Comments on : Nonparametric Tail Risk, Stock Returns and the Macroeconomy

Scaillet, Olivier; Trojani, Fabio; Camponovo, Lorenzo

Comments on : Nonparametric Tail Risk, Stock Returns and the Macroeconomy

Lorenzo Camponovo, Olivier Scaillet and Fabio Trojani

University of Surrey, University of Geneva & Swiss Finance Institute

Address correspondence to Olivier Scaillet, University of Geneva & Swiss Finance Institute, Unimail, 40 Bd du Pont d'Arve, 1211 Geneva 4, Switzerland, or email: olivier.scaillet@unige.ch

1 Motivation and overview

Almeida, Ardison, Garcia and Vicente (2016) suggest to use the excess expected shortfall as a tail risk measure:

$$TR_{i,t,h} = E^Q \left[(R_{i,t,h} - z_{i,h,\alpha}) \, | R_{i,t} \leq z_{i,h,\alpha} \right], \tag{1}$$

where $R_{i,t,h}$ is the spot (or forward) return of asset i at date t for the horizon h, $z_{i,h,\alpha}$ is the α-quantile of the return distribution, and Q is a risk neutral (or forward neutral) probability. In the Basel terminology, $-z_{i,h,\alpha}$ is called the Value-at-Risk at $Q-$probability level α and for horizon h, so that losses receive a positive sign, and we can interpret the risk measure as a capital buffer. Typical h are one-day, ten-day, or one-month horizons. If log-returns are Gaussian with volatility parameter σ_i, it follows (see e.g., Scaillet (2004), Fermanian and Scaillet (2005)):

$$TR_{i,t,h} = \sigma_i \sqrt{h} \left(\frac{\varphi(z_\alpha)}{\alpha} - z_\alpha \right), \tag{2}$$

where $\varphi(z)$ is the density of a standard normal distribution at point z and z_α the $\alpha-$quantile of this distribution. As $TR_{i,t,h}$ is proportional to the integrated implied volatility in such a setting, these two measures of risk are perfectly correlated whenever log-returns are Gaussian under the pricing measure. In contrast, in presence of departures from log-normality, $TR_{i,t,h}$ incorporates information about Q which is distinct from the one generated by model-based or model-free measures of implied stock volatility like, e.g., $VIX_{i,t}$; see Schneider and Trojani (2014, 2015), among others.

Almeida et al. (2016) obtain a measure of aggregate tail risk by averaging tail risk measure (1) across a set of benchmark returns. As their estimation approach relies on the nonparametric estimation of a pricing probability consistent with the joint distribution of benchmark returns, they propose for parsimony to summarize the information in the cross-section of individual stock returns using five principal components of size and book-to-market returns. Their measure or aggregate tail risk is the average expected shortfall of the first five principal components of size and book-to-market returns:

$$TRM_{t,h} := \frac{1}{5}\sum_{k=1}^{5} TR_{pc_k,t,h}, \tag{3}$$

where $TR_{pc_k,t,h}$ is tail measure (1) for principal component return $R_{pc_k,t,h}$.

Almeida et al. (2016) apply predictive regression methods based on standard asymptotics and find that $TRM_{t,h}$ is a powerful predictor for market returns and a set of important macro variables. The tail risk proxy (3) is different from the aggregate market excess expected shortfall:

$$TR_{m,t,h} = E^Q\left[(R_{m,t,h} - z_{m,h,\alpha})\,|R_{m,t} \leq z_{m,h,\alpha}\right], \tag{4}$$

and similar measures of implied market tail risk (e.g., Bollerslev et al. 2015), as well as from model-free proxies of implied market volatility, such as $VIX_{m,t}$. Therefore, we expect an imperfect empirical comovement of $TRM_{t,h}$, $TRM_{m,t,h}$, and $VIX_{m,t}$, e.g., in presence of

stochastic return correlations or a time-varying conditional return nonnormality.[1] While we can directly extract empirical proxies for tail risk measure (1) from the prices of individual stock options, a key insight of Almeida et al. (2016) is to avoid the use of options in order to obtain a longer time series of implied tail risk measures. They attain this goal by estimating risk measure (3) nonparametrically for an horizon of $h = 1$ days, based on a monthly window of n past daily principal component returns.

Let for brevity $R_{k,s} := R_{pc_k,s-1,1}$ be the daily (forward) return of principal component $k = 1, \ldots, 5$ in day s and denote by $\{R_{k,s} : k = 1, \ldots, 5, s = t - n + 1, \ldots, t\}$ the sample of observed daily returns in the monthly window before time t. We denote by $E_n[\cdot]$ expectations under the joint empirical distribution P_n of past principal component forward returns and define the empirical forward-neutral measure $Q_n(A) := E_n[M_n \mathbf{1}_A]$ for any measurable event A, where M_n is a normalized empirical pricing kernel that prices the risk-free return $R_0 := \mathbf{1}$ and the principal component returns:

$$E_n[M_n R_k] = 1, \;\; k = 0, \ldots, 5. \tag{5}$$

Given the inherent market incompleteness, Almeida et al. (2016) select the empirical forward-neutral measure that corresponds to a particular empirical minimum power divergence pricing kernel. Precisely, they solve for $p = 1/2$ the minimization problem:

$$M_n^\star(p) := \arg\min_{M_n} E_n \left[\frac{M_n^p - 1}{p(p-1)} \right], \tag{6}$$

s.t. (5) and positivity constraints. We can motivate the choice of power parameter $p = 1/2$ (Hellinger divergence) by the convenient robustness properties of the minimum Hellinger divergence pricing kernel $M_n^\star := M_n^\star(1/2)$; see, e.g., Kitamura, Otsu and Evdokimov (2013). Using the empirical forward neutral measure $Q_n^\star(\cdot) := E_n[M_n^\star \cdot]$, they compute an estimate of

