Pacific Economic Review

Pacific Economic Review, 15: 3 (2010) pp. 385–400
doi: 10.1111/j.1468-0106.2010.00509.x

COMOVEMENTS OF DIFFERENT ASSET CLASSES DURING MARKET STRESS

JAN PIPLACK *University of Utrecht*
STEFAN STRAETMANS* *Maastricht School of Business and Economics*

Abstract. This paper measures US financial asset class linkages (stocks, bonds, T-bills and gold) during crisis periods. We use extreme value analysis to assess the bivariate exposure of one asset class to extreme movements in the other asset classes. These bivariate co-crash probabilities can be interpreted as a measure of financial contagion. Statistical testing reveals that bivariate extreme linkage estimates exhibit time variation for certain asset pairs, possibly caused by exogenous factors like oil shocks or shifts in monetary policy. Our results have potentially important implications for long-run strategic asset allocation and pension fund management.

1. INTRODUCTION

Joint crashes across financial asset classes can have destabilizing effects on countries and the international financial architecture. They can severely increase the risk of bank failures through a joint deterioration of asset values possibly leading to domino effects in countries' financial systems, even when the banks' asset portfolio is well diversified. In general, extreme comovements of financial markets determine the systemic risk of these markets. The amount and size of jointly affected markets together with potential difficulties and bottlenecks in the financial system and payment process determine the severeness of any real effects that might follow. There have been periods of financial and political instability, like the oil crises in the 1970s and 1980s, the Asian and Russian financial crises (1997 and 1998, respectively) and, more recently, the subprime mortgage crisis in 2007, where such real effects have been witnessed. Therefore, the study of extreme (co-)movements in asset markets is not only important to investors but also to policy-makers and financial regulators that care about overall economic and financial stability.

Possibly the first systematic study of cross-country financial crisis spillovers is Morgenstern (1956, chapter X). He explicitly refers to 'statistical extremes' of the 23 stock markets and their effects on foreign stock markets. More recently, the econometric literature utilizes correlation analysis based, for example, on autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH (GARCH)-type models. Such contributions usually examine whether stock market comovements differ between crisis and non-crisis episodes and typically also try to determine the direction of possible spillover effects. Contributions

**Address for Correspondence*: Maastricht School of Business and Economics, PO Box 616, 6200 MD Maastricht, the Netherlands. E-mail: s.straetmans@maastrichtuniversity.nl. We benefited from suggestions and comments by participants at the Methods of International Finance network meetings (Barcelona 2008). We would like to thank Bertrand Candelon and Yin-Wong Cheung for helpful critiques and suggestions.

like King and Wadhwani (1990), Lin *et al.* (1994) and Engle and Susmel (1993) belong to this strand of the literature. Papers focusing on foreign exchange markets and currency crises include Eichengreen *et al.* (1996) and Kaminsky and Reinhart (2000). However, little work has been done on linkages across asset classes. Hartmann *et al.* (2004) constitutes a notable exception.

The present paper extends the published literature by increasing the amount of asset classes considered. This allows us to measure and compare the relative importance of asset substitution effects like 'flight to quality', and 'flight to liquidity'. We define 'flight to quality' as the simultaneous event of a stock market crash with a boom in either government bond or gold markets, whereas 'flight to liquidity' stands for a stock market crash coinciding with a boom in the market for T-bills. Compared to the scant existing published literature on cross-asset linkages, we use more assets and longer time series. This allows us to implement extreme value techniques and to apply tests for structural change on our linkage measures.

The methodology combines extreme value theory (EVT) with a structural stability test developed by Quintos *et al.* (2001). Contributions using similar approaches include Hartmann *et al.* (2004, 2006) and Straetmans *et al.* (2008). Bivariate extreme value theory captures the dependence structure in the tails of multivariate distributions by means of the so-called tail dependence parameter. This parameter is able to capture both linear and nonlinear dependence in the tails, whereas traditional correlation analysis only measures linear dependence and is predisposed toward the multivariate normal distribution. Another advantage constitutes the non-parametric character of the used methodology (i.e. we leave the joint asset return process unspecified and, thereby, limit the scope for mis-specification (model risk)).

The present paper is also connected to research on financial contagion. Loosely speaking, the main criteria proposed so far to measure contagion are that: (i) an adverse development in a market, company or asset affects other markets/companies/assets; (ii) the interdependencies between occurrences of financial distress and/or asset price declines during times of market stress must be different from those observed in nonvolatile times (regular 'interdependence'); (iii) the relations are in excess of what can be explained by underlying fundamentals; (iv) the events constituting contagion are negative 'extremes', such as company failures or market crashes, so that they reflect true crisis situations; and (v) the relationships are the result of transmissions through time, rather than being triggered by the simultaneous effects of common shocks.

Most empirical contagion work captures the first criterion, but there seems to be disagreement regarding which of the other criteria earn the contagion 'label'. The reason that we limit ourselves to criterion (iv) is that it enables one to focus on events that are sufficiently serious to be basically always of a concern for policy. Other criteria, definitions or classifications are also interesting and have their own ground, but more regular propagations or changes in them are not necessarily a concern for policies that aim at the stability of institutional investors with widely diversified portfolios or, more generally, the financial system as a whole. For a more elaborate discussion of the recent contagion literature and how it differs in philosophy along the abovementioned criteria, see, for example, Hartmann *et al.* (2006).

