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

Lorenzo Camponovo, Olivier Scaillet and Fabio Trojani

University of Essex, 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

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 $R_{i,t}$ is the 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 the risk neutral probability. In the Basel terminology, $-z_{i,h,\alpha}$ is called the Value-at-Risk at 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, 10-day, or one-month horizons. If the returns are i.i.d. Gaussian with volatility parameter σ_i under Q as in the Black-Scholes model, we can easily compute the tail risk measure (see e.g. Scaillet (2004)):

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

where $\varphi(z)$ denotes the probability density function of a standard normal distribution at point z, and z_α is the quantile of a standard normal distribution at probability level α. In a Black-Scholes world, the tail risk measure is simply the square root of the integrated Black-Scholes implied variance multiplied by the constant $\left(\frac{\varphi(z_\alpha)}{\alpha} - z_\alpha\right)$, and we get a perfect correlation between the two risk measures. Almeida et al. (2016) suggest to average the tail risk measures (1) obtained from single stocks to get an aggregate tail risk measure. Then, even in a Gaussian world, we do not get the same information as using the implied volatility of an index underlying the VIX construction for example. For i.i.d. jointly normal returns, we expect the aggregate measure to be higher than the measure directly computed from the corresponding portfolio or index. Indeed the volatility of an index is lower than the weighted sum of the volatility of its constituents due to the triangle inequality. The non-perfect correlation of 0.56% between the aggregate tail risk measure and the VIX might be explained by correlation effects and/or higher moment effects. By construction, we do not get a redundant information between the VIX and the aggregate tail risk measure. The question that Almeida et al. (2016) address then is the following: is that new tail measure useful for return predictability? They find using standard asymptotic methods in predictive regressions that the answer is positive. We revisit their claim by looking at robust predictive regressions.

Why going robust in predictive regressions?

A common feature of most approaches to test predictability hypotheses is their reliance on procedures that can be heavily influenced by a small fraction of anomalous observations in the data. For standard OLS estimators and t-test statistics, this problem is well-known since a long time; see, e.g., Huber (1981) for a review. More recent research has also shown that inference provided by bootstrap and subsampling tests may be easily inflated by a small fraction of anomalous observations. Intuitively, we explain this feature by the too high fraction of anomalous observations that is often simulated by standard bootstrap and subsampling procedures, when compared to the actual fraction of outliers in the original data. It is not possible to mitigate this problem simply by applying conventional bootstrap or subsampling methods to more robust estimators or test statistics. Resampling trimmed or winsorized estimators does not yield a robust resampling method; see, e.g., Camponovo, Scaillet and Trojani (2012) for detailed examples. Camponovo, Scaillet and Trojani (2015) have developed a new robust bootstrap and subsampling methodology for time series, which allows us to develop more robust tests of predictability hypotheses in predictive regression settings. Their robust predictive regression approach relies on robust weighted least-squares procedures that are fully data-driven and easily manageable. The approach relies on robust versions of the fast bootstrap and fast subsampling; see e.g., Goncalves and White (2004) and Hong and Scaillet (2006). This is motivated economically by the fact that time-varying ambiguity about predictive relations is consistently addressed by ambiguity averse investors only using robust estimators that bound the effects of anomalous data features. 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 for returns. The data-driven weights in the procedure dampen, where necessary, the few data points that are estimated as anomalous with respect to the postulated predictive link. This feature automatically avoids, e.g., arguing ex ante that a large value of the predicted or the predictive variables is per se an anomalous observation, which is not in general the case. Indeed, observations linked to large values of both the predictive and the predicted variables might be very informative about a potential predictability structure and discarding them in an ad hoc way might bias the inference. In a multivariate predictive regression setting, it is even more difficult to determine with an informal approach which subset of observations is potentially anomalous, 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.