[1] See Buraschi, Trojani and Vedolin (2014) for a related theoretical evidence in a general equilibrium model with heterogenous beliefs and Schneider and Trojani (2014) for corresponding empirical evidence based on tradable variance and skew swaps.

$TRM_{t,1}$ for a one-day horizon $h = 1$ as:

$$\widehat{TRM}_{t,1} := \frac{1}{5}\sum_{k=1}^{5}\widehat{TR}_{pc_k,t,1} := \frac{1}{5}\sum_{k=1}^{5} E^{Q_n^\star}[(R_k - \hat{z}_{k,1,\alpha})|R_k \leq \hat{z}_{k,1,\alpha}]. \tag{7}$$

Empirically, this proxy has an imperfect correlation with $VIX_{m,t}$, which is an indication that it contains non redundant information. We can explain this non redundancy by several features, including time-varying correlations among stock returns, time-varying conditional higher moment in market or individual stock returns, the conceptually different construction of these risk proxies and the different implicit horizons h, as well as the different information set and estimation risk implied by the computation of $VIX_{m,t}$ and $\widehat{TRM}_{t,h}$.

Remark 1 *We can use the Almeida et al. (2016) approach to compute additional interesting model-free proxies of implied aggregate tail risk in periods where no option information is available. For instance, if P_n is the empirical distribution of daily market returns $\{R_{m,s} : s = t-n+1,\ldots,t\}$ and $M_n^\star$ the solution of problem* (6) *under the pricing constraints $E_n[M_n R_0] = E_n[M_n R_m] = 1$, we can compute an estimate of a daily $VIX_{m,t}^2$ as twice the estimated forward neutral entropy of daily market returns:*

$$\widehat{VIX}_{m,t}^2 = -2E^{Q_n^\star}[\ln R_m], \tag{8}$$

see also Schneider and Trojani (2015). Similarly, a model-free estimate of the daily implied market excess expected shortfall is obtained as:

$$\widehat{TR}_{m,t,1} = E^{Q_n^\star}\left[(R_m - \hat{z}_{m,1,\alpha})\,|R_m \leq \hat{z}_{m,1,\alpha}\right]. \tag{9}$$

2 Why robust methods for predictive regression?

Using standard predictive regressions, Almeida et al. (2016) address the predictive properties of tail measure (7) for market returns and a number of important economic variables. We revisit their findings using robust resampling tests of predictive ability, developed in Camponovo,

Scaillet and Trojani (2015).

The motivation for our robust testing approach lies in that most approaches to test predictability hypotheses are based on procedures that can heavily depend on a small fraction of influential observations in the data. For standard asymptotic $t-$tests based on OLS or similar estimators, this problem is well-known since a long time; see, e.g., Huber (1981) for a review. Recent research has also shown that a small fraction of influential observations in the data may even more easily inflate inference based on bootstrap and subsampling tests. This feature is important for testing predictability hypotheses as well, because resampling methods are natural tools for producing tests with more reliable finite-sample accuracy in predictive regression settings with correlated innovations of endogenous and predictive variables and possibly persistent predictors.

Intuitively, the non robustness of standard resampling methods arises from the too high fraction of influential data points that is often simulated by standard bootstrap and subsampling procedures, when compared to the actual fraction of outliers in the original data. As it is not possible to fully mitigate this problem simply by applying conventional bootstrap or subsampling methods to more robust estimators or test statistics,[2] Camponovo, Scaillet and Trojani (2015) develop a general robust resampling methodology for time series, which allows us to obtain more robust tests of predictability hypotheses for predictive regression settings. This approach relies on robust weighted least-squares and resampling procedures that are fully data-driven and easily manageable, based on robust versions of fast bootstrap and subsampling methods; see e.g., Goncalves and White (2004) and Hong and Scaillet (2006).

Intuitively, robust predictive regression methods are likely even more important in settings where endogenous or predictive variables can feature a complex, potentially time-varying, tail behavior and observed data can include rare influential observations. Such infuential points may arise, for instance, in settings where some of these variables may actually be obtained from point estimates of corresponding measures of tail risk. Figure 1 illustrates the time series of estimated tail risk proxy (7) in the sample period from January 1926 to December 2014,

[2]Contrary to what is often thought, resampling trimmed or winsorized estimators does not yield a robust resampling method; see, e.g., Camponovo, Scaillet and Trojani (2012) for detailed examples.

where computations are based on CRSP data. It highlights rare large observations that may reflect both a sudden change in the underlying measure of tail risk or a large variation of the estimator precision over time. Given this evidence, robust predictive regression methods seem particularly appropriate for this kind of data. We can also motivate economically our robust testing approach by the fact that ambiguous time-varying predictive relations can be consistently addressed by ambiguity averse investors only using robust estimators that bound the effects of influential data points. Wrampelmeyer, Wiehenkamp and Trojani (2015) show that different specifications of aversion to ambiguity in the literature imply robust optimal estimator choices related to robust weighted least-squares. In this sense, a robust predictive regression testing approach is consistent with the preferences of investors that dislike a time-varying ambiguity in the data-generating processes.

