk model.

Further research may investigate whether data contamination also plagues the hedge fund industry. Indeed, VaR is often the preferred risk measure used by hedge fund managers to communicate about their risk taking behavior. If confirmed, data contamination would have even stronger implications for backtesting since hedge funds compute VaR with a one-month horizon and they rebalance their portfolio at a much higher frequency.

Appendix A: Presentation of the Backtesting Methodologies

Define the indicator variable I_t with *t* being a time subscript according to:

$$I_{t+1} = \begin{cases} 1 \ if \ VaR_{t+1|t} > PL_{t+1} \\ 0 \ otherwise \end{cases}. \tag{A1}$$

Berkowitz, Christoffersen and Pelletier (2009) note that a correctly specified VaR model implies:

$$E\left[I_{t+1} - p \middle| \Omega_t\right] = 0 \tag{A2}$$

with Ω_t being the information set available at time *t* and *p* being the VaR level. Further since the lagged values of the indicator series is in the information set:

$$E\left[\left(I_{t+1} - p\right)\left(I_{t-k}\right)\right] = 0, \text{ for all } k > 0. \tag{A3}$$

Christoffersen (1998) proposes three tests: the first is the same as in Kupiec (1995) and check for a correct number of exceptions (LR_{UC}), the second check for the independence of the exceptions (LR_{IND}), the third jointly test for a correct number of independent exceptions (LR_{CC}). These tests fit into the framework above when first order Markov dependence is used as the alternative hypothesis. Specifically the tests are calculated from:

$$LR_{UC} = 2\left(\log\left(\hat{\pi}_1^{T_1}\left(1-\hat{\pi}_1\right)^{T-T_1}\right) - \log\left(\left(1-p\right)^{T-T_1} p^{T_1}\right)\right) \tag{A4}$$

$$LR_{IND} = 2\left(\log\left(\left(1-\hat{\pi}_{01}\right)^{T_0-T_{01}} \hat{\pi}_{01}^{T_{01}} \left(1-\hat{\pi}_{11}\right)^{T_1-T_{11}} \hat{\pi}_{11}^{T_{11}}\right) - \log\left(\hat{\pi}_1^{T_1}\left(1-\hat{\pi}_1\right)^{T-T_1}\right)\right). \tag{A5}$$

The joint test statistic (LR_{CC}) is the sum of the two individual tests in Equations (A4) and (A5). The number of observations is given by *T*, the number of ones is given by T_1 and $\hat{\pi}_1 = T_1 / T$. The T_{ij} variable is the number of observations valued *i* followed by observations valued *j*. The maximum likelihood estimates of $\hat{\pi}_{ij}$ are $\hat{\pi}_{01} = T_{01} / T_0$ and $\hat{\pi}_{11} = T_{11} / T_1$.

Equation (A3) states that all autocorrelations of the mean adjusted indicator series should be equal to zero. Berkowitz, Christoffersen and Pelletier (2009) test this with a Ljung-Box test for one lag (*LB(1)*). The test statistic is given by:

$$LB(k) = T(T+2)\sum_{j=1}^{k}\frac{\gamma^2(j)}{T-j} \tag{A6}$$

with k being the number of lags, $\gamma(j)$ the autocorrelation at lag j and T is the number of observations. The last test we consider is based on the Caviar model of Engle and Manganelli (2004) that uses a lagged value of the indicator series and the VaR estimates from the model being evaluated as explanatory variables. The test consists of estimating the equation:

$$I_{t+1} = \alpha + \beta_1 I_t + \beta_2 VaR_{t+1|t} + u_t \tag{A7}$$

by logistic regression and comparing the unrestricted likelihood to the restricted likelihood by setting $\beta_1 = \beta_2 = 0$ and $\frac{e^{\alpha}}{1+e^{\alpha}} = p$.

All of the tests above have known asymptotic distributions but we rely instead on simulated critical values to correct for size distortions due to the small number of observations (Dufour, 2006). This guarantees that any differences in the number of rejections, from the correct ones, that we document are the results of data contamination only.

Appendix B: Additional Simulation Results

In addition to the results reported in Section 4.4 that use *iid* Gaussian contamination we generate contamination from the Generalized t-distribution of Hansen (1994) to accommodate for the excess kurtosis and non-zero skewness in the data. A stochastic variable Z with probability density function:

$$g(z\mid\eta,\lambda)=\begin{cases} bc\left[1+\dfrac{1}{\eta-2}\left(\dfrac{bz+a}{1-\lambda}\right)^2\right]^{-\frac{\eta+1}{2}} & for \quad z<-\dfrac{a}{b} \\ bc\left[1+\dfrac{1}{\eta-2}\left(\dfrac{bz+a}{1+\lambda}\right)^2\right]^{-\frac{\eta+1}{2}} & for \quad z\geq-\dfrac{a}{b} \end{cases} \tag{B1}$$

is said to be distributed as Generalized t. Further a, b, and c are defined as:

$$a=4\lambda c\left(\frac{\eta-2}{\eta-1}\right) \qquad b^2=1+3\lambda^2-a^2 \qquad c=\frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi(\eta-2)}\,\Gamma\left(\frac{\eta}{2}\right)}$$

The parameters η and λ are called the shape parameters of the distribution: λ controls the asymmetry of the distribution and η the tail fatness. In the simulations, we choose parameter values that correspond to a skewness of 0.5 or 1.5 and to a kurtosis of 30.

