

# Predictability Hidden by Anomalous Observations\*

Lorenzo Camponovo

*University of Lugano and University of St.Gallen*

Olivier Scaillet

*University of Geneva and Swiss Finance Institute*

Fabio Trojani

*University of Lugano and Swiss Finance Institute*

First Version: November 2007; This Version: March 15, 2013

---

\*We acknowledge the financial support of the Swiss National Science Foundation (NCCR FINRISK, grants 101312-103781/1, 100012-105745/1, and PDFM1-114533). We thank participants at the International Conference on Computational Management Science 2009 in Geneva, the International Conference on Robust Statistics 2009 in Parma, the International Conference on Computational and Financial Econometrics 2009 in Limassol, the Prospectus Workshop in Econometrics at Yale University 2010, the seminar at the University of Ghent 2010, the Conference on Resampling Methods and High Dimensional Data at Texas A&M University 2010, ECARES seminar at ULB 2010 in Bruxelles, the seminar at the Universities of Orleans 2012, the seminar at the University of Frankfurt 2012, the seminar at the University Bocconi 2012, the seminar at the University of St.Gallen 2012, the European Summer Symposium in Financial Markets 2012 in Gerzensee, the Econometric Society European Meeting 2012 in Malaga, the seminars at Northwestern University, Imperial College Business School and the Oxford Mann Institute 2012, and the SoFie Conference 2013 in Singapore for helpful comments. Part of this research was performed when the first author was visiting Yale University as a post doctoral research fellow, and the second author University of Cergy-Pontoise. Correspondence address: Fabio Trojani, Faculty of Economics, University of Lugano, Via Buffi 13, CH-6900 Lugano, e-mail: Fabio.Trojani@usi.ch.

## **Abstract**

Predictive relations motivated by financial theory should hold at least for the majority of the data, but they can be violated by a small fraction of the data. Conventional testing approaches can fail to detect predictive relations that hold for the vast majority of the data, because rare anomalous observations can have an excessive influence on the results. We propose a novel robust method for testing the null of no predictability, using predictive regression models that might be violated by a minority of anomalous observations. The Monte Carlo evidence demonstrates large improvements of our approach, while the empirical analysis produces a sharp evidence of market return predictability – using predictive variables like the dividend yield, the volatility risk premium, or labor income – which is valid for the vast majority of the data.

**Keywords:** Predictive Regression, Stock Return Predictability, Bootstrap, Subsampling, Robustness.

**JEL:** C12, C13, G1.

# 1 Introduction

Predictive relations implied by financial theories, such as, e.g., a simple present-value logic, are approximate descriptions of the reality that hopefully hold for the majority of the data, but which can be realistically violated by a small subset of the data. This paper proposes a novel robust approach for testing predictive relations that might be violated by a minority of anomalous observations for the postulated model. By design, our method aims at uncovering predictive relations satisfied by a sufficiently vast fraction of the data. At the same time, it is constructed to optimally filter the adverse effects that anomalous deviations from a predictive regression model can have on the results.

We develop a new family of robust tests of predictive relations based on a novel robust Monte Carlo resampling methodology for time series. In addition to controlling the effects of anomalous observations for the estimated theoretical relation, our approach simultaneously identifies observations that may produce a weak performance of conventional testing procedures, thus ensuring a robust inference with respect to local departures from the assumed predictive regression model. While it is not the purpose of our approach to try to explain the behavior of such anomalous observations, our method allows to isolate them reliably for further analysis, e.g., for cases where they might contain useful information for an improved re-specification of a predictive regression model.

We first find by Monte Carlo simulation that conventional hypothesis testing methods for predictive regressions are dramatically non-resistant to even small fractions of anomalous observations in the data. In particular, even though the test probability of rejecting a null by chance alone features some degree of resistance to anomalous data in our Monte Carlo experiments, the test ability to detect predictive relations satisfied by the vast majority of the data can be surprisingly low.

Second, we quantify theoretically the robustness properties of general resampling tests of predictability in a time series context, borrowing from the concept of breakdown point, which is a measure of the degree of resistance of a testing procedure to outliers; see, e.g., Hampel (1971), Donoho and Huber (1983), and Hampel, Ronchetti, Rousseeuw and Stahel (1986). Our

theoretical results confirm the dramatic non-robustness of conventional tests of predictability detected in our Monte Carlo findings.

Third, using a robust bootstrap and subsampling approach, we develop a novel class of robust resampling tests for times series, which are resistant to anomalous observations and are applicable in linear and nonlinear model settings at sustainable computational costs.<sup>1</sup> We use our robust resampling theory to propose a novel class of robust tests of predictability and confirm by Monte Carlo simulation that they preserve desirable finite sample properties, by filtering out the adverse effects of anomalous observations on the estimated economic relation.

Finally, we revisit the recent empirical evidence on stock return predictability for US stock market data. We study single-predictor and multi-predictor models, using several well-known predictive variables in the literature, such as the lagged dividend yield, the difference between option-implied volatility and realized volatility (Bollerslev, Tauchen and Zhou, 2009), the interest rate, and the share of labor income to consumption (Santos and Veronesi, 2006). We obtain the following novel empirical evidence.

First, we find that the dividend yield is a robust predictive variable of market returns, which is significant at the 5% significance in all our regressions, for each subperiod, sampling frequency and forecasting horizon considered. In univariate regressions with monthly data, the lagged dividend yield is significant at the 5% level according to the robust tests, in each window of 180 monthly observations from January 1980 to December 2010. In contrast, recent tests based on bias-corrected methods, near-to-unity asymptotics and standard resampling procedures produce a weaker and more ambiguous evidence overall, e.g., by not rejecting the null of no predictability at the 10% significance level in the subperiod from January 1994 to December 2009 (bias-corrected methods and near-to-unity asymptotics) or in the subperiod from January 1995 to December 2010 (resampling tests).