The data-driven weights in our robust procedure dampen, where necessary, the few data points that are estimated as influential with respect to the estimated predictive link. This feature automatically avoids arguing ex ante that, e.g., a large value of the predicted or the predictive variables is per se an anomalous observation, which is not the case in general. Indeed, large values of both the predictive and the predicted variables might obviously also be very informative about a potential predictability structure, and discarding them in an ad hoc way might bias the inference. In a truly multivariate predictive regression settings, it is even more difficult to precisely determine with an informal approach which subset of observations is potentially influential, for example by eyeballing the data. A useful property of our methodology it that it embeds a formal data-driven identification of observations that can be excessively influential for the resulting inference on predictive relations.

3 Empirical results

We revisit the predictive ability of aggregate tail risk proxy (7) for US stock market returns. Using our robust approach, we identify two most influential observations in October 1987 and November 1987, in concomitance and immediately after the Black Monday of October 19 1987. Additionally, we identify two clusters of infrequent influential data in the subperiods 1998-2000

and 2008-2010, which correspond to well-known hystorical periods of pronounced financial market turbulence and distress. Such influencial observations are reflected also in some of the particularly large values of tail proxy (7) in October-November 1987 and in the subperiods 1998-2000 and 2008-2010; see again Figure 1.

We study the predictive ability of lagged tail risk measure (7) for future monthly S&P 500 index returns, both in a single-predictor setting and in a two-predictor setting that additionally includes the dividend yield as a predictive variable.

Single-predictor model

We consider monthly S&P 500 index returns from Shiller (2000), $R_t = (P_t + d_t)/P_{t-1}$, where P_t is the end of month real stock price and d_t the real dividend paid during month t. We estimate the predictive regression model

$$\ln(R_t) = \alpha + \beta \cdot \widehat{TRM}_{t-1,1} + \epsilon_t, \quad t = 1, \ldots, T, \tag{10}$$

where $\widehat{TRM}_{t-1,1}$ is tail risk measure (7) in month $t-1$ for a one-day horizon $h = 1$, and test the null hypothesis of no predictability, $\mathcal{H}_0 : \beta_0 = 0$, where β_0 is the true value of the unknown parameter β. We collect monthly observations in the sample period 1980-2010 and estimate the predictive regression model using rolling windows of 180 monthly observations.

We first estimate the unknown parameter of interest using a least-squares estimator, and construct 90%-confidence intervals with the conventional subsampling and block bootstrap. Figure 2 reports the empirical results. Interestingly, we find that while in subperiod 1995-2005 both resampling approaches reject the null hypothesis of no predictability, in subperiod 2005-2010 the testing procedures do not detect predictability structures.[3] In a second step, we test the null hypothesis of no predictability $\mathcal{H}_0 : \beta_0 = 0$ using our robust fast resampling tests. We estimate the unknown parameter of interest using the robust Huber estimator instead of the least-squares estimator and construct 90%-confidence intervals with the robust fast subsampling

[3]Similar findings arise when computing confidence intervals with standard asymptotic theory, as in Almeida et al. (2016).

and robust fast bootstrap proposed in Camponovo, Scaillet and Trojani (2015). Figure 3 reports the empirical results. In this case, we always reject the null hypothesis of no predictability for the whole period under investigation.[4]

It is interesting to study to which extent influential observations might have caused the diverging conclusions of robust and conventional tests. We exploit the properties of our robust testing methods to identify such data points. Figure 4 plots the time series of Huber weights estimated by the robust Huber estimator. We find that subperiod 1998-2002 is characterized by a cluster of infrequent anomalous observations, which are likely related to the abnormal stock market performance during the NASDAQ bubble in the second half of the 1990s. Similarly, we find a second cluster of anomalous observations in subperiod 2008-2010, which is linked to the extraordinary events of the recent financial crisis. Finally, the most influential observation is November 1987, following the Black Monday on October 19 1987. Importantly, the total fraction of such influential observations is small and less than 3.3%.

To further illustrate the consequences of influential observations, we report in Figure 5 a scatter plot of tail risk measures (7) and returns. We find that the most influential observations correspond to tail risk proxies larger than 7.5 in October and November 1987. Since our testing results above are based on moving windows of 180 monthly data points, these influential observations have no impact on the confidence intervals computed in subperiod 2003-2010. This feature my explain the rejection (non-rejection) of the null hypothesis in subperiod 1995-2005 (2005-2010) using conventional tests.

[4]Somehow surprisingly, we obtain a weaker predictive evidence, both with conventional and robust methods, using the tail risk measure introduced in Kelly and Jiang (2014).