To account for possible heteroscedasticity in the contamination term we also simulate contamination from the GARCH model:

$$\varepsilon_t=\sigma_t z_t^* \tag{B2}$$

$$z^*\sim iid\,N(0,1) \quad \text{with} \quad \sigma_{t+1}^2=\omega+\delta_1\left(\sigma_t z_t\right)^2+\delta_t\sigma_t^2. \tag{B3}$$

We use the parameter values $\delta_1 = 0.15$, $\delta_2 = 0.80$ and vary ω so that the unconditional variance given by $\omega / [1-(\delta_1+\delta_2)]$ is equal to 0, 0.1 and 0.2. The results for the three data generating processes described above are displayed in Table B1. We can see that the results remain largely unchanged both in the presence of non-normality and heteroscedasticity.

Since it is possible that the clean P/L and the contamination term are dependent, we also generate data in which the random shocks (z and z^*) in Equations (5) and (B3) have correlation -0.2 and 0.2. The results are shown in Table B2. Here again the results from the main section are shown to be robust.

Informal visual inspection of the P/L and contamination data from La Caixa in Figure 2 suggests that the data could be described by a model that has one high and one low volatility regime. It also seems that the volatility regimes coincide for the clean P/L and for the contamination term indicating some second moment dependence (the correlation between the squared P/L and the squared contamination is 0.11) and that the regimes persist for a long period of time. To capture these features of the data, we use the Markov-Regime switching model of Klaassen (2002) to generate both the clean P/L and the contamination term. We use two states indexed by i. The model is given by:

$$Y_t = \sigma_t^{(i)} z_t \tag{B4}$$

$$z \sim iid\, N(0,1) \quad \text{with} \quad \sigma_{t+1}^{2(i)} = \omega^{(i)} + \delta_1 \left(\sigma_t^{(i)} z_t\right)^2 + \delta_2 E_t[\sigma_t^{2(i)} \mid s_t], \tag{B5}$$

with Y_t being either the clean P/L or the contamination term depending on the parameter values detailed below and s_t being the state at time t. The intercept $\omega^{(i)}$ is allowed to vary depending on what state (high or low volatility) we are in. The two states are determined by a Markov transition matrix that gives the four probabilities of remaining in the high volatility state, remaining in the low volatility state, moving from low to high volatility, and finally moving from high to low volatility. To mimic the apparently high persistence of the states in the La Caixa data, we set the probability for remaining in the same state to 0.98 and consequently the probability of moving from high to low and vice versa to 0.02. We impose the restriction that the P/L and contamination term always are in the same regime. The parameters δ_1 and δ_2 are equal to 0.15 and 0.80 and we standardize the P/L to have zero mean and unconditional variance of 0.5 in the low volatility regime and 1.5 in the high volatility regime, which gives an average unconditional volatility of 1. The unconditional variance of the contamination term is set to 0, 0.1 (0.05 in the low and 0.15 in the high regime), and 0.2 (0.15 and the low and 0.25 in the high regime) with a mean of 0, 0.25 and 0.50 that is independent of the regime. We show in Table B2 that we obtain in all cases qualitatively similar results.

References

Alexander, Carol (2008) *Value-at-Risk Models*, Wiley.

Allen, Linda, and Anthony Saunders (2004) Incorporating Systemic Influences Into Risk Measurements: A Survey of the Literature, *Journal of Financial Services Research* 26, 161-191.

Bank for International Settlements (2009) 79th Annual Report, Basel, Switzerland.

Barth, James R., Gerard Caprio Jr., and Ross Levine (2006) *Rethinking Bank Regulation*, Cambridge University Press.

Basak, Suleyman, and Alex Shapiro (2001) Value-at-Risk Based Risk Management: Optimal Policies and Asset Prices, *Review of Financial Studies* 14, 371-405.

Basel Committee on Banking Supervision (1996) Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements, Bank for International Settlements.

Basel Committee on Banking Supervision (2009a) Range of Practices and Issues in Economic Capital Frameworks, Bank for International Settlements.

Basel Committee on Banking Supervision (2009b) Strengthening the Resilience of the Banking Sector, Bank for International Settlements.

Berkowitz, Jeremy, Peter F. Christoffersen, and Denis Pelletier (2009) Evaluating Value-at-Risk Models with Desk-Level Data, *Management Science*, forthcoming.

Berkowitz, Jeremy, and James O'Brien (2002) How Accurate Are Value-At-Risk Models at Commercial Banks?, *Journal of Finance* 57, 1093-1111.

Berkowitz, Jeremy, and James O'Brien (2007) Estimating Bank Trading Risk: A Factor Model Approach. In: M. Carey, R.M. Stulz (Eds.), *The Risk of Financial Institutions*, University of Chicago Press.

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

Christoffersen, Peter F. (2003) *Elements of Financial Risk Management*, Academic Press.

Christoffersen, Peter F. (2009a) Value-at-Risk Models, *Handbook of Financial Time Series*, In: T.G. Andersen, R.A. Davis, J.-P. Kreiss, and T. Mikosch (Eds.), Springer Verlag.

Christoffersen, Peter F. (2009b) Backtesting, in *Encyclopedia of Quantitative Finance*, R. Cont (ed.), Wiley.

Cuoco, Domenico, and Hong Liu (2006) An Analysis of VaR-Based Capital Requirements, *Journal of Financial Intermediation* 15, 362-394.

