Open Access Repository
www.ssoar.info

Value-at-risk forecasts under scrutiny - the German experience

Jaschke, Stefan; Stahl, Gerhard; Stehle, Richard

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Jaschke, S., Stahl, G., & Stehle, R. (2007). Value-at-risk forecasts under scrutiny - the German experience. *Quantitative Finance*, *7*(6), 621-636. https://doi.org/10.1080/14697680600999104

Nutzungsbedingungen:
Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: http://www.peerproject.eu Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under the "PEER Licence Agreement ". For more Information regarding the PEER-project see: http://www.peerproject.eu This document is solely intended for your personal, non-commercial use.All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-220908

QUANTITATIVE FINANCE

Value-at-Risk Forecasts under Scrutiny -- The German Experience

Journal:	*Quantitative Finance*
Manuscript ID:	RQUF-2005-0048.R2
Manuscript Category:	Research Paper
Date Submitted by the Author:	06-Sep-2006
Complete List of Authors:	Jaschke, Stefan; BaFin, Q-RM Stahl, Gerhard; BaFin, Q-RM Stehle, Richard; Humboldt University, Institute of Banking, Stock Exchanges, and Insurance
Keywords:	banking supervision, Vakue at Risk, exploratory data analysis, backtesting
JEL Code:	G28 - Government Policy and Regulation < G2 - Financial Institutions and Services < G - Financial Economics

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

1intro.tex
2dataset.tex
3rovar.tex
6robust-scale-est.tex
8conclusion.tex
a4.aux
a4.bbl
a4.tex
abstract.tex
commands.tex
content.tex
cover.tex
credits.tex
mathcommands.tex
table1.tex

SCHOLARONE™
Manuscripts

Value-at-Risk Forecasts under Scrutiny – The German Experience[*]

Stefan Jaschke[a], Gerhard Stahl[a], Richard Stehle[b]

[a]Bundesanstalt für Finanzdienstleistungsaufsicht, Graurheindorfer Str. 108, 53117 Bonn
[b]Lehrstuhl für Bank- und Börsenwesen, Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin

Version 2.91, September 6, 2006

Abstract. We present an analysis of the VaR forecasts and the P&L-series of all 12 German banks that used internal models for regulatory purposes throughout the period from the beginning of 2001 to the end of 2004. One task of a supervisor is to estimate the "recalibration factor", i.e., by how much a bank over- or underestimates its VaR. The Basel traffic light approach to backtesting, which maps the count of exceptions in the trailing year to a multiplicative penalty factor, can be viewed as a way to estimate the "recalibration factor". We introduce techniques that provide a much more powerful inference on the recalibration factor than the Basel approach based on the count of exceptions. The notions "return on VaR (RoVaR)" and "well-behaved forecast system" are keys to linking the problem at hand to the established literature on the evaluation of density forecasts. We perform extensive bootstrapping analyses allowing (1) an assessment of the accuracy of our estimates of the recalibration factor and (2) a comparison of the estimation error of different scale and quantile estimators. Certain robust estimators turn out to outperform the more popular estimators used in the literature. Empirical results for the non-public data are compared to the corresponding results for hypothetical portfolios based on publicly available market data. While these comparisons have to be interpreted with care since the banks' P&L data tend to be more contaminated with errors than the major market indices, they shed light on the similarities and differences between banks' RoVaRs and market index returns.

Keywords: banking supervision, VaR, exploratory data analysis, backtesting
JEL Classification: K23, G28

---

1. Introduction

In an important regulatory innovation the Basel Committee on Banking Supervision has allowed banks to use their own internal models – so-called Value-at-Risk (VaR) models – to calculate the regulatory capital cushion needed to cover the market risk of open positions in their trading book. Compared with the standardized methods prescribed by the Basel regulation the internal models approach offers a number of important advantages within the process of risk management, i.e., in measuring, monitoring and managing market risk for trading portfolios. These advantages include the convergence of economic and regulatory capital (Matten; 2000), the avoidance of duplicated efforts for internal and regulatory risk measurement and the signaling of competence to the market, especially to rating agencies, by the regulatory approval of a bank's internal model.

---

[*] The first two authors point out that the views expressed herein should not be construed as being endorsed by the BaFin. We thank several anonymous referees for numerous suggestions as well as seminar participants from the Center of Financial Studies, Humboldt University, and the conference of the Swiss Society for Financial Market Research for helpful discussions. Special thanks go to Andreas Zapp for valuable pointers to literature, numerous discussions and support with regard to data and **R**-coding.

The Basel paper on backtesting describes in-depth the regulatory requirements on the forecast quality of the internal model in order to ensure an adequate calculation of regulatory capital (Basel Committee on Banking Supervision; 1996b). Numerous publications reflect on VaR forecast evaluation, starting with Kupiec (1995), who points out the lack of statistical power of a backtesting that is based on a binomial test statistic. Crnkovic and Drachman (1996) propose the use of the Kuiper statistic, which is a goodness-of-fit type statistic based on the whole forecast distribution. Christoffersen (1998) draws the attention to backtesting based on specification tests, focusing on the independence property to evaluate forecast quality. On a more general level, questions of forecast evaluations are studied by Dawid (Dawid; 1982a,b, 1984, 1986; Seillier-Moiseiwitsch and Dawid; 1993). These papers are partly influenced by the literature on weather forecasting (Murphy and Winkler; 1987, 1992). Diebold, Gunther and Tay (1998) study the evaluation of density forecasts in general, making use of the Rosenblatt-transform (Rosenblatt; 1952). In a similar spirit, Berkowitz (2000) proposes an approach to the evaluation of VaR forecasts that is based on conditioned forecast distributions. Lopez (2001) studies the evaluation of volatility forecasts under economic loss functions. He establishes the link to the weather forecasting literature by transforming volatility forecasts to probability forecasts of specific events, which are then subjected to probability scoring rules. A more detailed overview of backtesting issues is given by Overbeck and Stahl (2000). More recent reviews are given by Finger (2005) and Campbell (2005).

Before describing the contribution of this paper, let us introduce some notation. Consider the hypothetical, or "clean", P&L for the time period $[t-1, t]$:

$$C_t = v_t(\pi_{t-1}) - v_{t-1}(\pi_{t-1}), \tag{1}$$

where $v_t(\pi)$ is the value of a given portfolio π at time t. The random variable $v_t(\pi_{t-1})$ denotes the at time $t-1$ frozen portfolio, evaluated at prices of time t. The *Value-at-Risk* of a portfolio π_{t-1} at the confidence level α is the $\alpha-$quantile of the distribution of losses $-C_t$ during time period $[t-1, t]$. This is interpreted as an upper bound of losses that might be surpassed only with probability $1-\alpha$. From now on we will use the shorthand notation V_t to denote the VaR forecast made at time $t-1$ for the distribution of C_t, at the confidence level 99%, which is the confidence level used in the Basel framework.

Within an observation period of 250 trading days 2.5 violations of the forecasts V_t are to be expected on the average. Hence, if too many violations occur there is good reason to doubt that the internal model's level of significance is correctly covered. In order to ensure a sufficient forecast quality, the Basel Committee tied the capital requirement to the number of VaR exceptions of the bank's model (Basel Committee on Banking Supervision; 1996a). The empirical analysis of the forecast quality conducted in this paper is in the spirit of Dawid's forecast evaluation, which is based on the whole forecast distribution.

It is of great practical importance to note that freezing the portfolio – as in (1) – is not imperative. The Basel Committee also acknowledges backtesting VaR that is based on actual trading outcomes. In that case changes of the portfolio composition during the holding period, fees etc. are superimposed on the hypothetical P&L. From a statistician's point of view, the judgment of forecast quality should be based on the hypothetical P&L, (1), because the VaR forecast assumes a static portfolio by construction. From a risk manager's point of view, however, the actual or economic P&L is actively managed and reported. Obviously both ways to tackle the backtesting problem have their intrinsic merits. German legislation prescribes a backtesting based on (1), whereas the US legislation admits to base the backtesting on the actual P&L, see (Berkowitz and O'Brien; 2002).

This is the first article to provide a detailed empirical analysis of the performance of the actual VaR forecasts of all German banks that used internal models for regulatory purposes in the year 2001. Insofar, it is comparable to the empirical analysis of similar data from six US banks by Berkowitz and O'Brien (2002). The methodology employed, however, differs significantly. While Berkowitz and O'Brien fit an ARMA-GARCH model directly to the P&L time series, we focus on the *return on VaR* time series C_t/V_t. The P&L time series show pronounced autocorrelations and volatility clustering. On the other hand, most return on VaR time series show much less autocorrelations and volatility clustering. (See section 3 for a more detailed explanation why the ratio C_t/V_t has similarities with "usual" returns.)

The theoretical contribution of this paper is to show under what conditions estimators of scale – like the empirical standard deviation – can be used to draw conclusions about the conservativeness of VaR forecasts. The theoretical results (in the appendix) can be viewed as a generalization of the "conditions producing an iid z [the probability integral transform] series" in (Diebold, Hahn and Tay; 1998). The fact that the whole forecast distribution is not available, is bridged by the empirical observation that the forecasts are relatively "well behaved". We call a time series of VaR forecasts *well-behaved* if the return on VaR comes from a scale family and is an independent series. This suggests to separate the estimation of the shape of the scale family on the one hand and the estimation of the scale of the distribution on the other hand.

The practical contribution of this paper is to compare the estimation errors of estimators of scale and shape using an extensive bootstrap analysis on our data. Certain robust methods turn out to provide the lowest estimation errors across the banks in our sample. The estimator of the shape factor exploits the relationship between M-estimators and quantile estimation, as described by Kozek (2002). This separation of the estimation of scale and shape is similar in effect though different from the techniques used by Diebold, Hahn and Tay (1998).

2. Description of the Data Set

The data set considered here contains data from all thirteen German banks that used internal models for regulatory purposes in the year 2001. The data set for each bank consists of daily VaR forecasts V_t and the corresponding daily hypothetical P&L C_t for the period 2001-2004. Most of the following figures and tables are based on the normalized P&L and VaR time series. I.e., they are divided by the banks' full sample standard deviations of P&L to insure confidentiality.

Table 1 shows summary statistics of each bank's data for the period 2001-2004. The coefficient of variation (i.e., the ratio of the standard deviation and the mean) of VaR

bank name	kurtosis of P&L	skewness of P&L	99%-quantile of losses	average VaR	coefficient of variation of VaR	average of the loss exceeding VaR
A	22.81	0.75	2.26	2.28	0.68	0.91
B	9.63	−0.14	2.92	3.31	0.69	1.01
C	12.29	−0.38	2.91	2.43	0.66	0.64
D	5.47	−0.26	2.81	3.59	0.32	0.00
E	10.44	0.25	3.29	1.92	1.06	0.35
F	33.41	−2.25	2.74	2.05	0.73	2.15
G	4.77	−0.13	2.35	2.50	0.34	0.52
H	9.80	−0.23	2.64	2.57	0.36	0.86
I	28.61	−1.77	2.93	1.58	0.72	1.06
J	9.90	−0.78	3.26	2.86	0.40	1.06
K	5.21	0.15	2.31	2.85	0.39	0.73
L	6.19	−0.36	2.79	3.89	0.24	1.61

Table 1: **Summary statistics of P&L and VaR for the period 2001-2004.** Both P&L and VaR are divided by the empirical standard deviation of the P&L to protect confidentiality.

in column 5 shows that for the majority of the banks (except banks C, E, and I) the variability of the VaR is relatively small compared to its mean. The last column reports the average loss exceeding VaR, i.e., the estimate of the expected shortfall $E[-C_t - V_t| - C_t > V_t]$. Note that for the standard normal distribution, the expected shortfall is approximately 0.34 for the 99% confidence level. The comparatively large average losses exceeding VaR indicate the presence of outliers or fat tails. Three out of the thirteen banks had more than four violations in 2001 and only four had no violations at all. For reasons of confidentiality, the individual numbers of violations are not reported here.

Figure 1 shows the time series C_t and $-V_t$ of selected banks. As can be seen from banks B and F, for example, it is not reasonable to assume stationarity of the P&L and VaR time series. Hence, the summary statistics given in table 1 are to be interpreted with care. Banks D and L give an example of relatively conservative VaR forecasts.

As the descriptive statistics show, banks tend to be conservative in the sense that they overestimate their VaR. The traffic light approach implies a one-sided loss function. Hence, this backtest does not recover the information in the data about forecast quality of a VaR-model as a whole. In the following sections we will have a closer look at forecast quality using more powerful tools.