We find relatively small 'tail indexes' (measures of univariate tail 'thickness') for gold and T-bills as compared to stocks and bonds. Bivariate results indicate that the likelihood of co-crashes dominates flight to quality and flight to liquidity phenomena. Both univariate and bivariate tails are found to be nonstable over time for a limited number of assets and asset pairs. The breaks suggest a mean reverting pattern in the amount of tail thickness and tail dependence: if the degree of (univariate) tail fatness or (bivariate) tail dependence initially increases (decreases) it tends to decrease (increase) later on. However, upon comparing univariate tail risk and bivariate tail dependence across assets, we find little evidence of any significant difference.

The paper is organized as follows. Section 2 introduces two alternative measures for extreme asset linkages. Section 3 sketches the estimation procedure for these linkage measures as derived from bivariate extreme value analysis. In Section 4, we introduce the statistical tests used to test for time variation and cross-sectional differences in the proposed linkage measures. Section 5 contains empirical results and Section 6 concludes.

2. TWO MEASURES OF ASSET LINKAGES DURING MARKET STRESS

Consider, for example, a pair of stocks and bonds. Denote the return of stocks and bonds by the random variables X_i ($i = 1, 2$), respectively. Each series X_i is assumed to have n observations. To simplify the notation, we take the negative of returns so that crashes are situated in the upper tail of returns. Crisis levels or extreme percentiles $Q_i(i = 1; 2)$ are chosen such that the tail probabilities are equal across assets; that is, $P\{X_1 > Q_1\} = P\{X_2 > Q_2\} = p$. The quantile levels ($Q_1$, Q_2) can be interpreted as 'barriers' that will, on average, only be broken once in $1/p$ time periods (e.g. p^{-1} days in case of daily data frequency). We want to measure the cross-asset dependence beyond the crisis levels (Q_1, Q_2). A natural measure is the conditional tail probability:

$$\beta_\tau := P\{X_1 > Q_1(p) | X_2 > Q_2(p)\} = \frac{P\{X_1 > Q_1(p), X_2 > Q_2(p)\}}{p}, \tag{1}$$

which measures the likelihood that an asset's value falls sharply conditional upon an extreme negative shock to a second asset. Hartmann *et al.* (2006) and Straetmans *et al.* (2008) previously used this conditional tail probability to measure spillover risk in the banking sector and the US stock market, respectively. In the case of statistical independence, the conditional tail probability reduces to $p^2/p = p$, which constitutes a lower bound on the 'strength' of asset pairs' tail dependence.

Alternatively, suppose we would like to find the expected number of assets' extremes (booms or busts) given that one observes a boom or bust in at least one asset class. Let κ stand for the number of assets with extreme returns; that is, κ takes on the value of 0, 1 or 2. Our extreme linkage indicator is the conditional expectation $E[\kappa|\kappa \geq 1]$. From elementary probability theory (starting from the standard definition of conditional probability) we can state that:

$$E[\kappa|\kappa \geq 1] := \frac{E[\kappa]}{P\{\kappa \geq 1\}} = \frac{2p}{P\{X_1 \geq Q_1 \text{ or } X_2 \geq Q_2\}}, \tag{2}$$

with $P\{X_1 \geq Q_1 \text{ or } X_2 \geq Q_2\} = 1 - P\{X_1 \leq Q_1,\ X_2 \leq Q_2\}$ (see e.g. Hartmann *et al*., 2004). Notice that the conditional expectation reduces to $2/(2 - p)$ under the benchmark of full statistical independence. It is also easily observed that $E[\kappa|\kappa \geq 1] = P\{\kappa = 2|\kappa \geq 1\} + 1$, so that an alternative interpretation of our extreme linkage indicator is in terms of (1 plus) the conditional probability that both assets simultaneously boom or bust given that at least one asset exhibits extreme behaviour. For higher dimensions than two, $E[\kappa|\kappa \geq 1]$ is still equal to the ratio of the sum of the marginal excess probabilities divided by the joint failure probability. The relation between equations 1 and 2 follows from the following chain of equalities:

$$\begin{aligned} E[\kappa|\kappa \geq 1] &= \frac{2p}{P\{X_1 \geq Q_1 \text{ or } X_2 \geq Q_2\}} \\ &= \frac{2p}{2p - P\{X_1 \geq Q_1, X_2 \geq Q_2\}} \\ &= \frac{2}{2 - \beta_\tau}. \end{aligned}$$

Clearly, $1 \leq E \leq 2$ corresponds with $0 \leq \beta_\tau \leq 1$.

3. ESTIMATION OF THE LINKAGE INDICATORS

The estimation problem for equations 1 and 2 reduces to the estimation of the joint probability $P\{X_1 \geq Q_1,\ X_2 \geq Q_2\}$. Within the framework of a parametric probability law, the above linkage measures are easily estimable by maximum likelihood techniques. However, as there is no clear evidence that all asset returns follow the same distribution we avoid specific distributional assumptions and opt for the semi-parametric EVT approach proposed by Ledford and Tawn (1996); see also Poon *et al*. (2004) for other applications.