Dangl, Thomas, and Alfred Lehar (2004) Value-at-Risk vs. Building Block Regulation in Banking, *Journal of Financial Intermediation* 13, 96-131.

Daníelsson, Jon, Bjorn N. Jorgensen, and Casper G. de Vries (2002) Incentives for Effective Risk Management, *Journal of Banking and Finance* 26, 1407-1425.

Daníelsson, Jon, Hyun Song Shin, and Jean-Pierre Zigrand (2009) Risk Appetite and Endogenous Risk, Working Paper, London School of Economics and Princeton University.

Dufour, Jean-Marie (2006) Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics, *Journal of Econometrics* 133, 443-477.

Engle, Robert, and Simone Manganelli (2004) CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business & Economic Statistics* 22, 367-381.

Hansen, Bruce E. (1994) Autoregressive Conditional Density Estimation, *International Economic Review* 35, 705-730.

Hendricks, Darryll, and Beverly Hirtle (1997) Bank Capital Requirements for Market Risk: The Internal Models Approach, *Federal Reserve Bank of New York Economic Policy Review* (December), 1-12.

Hirtle, Beverly (2003) What Market Risk Capital Reporting Tells us about Bank Risk. *Federal Reserve Bank of New York Economic Policy Review* (September), 37-54.

Jorion, Philippe (2006) *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, 3rd Edition.

Klaassen, Franc (2002) Improving GARCH Volatility Forecasts with Regime-Switching GARCH, *Empirical Economics* 27, 363-394.

Kupiec, Paul H. (1995) Techniques for Verifying the Accuracy of Risk Measurement Models, *Journal of Derivatives* 3, 73-84.

Leippold, Markus, Fabio Trojani, and Paolo Vanini (2006) Equilibrium Impacts of Value-at-Risk Regulation, *Journal of Economic Dynamics and Control* 30, 1277-1313.

Lucas, André (2001) An Evaluation of the Basle Guidelines for Backtesting Banks' Internal Risk Management Models, *Journal of Money, Credit and Banking* 33, 826-846.

Morrison, Alan, and Lucy White (2005) Crises and Capital Requirements in Banking, *American Economic Review* 95, 1548-1572.

Pérignon, Christophe, Zi Yin Deng, and Zhi Jun Wang (2008) Do Banks Overstate their Value-at-Risk?, *Journal of Banking and Finance* 32, 783-794.

Pérignon, Christophe, and Daniel R. Smith (2010a) Diversification and Value-at-Risk, *Journal of Banking and Finance* 34, 55-66.

Pérignon, Christophe, and Daniel R. Smith (2010b) The Level and Quality of Value-at-Risk Disclosure at Commercial Banks, *Journal of Banking and Finance* 34, 362-377.

Stulz, René (2008) Risk Management Failures: What Are They and When Do they Happen?, *Journal of Applied Corporate Finance* 20, 39-48.

Stulz, René (2009) Six Ways Companies Mismanage Risk, *Harvard Business Review* 87, 86-94.

Wells Fargo (2007) Wells Fargo Comments on the Market Risk Amendment. Available at: http://www.federalreserve.gov/SECRS/2007/February/20070201/R-1266/R-1266_14_1.pdf

Table 1: VaR and Backtesting Disclosure

	Bank	Bank-year	VaR Discl.	Backtest Discl.	[1] Clean & Hyp. P/L	[2] Clean & Act. P/L	[3] Dirty & Hyp. P/L	[4] Dirty & Act. P/L
					Uncontaminated		**Contaminated**	
2005-2008	**189**	**714**	**88.38%**	**43.74%**	**5.71%**	**1.43%**	**22.50%**	**70.36%**
			(631)	(276)	(36)	(9)	(142)	(444)
2005	182	182	82.97%	41.72%	5.30%	0.66%	21.19%	72.85%
2006	187	187	87.17%	44.17%	5.52%	1.23%	20.25%	73.01%
2007	187	187	91.44%	41.52%	5.85%	1.17%	21.64%	71.35%
2008	158	158	92.41%	47.95%	6.16%	2.74%	27.40%	63.70%
Size quartile 1	48	187	97.33%	62.09%	14.84%	3.85%	18.13%	63.19%
Size quartile 2	49	192	89.58%	47.67%	5.23%	0.58%	34.30%	59.88%
Size quartile 3	46	172	84.88%	30.14%	0.00%	0.68%	29.45%	69.86%
Size quartile 4	46	163	80.37%	28.24%	0.00%	0.00%	5.34%	94.66%
Europe	108	410	95.37%	49.10%	9.21%	1.79%	25.32%	63.68%
North America	23	89	84.27%	57.33%	0.00%	1.33%	21.33%	77.33%
Asia	44	161	75.16%	28.93%	0.00%	0.83%	10.74%	88.43%
Pacific	7	28	96.43%	3.70%	0.00%	0.00%	37.04%	62.96%
Others	7	26	64.29%	62.50%	0.00%	0.00%	50.00%	50.00%

Notes: This table presents our sample and reports descriptive figures on the use of contaminated data by commercial banks. The sample is obtained directly from the annual reports of the 200 largest US and international commercial banks (based on total asset in USD as of fiscal year end 2006) between 2005 and 2008. *VaR Discl.* denotes the proportion of available annual reports that contain information about VaR. *Backtest Discl.* denotes the proportion of available annual reports that contain quantitative information about VaR backtesting. We distinguish between four types of profit and loss (P/L) data. [1] refers to Clean Hypothetical P/L, [2] refers to Clean Actual P/L, [3] refers to Dirty Hypothetical P/L, and [4] refers to Dirty Actual P/L. These types of data are described in Section 2. Numbers in parentheses are the number of bank-year observations.