3. Return on VaR

In their paper on backtesting, the Basel Committee on Banking Supervision (1996b) encourages banks to apply backtesting procedures beyond the so-called traffic light approach. In the next three sections, we make proposals for possible refinements. The proposed tools have been used by the *Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin)*, the German single regulator for integrated financial services supervision, for a

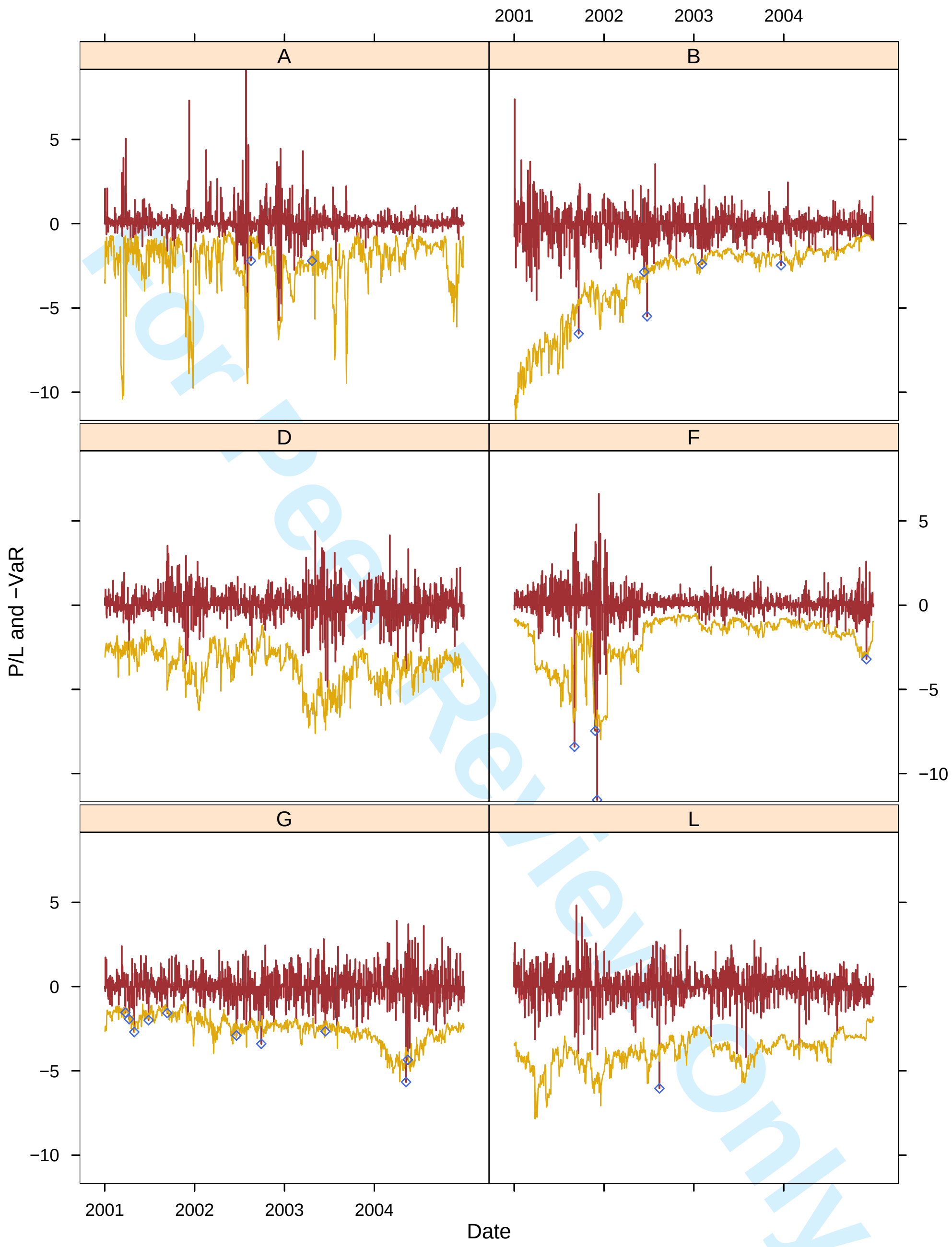


Figure 1: **Time series of P&L and -VaR.** Diamonds are added to mark the violations.

couple of years.

The key concept in the following analyses is the ratio of clean P&L over VaR

$$R_t := \frac{C_t}{V_t}, \tag{2}$$

which we call the *return on VaR (RoVaR)*. Note that both C_t, the P&L of period $[t-1,t]$, and the VaR V_t, estimated at time $t-1$, are expressed in level terms, i.e., denominated in Euros. V_t predicts the variability of C_t and is thus based on information available at $t-1$.

There are several reasons why the RoVaR concept is important:

economic interpretation The numbers R_t can be interpreted as a kind of (rate of) *return* and are thus – after proper rescaling – in principle comparable to returns from other investments like stocks and bonds.

The (rate of) return of an investment over a time period $[t-1,t]$ is usually defined as

$$R_t = \frac{\text{profit over the period } [t-1,t]}{\text{capital invested at the beginning of the period}}.$$

For portfolios π of standard products like stocks and bonds this makes

$$R_t = \frac{v_t(\pi_{t-1}) - v_{t-1}(\pi_{t-1})}{v_{t-1}(\pi_{t-1})}.$$

This does not extend to products like swaps and futures, however. The present value of swaps is zero at inception, v_t fluctuates around zero, and it has no relation to the capital invested. For the swap trading desk at an investment bank, for example, the "capital invested" is the economic capital that the bank has assigned to the trading desk. If regulatory and economic risk capital measures coincide, then the "capital invested" is the desk's VaR-limit. In fact, some banks use the VaR-limit to quantify the economic capital used by a trading unit. Otherwise economic and regulatory capital should at least be related. The definition and the utilization of the VaR-limit is, however, very different across banks and thus the ratio P&L over VaR-limit cannot be easily used for comparisons across banks. While the ratio P&L/VaR-limit is the return from a shareholder view, the ratio P&L/VaR is the return from the risk manager's view. Both behave similarly over time if VaR-limit utilization is nearly constant.

weather forecasting The literature on weather forecasting developed both theory and statistical methodology for the evaluation of forecasts. The next section shows how the RoVaR time series play a key role of linking VaR forecasts to forecasts of the whole P&L-distribution, which are at the center of the general forecasting literature.

statistical tractability The most important source of non-stationarity for both time series C_t and V_t is the deliberate decision by banks to decrease or increase their risk exposure. This can be seen from the time plots displayed in figure 1 for bank F, for example. An even more pronounced case of a bank that transferred its portfolio to

> London in three steps is not shown. Both VaR and P&L volatility were reduced from a few million EUR to some ten thousand EUR, and finally to 0. The assumption of stationarity seems much more sensible for R_t than for C_t. Moreover, RoVaRs of most banks display relatively little serial dependence, depicted by the cumulative periodograms of the RoVaRs (figure 2) and their absolute values (figure 3).

Figure 2 shows that there is no significant autocorrelation in the RoVaR series, as expected. The excess weight at small frequencies, corresponding to positive autocorrelation, of RoVaRs of banks A and G is barely significant and may point to data errors in the computation of the P&L. Another explanation is that less liquid (time-lagged) prices are used in the computation of C_t, thereby introducing positive autocorrelation in the portfolio returns.

Unlike the RoVaRs of stocks (MSGERM. denotes the MSCI Germany stock index), the RoVaRs of banks D, F, G, and L show no marked stochastic volatility (heteroscedasticity), see figure 3. One possibly reason is that the risk control limit system filters out some of the heteroscedasticity of the original returns. Another possible reason is that those banks were not exposed to risk factors with high heteroscedasticity – like MSGERM. – but to risk factors with low heteroscedasticity – like ICEUL5Y (denoting bond returns implied by the 5-year EUR swap rate).

Due to the Basel requirement of an “effective observation period of at least one year” almost all of the 12 banks use equally weighted estimators for covariance matrices or for empirical distribution functions (“historical simulation”) based on one year of daily data for regulatory purposes. Estimators that are based on exponential weights, which adapt more quickly to different volatility regimes can be used and are used for internal control purposes, however. The VaR numbers used here are based on the equally-weighted estimators. Consequently, the RoVaRs of the banks are compared to the appropriate RoVaRs of bonds and stocks. For this reason, the bond returns implied by the 5-year-EUR swap rate (ICEUL5Y) and the MSCI Germany index returns (MSGERM.) are divided by their empirical standard deviation estimated from the last 250 daily returns. We will later also consider the returns from the USD-EUR exchange rate (USEURSP) and Brent oil (OILBREN).

An explanation for the excess weight at low frequencies for bank B may be the decreasing conservativeness of bank B’s VaR forecast, as seen from figure 1. The heteroscedasticity of the RoVaRs from bank A is explained by the fact that (longer) time periods with relatively well diversified risks are interlaced with (shorter) time periods with a high concentration of risks. Since the model is partial in the sense that it only models the “general risk”, it cannot predict the higher variability of returns in those time periods with higher concentration of risks (“specific risk”).

4. Well-Behaved Forecasts

The literature on weather forecasting has developed an elaborate set of concepts and diagnostics for the evaluation of probability forecasts (Murphy and Winkler; 1987, 1992). The two main concepts are *calibration* and *refinement*.

Given the joint distribution of an event a and a probability forecast p for this event, the forecast p is called *well-calibrated*, if the probability of a conditional on the fact that

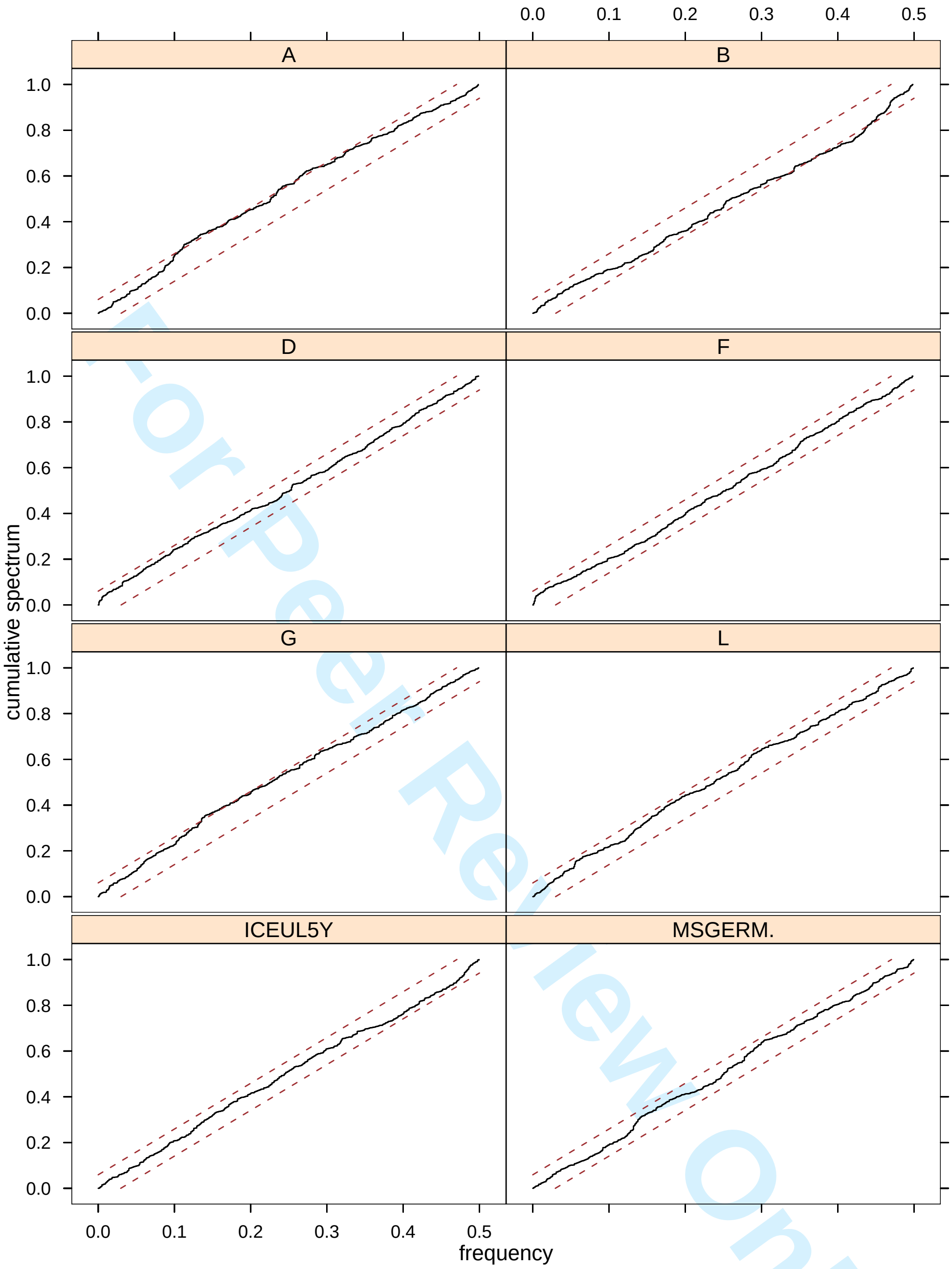


Figure 2: **Cumulative periodograms of RoVaRs for the period 2001-2004.** White noise is characterized by equal probabilities for all frequencies, which shows as a straight line in the cumulative periodogram. 95%-confidence bands are given for the null hypothesis of Gaussian white noise. ICEUL5Y and MSGERM. denote RoVaRs implied by the 5-year-EUR swap rate and the MSCI Germany index, respectively.

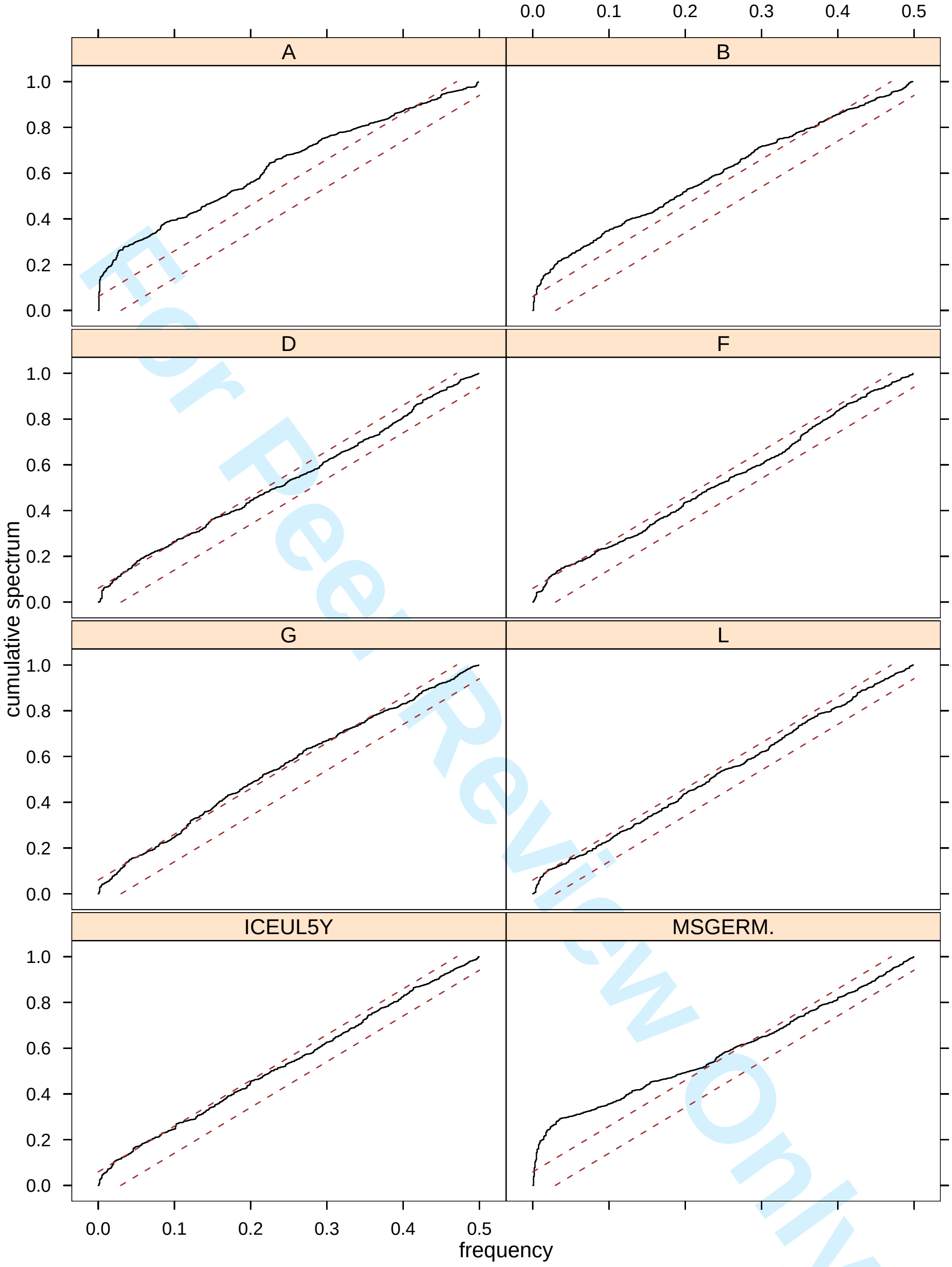


Figure 3: **Cumulative periodograms of absolute values of RoVaRs for the period 2001-2004.**