Before proceeding with the modelling of the extreme dependence structure, it is worthwhile eliminating any possible influence of marginal aspects on the joint tail probabilities by transforming the original variables to a common marginal distribution. After such a transformation, differences in joint tail probabilities can be solely attributed to differences in the tail dependence structure of the extremes. Therefore, our dependence measures, unlike, for example, correlation, are no longer influenced by the differences in marginal distribution shapes. To this aim we transform using the asset returns' ranking numbers:

$$\tilde{X}_i = \frac{1}{1 - R_{X_i}/(n+1)}, i = 1, 2, \quad (3)$$

with R_{X_i} mapping each element of X_i on its rank. This variable transform leaves the joint tail probability in the numerator of equation 1 invariant; that is,

$$P\{X_1 > Q_1(p), X_2 > Q_2(p)\} = P\{\tilde{X}_1 > s, \tilde{X}_2 > s\},$$

with $s = 1/p$. For further details on this transform, see Hartmann *et al.* (2006). The estimation problem can now be simplified toward estimating a univariate exceedance probability for the cross-sectional minimum of the two return series:

$$P\{\tilde{X}_1 > s, \tilde{X}_2 > s\} = P\{\min(\tilde{X}_1 \; \tilde{X}_2) > s\} = P\{Z_{\min} > s\}. \quad (4)$$

The marginal tail probability at the right-hand side can now be easily calculated by making an additional assumption on the univariate tail behaviour of Z_{min}. We assume that the auxiliary variable Z_{min} is nonnormally distributed or 'heavy tailed'. This is a reasonable assumption given the heavy tail feature of the original and transformed return pairs from which the cross-sectional minimum is derived. Assuming that Z_{min} exhibits a regularly varying tail implies:

$$P\{Z_{min} > s\} \approx L(s)s^{-\alpha}, \alpha \geq 1 \quad (5)$$

for large s ($p = 1/s$ small). The function $L(s)$ stands for a regularly varying function.[1] This is a reasonable assumption given the generally observed heavy tail feature of the original return pairs (and the transformed pairs from which the cross-sectional minimum is derived). Distributions with a Pareto-type tail decline have bounded moments only up to α, where α is the 'tail index' of Z_{min}. In contrast, distributions with exponentially decaying tails (e.g. the normal distribution function) or with finite endpoints have all moments bounded. Therefore, the larger α is, the thinner the tail of a distribution is. We can now distinguish two cases in which the $\tilde{X}_i (i = 1, 2)$ are either tail dependent or independent. In the former case, $\alpha = 1$ and

$$\lim_{s \to \infty} P\{\tilde{X}_1 > s | \tilde{X}_2 > s\} > 0.$$

Stated otherwise, the tail probability defined on the pair of random variables (X_1, X_2) does not vanish in the bivariate tail. Examples of asymptotically-dependent random variables include the multivariate student-*t* distribution and the multivariate logistic distribution (see e.g. Longin and Solnik, 2001; Poon *et al.*, 2004). For asymptotic independence of the random variables $\alpha > 1$, we have that:

$$\lim_{s \to \infty} P\{\tilde{X}_1 > s | \tilde{X}_1 > s\} = 0.$$

Examples of this class of distributions include the bivariate standard normal distribution.

[1] That is, $lim_{s \to \infty} L(ts)/L(s) = 1$ for all fixed $t > 0$.

Equations 3–5 show that the estimation of joint probabilities like that in equation 4 can be reduced to a univariate estimation problem. Univariate excess probabilities can be estimated using the semi-parametric probability estimator from de Haan *et al*. (1994):

$$\hat{p}_s = \frac{m}{n}(Z_{n-m,n})^{\alpha} s^{-\alpha}, \tag{6}$$

where the 'tail cut-off point' $Z_{n-m,n}$ is the $(n - m)$th ascending order statistic (or, loosely speaking, the mth smallest return, with m being the amount of returns belonging to the tail of the distribution) of the auxiliary variable Z_{min}.

We estimate the tail index, α, of the Z_{min} series by means of the popular Hill (1975) statistic:

$$\hat{\alpha} = \left(\frac{1}{m}\sum_{j=0}^{m-1}\ln\left(\frac{Z_{n-j,n}}{Z_{n-m,n}}\right)\right)^{-1}, \tag{7}$$

where m has the same value and interpretation as in equation 6. For further details on the Hill estimator and related procedures to estimate the tail index, see, for example, Embrechts *et al*. (1997).

The above discussion demonstrates that the pair of estimators in equations 6–7 both characterize univariate and multivariate tail behaviour. This is because the estimation of a joint exceedance probability can be reduced to estimating a univariate exceedance probability. In the latter case, the tail index, α, not only signals the tail thickness of the auxiliary variable Z_{min} but it also reflects the strength of the dependence in the tails of the original return pair (X_1, X_2) in the tail area $[Q_1, \infty\rangle \times [Q_2, \infty\rangle$. The smaller the value of α, the higher the probability mass in the tail of Z_{min} and, therefore, the higher the value of the joint probability in (1). Hence, one often calls the inverse parameter $\eta = 1/\alpha$ the tail dependence coefficient. An estimator of the bivariate tail probability measure in equation 1 now easily follows by combining equations 6 and 7:

$$\begin{aligned}\hat{\beta}_\tau &= \frac{\hat{p}_s}{p} \\ &= \frac{m}{n}(Z_{n-m,n})^{1/\hat{\eta}} s^{1-1/\hat{\eta}}\end{aligned} \tag{8}$$

for large but finite $s = 1/p$. When the original pair of returns exhibit asymptotic independence ($\eta < 1$), the tail probability is a declining function of the threshold s and converges to zero if $s \to \infty$. In the polar case of asymptotic or tail dependence ($\eta = 1$), the tail probability will always be above zero (regardless of the value of the conditioning percentile). However, in this paper we will not focus on the asymptotic dependence versus independence debate and will leave the tail dependence coefficient unrestricted.