Table 2: Profit-and-Loss Data and Backtesting Results

	Sample	**[1]** Clean & Hyp. P/L	**[2]** Clean & Act. P/L	**[3]** Dirty & Hyp. P/L	**[4]** Dirty & Act. P/L
		Uncontaminated		**Contaminated**	
Days with negative P/L					
Mean	83.67	121.81	124.00	97.13	64.82
Median	93	120	125	109	65
Standard-Dev	42.39	27.69	13.34	40.88	36.55
Min	5	74	112	5	5
Max	215	181	143	215	151
Bank-year	196	21	4	70	101
VaR Exceptions					
Mean	3.18	6.12	4.75	3.28	2.14
Median	1	1	2	1	1
Standard-Dev	6.06	11.40	6.95	5.10	3.72
Min	0	0	0	0	0
Max	50	50	15	35	29
Bank-year	235	34	4	86	111
Backtesting Results					
Exp. # of Exceptions	587.5	85	10	215	277.5
Actual # of Exceptions	747	208	19	282	238
Reject Basel green light	15.30%	23.50%	25.00%	17.40%	10.80%

Notes: This table presents the effect of P/L contamination on banks' backtesting results. The sample is obtained directly from the annual reports of the 200 largest US and international commercial banks (based on total asset in USD as of fiscal year end 2006) between 2005 and 2008. We distinguish between four types of profit and loss (P/L) data. [1] refers to Clean Hypothetical P/L, [2] refers to Clean Actual P/L, [3] refers to Dirty Hypothetical P/L, and [4] refers to Dirty Actual P/L. These types of data are described in Section 2. *Exp. # of Exceptions* denotes the expected number of annual exceptions which is computed as the number of observations (T) multiplied by 2.5 (250 trading days times p). *Actual # of Exceptions* is the sum of all the sample exceptions. *Reject Basel green light* reports the fraction of observations for which the annual number of exceptions exceeds 4 and is thus outside of the Basel Committee so-called "green light" zone.

Table 3: VaR Exceptions

	Poisson Regressions				
Variables	(1)	(2)	(3)	(4)	(5)
Clean Hypothetical [1]	0.711**	1.139**	1.141**	1.128**	1.265**
	[6.32]	[8.14]	[8.16]	[8.08]	[7.57]
Clean Actual [2]		0.829**	0.865**	0.834**	1.194**
		[4.64]	[4.51]	[4.67]	[5.40]
Dirty Hypothetical [3]		0.582**	0.580**	0.584**	0.658**
		[5.44]	[5.41]	[5.49]	[5.15]
Log (Assets)	-0.041	-0.051	-0.05	-0.043	0.002
	[1.27]	[1.55]	[1.50]	[1.32]	[0.04]
Loan / Assets	0.395	0.930**	0.910**	0.921**	1.383**
	[1.39]	[3.09]	[3.00]	[3.07]	[3.62]
Deposits / Assets	-0.754**	-1.164**	-1.152**	-1.148**	-1.561**
	[3.28]	[4.75]	[4.66]	[4.69]	[5.00]
Assets / Equity	-0.009**	-0.011**	-0.011**	-0.011**	-0.020**
	[4.02]	[4.52]	[4.49]	[4.41]	[5.98]
Securities / Assets	1.630**	1.566**	1.582**	1.548**	1.577**
	[4.87]	[4.53]	[4.55]	[4.49]	[3.77]
Public Status	0.622**	0.644**	0.643**	0.627**	0.708**
	[6.42]	[6.22]	[6.22]	[6.09]	[5.69]
Historical Simulation			-0.045		
			[0.51]		
S&P500 Volatility				1.320**	
				[6.31]	
Capital Requirement Index					-0.690**
					[3.58]
Country Effects	Yes	Yes	Yes	Yes	No
Year Effects	Yes	Yes	Yes	No	Yes
Observations	222	222	222	222	165
Pseudo R^2	0.38	0.40	0.40	0.40	0.38

Notes: This table presents the results of Poisson regressions of the number of VaR exception on the type of data contamination. The number of exceptions are obtained directly from the annual reports of the 200 largest US and international commercial banks between 2005 and 2008. *Clean Hypothetical* is a dummy variable that equals one if a bank-year uses clean and hypothetical P/L (Type [1]), and zero otherwise. Similarly, *Clean Actual* and *Dirty Hypothetical* are dummies that equals one if a bank-year used clean and actual (Type [2]), respectively dirty and hypothetical (Type [3]) P/L. Bank-level variables are from Bankscope and include the log of total assets, loans on total assets, deposits on total assets, total assets on equity, securities on total assets, and a dummy variable Public that equals one for public banks. *Historical Simulation* is a dummy equal to one for firms that use the historical simulation method to compute their VaR (information obtained from annual reports). *S&P500 Volatility* is the annual standard deviation of the S&P500 returns computed from CRSP. The *Capital Requirement Index* is an aggregate measure of regulatory oversight of bank capital from Barth, Caprio and Levine (2006). The standard errors are adjusted for heteroskedasticity and within-bank-year clustering. The t-statistics are in brackets. * means significant at the 5% confidence level, and ** means significant at the 1% confidence level.