Two-predictor model

In this section, we study the joint predictive ability of tail risk measure (7) and the dividend yield for future monthly S&P 500 index returns, using the two-predictor regression model:[5]

$$\ln(R_t) = \alpha + \beta_1 \cdot \widehat{TRM}_{t-1,1} + \beta_2 \cdot \ln\left(\frac{D_{t-1}}{P_{t-1}}\right) + \epsilon_t,\ t = 1, \ldots, T. \tag{11}$$

Let β_{01} and β_{02} denote the true values of parameters β_1 and β_2, respectively. Using conventional and robust bootstrap and subsampling tests, we first test the null hypothesis of no return predictability by tail risk measure (7), $\mathcal{H}_{01} : \beta_{01} = 0$. Figure 6 reports the 90%-confidence intervals for parameter β_1, based on rolling windows of 180 monthly observations in sample period 1980-2010. We find again that the robust tests always clearly reject the null of no predictability. In contrast, the conventional bootstrap and subsampling tests do not detect predictability structures for the subperiod 2003-2010. We also test the hypothesis of no predictability by the dividend yield, $\mathcal{H}_{02} : \beta_{02} = 0$. Figure 7 reports the resulting confidence intervals for parameter β_{02}. Also in this case and in line with the empirical evidence in Camponovo, Scaillet and Trojani (2015), the robust procedures always reject the hypothesis of no predictability. In contrast, the conventional bootstrap and subsampling tests produce a weaker and more ambiguous predictability evidence. By inspecting the Huber weights in Figure 8, implied by the robust estimation of the predictive regression model (11), we find again a cluster of infrequent anomalous observations, during the Black Monday on October 1987, the NASDAQ bubble, and the recent financial crisis.

Time-varying predictability?

The evidence of influential observations in the previous section might suggest a broader misspecification of predictive relations for market returns, which might be captured by time-varying

[5]Consistent with the literature, the annualized dividend series D_t is defined as

$$D_t = d_t + (1 + r_t)d_{t-1} + (1 + r_t)(1 + r_{t-1})d_{t-2} + \cdots + (1 + r_t)\ldots(1 + r_{t-10})d_{t-11},$$

where r_t is the one-month maturity Treasury-bill rate.

parameters. We test for the presence of time-varying parameters in predictive regression model (11), using the standard Wald statistic for endogenous breaks in Andrews (1993) and its robust version introduced in Gagliardini, Trojani and Urga (2005). Using both statistics, we never reject the null hypothesis of no structural break at the 10% significance level in our sample period. Therefore, we cannot explain the lack of predictability produced by classical tests in some cases by a structural break in a significant subset of the data. We conclude that the presence of a small fraction of influential observations is a plausible explanation for the diverging conclusions of standard and robust predictive regression methods.

Out-of-sample predictability?

We close our analysis, by quantifying the out-of-sample predictive accuracy of the predictive regression model (10) estimated by our robust approach. We follow Goyal and Welsh (2003) and Campbell and Thompson (2008), and consider the out-of-sample $R^2_{OS,ROB}$ statistic:

$$R^2_{OS,ROB} = 1 - \frac{\sum_{t=t_1+1}^{t_2}(y_t - \hat{y}_{t,ROB})^2}{\sum_{t=t_1+1}^{t_2}(y_t - \bar{y}_t)^2}, \tag{12}$$

where $\hat{y}_{t,ROB}$ is the fitted value from a predictive regression estimated with data up to time t for the out-of-sample forecast period $t+1$, using the robust Huber estimator, $\bar{y}_t$ is the historical average return estimated through period up to time t, $t_1 = 1980$, and $t_2 = 2010$. Whenever statistic $R^2_{OS,ROB}$ is positive, the robust estimation of predictive regression model (10) provides more accurate out-of-sample predictions than simple forecasts based on the sample mean of market returns. For the period under investigation 1980-2010, we obtain $R^2_{OS,ROB} = 0.90\%$. Similar empirical findings also arise by estimating the predictive regression model using the nonrobust least-squares estimator. Therefore, in our data, nonrobust and robust methods provide more accurate out-of-sample predictions than simple forecast based on the sample mean of market returns.

References

[1] Andrews, D., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica, 61, 821-856.

[2] Almeida, C., Ardison, K., Garcia, R., and J. Vicente, 2016. Nonparametric tail risk, stock returns and the macroeconomy, forthcoming Journal of Financial Econometrics.

[3] Buraschi, A.,Trojani, F., and A. Vedolin, 2014. When uncertainty blows in the orchard: comovement and equilibrium volatility risk premia. Journal of Finance, Volume 69, Issue 1, February, 101-137.

[4] Bollerslev, T., Todorov, V., and L. Xu, 2015. Tail risk premia and return predictability. Journal of Financial Economics, 118, 1, 113-134.

[5] Campbell, J.Y., and S.B. Thompson, 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Review of Financial Studies, 21, 1509-1531.

[6] Camponovo, L., Scaillet, O., and F. Trojani, 2012. Robust subsampling. Journal of Econometrics, 167, 197-210.

[7] Camponovo, L., Scaillet, O., and F. Trojani, 2015. Predictability hidden by anomalous observations. Swiss Finance Institute working paper.