Notice that the Hill statistic (eqn 7) still requires the choice of a nuisance parameter m (i.e. where do we let the tail start?). Clearly, shifting this threshold should have opposing effects on bias and variance. Balancing bias and variance to minimize the asymptotic mean squared error of $\hat{\alpha}$ underpins most empirical techniques to determine m. We opted for the Beirlant *et al.* (1999) algorithm. Beirlant *et al.* propose using an exponential regression model (ERM) on the basis of scaled log-spacings between subsequent extreme order statistics from a Pareto-type distribution.

4. HYPOTHESIS TESTING

In this section we introduce tests that can be used to assess various hypotheses regarding the temporal stability and cross-sectional equality of the considered asset linkage indicators.

From, for example, a strategic asset allocation perspective, it is important to know whether the extreme linkage indicators (1) and (2) stay constant over time. As the discussion of the Ledford and Tawn (1996) approach toward estimating equation 1 has shown, the structural (in)stability of the indicators will critically depend on whether the tail dependence parameter η is constant or not. Quintos *et al.* (2001) present several tests for identifying a single unknown break in the estimated tail index $\hat{\alpha} = 1/\hat{\eta}$. Balancing the prevention of type I and type II errors, we opt for their recursive test.

Let t denote the endpoint of a subsample of size $w_t < n$. The recursive estimator for the tail dependence parameter η is calculated from equation 7 for subsamples $[1; t] \subset [1; n]$:

$$\hat{\eta}_t = \frac{1}{m_t} \sum_{j=0}^{m_t - 1} \ln\left(\frac{Z_{t-j,t}}{Z_{t-m_t,t}}\right), \tag{9}$$

with $m_t = \kappa t^{2/3}$ (see e.g. Hartmann *et al.* (2006) for further details).

The value of the recursive test statistic equals the supremum of the following series:

$$Y_n^2(t) = \left(\frac{tm_t}{n}\right)\left(\frac{\hat{\eta}_n}{\hat{\eta}_t} - 1\right)^2. \tag{10}$$

Equation (10) compares the recursive value of the estimated tail parameter (eqn 7) with its full sample counterpart $\hat{\eta}_n$. The null hypothesis of interest is that the tail dependence parameter does not exhibit time variation. The null hypothesis of constancy then takes the form

$$H_0 : \eta_{[nr]} = \eta, \quad \forall r \in R_\varepsilon = [\varepsilon; 1-\varepsilon] \subset [0; 1], \tag{11}$$

where [·] is the integer part operator. Without prior knowledge about the direction of a break, one is interested in testing the null against the two-sided alternative hypothesis H_A: $\eta_{[nr]} \neq \eta$. As the test becomes unstable and lacks power towards the boundaries of the sample, it is evaluated over compact subsets of [0; 1]; that is, t equals the integer part of nr for $r \in R_\varepsilon = [\varepsilon; 1 - \varepsilon]$ and for small $\varepsilon > 0$. Sets like R_ε are often used in the construction of parameter constancy tests (see e.g. Andrews, 1993). In line with Quandt's (1960) pioneering work on endogenous breakpoint determination in linear time-series models, the candidate break date r maximizes test statistic (10), because at this point in time the constancy hypothesis is most likely to be violated.

Quintos *et al.* (2001) derive asymptotic critical values for the sup-value of equation 10, but these are not applicable in our framework. First, Quintos *et al.* (2001) assume that m is selected in such a way that the Hill estimator, the stability test and the resulting critical values are not marred by asymptotic bias. In practice, however, nearly all algorithms based on asymptotic mean squared error minimization (including the Beirlant *et al.* (1999) algorithm that we implement) induce an asymptotic bias term in the critical values. In addition, the critical values can be further biased by temporal dependencies like, for example, serial correlation or ARCH effects.

We decided to determine the critical values by means of a parametric bootstrap of the recursive test, while m and its subsample counterpart m_t are chosen by means of the Beirlant algorithm. The parametric bootstrap takes account of the temporal dependence in the data and the possibility of volatility spillovers from one series to another by using a bivariate GARCH model as the basis for the bootstrap. To limit the parameter space to be estimated we opted for a diagonal BEKK(1,1,1) model (see e.g. Engle and Kroner, 1995).

Quintos *et al.* (2001) describe a Monte Carlo study that indicates good small sample power, size and bias properties of the recursive break test. Only in the case of a decrease of extreme tail dependence under the alternative hypothesis ($\eta_1 > \eta_2$) do they detect poor power properties. We solve this problem by executing the recursive test both in a 'forward' version and a 'backward' version. The forward version calculates the subsample ηs in calendar time, and the backward version in reverse calendar time. If a downward break in η occurs and the forward test does not pick it up, then the backward test corrects for this.

Next to temporal (in)stability, we would also like to know whether cross-sectional differences in linkage indicators for various asset pairs are statistically and economically significant. The asymptotic normality of $\hat{\eta}$ enables some straightforward hypothesis testing. However, full sample equality tests for the tail dependence parameter η will be distorted if η values exhibit structural breaks. Therefore, a test for the cross-sectional equality of tail dependence parameters (null hypothesis) over time seems more appropriate:

$$Q_t = \frac{\hat{\eta}_{1,t} - \hat{\eta}_{2,t}}{s.e.(\hat{\eta}_{1,t} - \hat{\eta}_{2,t})}, \tag{12}$$

with $\hat{\eta}_{1,t}$ and $\hat{\eta}_{2,t}$ standing for recursive estimates of the tail dependence of asset pairs to be compared. The test statistic should be close to normality provided t is sufficiently large. In the empirical applications below, the asymptotic standard error in the test's denominator (eqn 12) is estimated using a non-parametric asymptotic variance estimator proposed by Drees (2002) that is robust for general nonlinear temporal dependence in the data.