Multi-predictor regressions including variance risk premium and labor income proxies con-

---

<sup>1</sup>Our robust resampling approach relies on the fast resampling idea put forward, among others, in Shao and Tu (1995), Davidson and McKinnon (1999), Hu and Kalbfleisch (2000), Andrews (2002), Salibian-Barrera and Zamar (2002), Goncalves and White (2004), Hong and Scaillet (2006), Salibian-Barrera, Van Aelst and Willems (2006, 2007), and Camponovo, Scaillet and Trojani (2012). The methodology is applicable to a wide set of bootstrap and subsampling simulation schemes in the literature.

firm the significant predictive power of the dividend yield. While the dividend yield is again significant at the 5% level in all cases using the robust tests, it is significant only at the 10% level using the conventional tests in the sample period 1994-2000, within monthly predictive regressions including the difference of implied and realized volatility as a predictive variable. It is not significant using conventional tests in the sample period 1955-2010 within quarterly predictive regressions based on the share of labor income to consumption.

Second, we find that the difference between option-implied volatility and realized volatility is a robust predictive variable of future market returns at quarterly forecasting horizons. It is always significant at the 5% significance level in each window of 180 observations, using both robust and nonrobust testing approaches. This finding supports the remarkable market return forecasting ability of the variance risk premium, noted in Bollerslev, Tauchen and Zhou (2009) and confirmed in Bollerslev, Marrone and Zhou (2012) in an international context.

Third, using conventional testing approaches, we find that the evidence of return predictability associated with the ratio of labor income to consumption, documented in Santos and Veronesi (2006), is either absent or weak in the sample periods 1955-2000 and 1965-2010, respectively. In contrast, the null of no predictability is always rejected at the 5% significance level by our robust testing method, indicating that the weak and ambiguous evidence produced by the conventional tests is likely a consequence of their low power in presence of anomalous observations.

Fourth, we exploit the properties of our robust testing method to identify potential anomalous observations that might excessively influence the diverging conclusions of conventional testing approaches. We find a fraction of less than about 5% of anomalous observations in the data, which tend to be more frequent during the NASDAQ bubble and the more recent financial crisis. A small subset of these influential data points, including the Lehman Brothers default on September 2008, the terrorist attack of September 2001, the Black Monday on October 1987, and the Dot-Com bubble collapse in August 2002, is largely responsible for the failure of conventional testing methods in uncovering the hidden predictability structures.

Overall, these findings indicate that several key predictive relations implied by finance theory are well supported by the vast majority of the data. At the same time, unfrequent

anomalous deviations from the assumed predictive regression specification can have a substantial distorting impact on the conclusions derived from conventional testing methods. Such anomalous data points also appear to be extraordinarily difficult to model in a predictive regression context, as they seem to be mostly related to sporadic and very turbulent market events. Those events are unlikely to be predictable in a reliable way using conditioning variables typically available for predictability studies.

Our work borrows from a large literature that has investigated whether economic variables such as, e.g., the price-dividend ratio, proxies of labour income, or the interest rate can predict stock returns.<sup>2</sup> The econometric approach to test for predictability is mostly based on a predictive regression of stock returns onto a set of lagged financial variables; see, e.g., Stambaugh (1999). Important differences between testing approaches in the literature arise because of the different test statistics, asymptotic theories or resampling approaches used to test the null hypothesis of no predictability. These differences lead in a number of cases to diverging results and conclusions.

Mankiw and Shapiro (1986) and Stambaugh (1986) note that in a setting with endogenous predictor and correlated innovations standard asymptotic theory causes small sample biases that may imply an overrejection of the hypothesis of no predictability. To mitigate the problem, recent studies propose tests based on bias-corrected estimators of predictive regressions. For instance, Stambaugh (1999), and Amihud, Hurvich and Wang (2008) introduce bias-corrected OLS estimators for the univariate and the multi-predictor setting, respectively. Recent work has also considered the issue of endogenous integrated or nearly integrated predictors, following the evidence in Torous, Valkanov and Yan (2004) that various variables assumed to predict stock returns follow a local-to-unit root autoregressive process. Lewellen (2004), Torous, Valkanov and Yan (2004), and Campbell and Yogo (2006) introduce new testing procedures and more accurate unit-root and local-to-unit root asymptotics for predictive regression models with a single persistent predictor and correlated innovations.

---

<sup>2</sup>See Rozeff (1984), Fama and French (1988), Campbell and Shiller (1988), Nelson and Kim (1993), Goetzmann and Jorion (1995), Kothari and Shanken (1997), Campbell and Yogo (2006), Jansson and Moreira (2006), Polk, Thompson and Vuolteenaho (2006), Santos and Veronesi (2006), Bollerslev, Tauchen and Zhou (2009), among others. Koijen and Van Nieuwerburgh (2011) provide an excellent recent review of part of this literature.

A general approach to obtain tests that are less susceptible to finite sample biases or assumptions on their asymptotic distribution relies on Monte Carlo simulation methods, such as the bootstrap or the subsampling. Ang and Bekaert (2007) use the bootstrap to quantify the bias of parameter estimation in a regression of stock returns on the lagged dividend yield and the interest rate. In a multi-predictor setting with nearly integrated regressors, Amihud, Hurvich and Wang (2008) compare the results of bootstrap tests to bias-corrected procedures and find the latter to have accurate size and good power properties. Wolf (2000) introduces subsampling tests of stock return predictability in single-predictor models.

As shown in Hall and Horowitz (1996) and Andrews (2002), among others, a desirable property of bootstrap tests is that they may provide asymptotic refinements of the sampling distribution of standard  $t$ -test statistics for testing the hypothesis of no predictability.<sup>3</sup> Moreover, as shown in Romano and Wolf (2001), Choi and Chue (2007), and Andrews and Guggenberger (2010), subsampling methods produce reliable inference also in predictive regressions models with multiple nearly-integrated predictors.

A common feature of all above approaches to test predictability hypotheses is their reliance on procedures that can be heavily influenced by a small fraction of anomalous observations. Therefore, they are inappropriate to test economic relations valid for the vast majority of the data, but which might be violated by a small fraction of the data. For standard Ordinary Least Squares 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.<sup>4</sup> Intuitively, this feature is explained by the too high fraction of anomalous data that is often simulated by standard bootstrap and subsampling procedures, when compared to the actual fraction of anomalous observations in the original data. Since it is not possible

---

<sup>3</sup>In the sense that the errors made in approximating the true finite-sample distribution of the  $t$ -test statistic are of lower order with respect to the sample size than those implied by the conventional asymptotics.