Table 4: Descriptive Statistics and Backtesting Results for La Caixa

	2005-2008				**2005**	**2006**	**2007**	**2008**
	VaR	Clean P/L	Dirty P/L	Contamination (ε)	ε	ε	ε	ε
Profit and Loss (P/L)								
Mean	-2,073	3	194	191	92	96	273	299
Median	-1,790	50	170	110	90	70	140	215
Standard-Deviation	1,180	964	1,153	690	143	192	615	1,196
Skewness	1.99	-2.80	-0.98	1.99	-0.83	1.94	2.26	0.90
Kurtosis	-2.29	34.95	26.12	29.52	8.44	14.40	20.30	11.21
Autocorrelation	0.98	0.136	0.126	0.105	-0.122	0.014	0.266	0.043
Min	-5,000	-12,410	-12,700	-4,890	-770	-370	-2,820	-4,890
Max	-340	4,000	8,140	7,710	490	1,220	4,860	7,710
Days with negative P/L		431	334		102 (clean) 68 (dirty)	98 (clean) 83 (dirty)	136 (clean) 103 (dirty)	96 (clean) 81 (dirty)
VaR Exceptions		13	12		0 (clean) 0 (dirty)	0 (clean) 0 (dirty)	8 (clean) 3 (dirty)	5 (clean) 9 (dirty)

Notes: This table presents some descriptive statistics on the one-day ahead 99% VaR, clean and dirty P/L, and contamination term (ε) for La Caixa between 2005 and 2008. The contamination term is obtained by taking the daily difference between the clean and dirty P/L.

Table 5: Effect of Contamination on VaR Evaluation Tests

Panel A: Basel Traffic Light

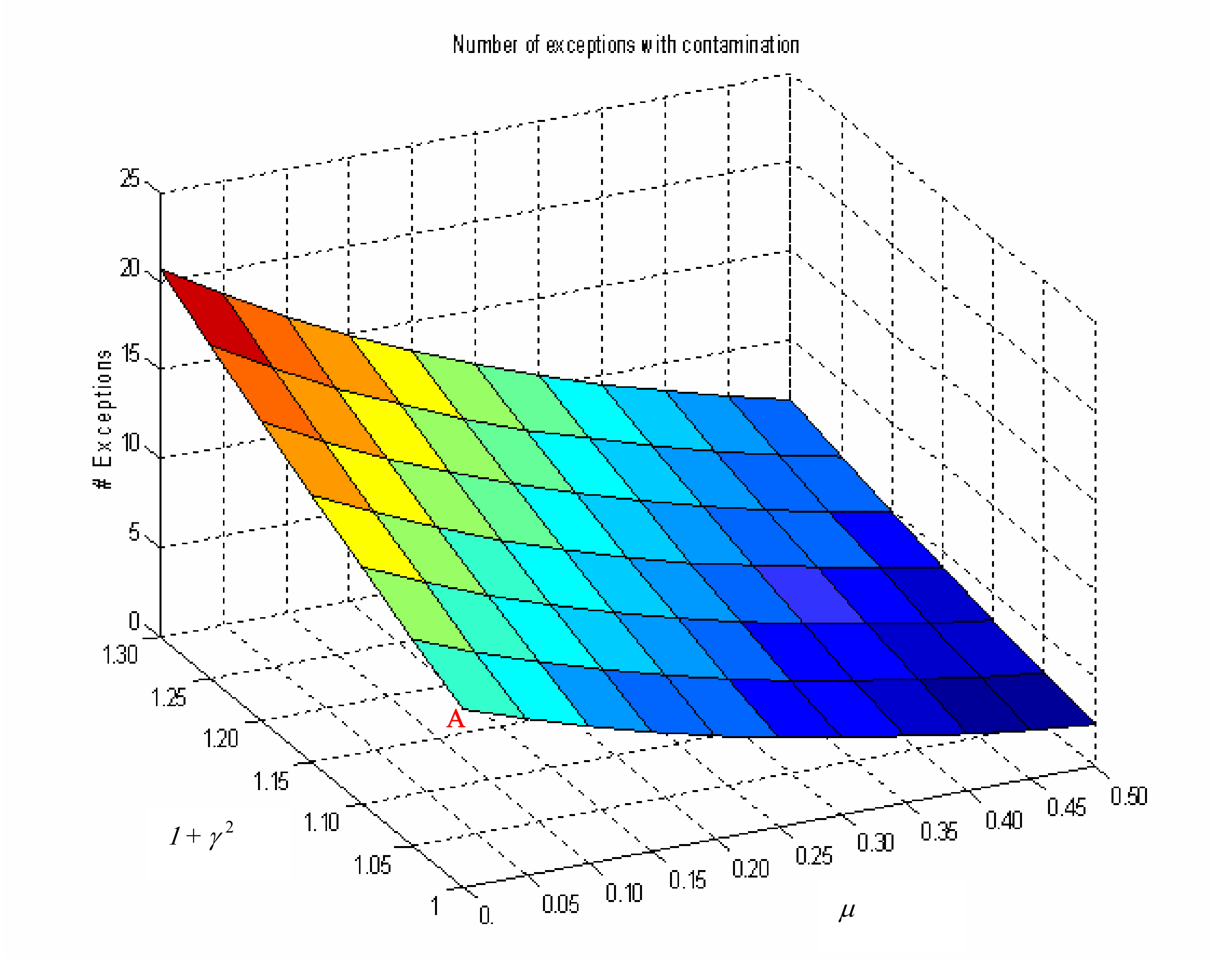
	mean 0			mean 0.25			mean 0.50		
	Green	Yellow	Red	Green	Yellow	Red	Green	Yellow	Red
var 0.0	89.2	10.8	<0.1	99.1	0.9	<0.1	99.9	0.1	<0.1
var 0.1	70.3	29.4	0.3	96.6	3.4	<0.1	99.7	0.3	<0.1
var 0.2	44.8	53.1	2.1	88.7	11.3	<0.1	98.9	1.1	<0.1