5. EXTREME ASSET LINKAGE RESULTS: STOCKS, BONDS, T-BILLS, GOLD

In this section we assess the likelihood of extreme return exceedances and co-exceedances for different US asset classes. We also assess whether the exceedance and co-exceedance measures (1) and (2) are stable over time and are equal across asset classes by using the earlier proposed cross-sectional and stability tests.

The data consist of 11 327 daily log price differences for stocks, bonds and T-bills and 8 480 daily observations for gold. Stock, T-bill and bond price series start on 2 February 1962, whereas gold starts on 2 January 1973. The sample endpoint is 5 July 2005. We chose the Dow Jones Industrial Average as the stock price index and extracted it from Datastream. 10-year government bond and 3-month T-bill returns were calculated from the corresponding yield to maturity data (Federal Reserve Board) by means of an approximation formula (see e.g. Campbell *et al.* (1997)). Gold prices were extracted from www.usagold.com. We did not include corporate bond indices, because of our particular interest in the flight to quality phenomenon. The stock and bond returns are not compensated for dividends and coupon payments, respectively.

5.1. *Downside risk for investments in different assets*

Table 1 summarizes the magnitude and timing of the two most extreme in-sample events together with the tail index and the percentile estimates based on equations 6 and 7, respectively. Panels A and B distinguish between the left and right tail of the unconditional return distributions in order to account for possible asymmetries. The table reveals that extreme losses and gains for stocks and gold are generally much higher than for bonds and T-bills. Even excluding the most extreme stock returns in October 1987 would not change this result. Moreover, for stocks and gold the historical extremes point toward tail asymmetries. The extreme negative returns are much larger in absolute value than the respective positive returns. For bonds and T-bills this is not so clear-cut and tends to be the other way around.

A somewhat different picture emerges when we consider the estimated tail indices, $\hat{\alpha}$. Stock returns seem to be asymmetric, but the left tail index, $\hat{\alpha} = 3.42$, only falls slightly below its right tail counterpart $\hat{\alpha} = 3.69$. Bond and gold tail behaviour suggest a fatter right tail. T-bills seem to exhibit symmetric tails. The left-tail index estimates are highest in the case of bond returns. Otherwise stated, long-term government-bond investments exhibit smaller downside risk than stocks, which confirms earlier research by de Haan *et al.* (1994).

The table also provides preliminary evidence for cross-asset linkages during crisis periods. The extreme events' calendar dates suggest the presence of a flight to quality effect from stocks into bonds after 'Black Monday' in 1987. Indeed, stocks crashed on 19 October 1987 and bonds boomed on 20 October 1987. Notice also that the US stock market showed a strong technical upward correction on 21 October 1987, partly offsetting the sharp drop from 2 days previously.

Table 1. Minima, maxima, tail index, and univariate tail estimates for daily US asset returns

Panel A: Returns left tail

Asset	Min 1 (%)	Min 2 (%)	Optimal *m*	$\hat{\alpha}$	Percentile 1/10 000	Percentile 1/1 000
Stocks	–25.63 19 October 1987	–8.38 26 October 1987	287	3.42	–9.24	–4.71
Bonds	–3.59 19 February 1980	–2.74 4 April 1994	158	3.84	–4.01	–2.20
T-bills	–0.23 4 May 1981	–0.20 9 October 1979	51	2.64	–0.27	–0.11
Gold	–14.20 22 January 1980	–12.89 28 February 1983	103	3.12	–16.22	–7.76

Panel B: Returns right tail

Asset	Max 1 (%)	Max 2 (%)	Optimal *m*	$\hat{\alpha}$	Percentile 1/10 000	Percentile 1/1 000
Stocks	9.67 21 October 1987	6.15 24 July 2002	189	3.69	8.76	4.69
Bonds	4.63 20 October 1987	3.90 16 April 1980	145	3.68	4.37	2.34
T-bills	0.32 19 December 1980	0.29 5 January 1981	781	2.64	0.61	0.25
Gold	12.50 3 January 1980	11.21 3 November 1976	191	2.73	20.92	9.00

Panel C: Test for tail index equality

Asset	Test statistic	*p*-value in %
Stocks	–0.27	39.36
Bonds	0.10	45.84
T-bills	–0.00	50.07
Gold	0.31	37.81

Notes: $\hat{\alpha}$ is the reciprocal of the Hill estimator in equation 7. The 'percentile' columns show the percentiles with marginal probabilities of $p = 1/10.000$ and $p = 1/1.000$, respectively. Max 1 and 2 and Min 1 and 2 are the two most extreme positive and negative return observations in the sample, respectively. Panel A shows estimation results for the left and right tail. Panel B gives test statistics and *p*-values for the test for equality of the left and right tail indexes, as shown in equation 12.

Another interesting observation is that of the 12 most extreme events in the case of bonds, T-bills and gold, 8 fall in the years between 1979 and 1981, which probably reflects that extreme volatility was at its highest around the second oil crisis. The question remains whether the observed differences in $\hat{\alpha}$ are statistically significant across tails. Test statistics and corresponding *p*-values for the tail asymmetry test are reported in Panel C of Table 1. Additionally, we performed the (subsample) recursive equality test (eqn 12) for the left and right tails of all four asset classes.[2] Neither the (full sample) test statistics in the table nor the recursive statistics reveal significant tail asymmetries.