[8] Fermanian, J.D., and O. Scaillet, 2005. Sensitivity Analysis of VaR and Expected Shortfall for Portfolios under Netting Agreements. Journal of Banking and Finance, 29, 927-958.

[9] Gagliardini, P., Trojani, F., and G. Urga, 2005. Robust GMM tests for structural breaks. Journal of Econometrics, 129, 139-182.

[10] Goncalves, S., and H. White, 2004. Maximum likelihood and the bootstrap for nonlinear dynamic models. Journal of Econometrics, 119, 199-220.

[11] Goyal, A., and I. Welch, 2003. Predicting the equity premium with dividend ratios. Management Science, 49, 639-654.

[12] Hong, H., and O. Scaillet, 2006. A fast subsampling method for nonlinear dynamic models. Journal of Econometrics, 133, 557-578.

[13] Huber, P. J., 1981. *Robust Statistics*. Wiley, New York.

[14] Huber, P.J., 1984. Finite sample breakdown point of M- and P-estimators. Annals of Statistics, 12, 119-126.

[15] Kelly, B., and H. Jiang, 2014. Tail risk and asset prices. Review of Financial Studies, 27, 2841-2871.

[16] Kitamura, Y., Otsu, T., and K. Evdokimov, 2013. Robustness, infinitesimal neighborhoods, and moment restrictions. Econometrica, Volume 81, 3, 1185-1201.

[17] Scaillet, O., 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance, 14, 115-129.

[18] Schneider, P. and F. Trojani, 2014. Fear trading. Swiss Finance Institute working paper.

[19] Schneider, P. and F. Trojani, 2015. Divergence and the price of uncertainty. Swiss Finance Institute working paper.

[20] Shiller, R.J., 2000. *Irrational Exuberance*. Princeton University Press, Princeton, NJ.

[21] Wrampelmeyer, J., Wiehenkamp, C., and F. Trojani, 2015. Ambiguity and reality, Swiss Finance Institute working paper.

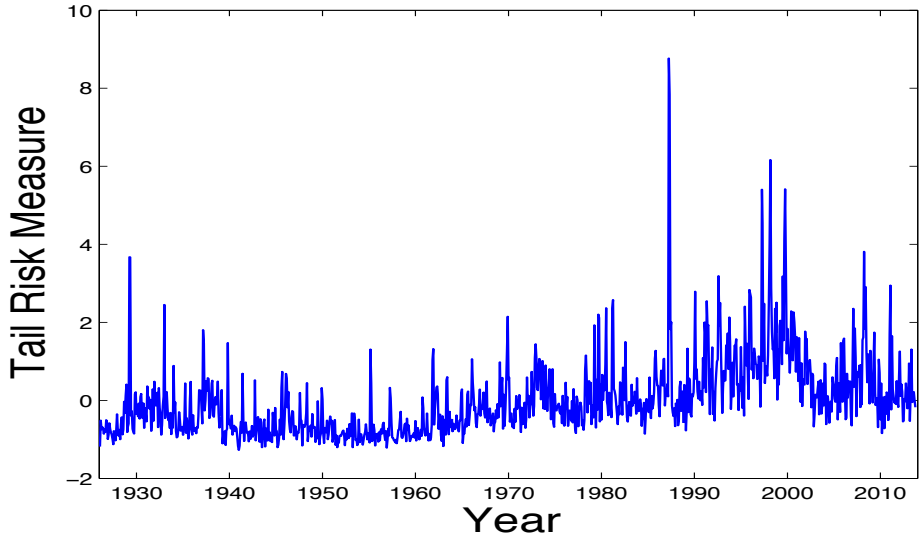

Figure 1: **Tail risk measure.** We plot the tail risk measure (7) for the period from 1926 to 2014.

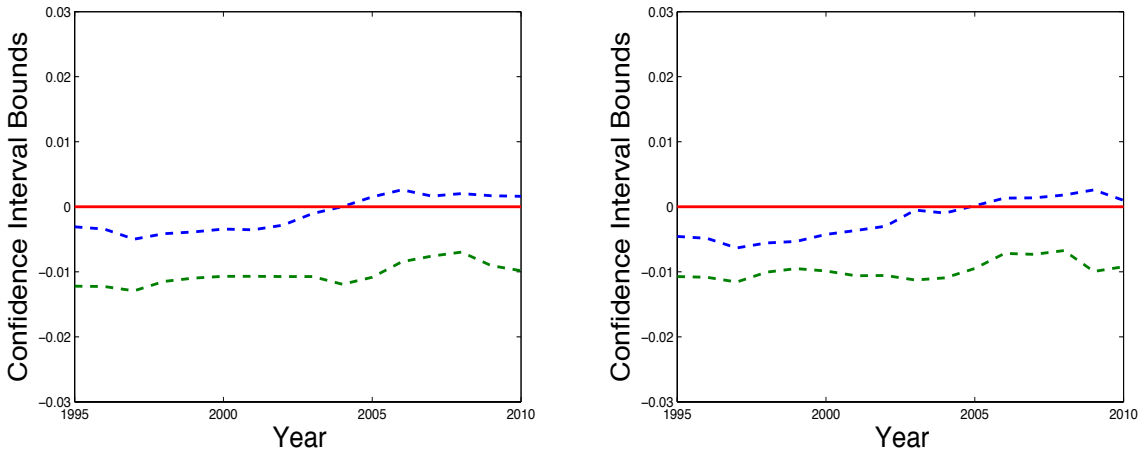

Figure 2: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_0 in the predictive regression model (10). We consider rolling windows of 180 observations for the period 1980-2010. We present the conventional subsampling (left panel) and block bootstrap (right panel).