Empirical results

In this empirical analysis, we study the forecast ability of the aggregate tail risk measure introduced in Almeida et al. (2016) for US stock market. Figure 1 reports the tail risk measure for the period from 1926 to 2014 calculated using CRSP data. The most influential observation is

November 1987, following the Black Monday on October 19 1987. Furthermore, the subperiods 1998-2000 and 2008-2010 are characterized by a cluster of infrequent anomalous observations. This latter subperiod is obviously linked to the extraordinary events of the recent financial crisis. The empirical study is articulated in two parts. First, we study the forecast ability of lagged tail risk measure for explaining monthly S&P 500 index returns, in a predictive regression model with a single predictor. Finally, we also study the predictive power of the tail risk measure and dividend yield in a two-predictor model.

Single-Predictor Model

In our predictability analysis, 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

$$\ln(R_t) = \alpha + \beta \cdot TRM_{t-1} + \epsilon_t, \quad t = 1, \ldots, n, \tag{2}$$

where TRM_{t-1} denote the tail risk measure at time $t-1$, and test the null hypothesis of no predictability, $\mathcal{H}_0 : \beta_0 = 0$, where β_0 denote 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 observations. More precisely, we estimate the unknown parameter of interest using the least squares estimator, and construct 90%-confidence intervals with the conventional subsampling and block bootstrap. Figure 2 reports the empirical results. It is interesting to highlight that for the subperiod 1995-2005 both resampling approaches reject the null hypothesis of no predictability. However, for the last subperiod 2005-2010, the testing procedures do not detect predictability structures.[1]

As a robustness check, we also test the null hypothesis of no predictability $\mathcal{H}_0 : \beta_0 = 0$ using the robust fast resampling procedures proposed in Camponovo, Scaillet and Trojani (2015). More precisely, first we estimate the unknown parameter of interest using the robust

[1]Similar empirical findings also arise by computing confidence intervals with standard asymptotic theory as chosen by Almeida et al. (2016) instead of subsampling or block bootstrap methods.

Huber estimator instead of the least squares estimator. Finally, we construct 90%-confidence intervals with the robust fast subsampling and robust fast bootstrap proposed in Camponovo, Scaillet and Trojani (2015). Figure 3 reports the empirical results. In this case, when we use the robust fast resampling procedures, we always reject the null hypothesis of no predictability for the whole period under investigation.[2]

It is interesting to study to which extent anomalous observations in sample period might have caused the diverging conclusions of robust and conventional resampling methods. We exploit the properties of the robust testing methods proposed in Camponovo, Scaillet and Trojani (2015) 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. Overall, anomalous observations are less than 3.3% of the whole data sample. To further investigate the features of these anomalous observations, we present the scatter plot of tail risk measure and returnsin Figure 5. The most influential observations with tail risk measure larger than 7.5 are October and November 1987. Since we are using moving windows of size 180 in our predictability analysis, these observations are excluded for the construction of confidence intervals for the subperiod 2003-2010. The absence of these influential observations may explain the non-rejection of the null hypothesis for the subperiods 2005-2010 using conventional procedures.

Two-Predictor Model

In this section, we extend our empirical study to two-predictor regression models. More precisely, we study the joint predictive ability of tail risk measure and dividend yield and for future

[2]Empirical finding based on the tail risk measure computed using the first five principal components of 25 Fama and French size and book-to-market portfolios are very similar. On the other hand, surprisingly, the predictive power of the tail risk measure introduced in Kelly and Jiang (2014) is quite weak using both conventional and robust methods.

monthly S&P 500 index returns.

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. Finally, we focus on the two-predictor regression model

$$\ln(R_t) = \alpha + \beta_1 \cdot TRM_{t-1} + \beta_2 \cdot \ln\left(\frac{D_{t-1}}{P_{t-1}}\right) + \epsilon_t, \ ; \ t = 1, \ldots, n. \tag{3}$$

Let β_{01} and β_{02} denote the true values of parameters β_1 and β_2, respectively. Using the conventional bootstrap and subsampling tests, as well as our robust bootstrap and subsampling tests, we first test the null hypothesis of no return predictability by tail risk measure, $\mathcal{H}_{01} : \beta_{01} = 0$. Figure 6 plots 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 dividend yield, $\mathcal{H}_{02} : \beta_{02} = 0$. Figure 7 plots the resulting confidence intervals for parameter β_{02}, using the conventional bootstrap and subsampling tests, as well as our robust bootstrap and subsampling tests. Also in this case, in line with the empirical results provided 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 (3), we find again a cluster of infrequent anomalous observations, during the Black Monday on October 1987, the NASDAQ bubble and the recent financial crisis.