The economic issue of interest, however, for the general assessment of financial market stability, stress testing and risk management, is the likelihood and

[2] Graphs of the recursive equality *t*-tests are available upon request from the authors.

size of extreme returns, as reflected by the tail probabilities and corresponding percentiles. The percentiles stand for possible extreme events or scenarios whose expected waiting time to occur equals the inverse of the corresponding marginal excess probability, p. For example, a daily meltdown in the Dow Jones Industrial Average of –9.24% or more is expected to occur only once every 10 000 (trading) days, or 38.5 years. Hence, the reported values can be interpreted as (extreme) Value-at-Risk (VaR) estimates for given marginal significance levels, p.

To check whether the downside risk outcomes of Table 1 exhibit time variation we implement the structural change test described in Section 4. Results of the structural change test are summarized in Table 2. We separately test for structural change in the left and right tails. We also test against two alternative hypotheses: drops or increases in the tail index, if present in the data, are detected by the forward or backward recursive test (Panel A and Panel B, respectively). For each performed break test, the test statistic, the bootstrap simulated critical value and (statistically significant) break dates are reported. The evidence for breaks in the univariate tail behaviour is not so strong. The null hypothesis of tail index stability can only be rejected at the 1% level for the right tail of bonds and T-bills, whereas left tail breaks for stocks and T-bills are significant at the 5% level but not at the 1% level. A decrease in the right tail index for bonds and T-bills (forward test) seems to precede an increase in the tail index of the respective series. This U-shaped pattern in the tail index α (and, therefore, an inverted U-shape in the downside risk as measured by the percentiles) for bond and T-bill return tails is probably related to the shifts in US

Table 2. Univariate results for stability test for daily US asset returns

	Left tail			Right tail		
Asset	Test statistic	BT critical values	Break date	Test statistic	BT critical values	Break date
Panel A: Forward recursive test						
Stocks	6.12*	4.59	14 April 1986	1.39	3.58	–
Bonds	2.44	2.93	–	4.79**	2.11	18 February 1981
T-bills	1.20	2.24	–	24.83**	19.36	22 September 1982
Gold	0.53	3.68	–	0.27	2.12	–
Panel B: Backward recursive test						
Stocks	1.42	3.90	–	1.49	3.90	–
Bonds	1.63	3.02	–	3.43**	1.93	5 June 1989
T-bills	3.43*	2.62	29 September 1980	102.40**	19.36	8 July 1985
Gold	1.92	4.83	–	2.87*	2.35	14 June 2000

Notes: Test statistics are based on equation 10. The bootstrap (BT) critical values were simulated as described in Section 4 in the text. * and ** indicate rejection of the null hypothesis of tail index constancy at the 5% and 1% significance levels, respectively.

monetary policy that took place during the 1980s (Volcker's monetarism *vs* Greenspan's quantitative easing).

5.2. *Spillover risk across asset classes: co-crashes, flight to quality, flight to liquidity*

In this section we calculate the extreme comovement measures (eqns 1,2) introduced earlier. The results enable one to determine whether co-crashes across asset classes tend to occur more often than flight to quality or flight to liquidity phenomena.

Table 3 reports full sample estimates for the conditional probability measure (eqn 1), the conditional expectation measure (eqn 2) and for all possible US asset combinations. The extreme measures are conditioned on different marginal

Table 3. Bivariate results: full sample

Panel A1: Stocks and bonds, stocks and t-bills, stocks and gold

	Stocks and bonds		Stocks and T-bills		Stocks and gold	
	Co-crash	FtQ S/B	Co-crash	FtL S/TB	Co-crash	FtQ S/G
Tail index	1.45	1.622	1.52	1.84	1.82	1.79
Optimal *m*	290	404	397	424	371	437
Conditional probability in %						
$p = 0.1$	2.33	0.68	1.54	0.23	0.29	0.37
$p = 0.05$	1.68	0.44	1.07	0.13	0.16	0.22
$p = 0.01$	0.79	0.16	0.47	0.03	0.04	0.06
E-values						
$p = 0.1$	1.0118	1.0034	1.0082	1.0011	1.0014	1.0019
$p = 0.05$	1.0085	1.0022	1.0051	1.0009	1.0008	1.0011
$p = 0.01$	1.004	1.0008	1.0022	1.0004	1.0002	1.0003

Panel A2: Bonds and t-bills, bonds and gold, gold and t-bills

	Bonds and T-bills		Bonds and Gold		Gold and T-bills	
	Co-crash	FtL B/TB	Co-crash	FtQ B/G	Co-crash	FtL G/TB
Tail index	1.18	1.97	1.58	1.33	1.59	1.48
Optimal m	397	477	296	121	420	499
Conditional probability in %						
$p = 0.1$	14.61	0.07	0.89	2.89	0.92	2.08
$p = 0.05$	12.83	0.03	0.61	2.31	0.61	1.51
$p = 0.01$	9.60	0.00	0.24	1.35	0.23	0.72
E-values						
$p = 0.1$	1.0792	1.0003	1.0045	1.0146	1.0046	1.0105
$p = 0.05$	1.0691	1.0002	1.0031	1.0116	1.0031	1.0075
$p = 0.01$	1.0523	1.0000	1.0012	1.0068	1.0012	1.0035

Notes: FtQ S/B stands for 'flight-to-quality' from stocks into bonds, for example, as defined in the text. S stands for stocks, B for bonds, TB for T-bills and G for gold. *E*-values stand for the expected amount of co-events conditioned on an extreme percentile given by *p*.

excess probabilities, allowing us to evaluate the extreme dependence measures for different crisis levels. The table further reports the values for the tail index, α, and the number of extremes, m.