<sup>4</sup>Following Huber seminal work, several authors have emphasized the potentially weak robustness features of many standard asymptotic testing procedures; see Heritier and Ronchetti (1994), Ronchetti and Trojani (2001), Mancini, Ronchetti and Trojani (2005), and Gagliardini, Trojani and Urga (2005), among others. Singh (1998), Salibian-Barrera and Zamar (2002), and Camponovo, Scaillet and Trojani (2012), among others, highlight the failing robustness of the inference implied by bootstrap and subsampling tests.

to mitigate this problem simply by applying conventional bootstrap or subsampling methods to more robust estimators or test statistics, we consider in this paper a new robust bootstrap and subsampling methodology that allows us to develop more robust tests of predictability hypotheses in predictive regression settings.

The rest of the paper is organized as follows. In Section 2, we introduce the usual predictive regression model, and we illustrate by simulation the robustness problem of some of the recent tests of predictability proposed in the literature. In Section 3, we study theoretically the robustness properties of tests based on resampling procedures in general time series settings. In Section 4, we introduce our robust approach, and develop robust bootstrap and subsampling tests of predictability. In Section 5, we apply our robust testing procedure to US equity data and reconsider some of the recent empirical evidence on market return predictability. Section 6 concludes.

## 2 Predictability and Anomalous Observations

In this section, we introduce the benchmark predictive regression model and a number of recent methods proposed for testing the predictability of stock returns. Through Monte Carlo simulations, we study the finite-sample properties of these testing procedures both in presence and absence of anomalous observations. In Sections 2.1 and 2.2, we focus on bias-corrected methods and testing procedures based on local-to-unity asymptotics. In Section 2.3, we consider more broadly testing approaches based on Monte Carlo simulation methods, such as the bootstrap and the subsampling.

### 2.1 The Predictive Regression Model

We consider the predictive regression model:

$$y_t = \alpha + \beta x_{t-1} + u_t, \tag{1}$$

$$x_t = \mu + \rho x_{t-1} + v_t, \tag{2}$$



where  $y_t$  denotes the stock return at time  $t$ , and  $x_{t-1}$  an economic variable observed at time  $t - 1$ , predicting  $y_t$  for  $t = 1, \dots, n$ . The parameters  $\alpha \in \mathbb{R}$  and  $\mu \in \mathbb{R}$  are the unknown intercepts of the linear regression model and the autoregressive model, respectively,  $\beta \in \mathbb{R}$  is the unknown parameter of interest,  $\rho \in \mathbb{R}$  is the unknown autoregressive coefficient, and  $u_t \in \mathbb{R}$ ,  $v_t \in \mathbb{R}$  are error terms with  $u_t = \phi v_t + e_t$ ,  $\phi \in \mathbb{R}$ , and  $e_t$  is a scalar random variable.

In this setting, it is well-known that inference based on standard asymptotic theory suffers from small sample biases, which may imply an overrejection of the hypothesis of no predictability,  $\mathcal{H}_0 : \beta_0 = 0$ , where  $\beta_0$  denotes the true value of the unknown parameter  $\beta$ ; see Mankiw and Shapiro (1986), and Stambaugh (1986), among others. Moreover, as emphasized in Torous, Valkanov, and Yan (2004), various state variables considered as predictors in the model (1)-(2) are well approximated by a nearly integrated process. Consequently, this suggests a local-to-unity framework  $\rho = 1 + c/n$ ,  $c < 0$ , for the autoregressive coefficient of model (2), which may imply a nonstandard asymptotic distribution for the OLS estimator  $\hat{\beta}_n$  of parameter  $\beta$ .

Several recent testing procedures have been proposed in order to overcome these problems. Stambaugh (1999), Lewellen (2004), Amihud and Hurvich (2004), Polk, Thompson and Vuolteenaho (2006), and Amihud, Hurvich and Wang (2008, 2010), among others, propose bias-corrected procedures that correct the bias implied by the OLS estimator  $\hat{\beta}_n$  of parameter  $\beta$ . Cavanagh, Elliott and Stock (1995), Torous, Valkanov and Yan (2004), and Campbell and Yogo (2006), among others, introduce testing procedures based on local-to-unity asymptotics that provide more accurate approximations of the sampling distribution of the  $t$ -statistic  $T_n = (\hat{\beta}_n - \beta_0)/\hat{\sigma}_n$  in nearly integrated settings, where  $\hat{\sigma}_n$  is an estimate of the standard deviation of the OLS estimator  $\hat{\beta}_n$ .

## 2.2 Bias Correction Methods and Near-to-Unity Asymptotic Tests

A common feature of bias-corrected methods and inference based on local-to-unity asymptotics is a non-resistance to anomalous observations, which may lead to conclusions determined by the particular features of a small subfraction of the data. Intuitively, this feature emerges because these approaches exploit statistical tools that can be sensitive to outliers or more

general small deviations from the predictive regression model (1)-(2). Consequently, despite the good accuracy under the strict model assumptions, these testing procedures may become less efficient or biased even with a small fraction of anomalous observations in the data.

To illustrate the lack of robustness of this class of tests, we analyze through Monte Carlo simulation the bias-corrected method proposed in Amihud, Hurvich and Wang (2008) and the Bonferroni approach for the local-to-unity asymptotic theory introduced in Campbell and Yogo (2006). We first generate  $N = 1,000$  samples  $z_{(n)} = (z_1, \dots, z_n)$ , where  $z_t = (y_t, x_{t-1})'$ , of size  $n = 180$  according to model (1)-(2), with  $v_t \sim N(0, 1)$ ,  $e_t \sim N(0, 1)$ ,  $\phi = -1$ ,  $\alpha = \mu = 0$ ,  $\rho = 0.9$ , and  $\beta_0 \in [0, 0.15]$ .<sup>5</sup> In a second step, to study the robustness of the methods under investigation, we consider replacement outliers random samples  $\tilde{z}_{(n)} = (\tilde{z}_1, \dots, \tilde{z}_n)$ , where  $\tilde{z}_t = (\tilde{y}_t, x_{t-1})'$  is generated according to,

$$\tilde{y}_t = (1 - p_t)y_t + p_t \cdot y_{3max}, \quad (3)$$

with  $y_{3max} = 3 \cdot \max(y_1, \dots, y_n)$  and  $p_t$  is an iid 0 – 1 random sequence, independent of process (1)-(2) such that  $P[p_t = 1] = \eta$ . The probability of contamination by outliers is set to  $\eta = 4\%$ , which is a small contamination of the original sample, compatible with the features of the real data set analyzed in the empirical study in Section 5.1.<sup>6</sup>

We study the finite sample properties of tests of the null hypothesis  $\mathcal{H}_0 : \beta_0 = 0$  in the predictive regression model. Figure 1 plots the empirical frequency of rejection of null hypothesis  $\mathcal{H}_0$  for the different testing methods, with respect to different values of the alternative hypothesis  $\beta_0 \in [0, 0.15]$ . The nominal significance level of the test is 10%.