To investigate the potential presence of time-varying parameters in the predictive regression model (3), we test formally for the presence of structural breaks. We apply both the standard Wald test statistic proposed 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 the lack of predictability produced in some cases by the standard approach by a structural break in a significant subset of the data. This evidence supports the presence of a small subset of influential anomalous observations as a plausible explanation for the diverging conclusions of standard and robust predictive regression methods.

Out-of-Sample Analysis

Finally, we study the out-of-sample accuracy of the predictive regression model (2) estimated by the robust Huber estimator. Borrowing from Goyal and Welsh (2003) and Campbell and Thompson (2008), we introduce the out-of-sample $R^2_{OS,ROB}$ statistics, defined as

$$R^2_{OS,ROB} = 1 - \frac{\sum_{t=t_1+1}^{t_2} (y_t - \hat{y}_{t,ROB})^2}{\sum_{t=t_1+1}^{t_2} (y_t - \bar{y}_t)^2}, \tag{4}$$

where $\hat{y}_{t,ROB}$ are the fitted values from a predictive regression estimated up to period t_1 for the out-of-sample forecast periods $t_1 + 1, \ldots, t_2$, using the robust Huber estimator, and $\bar{y}_t$ is the historical average return estimated through period t_1. Whenever statistic $R^2_{OS,ROB}$ is positive, the robust estimation of the predictive regression model (2) provides more accurate out-of-sample predictions than simple forecast 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, both 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] 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.

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

[5] Camponovo, L., Scaillet, O., and F. Trojani, 2015. Predictability hidden by anomalous observations, working paper.

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

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

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

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

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

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

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

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

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

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

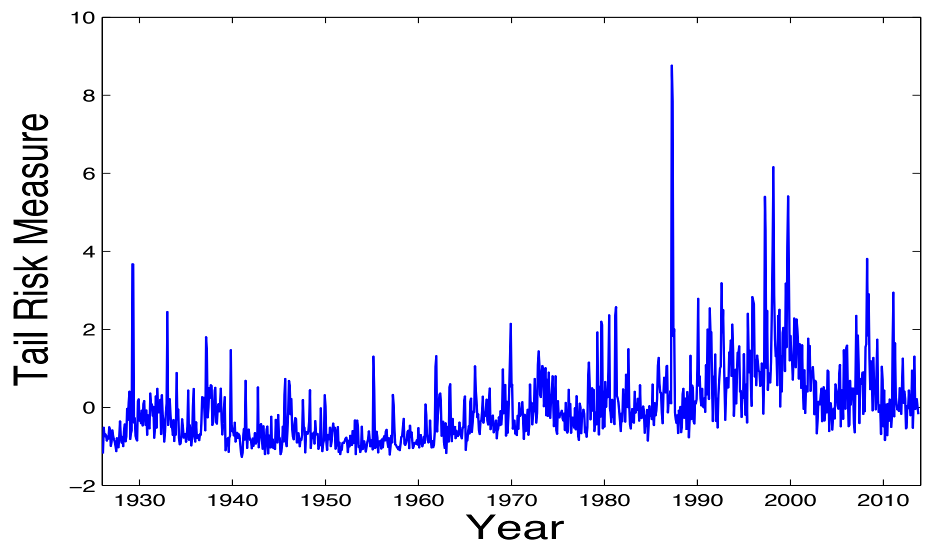

Figure 1: **Tail risk measure.** We plot the tail risk measure for the period from 1926 to 2014.