An economic interpretation of the conditional probabilities is straightforward. For example, the entry 2.33 for stock–bond co-crashes means that there is a 2.33% chance of a sharp joint drop in stock and bond returns. A 'sharp' drop in this context means that the crash levels correspond to 0.1% VaR for stock and bond tails (see Table 1 for the VaR levels for stocks and bonds). One might perceive this potential for stock–bond co-crashes as small. However, if the extreme events were truly independent, the conditional probability should equal 0.1%, which is the marginal exceedance probability, p. Therefore, if a bond market crash occurs, a strong correction in the value of the stock market becomes 23 times more likely as compared to the situation where we evaluate the univariate likelihood of a stock market crash without using information about the occurrence of simultaneous sharp movements in other asset classes.

The co-exceedance probabilities in Table 3 all exceed the benchmark level of 0.1%, implying that there is significant tail dependence. Stated otherwise, the probability of having an extreme gain or loss in one asset category suddenly becomes much higher once another 'domino stone' has fallen.

Upon comparing the linkage results in more detail, one observes that co-crashes between bonds and T-bills strongly dominate co-crashes between other asset classes, which should not surprise given their relation via the term structure. In addition, stock–bond co-crashes dominate flight to quality substitution from stocks into bonds. The same dominance of co-crashes versus flight to quality or flight to liquidity phenomena also holds for pairs of stocks and T-bills. Finally, gold seems to act as a 'safe haven' because co-crashes of any asset with gold seem less likely than a drop in asset value coinciding with sharp rises in the price of gold.

One should be aware that extreme asset linkage measures like the ones above are difficult to compare with traditional correlation analysis. Apart from being a linear dependence measure, correlation analysis is often used in conjunction with the normality assumption, whereas our dependence measures do not require any parametric assumptions whatsoever. Moreover, the normality assumption highly underestimates both the potential for univariate tail risk as well as the potential for extreme comovements (see Hartmann *et al.* (2004) for an extensive discussion on the pitfalls of correlation). To illustrate this, consider the bonds/T-bill return correlation, which equals 43.02%. Most people probably perceive this correlation as relatively high, but this can be rather illusory. Upon imposing the bivariate normal distribution to calculate the extreme bonds-T-bills market linkage, the conditional co-crash probability boils down to 0.000125% for the marginal distribution exceedance probability $p = 0.01\%$, whereas EVT estimation conditioned on the same marginal percentile rendered a conditional co-crash probability of 14.61% (see Table 3). Hence, the use of correlations in conjunction with the multivariate normality assumption leads to severe underestimation of extreme financial market linkages.

Table 4. Bivariate results: Stability test

	Stocks and Bonds		Stocks and T-bills		Stocks and Gold	
	Co-crash	FtQ S/B	Co-crash	FtL S/TB	Co-crash	FtQ S/G
Panel A1: Forward recursive test						
1969	22.3** 25 March		28.4** 10 June			
1973				25.35** 26 January		
1979		46.0** 22 February				
Panel A2: Backward recursive test						
1983						8.5* 11 March
1988				13.3** 24 May	10.0** 4 February	
1989			42.7** 13 April			

	Bonds and T-bills		Bonds and Gold		Gold and T-bills	
	Co-crash	FtL B/TB	Co-crash	FtQ B/G	Co-crash	FtL G/TB
Panel B1: Forward recursive test						
1968	38.97** 29 November					
1973		55.2** 5 December				
1979				8.8** 28 September		
1980			17.1** 16 January			
Panel B2: Backward recursive test						
1983	54.6** 12 October				50.2** 24 November	
1986			42.9** 15 October			
1987				46.0** 29 December		
1991		45.8** 7 June				408.8** 30 May

Notes: FtQ S/B stands for 'flight to quality' from stocks into bonds: for example, as defined in the text. S stands for stocks, B for bonds, TB for T-bills and G for gold. * and ** indicate rejection of the null hypothesis of tail index constancy at the 5% and 1% significance levels, respectively.

Finally, we also apply the earlier-used structural change test to the tail dependence parameters governing the conditional tail probabilities. Table 4 summarizes the structural stability results. Panels A1 and B1 show results for the forward recursive test (alternative hypothesis = increase in tail dependence), whereas Panel A2 and B2 contain results for the backward recursive test (alternative hypothesis = decrease in tail dependence). The forward recursive breaks fall in the early part of the sample (1968–1980), in contrast to the backward recursive breaks (1980s). This sequencing of breakpoints implies that there is an inverted U-shape in most asset pairs' tail dependence parameters. Notice that this runs parallel to the breakpoint analysis in the univariate section: tail index decreases (fatter tails) also preceded tail index increases (thinner tails).

The forward recursive test signals an increase in tail dependence either shortly before the first oil shock (early 1970s) or the second oil shock and the beginning of Paul Volcker's term at the Federal Reserve (end of the 1970s). In contrast, the degree of tail dependence seems to drop again (backward recursive test results) in the 1980s, coinciding with other extreme market events like the extreme appreciation of the US dollar in the mid-1980s or the 1987 stock market crash. As such, the full sample results in Table 3 represent an average across time.

6. CONCLUDING REMARKS

In this paper we have studied the linkages between four different US asset classes (US stocks, government bonds, T-bills and gold) in times of market turbulence. The linkages are characterized by their tail dependence.

We use a non-parametric multivariate measure to identify the tail dependence of the marginal return distributions and derive estimates for the expected conditional probabilities of return co-exceedances. As this approach does not rely on a particular probability law for the marginal return distributions, it circumvents the risk of (parametric) misspecification. We also tested whether the extreme linkage estimates differ across asset pairs and change over time.