Panel B: Other VaR Evaluation Tests

	mean 0	mean 0.25	mean 0.50
Kupiec (LR_{UC})			
var 0.0	5.0	16.0	30.4
var 0.1	8.0	8.6	21.9
var 0.2	21.8	4.9	13.8
Independence (LR_{IND})			
var 0.0	5.0	1.2	0.6
var 0.1	14.1	2.2	0.7
var 0.2	31.8	4.9	1.1
Markov (LR_{CC})			
var 0.0	5.0	1.6	0.8
var 0.1	15.4	2.9	1.1
var 0.2	33.6	6.0	1.6
LB(1)			
var 0.0	5.0	1.2	0.6
var 0.1	14.0	2.2	0.7
var 0.2	32.0	4.9	1.1
Caviar			
var 0.0	5.0	2.5	1.8
var 0.1	12.8	3.9	2.9
var 0.2	28.9	6.4	3.6

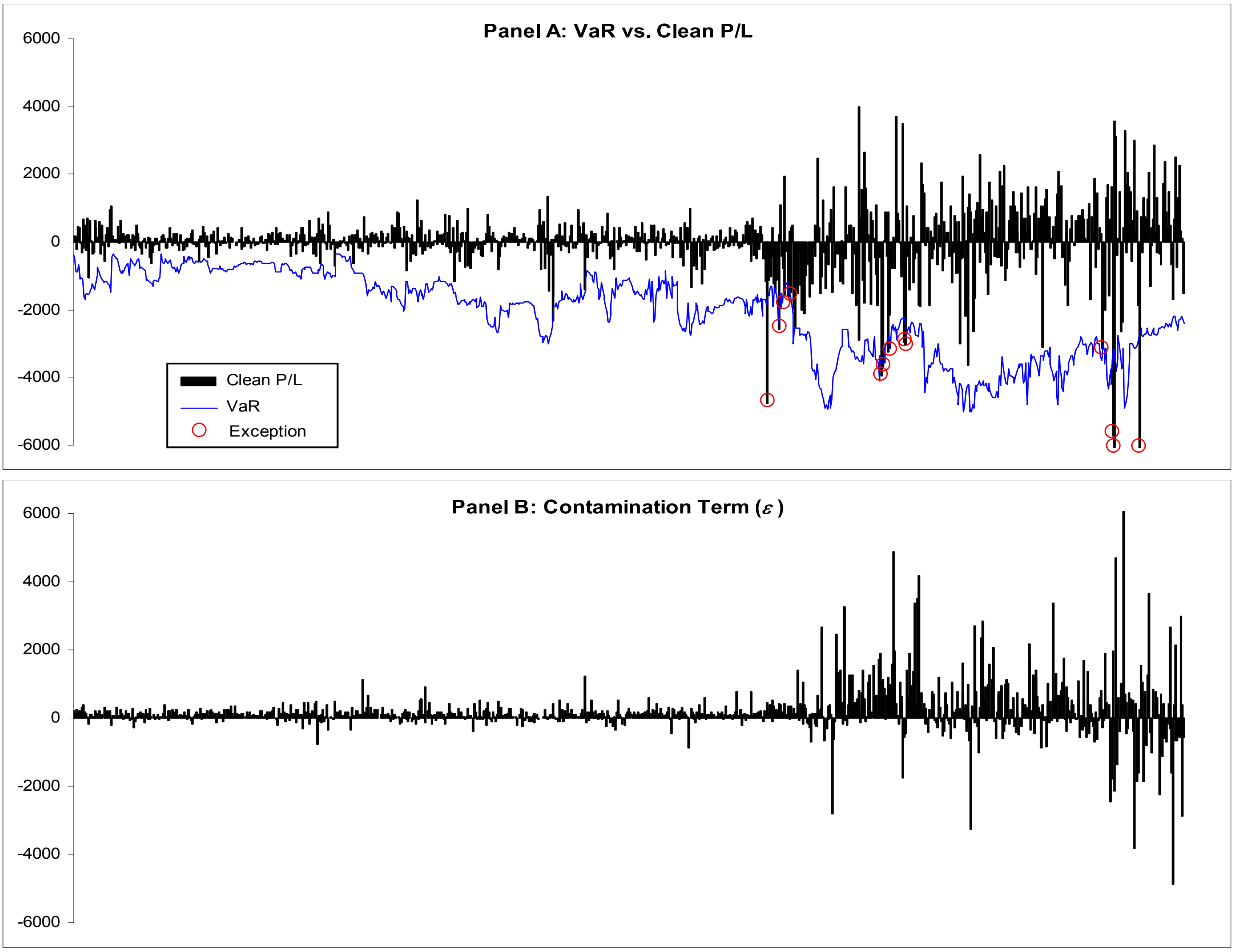
Notes: This table presents the results of popular VaR backtesting methodologies with and without contamination in the P/L. Panel A presents the simulation results of the Basel traffic light. The Green zone is between 0 and 4 exceptions, the Yellow zone is between 5 and 9 exceptions, and the Red zone is 10 and more exceptions. Panel B presents the results for the Kupiec test (LR_{UC}), the Independence test (LR_{IND}), the Markov test (LR_{CC}), the Ljung Box test with one lag (LB(1)), and the Caviar test. All tests are conducted at the 5% significance level using 250 observations and 100,000 simulation runs.