the forecast p has been made, is p: $E[a|p] = p$. The constant forecast $\tilde{p} := E[a]$ (the "climatological probability") is well-calibrated, but not "refined". The aspect how much information the forecast p contains about the event a is called *refinement* or *resolution* and is formalized and measured in different ways, see (Dawid; 1986). Partial orderings among forecasts are based on the notion that p_A is at least as refined as p_B, if the forecast p_B can be derived from p_A in a certain way. Complete orderings among forecasts can be defined by the expected loss $ES(a, p)$ for a loss function S, called *scoring rule* in this context. Another way to define refinement for a sequence of forecasts and events (a_i, p_i) is to say that the subsequence of events corresponding to a specific forecast p^*, $(a_i)_{\{i|p_i=p^*\}}$, should be stochastically independent. (Otherwise, it would be possible to improve the forecast.) In the context where we have a forecast $\hat{F}$ for the probability distribution F of a continuous random variable C, the concept of well-calibration becomes

$$P\{C \le x|\hat{F}\} = \hat{F}(x), \tag{3}$$

i.e., the forecast of each event based on C is well-calibrated (Dawid; 1984, p.281). If $\hat{F}$ is continuous, (3) implies that $\hat{F}(C)$ is uniformly distributed on $[0, 1]$. The requirement that a sequence of *realized percentiles* $\hat{F}_t(C_t)$ is stochastically independent is a kind of refinement requirement.

While it has been proposed (Berkowitz; 2000) that banks should report the realized percentiles $\hat{F}_t(C_t)$ to the supervisory authorities, the current rules only require the reporting of C_t and $V_t = -\hat{F}_t^{-1}(0.01)$. The forecast evaluation ("backtesting") as laid down in the Basel Amendment is defined in terms of the VaR-exceptions $\mathbf{1}_{\{V_t \ge C_t\}}$. The drawbacks of this approach are discussed by Kupiec (1995). See also Lopez (1999) and (Jorion; 2001, chapter 6).

Given that banks do not report the whole forecast distribution $\hat{F}_t$ but only a quantile $\hat{q}(0.01) := \hat{F}^{-1}(0.01) = -V_t$, the two key questions are:

1. Under what conditions can the "realized percentiles based on RoVaRs"

   $$p_t := F(C_t \frac{q(0.01)}{\hat{q}_t(0.01)}) = F(-R_t q(0.01)) \qquad (q(\alpha) := F^{-1}(\alpha)) \tag{4}$$

   be used as substitutes for the true realized percentiles $\hat{F}_t(C_t)$, given some fixed base distribution F?

2. Under what conditions can a scale measure $s(R_t)$ of the RoVaRs be interpreted as a *recalibration factor* that measures by how much the bank is over- or underestimating its VaR? In other words, under what conditions does the well-calibration of the forecast $\hat{F}$ imply that $s(R_t)/s_0 = 1$ for the appropriate constant s_0?

It turns out

1. that the "realized percentiles based on RoVaRs" are perfect substitutes for the true realized percentiles if and only if the banks' forecasts $\hat{F}$ come from a scale family $\hat{F}(x) = F(x/\hat{\sigma})$, see proposition 1 in the appendix;

2. that if the RoVaRs $R_t = C_t/\hat{q}_t(0.01)$ come from a scale family $R_t \sim F(./\sigma_t)$, then the well-calibration of the forecasts $\hat{F}_t$ implies $\sigma = q(0.01)$, see proposition 2 in the appendix. In other words, if R_t comes from a scale family, then any scale measure $s(R)$ can serve as "recalibration factor" for the VaR forecast.

The assumption that the banks' forecasts $\hat{F}$ come from a scale family is a rather strict one. We actually know from those banks using historical simulation and various delta-gamma-normal methods that the forecasts $\hat{F}$ do *not* in general come from such a family. This means that we *cannot* use the theory of the evaluation of distributional forecasts directly and rather base our analysis on the weaker, second result (proposition 2). We call a VaR forecast system *well-behaved*[1] if the RoVaRs R_t are an independent series and come from a scale family.

The practically needed implication is

> (I) If the estimated scale $\hat{s}(R)$ differs too much from its theoretical value s_0, then the supervisor can conclude with high confidence that the VaR forecast is not well calibrated.

This implication is possible if (1) the implication would be possible with the absence of statistical measurement error: "$s(R) \neq s_0$ implies that the VaR forecast is not well calibrated", (2) a consistent estimator for the scale $s(R)$ exists, and (3) the base distribution F of the scale family can be identified – either theoretically or statistically. Thus, the implication (I) is valid for well-behaved VaR forecasts. The assumption that the RoVaRs come from a scale family takes care of (1) and the independence assumption takes care of (2) and (3).

The previous section showed that the independence assumption for RoVaRs is quite natural for the majority of banks. The next section will show that the assumption that the RoVaRs come from a scale family is quite natural as well.

5. The Scale Families

The boxplot of RoVaRs (figure 4) gives a first glimpse of the empirical distributions of the RoVaRs. It shows that the RoVaRs are located relatively symmetrically around zero. While the mean's difference from zero is statistically significant for some banks, it is certainly not economically relevant.[2] Skewness is not significantly different from 0 for all banks.

The Q-Q plot against normal (figure 5) allows a comparison of the empirical distribution of RoVaRs with the normal distribution. The observed distributions of the RoVaRs are in line with the following "theoretical" considerations:

normality Bank G provides a striking picture: the RoVaRs are almost perfectly normal (the small circles lie on the solid line) and the VaR forecast is well-calibrated (the solid line and the dotted line almost coincide). Bank M's RoVaRs are very close to normal as well (not shown). Both of these are large, well-diversified banks, so that it is plausible that the central limit theorem "works" here.

---

[1] Note that "well-behaved" is not to be interpreted as "favorable from a supervisory point of view". It rather stands for "tractable from a statistical point of view".

[2] This is one of the important differences between the hypothetical and the economic P&L. While the mean of the economic P&L is highly significant and economically relevant, differences among banks in the mean of the hypothetical P&L are almost solely explained by differences in the treatment of theta.

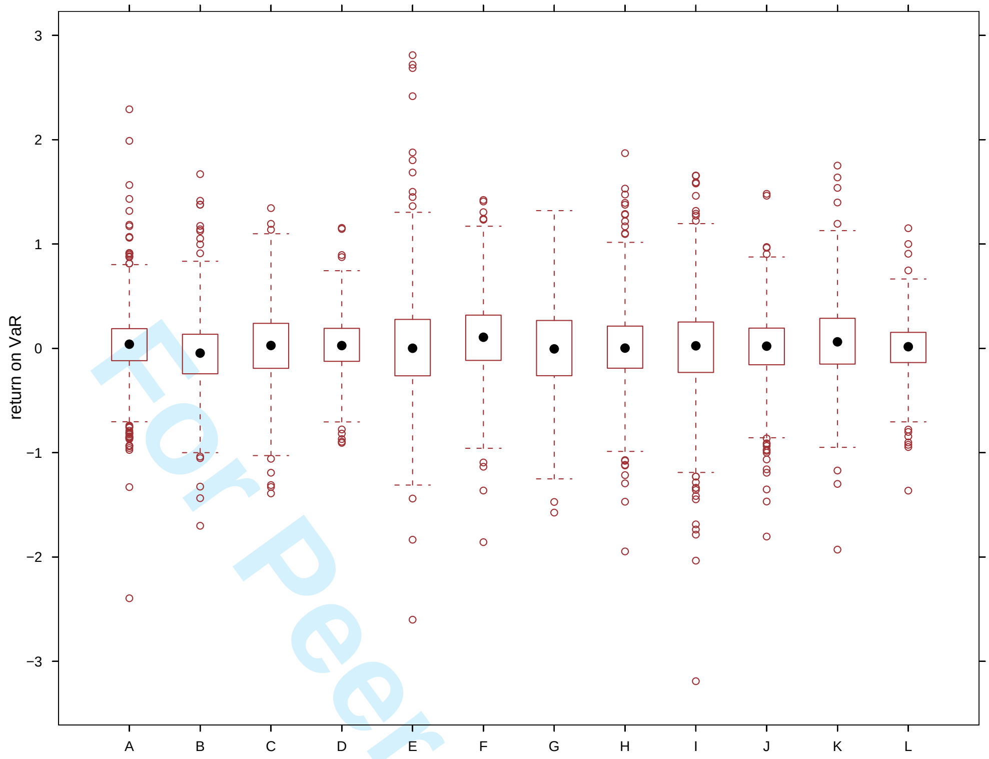


Figure 4: **Boxplot of RoVaRs.** The box shows the interquartile range and the bullet inside shows the median. The "whiskers" are drawn 2 times the interquartile range away from the box. Any point further away is shown individually.

fat tails Both swap returns (ICEUL5Y) and stock returns (MSGERM.) show fatter tails than normal, as expected. A diversified portfolio of heavy-tailed investments is still heavy-tailed, though with a faster rate of decay (larger tail index) and the center of the distribution closer to the Gaussian. This is exactly the picture for most of our banks: relatively close to normal at the center, but the tails are a bit heavier.

contamination The RoVaRs of some banks are contaminated by P&L-values that a closer inspection reveals as being economically inconsequential (plain errors, data feed inconsistencies, adjustment of illiquid prices, . . .). This also means that the RoVaRs from banks are not directly comparable to the returns from high quality stock indexes. Since VaR exceptions ($R_t < -1$) are reported to both senior management and supervisors, however, these large losses are usually well understood and free of simply correctable errors. The three largest losses of bank F, for example, correspond exactly to the three largest losses (interest rate increases) in ICEUL5Y. On the other hand, the 6σ-loss of bank A was economically irrelevant. The actual loss $-C_t$ was not exceptionally large, but the forecast V_t was "too low". The reasons for the latter are well understood by the bank's management and supervisors.

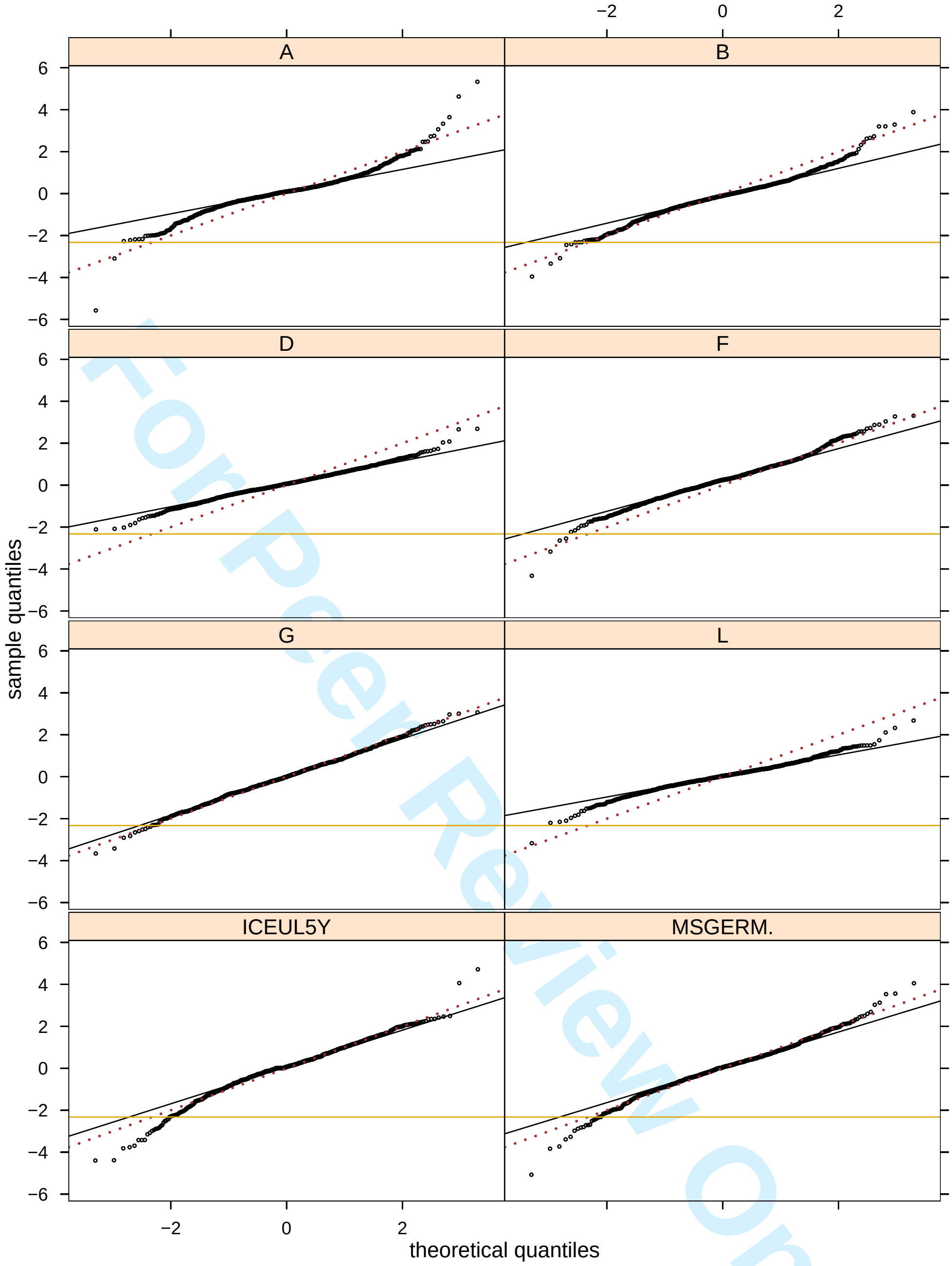


Figure 5: **Q-Q plot of the standardized RoVaR $\Phi^{-1}(0.99)R_t$.** The small circles represent the empirical distribution, which is compared to the normal distribution with mean and variance estimated from the sample, represented by the solid line. The dotted line denotes the standard normal distribution, which represents the case when the standardized RoVaRs come from the normal scale family and are well-calibrated. All points below the horizontal line at $\Phi(0.01)$ are VaR exceptions, corresponding to $R_t < -1$.