In the Monte Carlo simulation with non-contaminated samples (straight line), we find that the fraction of null hypothesis rejections of all procedures is quite close to the nominal level 10% when  $\beta_0 = 0$ . As expected, the power of the tests increases for increasing values of  $\beta_0$ . For  $\beta_0 = 0.1$ , both methods have a frequency of rejection close to 70%, and for  $\beta_0 = 0.15$  a frequency larger than 95%. In the simulation with contaminated samples (dashed line), the

---

<sup>5</sup>These parameter choices are in line with the Monte Carlo setting studied, e.g., in Choi and Chue (2007).

<sup>6</sup>For the monthly data set in Section 5.1, the estimated fraction of anomalous observations in the sample period 1980-2010 is less than about 4%.

size of all tests remains quite close to the nominal significance level. In contrast, the presence of anomalous observations dramatically deteriorates the power of both procedures. Indeed, for  $\beta_0 > 0$ , the frequency of rejection of the null hypothesis for both tests is much lower than in the non-contaminated case. The power of both tests is more flat and below 55% even for large values of  $\beta_0$ . Unreported Monte Carlo results for different parameter choices and significance levels produce similar findings.

The results in Figure 1 highlight the lack of resistance to anomalous data of bias-corrected methods and inference based on local-to-unity asymptotics. Because of a small fraction of anomalous observations, the testing procedures become unreliable, and are unable to reject the null hypothesis of no predictability, even for large values of  $\beta_0$ . This is a relevant aspect for applications, in which typically the statistical evidence of predictability is weak.

To overcome this robustness problem, a natural approach is to develop more resistant versions of the nonrobust tests considered in our Monte Carlo exercise. However, this task may be hard to achieve in general.<sup>7</sup> A more general approach to obtain tests that are less susceptible to finite sample biases or assumptions on their asymptotic distribution can rely on nonparametric Monte Carlo simulation methods, such as the bootstrap or the subsampling. We address these methods in the sequel.

### 2.3 Bootstrap and Subsampling Tests

Nonparametric Monte Carlo simulation methods, such as the bootstrap and the subsampling, may provide improved inferences in predictive regression model (1)-(2) both in stationary or nearly integrated settings. As shown in Hall and Horowitz (1996) and Andrews (2002), for stationary data the block bootstrap may yield improved approximations to the sampling distribution of the standard  $t$ -statistics for testing predictability, having asymptotic errors of lower order in sample size. Moreover, as shown in Choi and Chue (2007) and Andrews and Guggen-

---

<sup>7</sup>To robustify the bias-corrected procedure in Amihud, Hurvich and Wang (2008), we would need to derive an expression for the bias of robust estimators of regressions, and then derive the asymptotic distribution of such bias-corrected robust estimators. For nearly-integrated settings, a robustification of the procedure proposed in Campbell and Yogo (2006) would require a not obvious extension of the robust local-to-unity asymptotics developed in Lucas (1995, 1997) for the predictive regression model.

berger (2010), we can use the subsampling to produce correct inferences in nearly integrated settings. We first introduce in more detail block bootstrap and subsampling procedures. We then focus on predictive regression model (1)-(2) and study by Monte Carlo simulation the degree of resistance to anomalous observations of bootstrap and subsampling tests of predictability.

Consider a random sample  $X_{(n)} = (X_1, \dots, X_n)$  from a time series of random vectors  $X_i \in \mathbb{R}^{d_x}$ ,  $d_x \geq 1$ , and a general statistic  $T_n := T(X_{(n)})$ . Block bootstrap procedures split the original sample  $X_{(n)}$  into overlapping blocks of size  $m < n$ . From these blocks, bootstrap samples  $X_{(n)}^*$  of size  $n$  are randomly generated.<sup>8</sup> Finally, the empirical distribution of statistic  $T(X_{(n)}^*)$  is used to estimate the sampling distribution of  $T(X_{(n)})$ . Similarly, the more recent subsampling method applies statistic  $T$  directly to overlapping random blocks  $X_{(m)}^*$  of size  $m$  strictly less than  $n$ .<sup>9</sup> Then, we use the empirical distribution of statistic  $T(X_{(m)}^*)$  to estimate the sampling distribution of  $T(X_{(n)})$ , under the assumption that the impact of the block size is asymptotically negligible ( $m/n \rightarrow 0$ ).

In the predictive regression model (1)-(2), the usual  $t$ -test statistic for testing the null of no predictability is  $T_n = (\hat{\beta}_n - \beta_0)/\hat{\sigma}_n$ . Therefore, we can define a block bootstrap test of the null hypothesis with the block bootstrap statistic  $T_{n,m}^{B*} = (\hat{\beta}_{n,m}^{B*} - \hat{\beta}_n)/\hat{\sigma}_{n,m}^{B*}$ , where  $\hat{\sigma}_{n,m}^{B*}$  is an estimate of the standard deviation of the OLS estimator  $\hat{\beta}_{n,m}^{B*}$  in a random bootstrap sample of size  $n$ , constructed using blocks of size  $m$ . Similarly, we can define a subsampling test of the same null hypothesis with the subsampling statistic  $T_{n,m}^{S*} = (\hat{\beta}_{n,m}^{S*} - \hat{\beta}_n)/\hat{\sigma}_{n,m}^{S*}$ , where  $\hat{\sigma}_{n,m}^{S*}$  is now an estimator of the standard deviation of the OLS estimator  $\hat{\beta}_{n,m}^{S*}$  in a random overlapping block of size  $m < n$ .