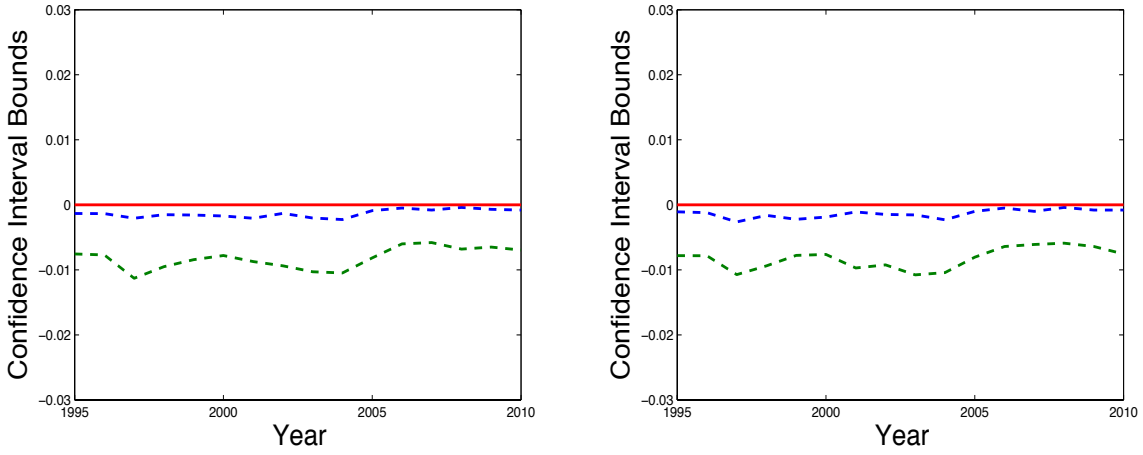

Figure 3: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_0 in the predictive regression model (10). We consider rolling windows of 180 observations for the period 1980-2010. We present the robust fast subsampling (left panel) and robust fast bootstrap (right panel).

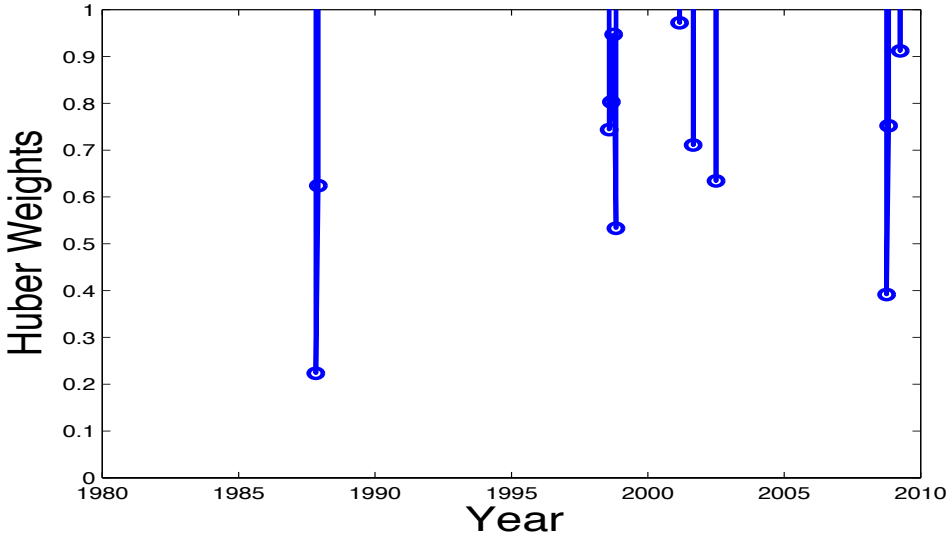

Figure 4: **Huber weights under the predictive regression model (10).** We plot the Huber weights for the predictive regression model (10) in the period 1980-2010.