Figure 1: Number of Exceptions for Different Levels of Contamination



Notes: This figure shows the effects on the VaR exceptions of P/L contamination with varying mean and variance. The VaR level is $p = 0.01$, the number of observations is 1,000, and the correct (uncontaminated) number of exceptions is 10 (**A** on the graph). We assume that the clean P/L is *iid* standard normal and that the contamination term ε is *iid* normally distributed with parameters μ and γ^2. Under this set of distributional assumptions, the contaminated P/L is *iid* normally distributed with parameters μ and $1+\gamma^2$ and the probability of getting a VaR exception (at the 1% level) with contaminated data is given in Equation (1).

Figure 2: Value-at-Risk, Profit-and-Loss, and Contamination for La Caixa



Notes: Panel A displays the daily VaR (line) and the clean trading revenues (vertical bars) of La Caixa between January 1, 2005 and December 31, 2008. All values are in thousands of euros. The circles represent days on which the trading loss exceeds the VaR. Two trading losses have been capped at -6,000 to ease readability (actual values are -12,410 and -6,320). Panel B displays the contamination term (ε) of La Caixa between January 1, 2005 and December 31, 2008. The contamination term includes intraday trading revenues as well as fees and commissions. It is obtained by subtracting the clean P/L from the dirty P/L of the bank. One observation has been capped at 6,000 (actual value is 7,710).

Table B1: Effect of Contamination on VaR Evaluation Tests

Panel A: Basel Traffic Light

	Skewness 0.5, Kurtosis 30									Skewness 1.5, Kurtosis 30									GARCH								
	mean 0			mean 0.25			mean 0.50			mean 0			mean 0.25			mean 0.50			mean 0			mean 0.25			mean 0.50		
	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R
var 0.0	89.2	10.8	<0.1	99.2	0.8	<0.1	99.9	0.1	<0.1	89.2	10.8	<0.1	99.2	0.8	<0.1	99.9	0.1	<0.1	89.2	10.8	<0.1	98.1	1.9	0.1	99.4	0.6	<0.1
var 0.1	75.2	24.6	0.2	97.4	2.6	<0.1	99.8	0.2	<0.1	76.1	23.7	0.2	97.5	2.5	<0.1	99.9	0.1	<0.1	69.3	29.9	0.8	93.7	6.2	0.1	98.8	1.1	0.1
var 0.2	57.6	41.6	0.8	93.6	6.5	<0.1	99.5	0.5	<0.1	60.8	38.6	0.6	94.7	5.3	<0.1	99.6	0.4	<0.1	47.1	49.3	3.6	83.7	15.8	0.4	96.3	3.6	0.2

Panel B: Other VaR Evaluation Tests

	Skewness 0.5, Kurtosis 30			Skewness 1.5, Kurtosis 30			GARCH		
	mean 0	mean 0.25	mean 0.50	mean 0	mean 0.25	mean 0.50	mean 0	mean 0.25	mean 0.50
	Kupiec (LR_{UC})			Kupiec (LR_{UC})			Kupiec (LR_{UC})		
var 0.0	5.0	16.1	30.1	5.0	16.2	30.2	5.0	3.7	3.5
var 0.1	6.8	9.9	23.7	6.3	10.4	24.8	15.0	5.1	4.6
var 0.2	13.2	6.3	18.2	12.0	6.8	19.4	29.3	10.2	6.4
	Independence (LR_{IND})			Independence (LR_{IND})			Independence (LR_{IND})		
var 0.0	5.0	1.2	0.6	5.0	1.1	0.6	5.0	2.0	1.9
var 0.1	11.8	1.8	0.7	10.7	1.9	0.7	15.5	3.5	2.2
var 0.2	21.9	3.4	1.0	19.8	2.6	0.9	31.5	9.2	4.0
	Markov (LR_{CC})			Markov (LR_{CC})			Markov (LR_{CC})		
var 0.0	5.0	1.8	0.8	5.0	1.7	0.7	5.0	13.8	25.8
var 0.1	13.0	2.5	1.0	12.2	2.5	1.0	9.8	8.0	17.9
var 0.2	23.6	4.3	1.4	21.4	3.4	1.2	22.7	7.6	12.7
	LB(1)			LB(1)			LB(1)		
var 0.0	5.0	1.2	0.6	5.0	1.1	0.6	5.0	2.1	1.9
var 0.1	11.3	1.8	0.7	10.6	1.8	0.7	16.1	3.8	2.3
var 0.2	21.3	3.3	0.9	19.4	2.5	0.8	32.3	9.4	4.1
	Caviar			Caviar			Caviar		
var 0.0	5.0	2.4	1.8	5.0	2.6	1.8	5.0	3.4	3.4
var 0.1	10.1	3.4	2.3	9.6	3.4	2.1	17.9	5.1	4.1
var 0.2	18.7	4.4	3.1	17.2	4.1	2.7	34.3	11.1	5.9