It is well-known that OLS estimators and empirical averages are very sensitive to even small fractions of anomalous observations in the data; see, e.g., Huber (1981). Since block bootstrap and subsampling tests rely on such statistics, inference based on resampling methods may inherit the lack of robustness. To verify this intuition, we study the finite-sample properties of

---

<sup>8</sup>See, e.g., Hall (1985), Carlstein (1986), Künsch (1989) and Andrews (2004), among others. Alternatively, it is possible to construct the bootstrap samples using nonoverlapping blocks.

<sup>9</sup>See Politis, Romano and Wolf (1999), among others.

block bootstrap and subsampling tests of predictability in presence of anomalous observations through Monte Carlo simulations. For comparison, we consider the same simulation setting of the previous section, and study tests of the null hypothesis  $\mathcal{H}_0 : \beta_0 = 0$ , using symmetric (bootstrap and subsampling) confidence intervals for parameter  $\beta$ .<sup>10</sup>

Figure 2 plots the empirical frequencies of rejection of null hypothesis  $\mathcal{H}_0$  for different values of the alternative hypothesis  $\beta_0 \in [0, 0.15]$ . The nominal significance level of the test is 10%, as before. With non-contaminated samples (straight line), we find for all values of  $\beta_0 \in [0, 0.15]$  that the frequency of rejection of block bootstrap and subsampling tests is close to the one of the bias-corrected method and the Bonferroni approach observed in the previous section. For  $\beta_0 = 0$ , the size of the tests is close to the nominal level 10%, while for  $\beta_0 = 0.15$  the power is larger than 95%. Contaminations with anomalous observations strongly deteriorate the power of the tests. With contaminated samples (dashed line), the frequency of rejection of the null hypothesis is always less than 55%, even for large values of  $\beta_0$ . In particular, when  $\beta_0 = 0.15$ , the difference in power for the subsampling applied to non-contaminated and contaminated samples is larger than 50%.

The results in Figure 2 show that bootstrap and subsampling tests inherit, and to some extent exacerbate, the lack of robustness of OLS estimators for predictive regressions. To robustify the inference produced by resampling methods, a natural idea is to apply standard bootstrap and subsampling simulation schemes to a more robust statistic, such as, e.g., a robust estimator of linear regression. Unfortunately, as shown in Singh (1998), Salibian-Barrera and Zamar (2002), and Camponovo, Scaillet and Trojani (2012) for iid settings, resampling a robust statistic does not yield a robust inference, because standard block bootstrap and subsampling procedures have an intrinsic non-resistance to outliers. Intuitively, this problem arises because the fraction of anomalous observations generated in bootstrap and subsampling blocks is often much higher than the fraction of outliers in the data. To solve this problem, it is necessary to analyze more systematically the robustness of bootstrap and subsampling methods for time

---

<sup>10</sup>Appendix A shows how to construct symmetric confidence intervals for the parameter of interest, based on block bootstrap and subsampling distributions. For the selection of the block size  $m$ , we use the standard data-driven methods described in Appendix B.

series.

### 3 Robust Resampling Inference and Quantile Breakdown Point

We characterize theoretically the robustness of bootstrap and subsampling tests in predictive regression settings. Section 3.1 introduces the notion of a quantile breakdown point, which is a measure of the global resistance of a resampling method to anomalous observations. Section 3.2 quantifies and illustrates the quantile breakdown point of standard bootstrap and subsampling tests in predictive regression models. Finally, Section 3.3 derives explicit bounds for quantile breakdown points, which quantify the intrinsic degree of resistance to outliers of bootstrap and subsampling tests for predictability.

#### 3.1 Quantile Breakdown Point

Given a random sample  $X_{(n)}$  from a sequence of random vectors  $X_i \in \mathbb{R}^{d_x}$ ,  $d_x \geq 1$ , let  $X_{(n,m)}^{B*} = (X_1^*, \dots, X_n^*)$  denote a block bootstrap sample, constructed using overlapping blocks of size  $m$ . Similarly, let  $X_{(n,m)}^{S*} = (X_1^*, \dots, X_m^*)$  denote an overlapping subsampling block. We denote by  $T_{n,m}^{K*} := T(X_{(n,m)}^{K*})$ ,  $K = S, B$ , the corresponding block bootstrap and subsampling statistics, respectively.<sup>11</sup> For  $t \in (0, 1)$ , the quantile  $Q_{t,n,m}^{K*}$  of  $T_{n,m}^{K*}$  is defined by

$$Q_{t,n,m}^{K*} = \inf\{x | P^*(T_{n,m}^{K*} \leq x) \geq t\}, \tag{4}$$

where  $P^*$  is the probability measure induced by the block bootstrap or the subsampling method and, by definition,  $\inf(\emptyset) = \infty$ .

Quantile  $Q_{t,n,m}^{K*}$  is effectively a useful nonparametric estimator of the corresponding finite-sample quantile of statistic  $T(X_1, \dots, X_n)$ . We characterize the robustness properties of block

---

<sup>11</sup>We focus for brevity on one-dimensional real-valued statistics. However, as discussed for instance in Singh (1998) in the iid context, our results for time series can be naturally extended to multivariate and scale statistics.

bootstrap and subsampling by the breakdown point  $b_{t,n,m}^K$  of the quantile (4), which is defined as the smallest fraction of outliers in the original sample such that  $Q_{t,n,m}^{K*}$  diverges to infinity.<sup>12</sup>

Intuitively, when a breakdown occurs, inference about the distribution of  $T(X_1, \dots, X_n)$  based on bootstrap or subsampling tests becomes meaningless. Estimated test critical values may be arbitrarily large and confidence intervals be arbitrarily wide. In these cases, the size and power of bootstrap and subsampling tests can collapse to zero or one in presence of anomalous observations, making these inference procedures useless. Therefore, quantifying  $b_{t,n,m}^K$  in general for bootstrap and subsampling tests of predictability, in dependence of the statistics and testing approaches used, is key in order to understand which approaches ensure some resistance to anomalous observations and which do not. This provides useful practical guidelines, even before looking at the data.