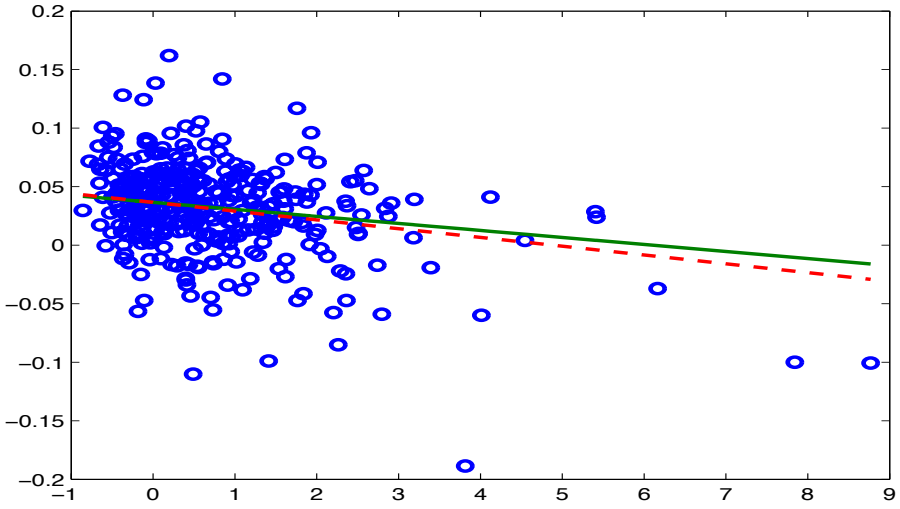

Figure 5: **Scatter Plot.** On the x-axis and y-axis are represented the tail risk measure and returns, respectively. The solid line is the robust linear regression computed with the Huber estimator, while the dashed line is the conventional linear regression computed with the least-squares estimator.

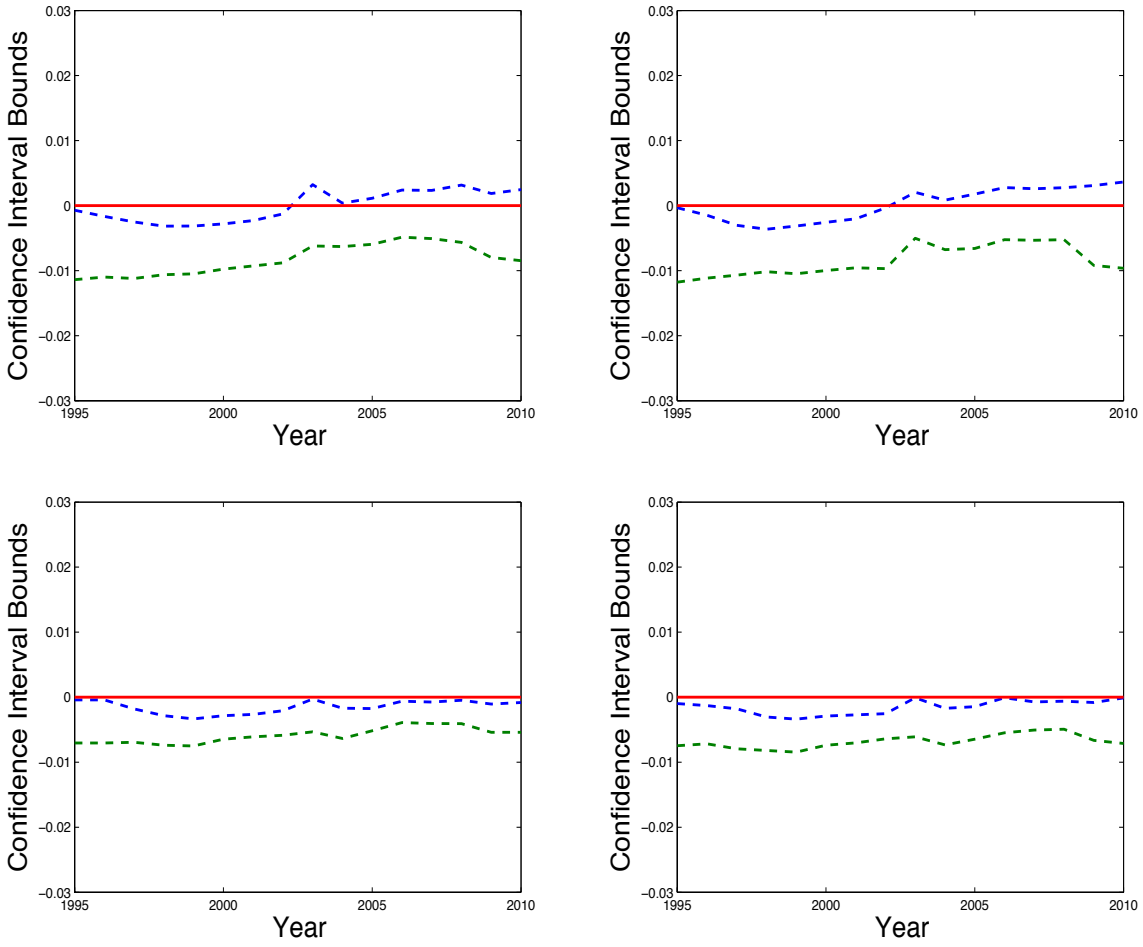

Figure 6: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_1 in the predictive regression model (11). We consider rolling windows of 180 observations for the period 1980-2010. In the top line, we present the conventional subsampling (left panel) and block bootstrap (right panel), while in the bottom line we consider the robust fast subsampling (left panel) and robust fast bootstrap (right panel).