The three extremes described above correspond to Tukey's "three corners". As nicely summarized by Randal (2002), Tukey criticized the usage of location and scale estimators that are optimal for the normal distribution, but very far from optimal for certain deviations from the normal distribution. The considered deviations are the "one-wild", where 1 out of the n-sample is drawn from $N(0,(10\sigma)^2)$ instead of $N(0,\sigma^2)$, representing "contamination". The other deviation from normality is the "slash distribution", which is quite similar to the normal at the center but has very fat tails[3]. An estimator's tri-efficiency is then defined as the minimal efficiency over the three "corner cases". Randal (2002) presents an overview of scale estimators that have high tri-efficiency in Tukey's sense and are thus robust w.r.t. deviations from normality in the direction of the two other corners. While our distributions are not as extreme as Tukey's two non-normal "corners", the analysis of robust estimators of scale in the next section is in the same spirit as Randal's.

6. Robust Estimation of Scale

This section provides a bootstrap analysis of the accuracy that different scale estimators provide across the different banks and years. Let $\hat{F}^{i,j}$ denote the empirical distribution function of the RoVaRs of bank i in year j. Given a scale estimator S that maps a sample $R=(R_1,\ldots,R_n)$ to $S_n(R)$, the distribution of $S_{n=125}(R_1^*,\ldots,R_n^*), R_k^* \sim \hat{F}^{i,j}$ is computed by re-sampling subsamples of size 125 from $\hat{F}^{i,j}$. This provides a picture of the variability of the scale estimate S when applied to half of a year of daily returns, provided the returns come from a distribution near the empirical distribution function of bank i's RoVaRs in year j.

A time window of half a year ($n=125$) was chosen for the scale estimate because half a year presents a practically useful time frame. Model changes or significant changes to the trading strategies, possibly implying a changing shape F^i, ask for shortest possible time windows. On the other hand, $n=125$ is long enough to allow a reasonably accurate scale estimate.

The bootstrap distributions of the log scale $\log(S(R^*))$ are very close to normal, which justifies to compute the two-sided confidence interval for the scale estimate $t_0 := S_n(\hat{F}^{i,j})$ as $[t_0/a, t_0 a]$, where $a=\exp\{z_{0.975}\hat{\sigma}(\log[S(R^*)])\}, R_k^* \sim \hat{F}^{i,j}$. We call $a-1$ the "relative accuracy at 95% of the scale estimator S at the distribution $\hat{F}^{i,j}$".

Figure 6 shows this "relative accuracy at 95%" for a series of scale estimators. The boxes and whiskers show the variability of the accuracy of a scale estimator across the 48 empirical distributions given by the 12 banks times the 4 years.

Figure 6 can be interpreted as follows. We do not intend to "identify" the "true" distributions of RoVaRs, but search for a scale estimator that performs acceptably regardless of what the "true" distribution is. For this purpose we consider 48 empirical distribution functions of RoVaRs (12 banks times 4 years). In the spirit of Randal (2002), we believe that these 48 empirical distributions span the area of RoVaR-distributions to be expected in the future.

The *length* of the boxes and whiskers in figure 6 show that the p-norm with $p=0.75$ (**pnorm**) is a reliable scale estimator in the sense that its accuracy does not depend a

---

[3] not even a first moment, like the Cauchy distribution

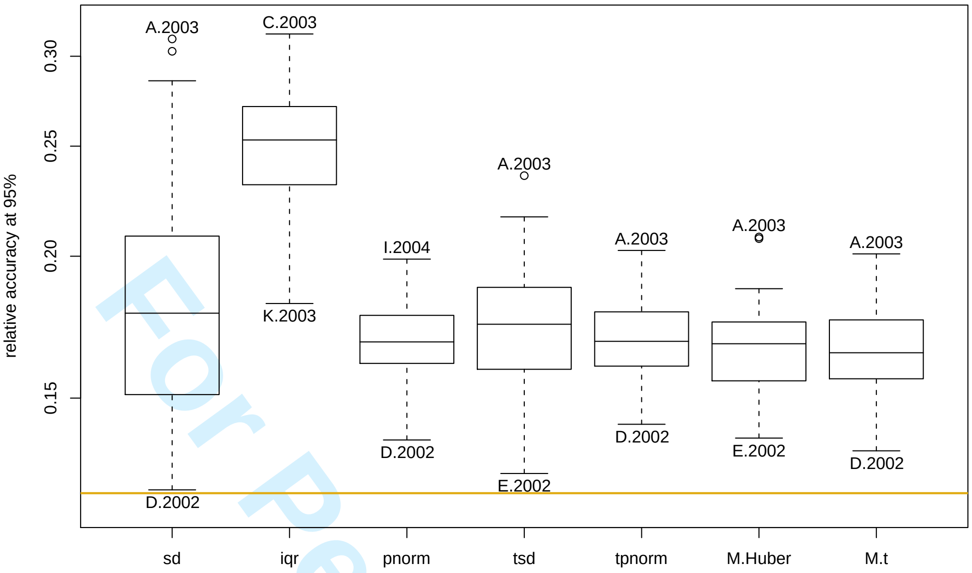


Figure 6: **Accuracy of scale estimators.** The bootstrap provides an estimate of the standard deviation of the log scale estimator $\hat{\sigma}(\log[S(R^*)])$ on samples of size $n = 125$. We translate this standard deviation to a "relative accuracy at 95%" of the scale estimator by computing $\exp\{z_{0.975}\hat{\sigma}(\log[S(R^*)])\} - 1$. The boxes and whiskers show the variability of this accuracy across banks and years. The horizontal line shows the accuracy achieved by the standard deviation at the normal distribution.

lot on the underlying distribution. On the opposite end, the accuracy of the standard deviation (**sd**) depends very much on the underlying distribution. The *location* of the boxes indicates the overall efficiency of the estimator. It shows the inefficiency of the interquartile range **iqr**, regardless of the underlying distribution.

The following scale estimators are considered:

sd the empirical standard deviation $s(R) = (\frac{1}{n}\sum_{k=1}^{n} R_k^2)^{1/2}$ (with the mean assumed to be zero),

iqr the interquartile range $s(R) = q_{0.75}(R) - q_{0.25}(R)$,

The following estimators can be viewed as generalizations of the standard deviation:

pnorm The empirical p-norm $s(x) = (\frac{1}{n}\sum_k |R_k|^p)^{(1/p)}$ reduces to **sd** for $p = 2$. $p^* = 0.75$ was chosen for figure 6.

tsd The trimmed standard deviation $s(R) = (\frac{1}{n-2k^*}\sum_{k=1+k^*}^{n-k^*} R_{(k)}^2)^{1/2}$ (with mean assumed to be zero), i.e., the k^* smallest and the k^* largest samples are dropped.

$k^*/n \approx 3\%$ was chosen for figure 6.

tpnorm The trimmed p-norm, mixing the previous two approaches. $k^*/n \approx 1\%$ and $p = 1$ were chosen for figure 6.

M-estimators for the scale parameter σ maximize

$$-n \log(\sigma) + \sum_{k=1}^{n} \log \rho(R_k/\sigma),$$

or equivalently, solve

$$\sigma^2 = \frac{1}{n} \sum_{k=1}^{n} w(R_k/\sigma) R_k^2$$

with $\psi := -[\log \rho]'$ and $w(x) := \psi(x)/x$. If $\rho = f$ is a probability density, then the M-estimator is a maximum likelihood estimator (MLE) for the scale family defined by f. For the standard normal distribution, the MLE is given by $\psi(x) = x$ and $w(x) = 1$.

M.Huber Huber's ψ-function is defined as $\psi(x, k) := \min(-k, \max(k, x))$. $k = 0.5$ was chosen for figure 6.

M.t This M-estimator is the MLE for the scale of the t-distribution. 5 degrees of freedom were chosen for figure 6.

The parameters p^* for **pnorm**, k^*/n for **tsd** and so on were chosen to provide a good median accuracy across the 64 distributions (the middle line in the boxes of figure 6), though no specific optimization was performed.

In summary, figure 6 shows that any of the estimators except **sd** and **iqr** do comparably well. In the following, we choose the trimmed mean absolute deviation (**tpnorm** with $p = 1$ and $k^*/n \approx 1\%$) as scale estimator across all banks for its conceptional and computational simplicitiy .

Figure 7 shows the scale estimates of the RoVaRs of selected banks, estimated from a moving window of length 125. The moving scale estimates of the RoVaRs of banks A and B show that the assumption of iid RoVaRs is inappropriate for some banks. Hence, we use a block bootstrap[4] to compute the confidence intervals in figure 7. Given the statistical uncertainty of the scale estimates, only banks A and B have significantly varying scale estimates over time: bank B reduced the conservativeness of its VaR forecast in 2002 and bank A had unusually large variability in returns in the second half of 2002. The latter is explained by the fact that A has significant exposure to equity. With the exception of A and B, variability of scale estimates tends to be larger across banks than across time.

7. Robust Estimation of the Shape Factor

Since the assumption of iid RoVaRs is inappropriate for some banks (A and B) we will instead assume that the standardized RoVaRs

$$\xi_t = R_t/\hat{\sigma}_t$$

---

[4] See (Bühlmann; 2002) for an overview of bootstrapping time series.

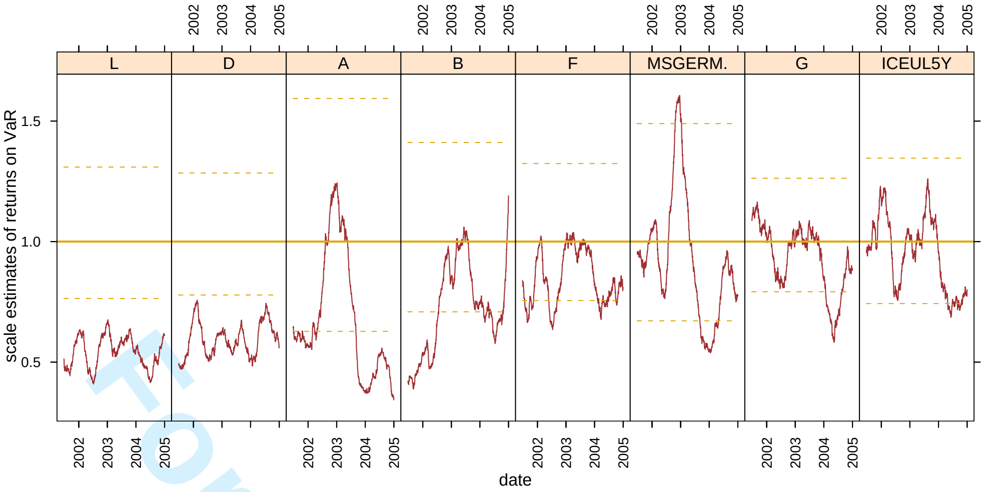


Figure 7: **Moving scale estimates for selected banks.** The scale of the RoVaRs is estimated from a moving window of length 125. The solid, horizontal line shows the scale value that is expected from a hypothetical bank with Gaussian RoVaRs and well-calibrated VaR-forecast. The dotted lines indicate the two-sided 95%-confidence intervals computed by block bootstrap.

are iid for a suitable estimate $\hat{\sigma}_t$ of the scale of the distribution of $R_t \sim F(./\sigma_t)$. We call minus the 1%-quantile of the distribution of ξ_t the *shape factor*. Note that if the RoVaRs R_t are iid, then $\hat{\sigma}_t$ can be assumed constant and the shape factor is simply the ratio between the 1%-quantile and the scale of the RoVaR distribution. In practice, risk measurement means estimation of the scale, while Basel looks in theory at the 1%-quantile. The shape factor relates both.

As in the previous section, we take as scale estimate $\hat{\sigma}_t$ the trimmed mean absolute deviation from a window of the last 125 values. This section uses a bootstrap analysis to compare different estimators of the shape factor. Figure 8 shows the accuracy of the following estimators:

linInt This estimator uses the linear interpolation of the EDF to estimate the quantile (and is the standard quantile estimator in **R**).