### 3.2 Quantile Breakdown Point and Predictive Regression

The quantile breakdown point of conventional block bootstrap and subsampling tests for predictability in Section 2.3 depends directly on the breakdown properties of OLS estimator  $\hat{\beta}_n$ . The breakdown point  $b$  of a statistics  $T_n = T(X_{(n)})$  is simply the smallest fraction of outliers in the original sample such that the statistic  $T_n$  diverges to infinity; see, e.g., Donoho and Huber (1983) for the formal definition. We know  $b$  explicitly in some cases and we can gauge its value most of the time, for instance by means of simulations and sensitivity analysis. Most nonrobust statistics, like OLS estimators for linear regression, have a breakdown point  $b = 1/n$ . Therefore, the breakdown point of conventional block bootstrap and subsampling quantiles in predictive regression settings also equals  $1/n$ . In other words, a single anomalous observation in the original data is sufficient to produce a meaningless inference implied by bootstrap or subsampling quantiles in standard tests of predictability.

It is straightforward to illustrate these features in a Monte Carlo simulation that quantifies the sensitivity of block bootstrap and subsampling quantiles to data contaminations by a single outlier, where the size of the outlier is increasing. We first simulate  $N = 1,000$  random samples

---

<sup>12</sup>In Appendix A, we provide the formal definition of the breakdown point  $b_{t,n,m}^K$ .

$z_{(n)} = (z_1, \dots, z_n)$  of size  $n = 120$ , where  $z_t = (y_t, x_{t-1})'$  follows model (1)-(2),  $v_t \sim N(0, 1)$ ,  $e_t \sim N(0, 1)$ ,  $\phi = -1$ ,  $\alpha = \mu = 0$ ,  $\rho = 0.9$ , and  $\beta_0 = 0$ . For each Monte Carlo sample, we define in a second step

$$y_{\max} = \arg \max_{y_1, \dots, y_n} \{w(y_i) | w(y_i) = y_i - \beta_0 x_{i-1}, \text{ under } \mathcal{H}_0 : \beta_0 = 0\} , \quad (5)$$

and we modify  $y_{\max}$  over the interval  $[y_{\max}, y_{\max} + 5]$ . This means that we contaminate the predictability relationship by an anomalous observation for only one single data point in the full sample. We study the sensitivity of the Monte Carlo average length of confidence intervals for parameter  $\beta$ , estimated by the standard block bootstrap and the subsampling. This is a natural exercise, as the length of the confidence interval for parameter  $\beta$  is in a one-to-one relation with the critical value of the test of the null of no predictability ( $\mathcal{H}_0: \beta_0 = 0$ ). For the sake of comparison, we also consider confidence intervals implied by the bias-corrected testing method in Amihud, Hurvich and Wang (2008) and the Bonferroni approach proposed in Campbell and Yogo (2006).

For all tests under investigation, Figure 4 and 5 plot the relative increase of the average confidence interval length in our Monte Carlo simulations, under contamination by a single outlier of increasing size. We find that all sensitivities are basically linear in the size of the outlier, confirming that a single anomalous observation can have an arbitrarily large impact on the critical values of those tests and make the test results potentially useless, as implied by their quantile breakdown point of  $1/n$ .

### 3.3 Quantile Breakdown Point Bounds

To obtain bootstrap and subsampling tests with more favorable breakdown properties, it is necessary to apply resampling procedures to a robust statistic with nontrivial breakdown point ( $b > 1/n$ ), such as, e.g., a robust estimator of linear regression. Without loss of generality, let  $T_n = T(X_{(n)})$  be a statistic with breakdown point  $1/n < b \leq 0.5$ .

In Theorem 2 in Appendix A, we compute explicit quantile breakdown point bounds, which characterize the resistance of bootstrap and subsampling tests to anomalous observations, in



dependence of relevant parameters, such as  $n$ ,  $m$ ,  $t$ , and  $b$ .<sup>13</sup> In particular, we show that  $b_{t,n,m}^S$  and  $b_{t,n,m}^B$  satisfy the following bounds,

$$\begin{aligned} \frac{\lceil mb \rceil}{n} &\leq b_{t,n,m}^S \leq \frac{1}{n} \cdot \left[ \inf_{\{p \in \mathbb{N}, p \leq r-1\}} \left\{ p \cdot \lceil mb \rceil \middle| p > \frac{(1-t)(n-m+1) + \lceil mb \rceil - 1}{m} \right\} \right], \\ \frac{\lceil mb \rceil}{n} &\leq b_{t,n,m}^B \leq \frac{1}{n} \cdot \left[ \inf_{\{p_1, p_2\}} \left\{ p = p_1 \cdot p_2 \middle| P\left( \text{BIN}\left( r, \frac{mp_2 - p_1 + 1}{n - m + 1} \right) \geq \left\lceil \frac{nb}{p_1} \right\rceil \right) > 1 - t \right\} \right], \end{aligned}$$

where  $p_1, p_2 \in \mathbb{N}$ , with  $p_1 \leq m, p_2 \leq r - 1$ .<sup>14</sup>

We quantify the implications of Theorem 2 by computing in Table 1 lower and upper bounds for the breakdown point of subsampling and bootstrap quantiles, using a sample size  $n = 120$  and a maximal statistic breakdown point ( $b = 0.5$ ), such as, e.g., the sample median. We find that even for a highly robust statistic with maximal breakdown point ( $b = 0.5$ ), the subsampling implies a very low quantile breakdown point, which increases with the block size but is also very far from the maximal value  $b = 0.5$ . For instance, for a block size  $m = 10$ , the 0.95-quantile breakdown point is between 0.0417 and 0.0833. In other words, even though a statistic is resistant to large fractions of anomalous observations, the implied subsampling quantile can collapse with just 5 outliers out of 100 observations.<sup>15</sup> Similar results arise for the bootstrap quantiles. Even though the bounds are less sharp than for the subsampling, quantile breakdown points are again clearly smaller than the breakdown point of the statistic used.<sup>16</sup>

Overall, the results in Theorem 2 imply that subsampling and bootstrap tests for time series

---

<sup>13</sup>Similar results can be obtained for the subsampling and the block bootstrap based on nonoverlapping blocks. The results for the block bootstrap can also be modified to cover asymptotically equivalent variations, such as the stationary bootstrap of Politis and Romano (1994).