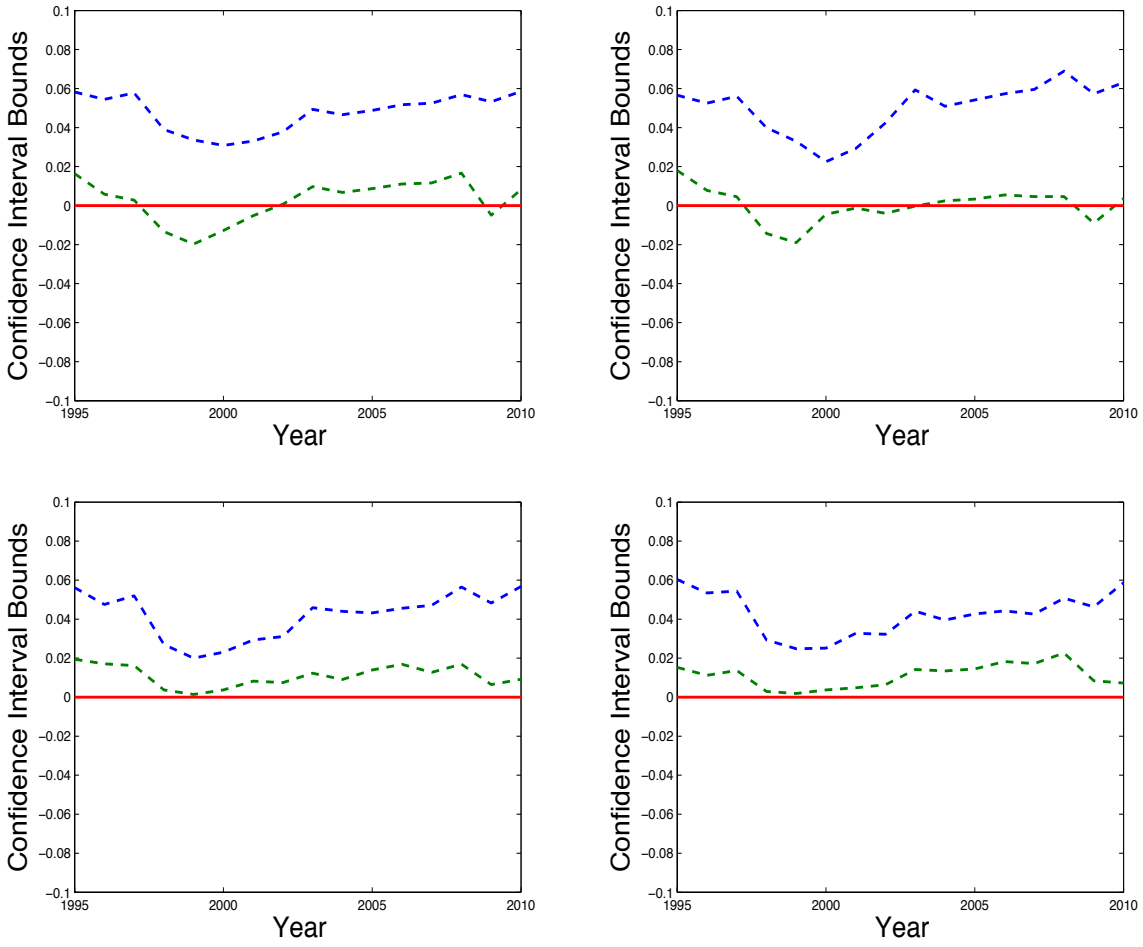

Figure 7: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_2 in the predictive regression model (11). We consider rolling windows of 180 observations for the period 1980-2010. In the top line, we present the conventional subsampling (left panel) and block bootstrap (right panel), while in the bottom line we consider the robust fast subsampling (left panel) and robust fast bootstrap (right panel).

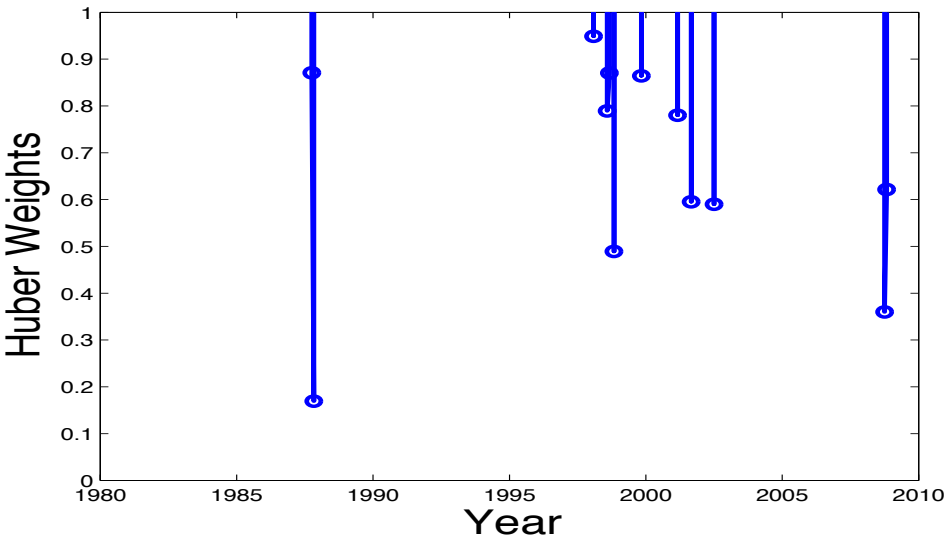

Figure 8: **Huber weights under the predictive regression model (11).** We plot the Huber weights for the predictive regression model (11) in the period 1980-2010.