As described by Kozek (2002), classical M-estimators of location can also be used to estimate quantiles. Interestingly, this corresponds to a smoothing (convolution) of the EDF, as follows. M-estimators of location are of the form

$$\hat{\theta} = \text{argmin} E_{\hat{F}}[M(R - \theta)] \tag{5}$$

where θ is the location parameter and $\hat{F}$ the EDF of R. If G is a probability distribution function and

$$M_G(y) := \int_0^y (2G(z) - 1)dz,$$
$$M_{p,G}(y) := M_G(y) - (2p - 1)y,$$

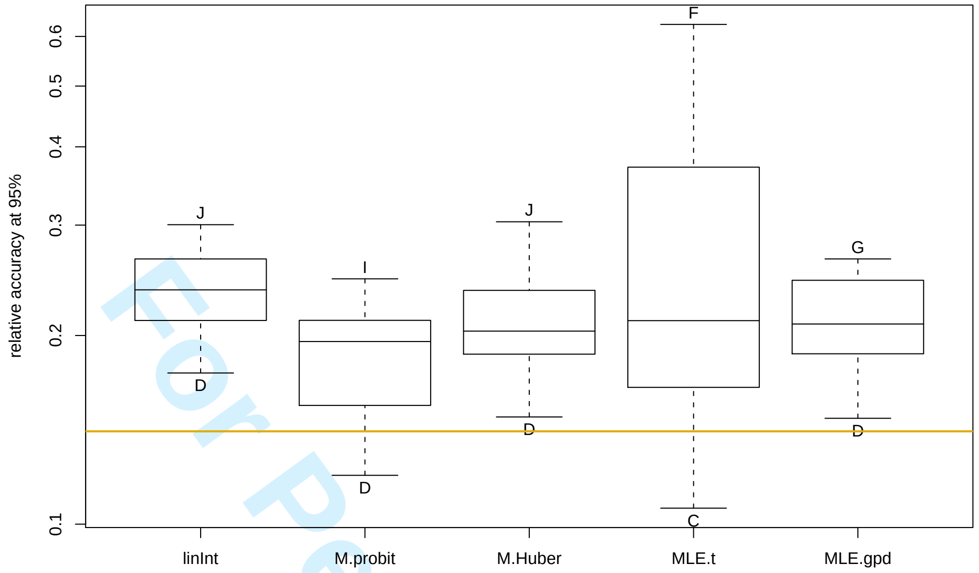


Figure 8: **Accuracy of estimators of the shape factor.** The bootstrap provides an estimate of the standard deviation of the log shape factor estimator $\hat{\sigma}(\log[S(R^*)])$ on samples of size $n = 500$ (2 years). We translate this standard deviation to a "relative accuracy at 95%" by computing $\exp\{z_{0.975}\hat{\sigma}(\log[S(R^*)])\} - 1$. The boxes and whiskers show the variability of this accuracy across banks. The horizontal line shows the accuracy achieved by the empirical quantile at the normal distribution.

then the M-estimator (5) applied to $M = M_{p,G}$ is the p-quantile of the convolution $\hat{F} * G$ (Kozek; 2002, lemma 1). I.e., the M-estimation of location (5) is equivalent to computing the quantile of a smoothed version of the EDF $\hat{F}$. Since the scale of $\hat{F} * G$ is larger than that of $\hat{F}$, the quantile of $\hat{F} * G$ is scaled back by $\sqrt{\frac{s^2(F)}{s^2(F)+s^2(G)}}$ to get an estimate of the quantile of F.

M.probit If the smoothing kernel G is Gaussian, then the resulting M-function is the probit M-function. The standard deviation $\sigma = 0.6$ is used for the smoothing kernel G in figure 8.

M.Huber If the smoothing kernel is rectangular (uniform distribution on $[-k, k]$), then the resulting M-function is Huber's M-function, already used in the previous section. $k = 0.8$ is used for figure 8.

MLE.t This uses maximum likelihood to fit a t-distribution and then infer the quantile from that.

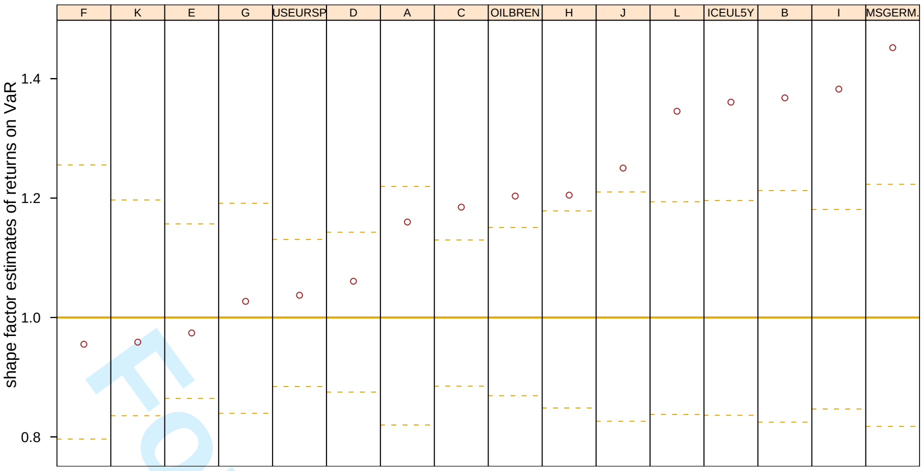


Figure 9: **Shape factor estimate per bank.** The solid, horizontal line shows the value that is expected from a hypothetical bank with Gaussian RoVaRs. The dotted lines show two-sided 95%-confidence intervals computed by block bootstrap. Also included are the shape factor estimates for the risk factors `USEURSP`, `OILBREN`, `MSGERM.` and `ICEUL5Y`.

MLE.gpd This uses the peaks over threshold (POT) method. I.e., the generalized Pareto distribution is fitted to the 6% most severe losses using maximum likelihood estimation. Alec Stephenson's EVD package is applied, see (Stephenson; 2004).

Suprisingly, **linInt**, **M.probit**, **M.Huber** and **MLE.gpd** achieve roughly the same accuracy on samples of size 500 near the 12 empirical distributions of the standardized RoVaRs $\xi_t = R_t/\hat{\sigma}_t$. **M.probit** achieves the highest accuracy, likely due to the fact that many banks' RoVaR distributions are quite close to normal. Hence, we choose **M.probit** to estimate the shape factor across all banks, which is shown in figure 9. Figure 10 shows the resulting "recalibration factor", which is the product of the moving scale estimator and the shape factor estimated from the corresponding standardized RoVaRs.

A sample size of $n = 500$ was chosen because it is the smallest sample size that achieves a relative accuracy of about 20%. This shows that much more data is needed to estimate the shape factor than the scale.

A few differences between the method as presented here and the methods actually employed should be noted. Both figures 9 and 10 assume that the standardized RoVaRs $\xi_t = R_t/\hat{\sigma}_t$ are iid and the shape factor is constant over the whole period 2001-2004. This assumption is not suitable for some banks (especially F and I) due to significant changes in the model or the business strategy, which affect the shape of the RoVaR distribution. In practice, we account for such significant model changes by assuming constancy of shape only over certain time intervals. All in all, there are 16 German banks with internal models that are approved for computing regulatory capital, some of which were approved more recently than others. Those have to be treated accordingly, but cannot be discussed

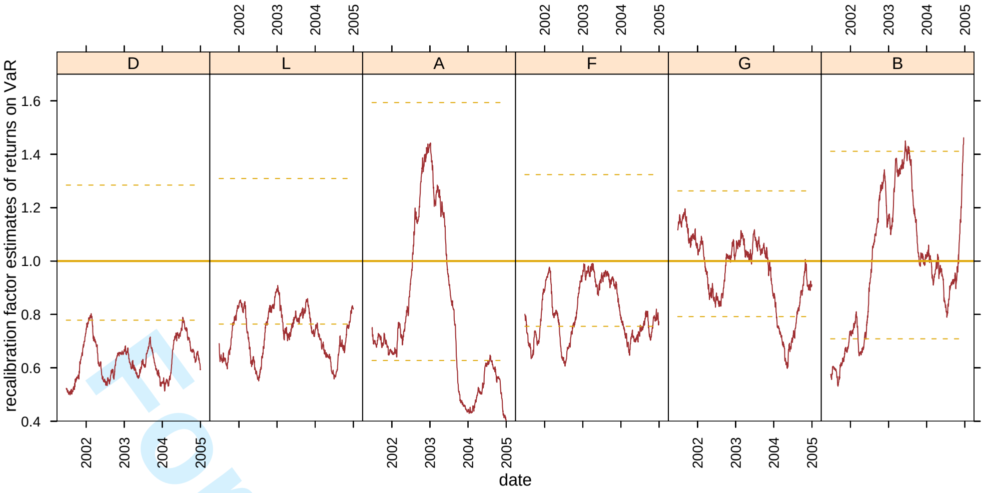


Figure 10: **Estimate of the recalibration factor.** The solid, horizontal line shows the value that is expected from a well-calibrated VaR forecast. The dotted lines show two-sided 95%-confidence intervals estimated by the block bootstrap.

here without exposing their identity.

Conclusion

1. It is useful to consider the *returns on VaR (RoVaRs)*, defined as the ratio P&L over VaR, for many economic and statistical reasons.
2. Both theoretical and practical considerations center on the question how and under what conditions a scale estimate of the empirical distribution of the RoVaRs of a bank can be used to estimate a bank's *recalibration factor*, i.e., by how much the bank is over- or underestimating its VaR. The proposals for the estimation of the recalibration factor made here can be viewed as a replacement for that table in the market risk Amendment to the Basel Accord that maps the number of VaR exceptions of the previous year to a multiplicative penalty factor.
3. If RoVaRs are "well-behaved", then scale and shape of the distribution of RoVaRs tell how well the VaR-forecast is calibrated. Most banks can be considered "relatively well-behaved".
4. Independently of whether banks' forecasts are well-behaved, it is useful to disaggreate the estimation of the recalibration factor into the estimation of scale and shape of RoVaRs. Scale can be estimated from shorter periods (like half a year in our case) or other methods like GARCH model estimators. The shape factor, on the other hand, should be estimated from longer periods (like 2 years) to achieve a similar accuracy.

5. Extensive bootstrapping analyses show that certain robust estimators of scale and shape perform well in comparison to other estimators used in the literature.

A. Appendix

If the forecast $\hat{F}$ comes from a scale family of distributions, i.e.,

$$\hat{F}(x) = F(x/\hat{\sigma}), \tag{6}$$

then the "realized percentiles based on RoVaRs" $F(-q_\alpha R_t)$ are obviously a perfect substitute for the "true" realized percentiles $\hat{F}(C)$:

$$\hat{F}(C) = F(C/\hat{\sigma}) = F(C\frac{q_\alpha}{\hat{q}_\alpha}) = F(-q_\alpha R_t) \qquad q_\alpha := F^{-1}(\alpha)$$

(The VaR forecast is $-\hat{q}_\alpha = -\hat{F}^{-1}(\alpha)$.) Interestingly, the converse also holds.

Proposition 1 *Let C be a random variable and F_0 its distribution under the true probability. Let $\hat{F}$ be a forecast for F_0 and F a fixed "benchmark" distribution. Assume F, F_0, and $\hat{F}$ are continuous and have support $(-\infty, \infty)$. Let q and $\hat{q}$ denote the inverses of F and $\hat{F}$, respectively. The value $U = F(C\frac{q_\alpha}{\hat{q}_\alpha})$ has the same distribution as the realized percentile $\hat{F}(C)$ under the true probability measure if and only if $\hat{F}$ comes from the scale-family $\hat{F}(x) = F(x\frac{q_\alpha}{\hat{q}_\alpha})$.*

Proof. We only have to show the non-trivial direction. Assume that

$$P_0\{U \le p\} = P_0\{\hat{F}(C) \le p\} \qquad \forall p \in [0, 1].$$

This implies

$$P_0\left\{C\frac{q_\alpha}{\hat{q}_\alpha} \le F^{-1}(p)\right\} = P_0\left\{C \le F^{-1}(p)\right\}$$

and

$$F_0\left(F^{-1}(p)\frac{\hat{q}_\alpha}{q_\alpha}\right) = F_0\left(\hat{F}^{-1}(p)\right).$$

Since F_0 is invertible, this equation also holds for the arguments of $F_0(.)$. Setting $x := F^{-1}(p)\frac{\hat{q}_\alpha}{q_\alpha}$, this leads to

$$\hat{F}(x) = p$$

which equals

$$= F\left(x\frac{q_\alpha}{\hat{q}_\alpha}\right)$$

by definition of x. □

An interesting question is now how the two concepts of well-calibration are related under the additional assumption

$$-q_\alpha R = C\frac{q_\alpha}{\hat{q}_\alpha} \sim F(./\sigma) \tag{7}$$

for some fixed benchmark distribution F and scale $\sigma > 0$.

Proposition 2 *Let $\hat{F}$ be a forecast for the distribution F_0 of a random variable C and F a fixed "benchmark" distribution. Assume F, F_0, and $\hat{F}$ are continuous and have support $(-\infty, \infty)$. Let q and $\hat{q}$ denote the inverses of F and $\hat{F}$, respectively. Assume that the standardized RoVaRs $-q_\alpha R = C\frac{q_\alpha}{\hat{q}_\alpha}$ are known to come from the scale family (7) and the forecast $\hat{F}$ is well-calibrated as in (3). Then $\sigma = 1$ and the forecast $\hat{F}$ "comes on average from a scale-family" in the sense of*

$$F(x) = E\left[\hat{F}\left(x\frac{\hat{q}_\alpha}{q_\alpha}\right)\right]. \tag{8}$$

In general it cannot be concluded, however, that each single forecast comes from a scale family $\hat{F}(x) = F(x/\hat{\sigma})$.

Proof. We first prove $\sigma = 1$. Since $\hat{F}$ is well-calibrated,

$$\begin{aligned}\alpha &= P\{C \leq \hat{q}_\alpha | \hat{F}\} \\ &= P\{-q_\alpha R \leq q_\alpha | \hat{F}\}.\end{aligned}$$

Taking expectation, this also holds unconditionally:

$$\alpha = P\{-q_\alpha R \leq q_\alpha\},$$

which equals $F(q_\alpha/\sigma)$ because of (7). But this implies $\sigma = 1$.

Furthermore,

$$F(x) = P\left\{C\frac{q_\alpha}{\hat{q}_\alpha} \leq x\right\} = E\left[P\{C\frac{q_\alpha}{\hat{q}_\alpha} \leq x|\hat{F}\}\right] = E\left[\hat{F}\left(x\frac{\hat{q}_\alpha}{q_\alpha}\right)\right].$$

□

References

Basel Committee on Banking Supervision (1996a). Amendment to the capital accord to incorporate market risks, `http://www.bis.org/publ/bcbs24.pdf`.

Basel Committee on Banking Supervision (1996b). Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements, `http://www.bis.org/publ/bcbs22.pdf`.

Berkowitz, J. (2000). Testing density forecasts, with applications to risk management, `http://www.uh.edu/~jberkowi/back.pdf`. appeared in the *Journal of Business and Economic Statistics* (October 2001).

Berkowitz, J. and O'Brien, J. (2002). How accurate are value-at-risk models at commercial banks?, *Journal of Finance* **57**: 1093–1112.

Bühlmann, P. (2002). Bootstraps for time series, *Statistical Science* **17**: 52–72. `ftp://ftp.stat.math.ethz.ch/Research-Reports/Other-Manuscripts/buhlmann/tsboot.rev2.ps.gz`.

Campbell, S. D. (2005). A review of backtesting and backtesting procedures, *Finance and Economics Discussion Series 2005-21*, Federal Reserve Board, Washington, D.C.

Christoffersen, P. (1998). Evaluating interval forecasts, *International Economic Review* **39**: 841–862.