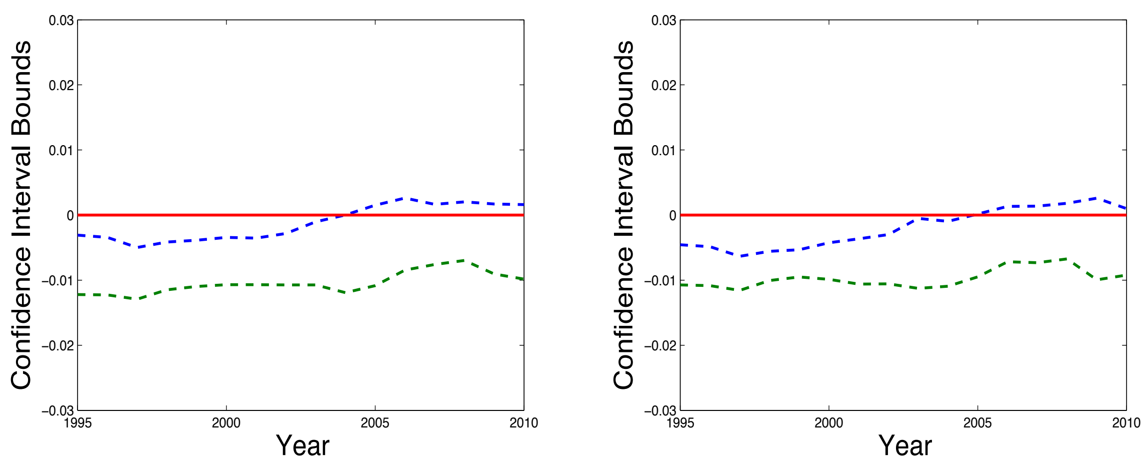

Figure 2: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_0 in the predictive regression model (2). We consider rolling windows of 180 observations for the period 1980-2010. We present the conventional subsampling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

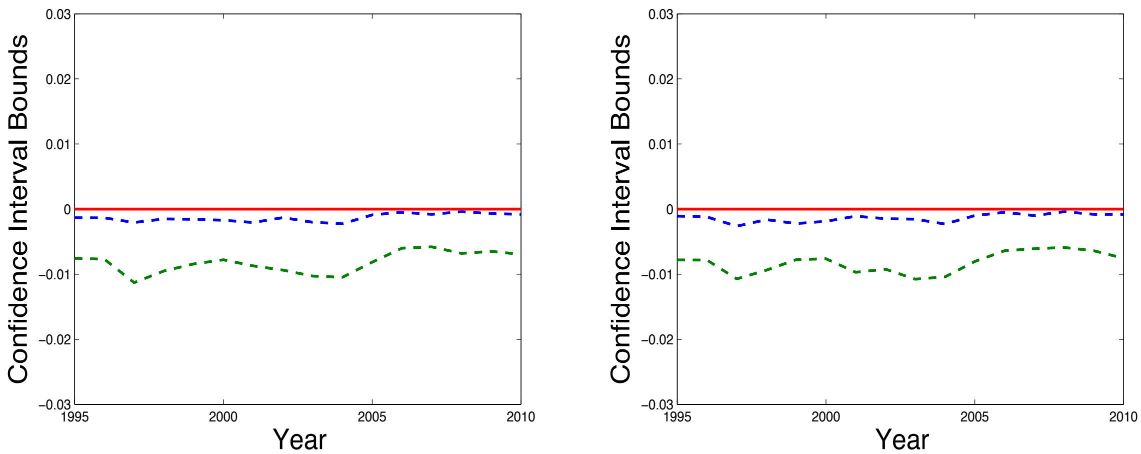

Figure 3: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_0 in the predictive regression model (2). We consider rolling windows of 180 observations for the period 1980-2010. We present the robust fast subsampling (left panel) and robust fast bootstrap (right panel).