<sup>14</sup>The term  $\frac{(1-t)(n-m+1)}{m}$  represents the number of degenerated subsampling statistics necessary in order to cause the breakdown of  $Q_{t,n,m}^{S*}$ , while  $\frac{\lceil mb \rceil}{n}$  is the fraction of outliers which is sufficient to cause the breakdown of statistic  $T$  in a block of size  $m$ . Note that the breakdown point formula for the iid bootstrap in Singh (1998) emerges as a special case of the formula (21), for  $m = 1$ .

<sup>15</sup>This breakdown point is also clearly lower than in the iid case; see Camponovo, Scaillet and Trojani (2012). For instance, for  $m = 10$ , the 0.95-quantile breakdown point of the overlapping subsampling is 0.23 in iid settings. Since in a time series setting the number of possible subsampling blocks of size  $m$  is typically lower than the number of iid subsamples of size  $m$ , the breakdown of a statistic in one random block tends to have a larger impact on the subsampling quantile than in the iid case.

<sup>16</sup>These quantile breakdown point bounds are again clearly lower than in the iid setting. For instance, for  $m = 30$ , the 0.95-quantile breakdown point for time series is less than 0.25, but it is 0.425 for iid settings, from the results in Camponovo, Scaillet and Trojani (2012).

feature an intrinsic non-resistance to anomalous observations. This robustness problem cannot be avoided, simply by applying conventional resampling approaches to more robust statistics.

## 4 Robust Bootstrap and Subsampling

When using a robust statistic with large breakdown point, the bootstrap and the subsampling still imply an important non-resistance to anomalous observations. This finding is consistent with our Monte Carlo results in the predictive regression model. To overcome the problem, it is necessary to introduce a novel class of more robust bootstrap and subsampling tests in the time series context.<sup>17</sup> Section 4.1 introduces our robust bootstrap and subsampling approach, and Section 4.2 demonstrates its favorable breakdown properties. Section 4.3 introduces new robust bootstrap and subsampling tests for predictability in predictive regression models.

### 4.1 Definition

Given the original sample  $X_{(n)} = (X_1, \dots, X_n)$ , we consider the class of robust M-estimators  $\hat{\theta}_n$  for parameter  $\theta \in \mathbb{R}^d$ , defined as the solution of the estimating equations

$$\psi_n(X_{(n)}, \hat{\theta}_n) := \frac{1}{n} \sum_{i=1}^n g(X_i, \hat{\theta}_n) = 0, \quad (6)$$

where  $\psi_n(X_{(n)}, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  depends on parameter  $\theta$  and a bounded estimating function  $g$ . Boundedness of function  $g$  is a characterizing feature of robust M-estimators, such as, e.g., robust estimators of linear regression; see Hampel, Ronchetti, Rousseeuw and Stahel (1986) for a review.

Conventional bootstrap (subsampling) methods solve equation  $\psi_k(X_{(n,m)}^{K*}, \hat{\theta}_{n,m}^{K*}) = 0$ , for each bootstrap (subsampling) random sample  $X_{(n,m)}^{K*}$ . This task can be computationally demanding.

---

<sup>17</sup>We develop such robust methods borrowing from the fast resampling approach considered, among others, in Shao and Tu (1995), Davidson and McKinnon (1999), Hu and Kalbfleisch (2000), Andrews (2002), Salibian-Barrera and Zamar (2002), Goncalves and White (2004), Hong and Scaillet (2006), Salibian-Barrera, Van Aelst and Willems (2006, 2007), and Camponovo, Scaillet and Trojani (2012).

Instead, we consider a standard Taylor expansion of (6) around the true parameter  $\theta_0$ ,

$$\hat{\theta}_n - \theta_0 = -[\nabla_{\theta}\psi_n(X_{(n)}, \theta_0)]^{-1}\psi_n(X_{(n)}, \theta_0) + o_p(1), \quad (7)$$

where  $\nabla_{\theta}\psi_n(X_{(n)}, \theta_0)$  is the derivative of function  $\psi_n$  with respect to parameter  $\theta$ . Based on this expansion, we can use  $-[\nabla_{\theta}\psi_n(X_{(n)}, \hat{\theta}_n)]^{-1}\psi_k(X_{(n,m)}^{K*}, \hat{\theta}_n)$  as an approximation of  $\hat{\theta}_{n,m}^{K*} - \hat{\theta}_n$  in the definition of the resampling scheme estimating the sampling distribution of  $\hat{\theta}_n - \theta_0$ . This approach avoids computing  $\hat{\theta}_{n,m}^{K*}$  and  $[\nabla_{\theta}\psi_k(X_{(n,m)}^{K*}, \hat{\theta}_n)]^{-1}$  in each bootstrap or subsampling sample. Therefore, it implies a markable computational advantage that produces a fast numerical procedure. This is an important improvement over conventional resampling schemes, which can easily become unfeasible when applied to robust statistics.

**Definition 1** *Given a normalization constant  $\tau_n$  such that  $\tau_n \rightarrow \infty$  as  $n \rightarrow \infty$ , a robust fast resampling distribution for  $\tau_n(\hat{\theta}_n - \theta_0)$  is defined by*

$$L_{n,m}^{K*}(x) = \frac{1}{N} \sum_{s=1}^N \mathbb{I} \left( \tau_k \left( -[\nabla_{\theta}\psi_n(X_{(n)}, \hat{\theta}_n)]^{-1}\psi_k(X_{(n,m),s}^{K*}, \hat{\theta}_n) \right) \leq x \right), \quad (8)$$

where  $\mathbb{I}(\cdot)$  is the indicator function and  $s$  indexes the  $N$  possible random samples generated by the bootstrap and subsampling procedures, respectively.