Crnkovic, C. and Drachman, J. (1996). Quality control, *RISK* **9**(9): 138–143.

Dawid, A. P. (1982a). Intersubjective statistical models, *Exchangeability in probability and statistics (Rome, 1981)*, North-Holland, Amsterdam, pp. 217–232.

Dawid, A. P. (1982b). The well-calibrated Bayesian, *J. Amer. Statist. Assoc.* **77**(379): 605–613.

Dawid, A. P. (1984). Statistical theory. The prequential approach, *J. Roy. Statist. Soc. Ser. A* **147**(2): 278–292.

Dawid, A. P. (1986). Probability forecasting, *Encyclopedia of Statistical Sciences*, John Wiley & Sons, New York, pp. 210–218.

Diebold, F., Hahn, J. and Tay, A. (1998). Real-time multivariate density forecast evaluation and calibration: Monitoring the risk of high-frequency returns on foreign exchange, `http://www.ssc.upenn.edu/~fdiebold/papers/paper24/dht.pdf`.

Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management, *International Economic Review* **39**(4): 863–883.

Finger, C. C. (2005). Back to backtesting, *Research Monthly 2005*, Risk Metrics Group.

Jorion, P. (2001). *Value at Risk: The New Benchmark for Managing Financial Risk*, 2 edn, McGraw-Hill, New York.

Kozek, A. S. (2002). On M-estimators and normal quantiles, `http://www.gloriamundi.org/picsresources/akqe.pdf`, published in Annals of Statistics 31 4 (2003) pp. 1170 - 1185.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives* **2**: 73–84.

Lopez, J. (1999). Regulatory evaluation of Value-at-Risk models, *Journal of Risk* **1**: 37–63.

Lopez, J. A. (2001). Evaluating the predictive accuracy of volatility models, *Journal of Forecasting* **20**: 87–109.

Matten, C. (2000). *Managing Bank Capital*, 2nd edn, John Wiley & Sons, Chichester.

Murphy, A. H. and Winkler, R. L. (1987). A general framework for forecast verification, *Monthly Weather Review* **116**: 1330–1338.

Murphy, A. H. and Winkler, R. L. (1992). Diagnostic verification of probability forecasts, *International Journal of Forecasting* **7**: 435–455.

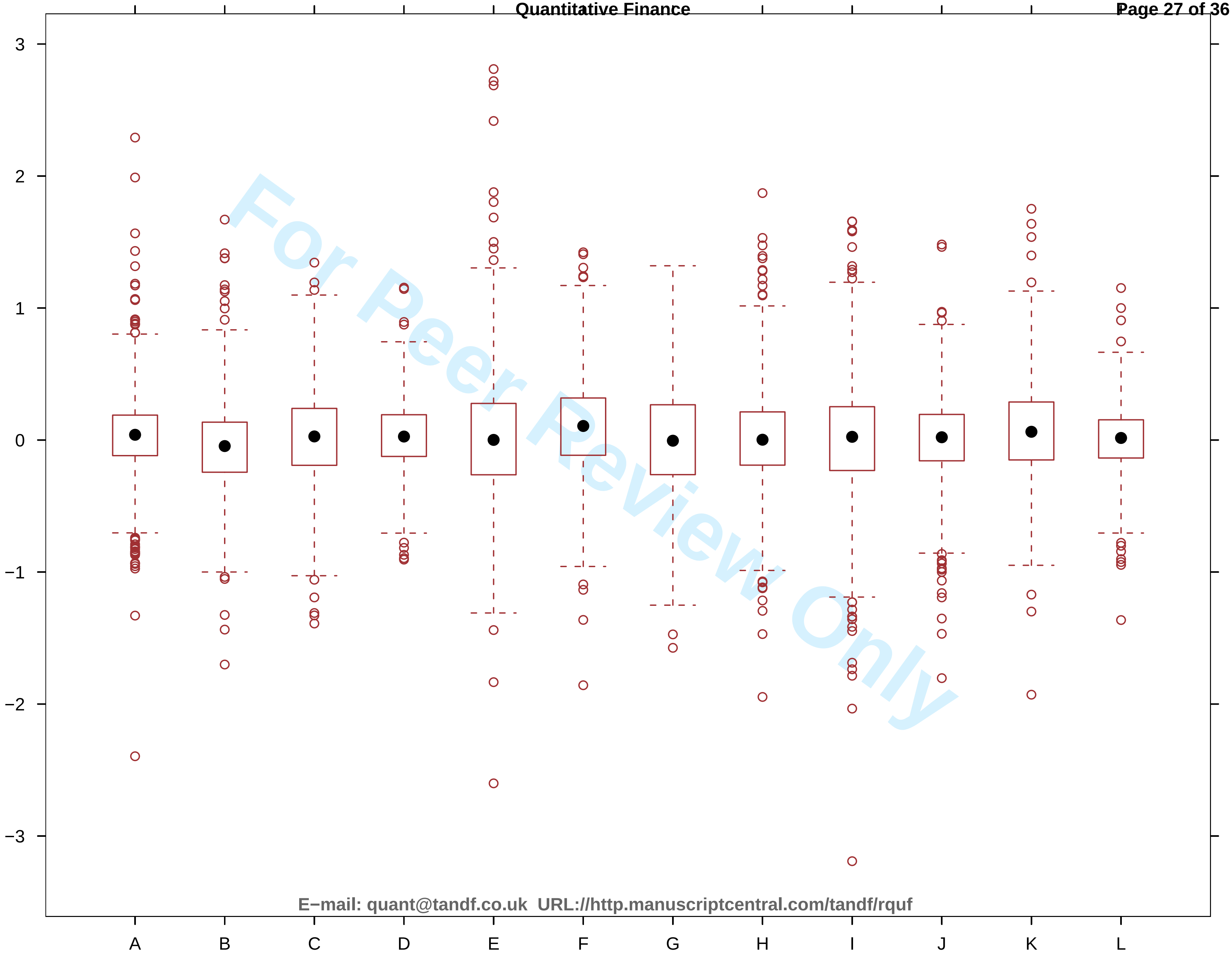
Overbeck, L. and Stahl, G. (2000). Backtesting: Allgemeine Theorie, Praxis und Perspektiven, *in* L. Johanning and B. Rudolph (eds), *Handbuch Risikomanagement*, Uhlenbruch Verlag, Bad Soden, pp. 289–320.

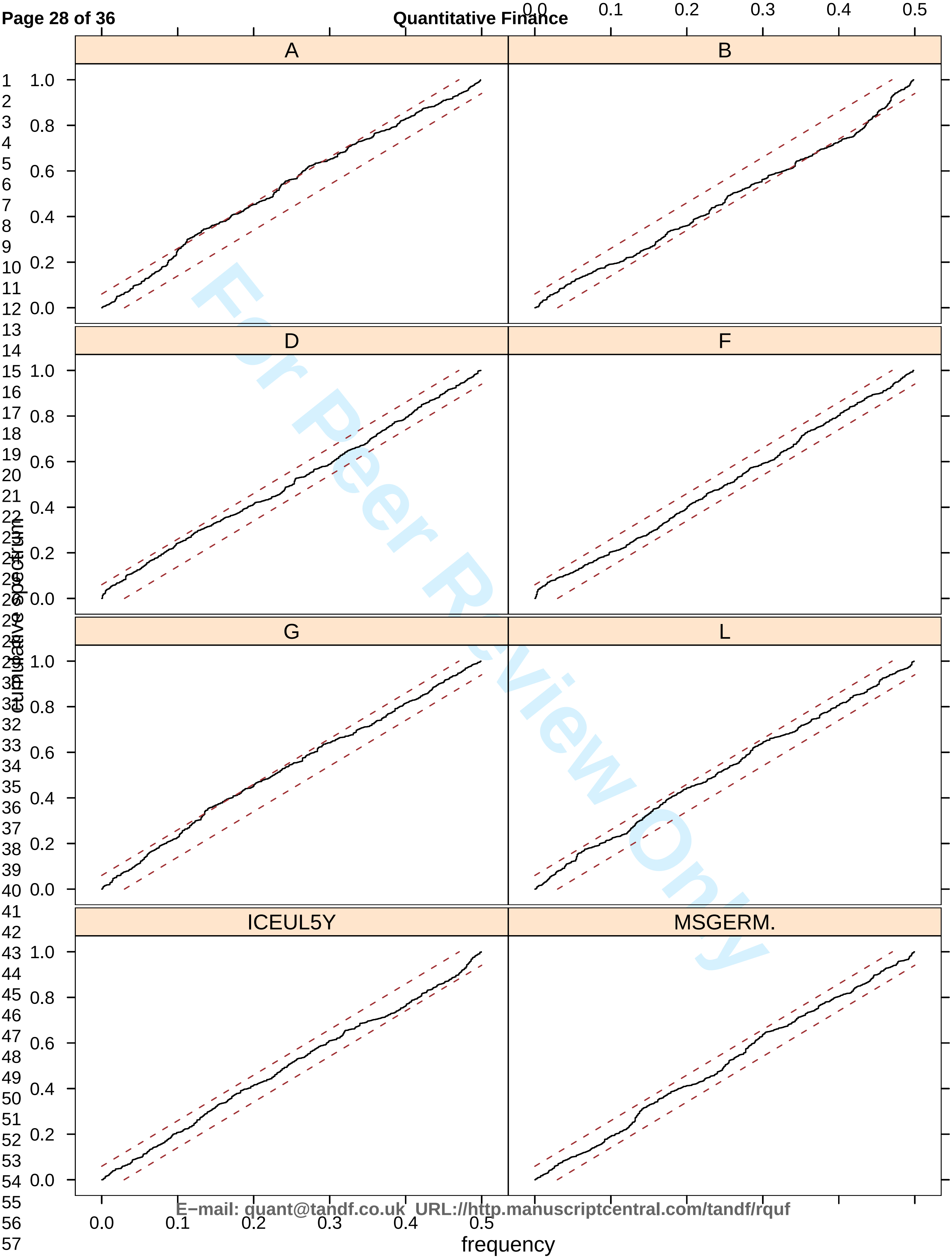
Randal, J. (2002). *Robust volatility estimation and analysis of the leverage effect*, PhD thesis, Victoria University of Wellington, Wellington. `http://www.vuw.ac.nz/staff/john_randal/Papers/phd.pdf`.

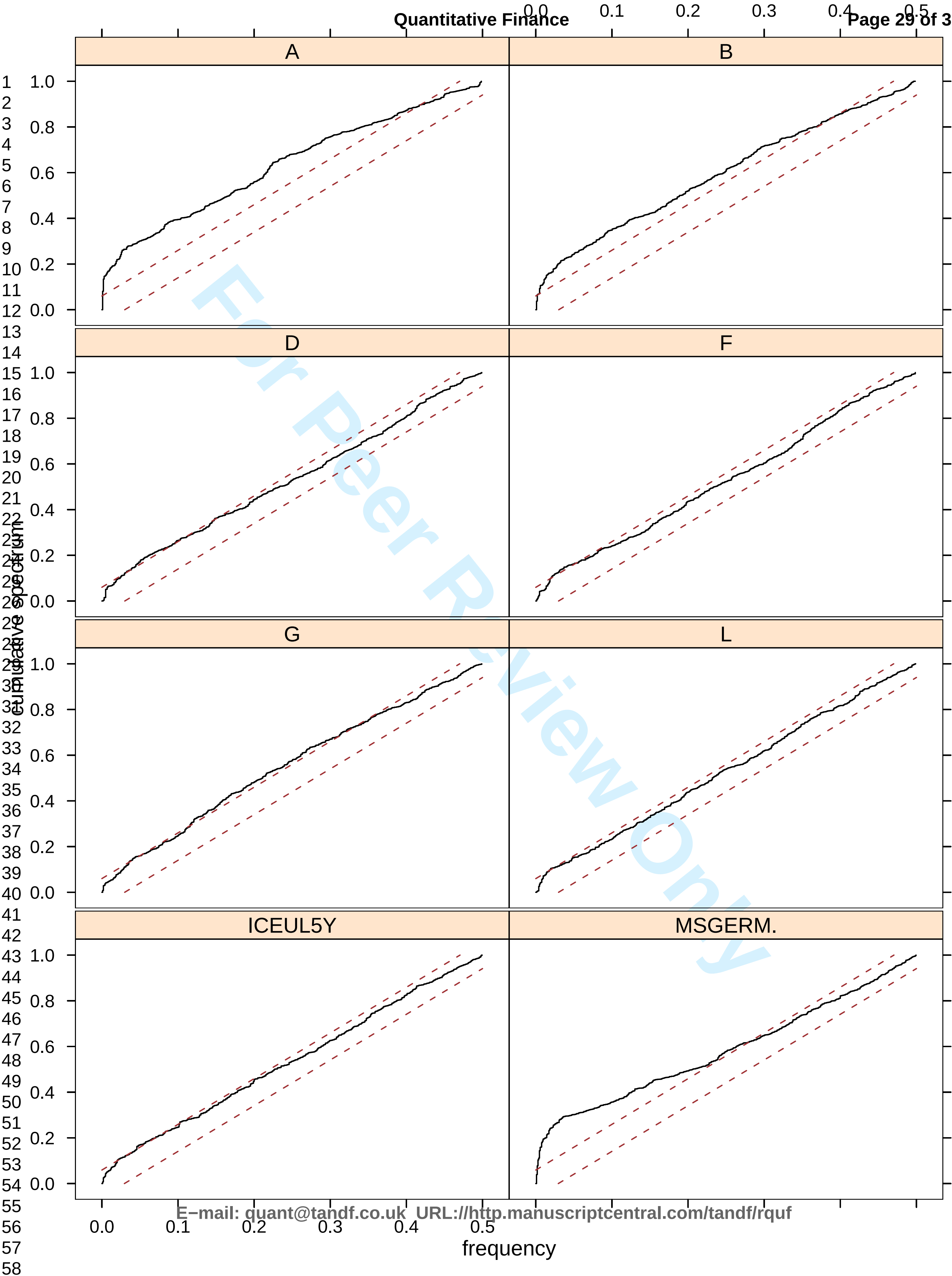
Rosenblatt, M. (1952). Remarks on a multivariate transformation, *Ann. Math. Statistics* **23**: 470–472.