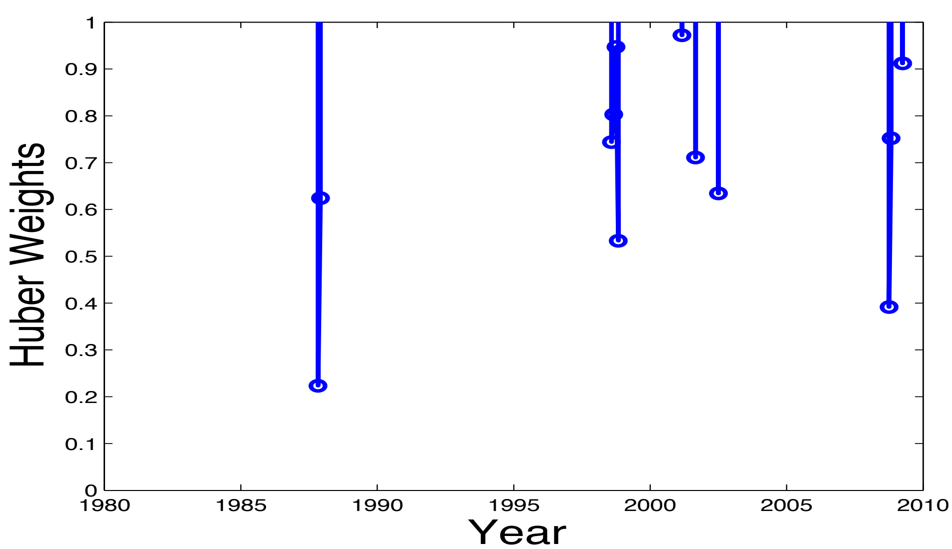

Figure 4: **Huber weights under the predictive regression model (2).** We plot the Huber weights for the predictive regression model (2) in the period 1980-2010.

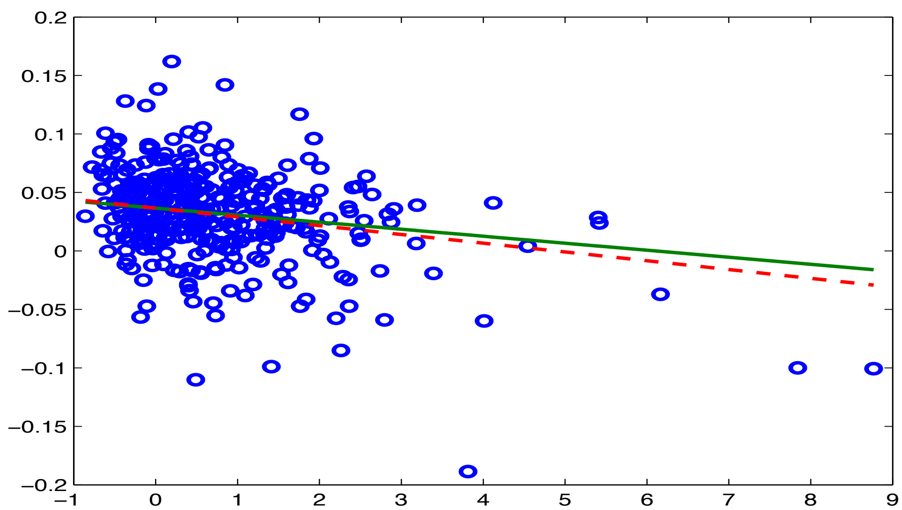

Figure 5: **Scatter Plot.** On the x-axis and y-axis are represented the tail risk measure and returns, respectively. The solid line is the robust linear regression computed with the Huber estimator, while the dashed line is the conventional linear regression computed with the least squares estimator.