Not surprisingly, downside risk, as measured by the extreme quantile or VaR level far into the distributional tail, is found to be highest for stocks and gold. However, the tail index, which is only an intermediate result for identifying tail risk, is surprisingly similar across asset classes and we were not able to reject the null of parameter equality across asset tails. Moreover, although left and right tails seem asymmetric upon looking to the point estimates of tail indices and corresponding quantiles, cross-sectional equality tests do not point to a statistically significant difference. However, tail indices seem to be changing over time for some asset classes.

Turning to the bivariate results, we find that co-crashes between stocks and bonds or stocks and T-bills are more likely than the flight to quality or flight to liquidity phenomena. Evidently, co-crashes between returns on long-term bonds and short-term T-bills are the strongest linkage encountered, which can be explained by their interdependencies via the term structure. As for the crisis relationship between gold and other assets, the results provide some evidence for a safe haven effect: the likelihood of a sharp rise in gold prices coincides more strongly with a sharp fall in other asset values than a co-crash of gold with other assets. We also performed some statistical testing on the cross-sectional equality and temporal stability of the linkage estimates. The bivariate testing results run parallel with the univariate outcomes: while cross-sectional equality cannot be rejected, we do find some temporal shifts in some of the linkage estimators.

Moreover, the dating of the shifts in both the univariate and bivariate tail behaviour seems to suggest that there may be a relationship with global shocks like oil crises, stock market crashes or fundamental shifts in US monetary policy.

Investigating the likelihood of co-exceedances across asset classes is of potential importance to institutional investors, like pension funds who try to invest in as many different asset classes as possible with an eye towards diversification. In addition, policy-makers and supervisory bodies are interested in the magnitude of these linkages because they can be interpreted as proxies of systemic risk in financial markets.

REFERENCES

Andrews, D. W. K. (1993) 'Tests for Parameter Stability and Structural Change with Unknown Change Point', *Econometrica* 59, 817–58.

Beirlant, J., G. Dierckx, Y. Goegebeur and G. Matthys (1999) 'Tail Index Estimation and An Exponential Regression Model', *Extremes* 2, 177–200.

Campbell, J. Y., A. W. Lo and A. C. McKinlay (1997) *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

de Haan, L., D. W. Jansen, K. G. Koedijk and C. G. de Vries (1994) 'Safety First Portfolio Selection, Extreme Value Theory and Long Run Asset Risks', in J. Galambos (ed.), *Proceedings from A Conference on Extreme Value Theory and Applications*, Dordrecht: Kluwer Press.

Drees, H. (2002) 'Extreme Quantile Estimation for Dependent Data with Applications to Finance', *Bernouilli* 9, 617–57.

Eichengreen, B., A. Rose and C. Wyplosz (1996) 'Contagious Currency Crises: First Tests', *Scandinavian Journal of Economics* 98, 463–84.

Embrechts, P., C. Klüppelberg and T. Mikosch (1997) *Modelling Extremal Events*. Berlin: Springer.

Engle, R. F. and K. F. Kroner (1995) 'Multivariate Simultaneous Garch', *Econometric Theory* 11, 122–50.

Engle, R. F. and R. Susmel (1993) 'Common Volatility in International Equity Markets', *Journal of Business and Economic Statistics* 86, 313–26.

Hartmann, P., S. Straetmans and C. G. de Vries (2004) 'Asset Market Linkages in Crisis Periods', *Review of Economics and Statistics* 86, 313–26.

Hartmann, P., S. Straetmans and C. G. de Vries (2006) 'Banking System Stability: A Cross Atlantic Perspective', in M. Carey and R. M. Stulz (eds), *The Risk of Financial Institutions*, Chicago, IL and London: The University of Chicago Press.

Hill, B. M. (1975) 'A Simple General Approach to Inference about the Tail of A Distribution', *The Annals of Statistics* 3, 1163–73.

Kaminsky, G. I. and C. M. Reinhart (2000) 'On Crises, Contagion and Confusion', *Journal of International Economics* 51, 145–68.

King, M. and S. Wadhwani (1990) 'Transmission of Volatility between Stock Markets', *Review of Financial Studies* 3, 5–33.

Ledford, A. and J. Tawn (1996) 'Statistics for Near Independence in Multivariate Extreme Values', *Biometrika* 83, 169–87.

Lin, W. L., R. Engle and T. Ito (1994) 'Do Bulls and Bears Move across Borders? International Transmission of Stock Returns and Volatility', *Review of Financial Studies* 7, 507–38.

Longin, F. and B. Solnik (2001) 'Extreme Correlation of International Equity Markets', *Journal of Finance* 56, 649–76.

Morgenstern, D. (1956) 'Einfache Beispiele Zweidimensionaler Verteilungen', *Mitteilungsblatt Fur Mathematische Statistik* 8, 234–5.

Poon, S. H., M. Rockinger and J. Tawn (2004) 'Extreme-Value Dependence Measures in Financial Markets: Diagnostics, Models and Financial Implications', *The Review of Financial Studies* 17, 581–610.

Quandt, R. (1960) 'Test of the Hypothesis That A Linear Regression Obeys Two Separate Regimes', *Journal of the American Statistical Association* 55, 324–30.

Quintos, C., Z. Fan and P. C. B. Phillips (2001) 'Structural Change Tests in Tail Behavior and the Asian Crisis', *Review of Economic Studies* 68, 633–63.

Straetmans, S., W. Verschoor and C. Wolff (2008) 'Extreme U.S. Stock Market Fluctuations in the Wake of 9/11', *Journal of Applied Econometrics* 23, 17–42.