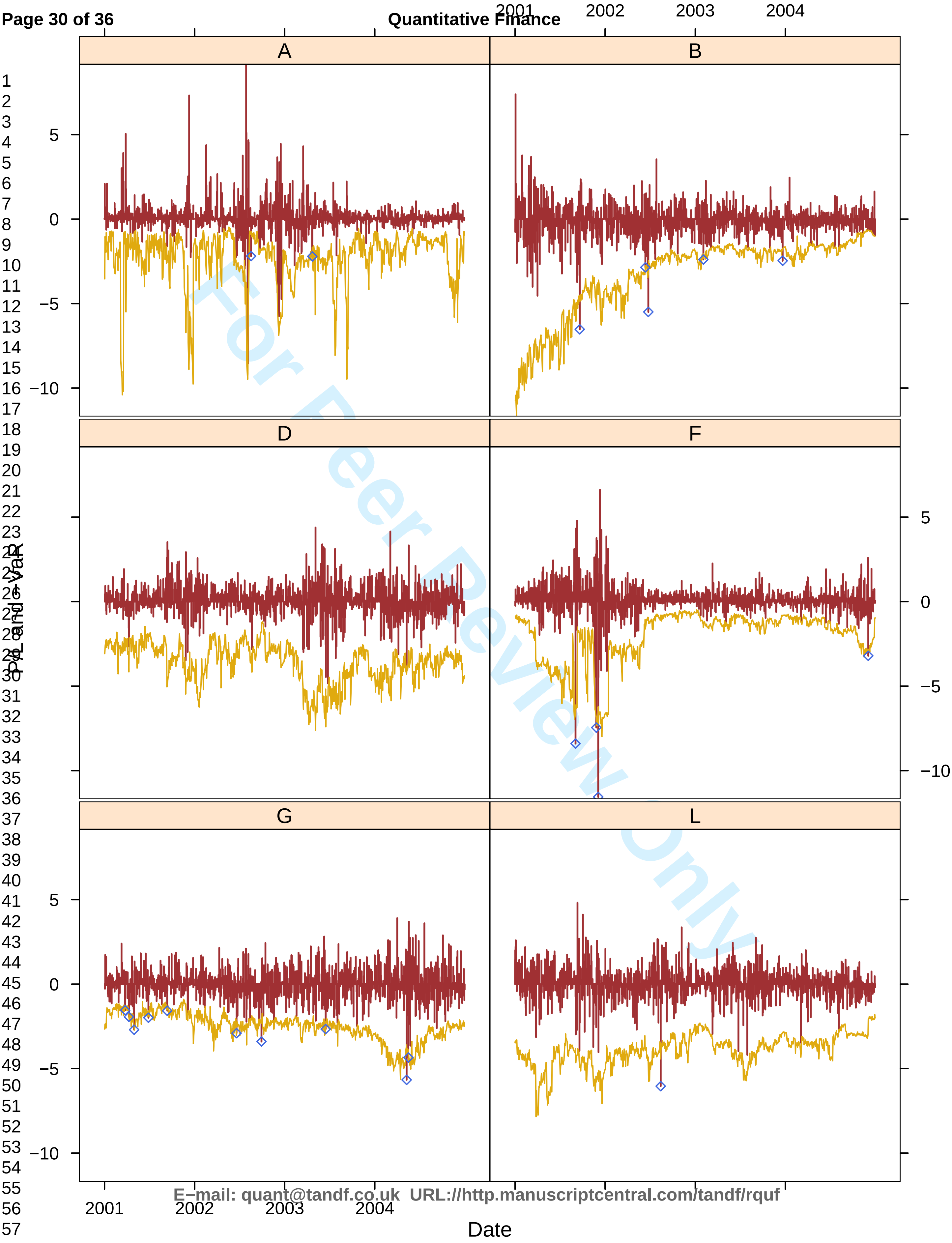
Seillier-Moiseiwitsch, F. and Dawid, A. P. (1993). On testing the validity of sequential probability forecasts, *J. Amer. Statist. Assoc.* **88**(421): 355–359.

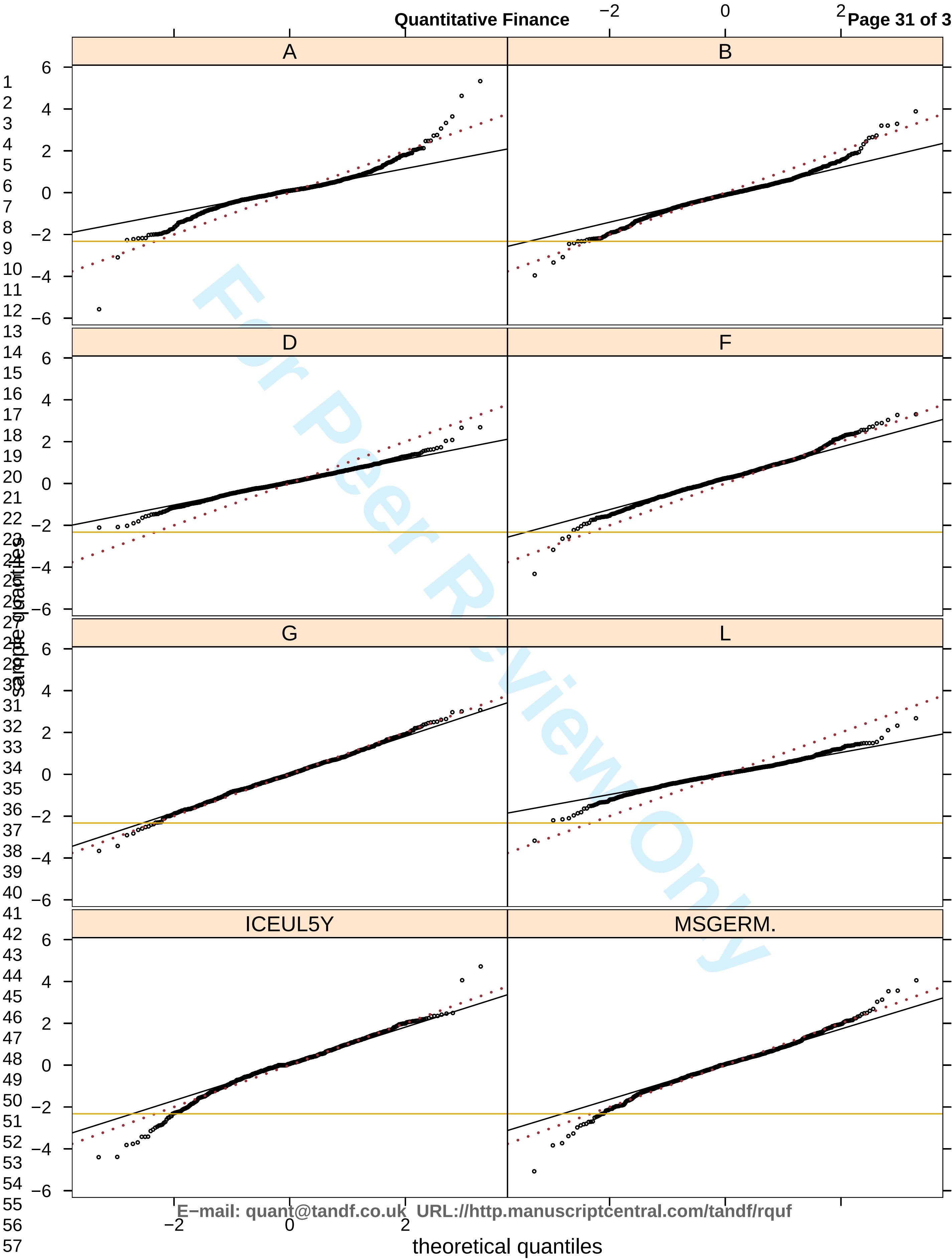
Stephenson, A. G. (2004). A user's guide to the `evd` package (version 2.1), `http://cran.r-project.org/`.

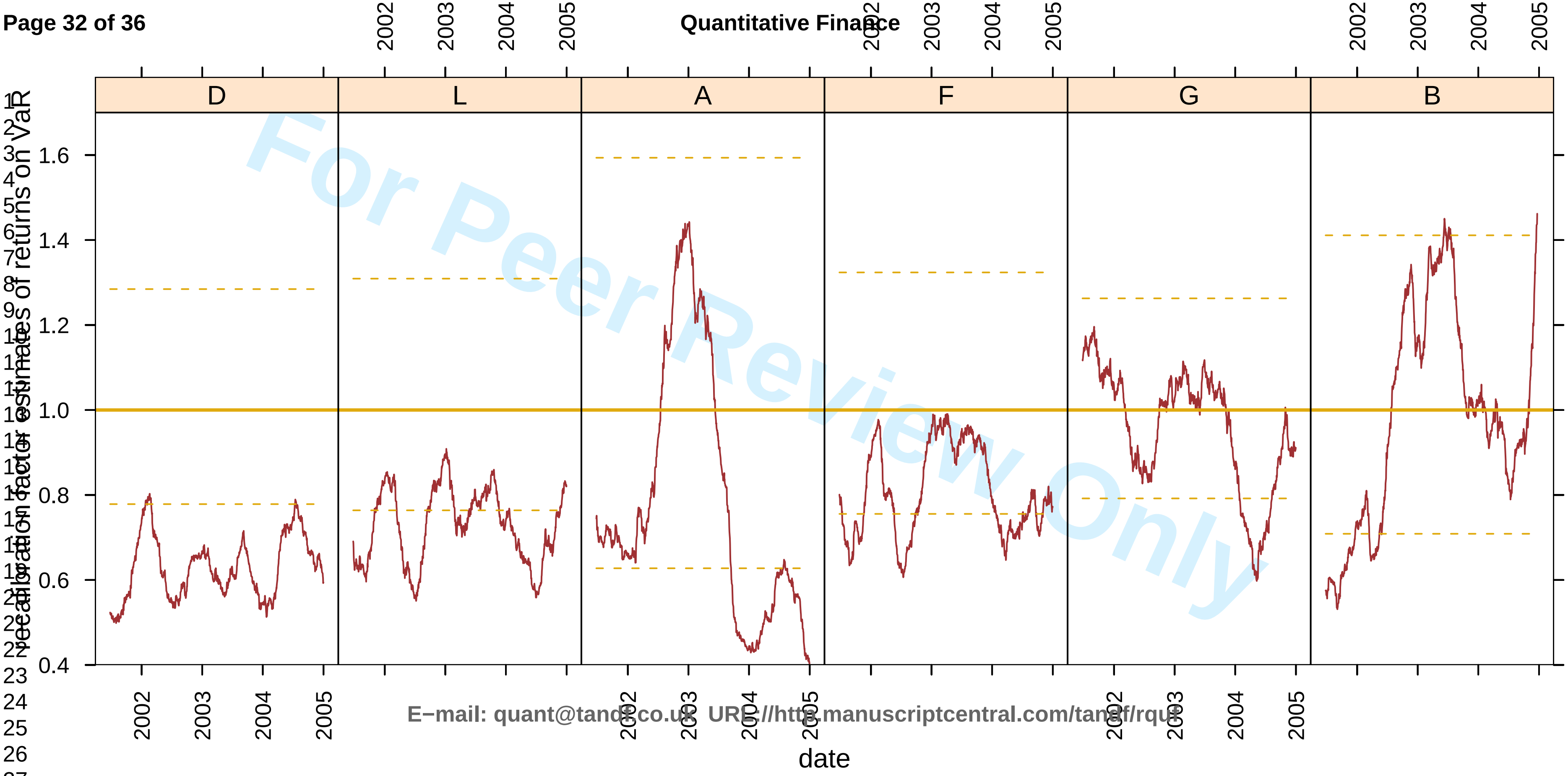

Return on VaR
3
2
1
0
−1
−2
−3
A
B
C
D
E
F
G
H
I
J
K
L

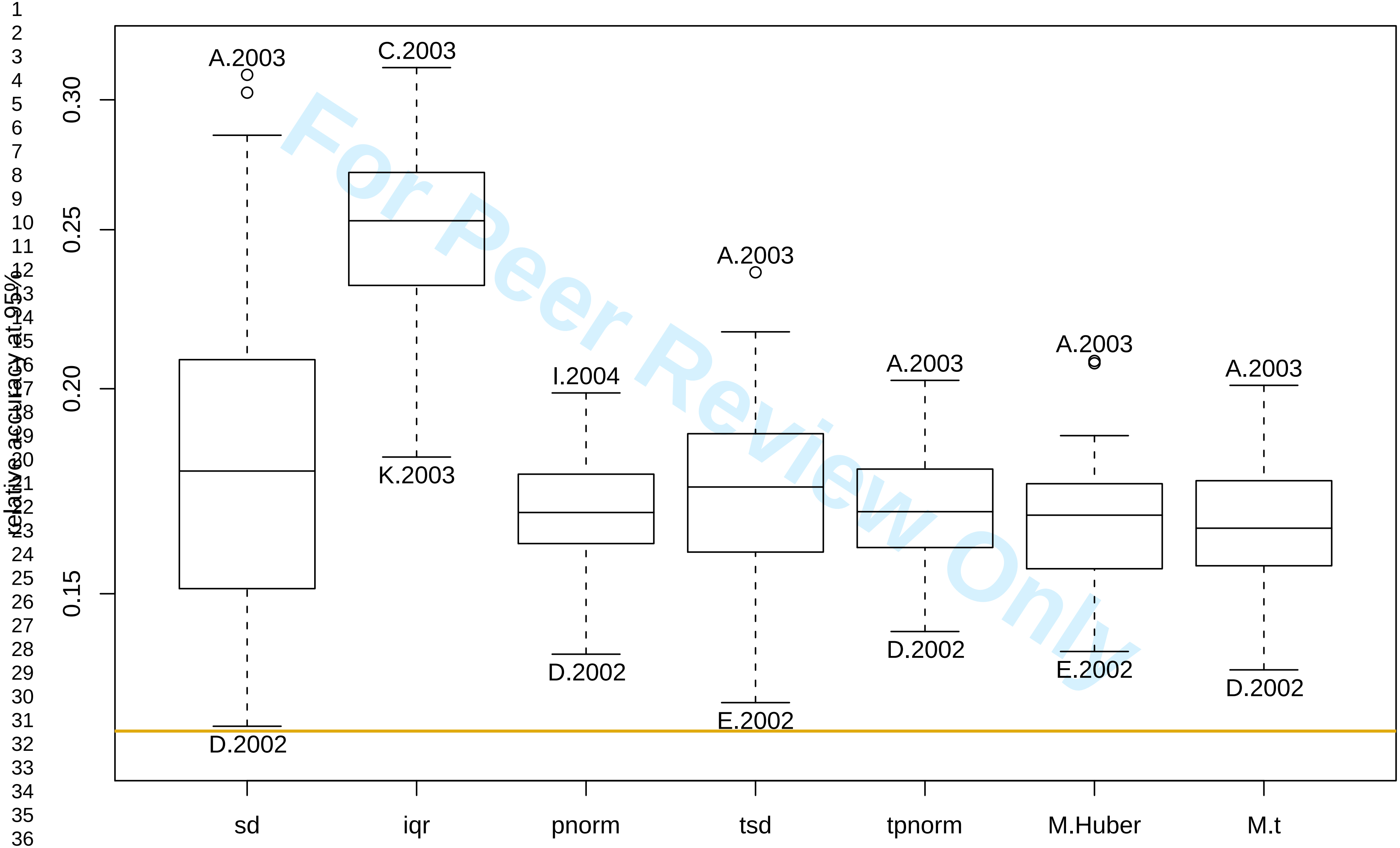

A
B
D
F
G
L
ICEUL5Y
MSGERM.
cumulative spectrum
0.0
0.2
0.4
0.6
0.8
1.0
frequency
0.0
0.1
0.2
0.3
0.4
0.5