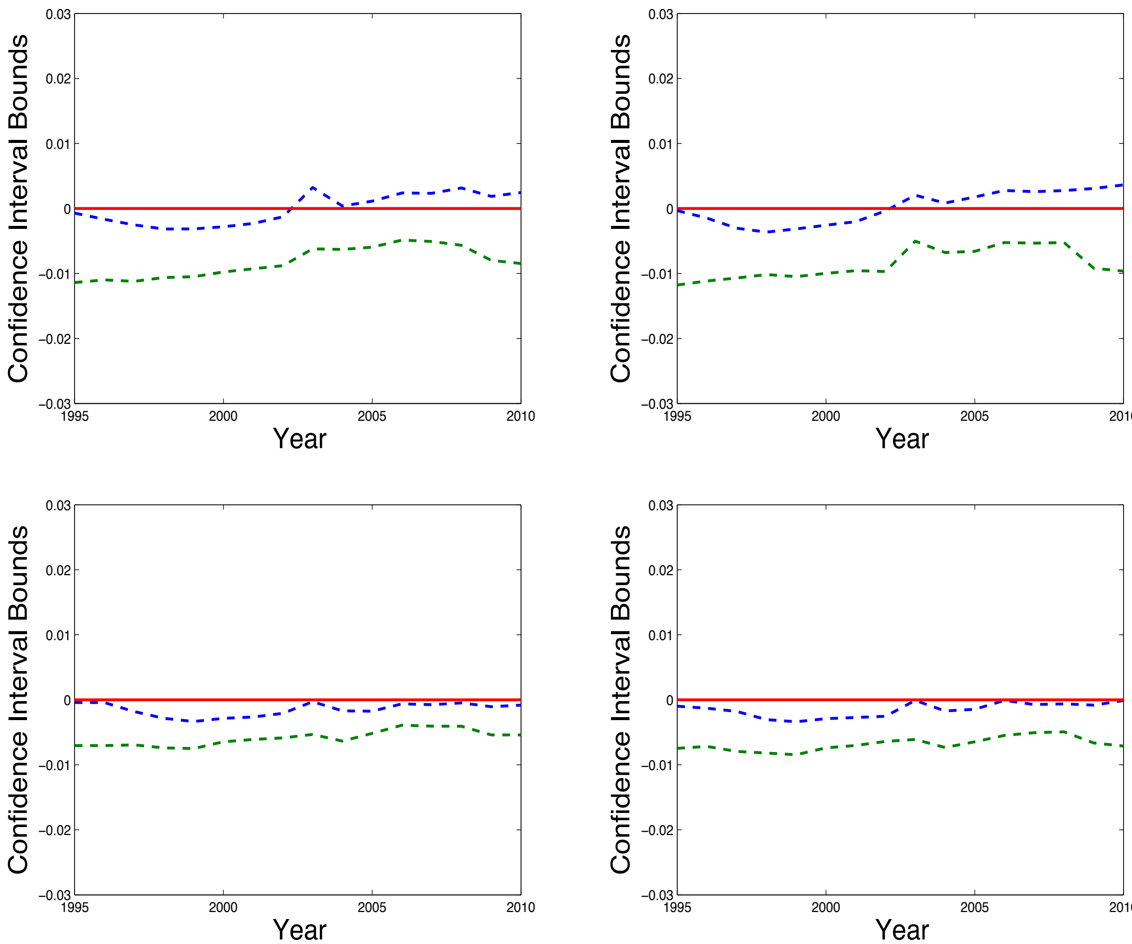

Figure 6: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_1 in the predictive regression model (3). We consider rolling windows of 180 observations for the period 1980-2010. In the top line, we present the conventional 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).