Notes: This table presents the results of popular VaR backtesting methodologies with and without contamination in the P/L. Panel A presents the simulation results of the Basel traffic light with the proportions of green (G), yellow (Y), and red (R) zones (see Table 5 for details). Panel B presents the results for the Kupiec test (LR_{UC}), the Independence test (LR_{IND}), the Markov test (LR_{CC}), the Ljung Box test with one lag (LB(1)), and the Caviar test. All tests are conducted at the 5% significance level using 250 observations and 25,000 simulation runs. The contamination component is generated from the skew student-t distribution of Hansen (1994) with skewness and kurtosis as given in the table (left and central panels) or as a GARCH model with mean and unconditional variance as given in the table (right panel).

Table B2: Effect of Contamination on VaR Evaluation Tests

Panel A: Basel Traffic Light

	GARCH, rho -0.2									GARCH, rho 0.2									MS-GARCH								
	mean 0			mean 0.25			mean 0.50			mean 0			mean 0.25			mean 0.50			mean 0			mean 0.25			mean 0.50		
	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R	G	Y	R
var 0.0	89.2	10.8	<0.1	99.2	0.8	<0.1	99.9	0.1	<0.1	89.2	10.8	<0.1	99.3	0.7	<0.1	99.9	0.0	<0.1	89.2	10.8	<0.1	99.3	0.7	<0.1	99.9	0.1	<0.1
var 0.1	70.4	29.2	0.5	95.8	4.2	<0.1	99.7	0.3	<0.1	70.8	28.7	0.5	95.6	4.3	0.1	99.7	0.3	<0.1	65.3	33.6	1.1	95.0	4.9	0.1	99.5	0.5	<0.1
var 0.2	46.6	49.5	3.8	86.7	13.0	0.3	97.7	2.2	0.1	47.0	49.2	3.8	86.7	13.1	0.3	98.0	2.0	<0.1	39.3	53.5	7.1	83.2	16.2	0.6	96.9	3.0	0.1

Panel B: Other VaR Evaluation Tests

	GARCH, rho -0.2			GARCH, rho 0.2			MS-GARCH		
	mean 0	mean 0.25	mean 0.50	mean 0	mean 0.25	mean 0.50	mean 0	mean 0.25	mean 0.50
	Kupiec (LR_{UC})			Kupiec (LR_{UC})			Kupiec (LR_{UC})		
var 0.0	5.0	15.7	30.3	5.0	15.9	31.0	5.0	17.0	31.6
var 0.1	8.3	8.6	21.2	8.6	9.4	22.0	11.3	9.2	21.7
var 0.2	22.3	7.1	14.5	23.6	6.9	14.3	30.3	8.1	14.2
	Independence (LR_{IND})			Independence (LR_{IND})			Independence (LR_{IND})		
var 0.0	5.0	1.3	0.8	5.0	1.0	0.6	5.0	1.1	0.4
var 0.1	14.8	2.7	1.2	14.6	2.6	1.0	18.7	3.3	1.1
var 0.2	31.8	7.7	2.6	32.7	7.4	2.5	39.7	9.6	3.0
	Markov (LR_{CC})			Markov (LR_{CC})			Markov (LR_{CC})		
var 0.0	5.0	1.7	0.9	5.0	1.4	0.8	5.0	1.5	0.6
var 0.1	16.3	3.4	1.3	12.1	2.8	1.1	16.0	3.2	1.2
var 0.2	33.6	8.7	3.2	28.7	6.7	2.7	36.1	9.1	3.2
	LB(1)			LB(1)			LB(1)		
var 0.0	5.0	1.3	0.7	5.0	0.9	0.6	5.0	1.1	0.4
var 0.1	15.0	2.8	1.2	13.5	2.4	1.0	17.6	3.2	1.1
var 0.2	32.1	7.9	2.7	30.5	6.9	2.5	38.9	9.7	3.1
	Caviar			Caviar			Caviar		
var 0.0	5.0	2.5	2.1	5.0	2.4	1.7	5.0	2.1	1.5
var 0.1	12.9	4.4	3.4	12.6	4.0	2.8	17.9	5.8	3.1
var 0.2	28.5	8.3	4.3	28.7	8.2	4.6	38.2	11.9	5.9

Notes: This table presents the results of popular VaR backtesting methodologies with and without contamination in the P/L. Panel A presents the simulation results of the Basel traffic light with the proportions of green (G), yellow (Y), and red (R) zones (see Table 5 for details). Panel B presents the results for the Kupiec test (LR_{UC}), the Independence test (LR_{IND}), the Markov test (LR_{CC}), the Ljung Box test with one lag (LB(1)), and the Caviar test. All tests are conducted at the 5% significance level using 250 observations and 25,000 simulation runs. The contamination is generated from a GARCH model with mean and unconditional variance as given in the table. The shock to the P/L series and the shock to the contamination term are correlated with correlation coefficient rho (left and central panels). MS-GARCH uses a Markov Switching GARCH model for both the clean P/L and the contamination (right panel).