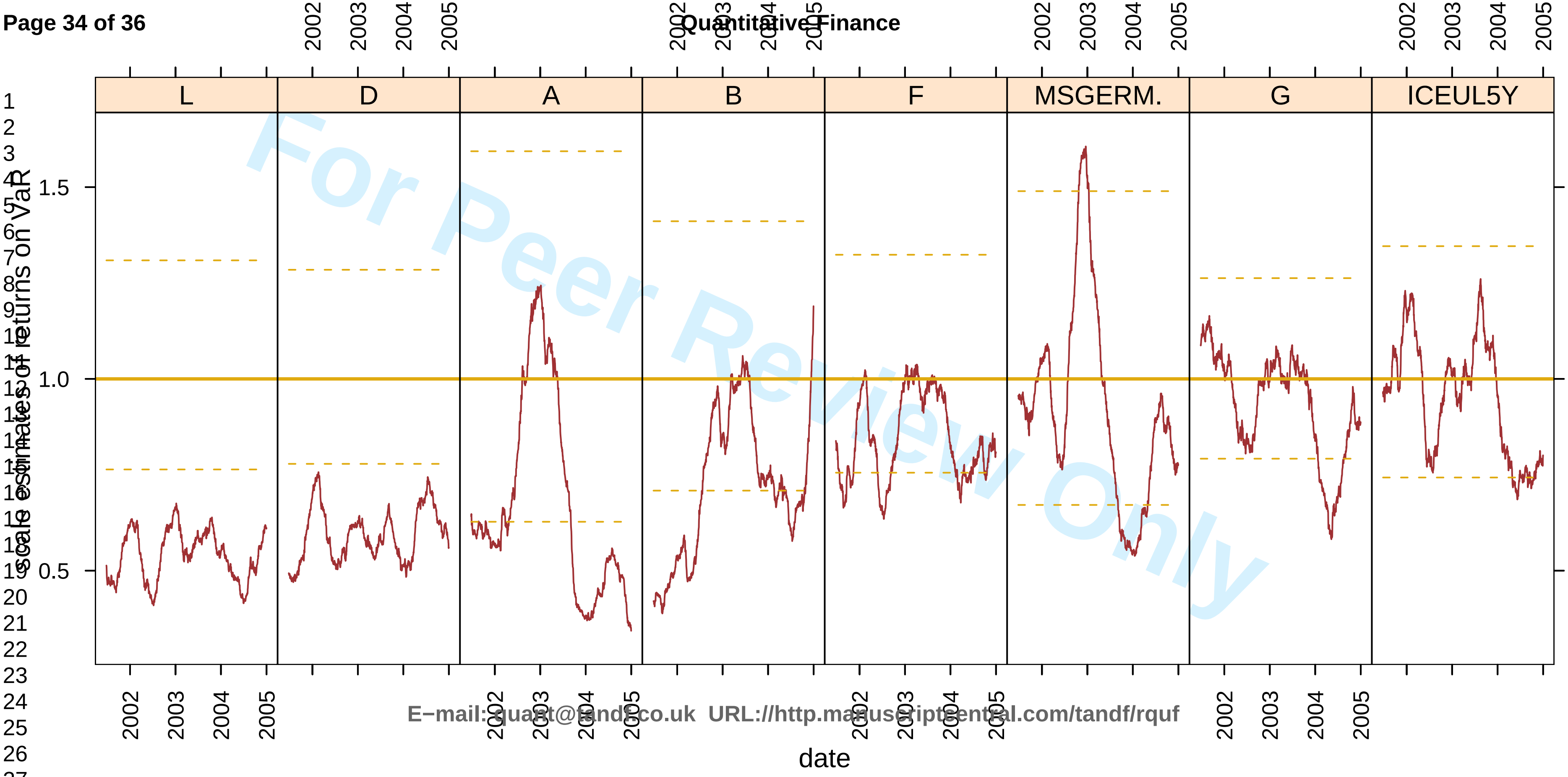

A
B
D
F
G
L
ICEUL5Y
MSGERM.
0.0
0.1
0.2
0.3
0.4
0.5
0.0
0.2
0.4
0.6
0.8
1.0
frequency
cumulative spectrum

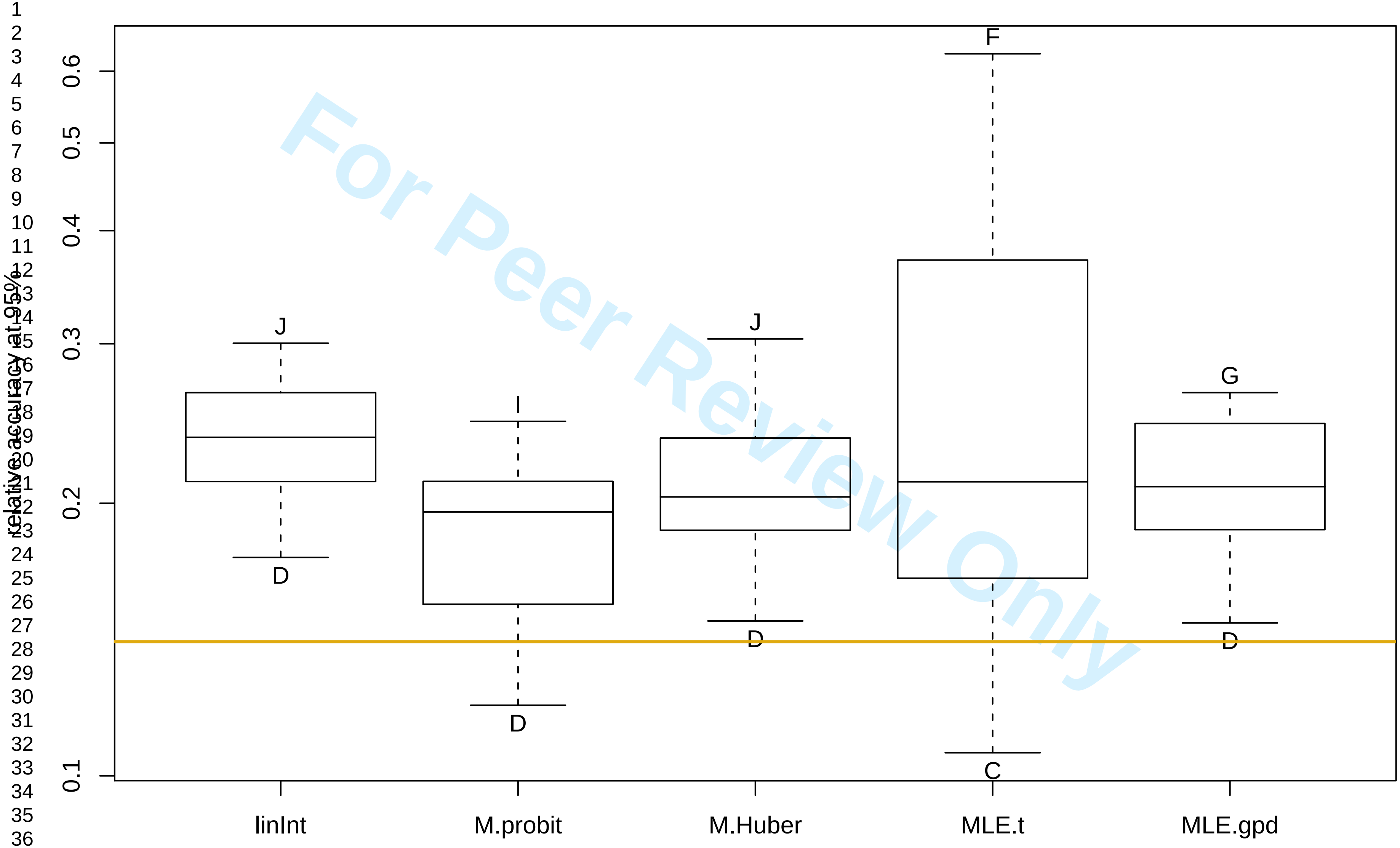

A
B
D
F
G
L
5
0
−5
−10
2001
2002
2003
2004
P/L and VaR
Date

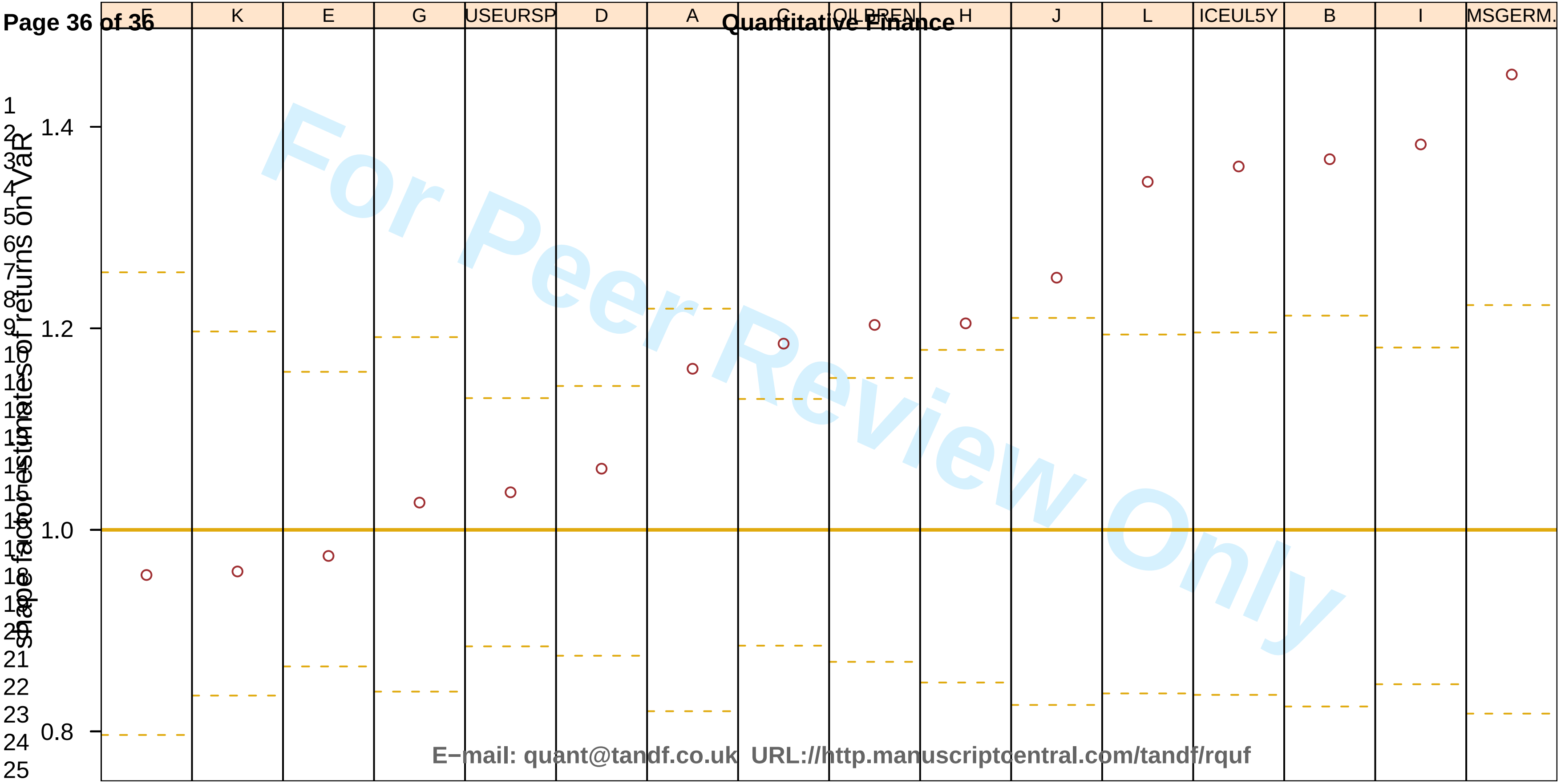

A
B
D
F
G
L
ICEUL5Y
MSGERM.
sample quantiles
theoretical quantiles
−6
−4
−2
0
2
4
6
−2
0
2


D
L
A
F
G
B
recalibration factor estimates of returns on VaR
1.6
1.4
1.2
1.0
0.8
0.6
0.4
2002
2003
2004
2005
date


relative accuracy at 95%
0.30
0.25
0.20
0.15
A.2003
C.2003
K.2003
I.2004
D.2002
A.2003
E.2002
A.2003
D.2002
A.2003
E.2002
A.2003
D.2002
D.2002
sd
iqr
pnorm
tsd
tpnorm
M.Huber
M.t


L
D
A
B
F
MSGERM.
G
ICEUL5Y
2002
2003
2004
2005
scale estimates of returns on VaR
0.5
1.0
1.5
date


relative accuracy at 95%
0.6
0.5
0.4
0.3
0.2
0.1
J
D
I
D
J
D
F
C
G
D
linInt
M.probit
M.Huber
MLE.t
MLE.gpd


F
K
E
G
USEURSP
D
A
C
OILBREN
H
J
L
ICEUL5Y
B
I
MSGERM.
shape factor estimates of returns on VaR
1.4
1.2
1.0
0.8