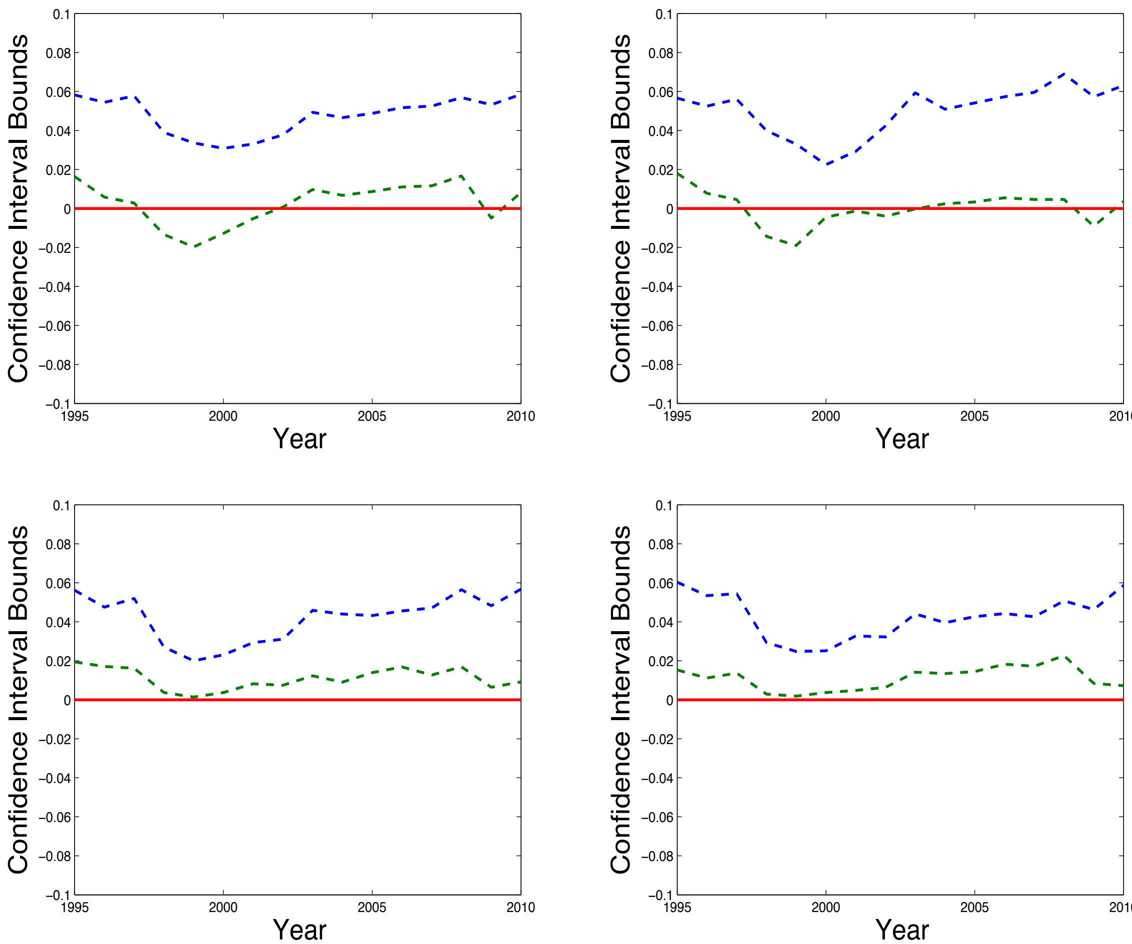

Figure 7: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_2 in the predictive regression model (3). We consider rolling windows of 180 observations for the period 1980-2010. In the top line, we present the conventional 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).

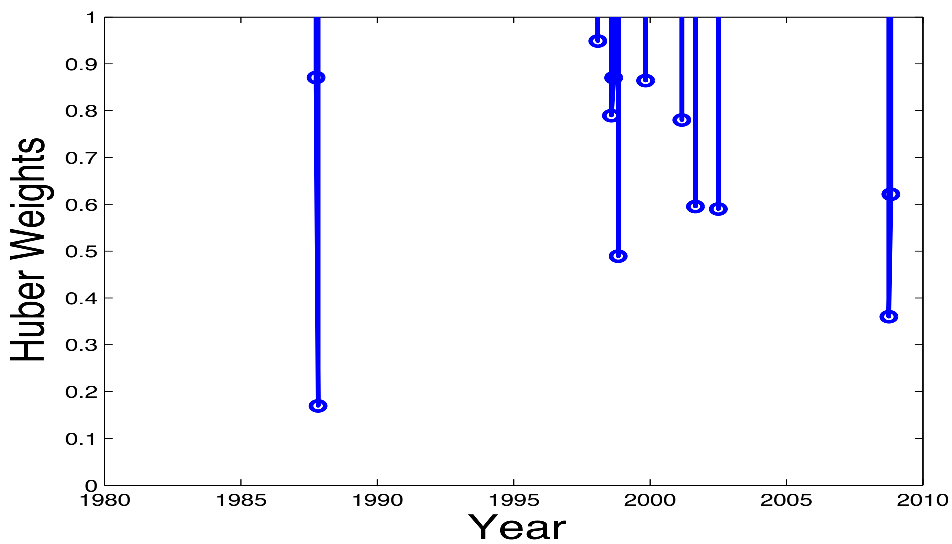

Figure 8: **Huber weights under the predictive regression model (3).** We plot the Huber weights for the predictive regression model (3) in the period 1980-2010.