General assumptions under which (8) consistently estimates the unknown asymptotic distribution of  $\tau_n(\hat{\theta}_n - \theta_0)$  in a time series context are given, e.g., in Hong and Scaillet (2006) for the subsampling (Assumption 1) and in Goncalves and White (2004) for the bootstrap (Assumption A and Assumptions 2.1 and 2.2).<sup>18</sup>

---

<sup>18</sup>Let  $L(\cdot)$  be the limit distribution function of  $\tau_n(\hat{\theta}_n - \theta_0)$ . Under regularity conditions, it follows, for any continuity point of  $L(\cdot)$ :  $L_{n,m}^{K*}(x) \rightarrow L(x)$  in probability, as  $N, n \rightarrow \infty$ , and additionally  $m \rightarrow \infty$ ,  $m/n \rightarrow 0$  in the subsampling case. If  $L(\cdot)$  is continuous, then:  $\sup_x |L_{n,m}^{K*}(x) - L(x)| \rightarrow 0$  in probability. Moreover, as noted, e.g., in Hall, Horowitz and Jing (1995), several standard resampling methods imply the distortion  $E^*[\hat{\theta}_{n,m}^{K*}] \neq E^*[\hat{\theta}_n]$ , where  $E^*[\cdot]$  denotes the expectation under the bootstrap or the subsampling distribution  $P^*$ , which decreases the rate of convergence. A solution to this problem relies on the recentered statistic  $\hat{\theta}_{n,m}^{K*} - E^*[\hat{\theta}_{n,m}^{K*}]$ , instead of  $\hat{\theta}_{n,m}^{K*} - \hat{\theta}_n$ , and we can use it in our robust approach as well. Indeed, as shown in

## 4.2 Robust Resampling Methods and Quantile Breakdown Point

In the computation of the resampling distribution (8), we only need consistent point estimates for parameter vector  $\theta_0$  and matrix  $-\left[\nabla_{\theta}\psi_n(X_{(n)},\theta_0)\right]^{-1}$ , based on the original sample  $X_{(n)}$ . These estimates are simply given by  $\hat{\theta}_n$  and  $-\left[\nabla_{\theta}\psi_n(X_{(n)},\hat{\theta}_n)\right]^{-1}$ , respectively.

A closer inspection of quantity  $-\left[\nabla_{\theta}\psi_n(X_{(n)},\hat{\theta}_n)\right]^{-1}\psi_k(X_{(n,m),s}^{K*},\hat{\theta}_n)$  in Definition 1 reveals important implications for the breakdown properties of the robust fast resampling distribution (8). Indeed, this quantity can degenerate only when either (i) matrix  $\nabla_{\theta}\psi_n(X_{(n)},\hat{\theta}_n)$  is singular or (ii) estimating function  $g$  is not bounded. However, since we are making use of a robust (bounded) estimating function  $g$ , situation (ii) cannot arise. Therefore, we intuitively expect the breakdown of the quantiles of robust fast resampling distribution (8) to arise only when condition (i) is realized.<sup>19</sup>

Based on this intuition, in Corollary 3 in Appendix A, we compute formally the quantile breakdown point  $b_{t,n,m}^K$  of the robust fast resampling distribution (8), which depends only on the breakdown properties of  $\hat{\theta}_n$  and  $-\left[\nabla_{\theta}\psi_n(X_{(n)},\hat{\theta}_n)\right]^{-1}$ . It turns out that given a concrete model setting, the characterization of the breakdown properties of our robust bootstrap and subsampling approaches is often straightforward.

## 4.3 Robust Predictive Regression and Hypothesis Testing

We develop a new class of easily applicable robust bootstrap and subsampling tests for the null hypothesis of no predictability in predictive regression models. To this end, consider the

---

Andrews (2002) the recentered robust fast distribution

$$L_{n,m}^{K*}(x) = \frac{1}{N} \sum_{s=1}^N \mathbb{I} \left( \tau_k \left( -\left[\nabla_{\theta}\psi_n(X_{(n)},\hat{\theta}_n)\right]^{-1}(\psi_k(X_{(n,m),s}^{K*},\hat{\theta}_n) - E^*[\psi_k(X_{(n,m),s}^{K*},\hat{\theta}_n)]) \right) \leq x \right),$$

implies a zero distortion under  $P^*$ . In particular, the computation of  $E^*[\psi_k(X_{(n,m)}^{K*},\hat{\theta}_n)]$  is much less demanding than the computation of  $E^*[\hat{\theta}_{n,m}^{K*}]$ .

<sup>19</sup>Unreported Monte Carlo simulations show that the application of our robust resampling approach to an  $M$ -estimator with nonrobust (unbounded) estimating function  $g$  does not solve the robustness problem, consistent with our theoretical results in Section 3.3.

predictive regression model

$$y_t = \theta' w_{t-1} + u_t, \quad t = 1, \dots, n, \quad (9)$$

with  $\theta = (\alpha, \beta)'$  and  $w_{t-1} = (1, x_{t-1})'$ , and denote by  $z_{(n)} = (z_1, \dots, z_n)$  an observation sample generated according to (9), where  $z_t = (y_t, w'_{t-1})'$ .

According to Definition 1, a robust estimator of predictive regression is needed, featuring a nontrivial breakdown point  $b > 1/n$  and a bounded estimating function  $g$ , in order to obtain robust bootstrap and subsampling tests with our approach. Several such estimators are available in the literature, which imply corresponding robust bootstrap and subsampling procedures.<sup>20</sup> Among those estimators, a convenient choice is a Huber-type estimator of regression. This estimator ensures together good robustness properties and moderate computational costs.

Given a positive constant  $c$ ,  $\hat{\theta}_n^R$  is the  $M$ -estimator that solves the equation

$$\psi_{n,c}(z_{(n)}, \hat{\theta}_n^R) := \frac{1}{n} \sum_{t=1}^n (y_t - w'_{t-1} \hat{\theta}_n^R) w_t \cdot h_c(z_t, \hat{\theta}_n^R) = 0, \quad (10)$$

where the function  $h_c$  is defined as

$$h_c(z_t, \theta) := \min \left( 1, \frac{c}{\|(y_t - w'_{t-1} \theta) w_{t-1}\|} \right). \quad (11)$$

In Equation (10), estimator  $\hat{\theta}_n^R$  is actually a Weighted Least Squares estimator with data-driven weights  $h_{ct} := h_c(z_t, \hat{\theta}_n^R)$  in Equation (11) that are jointly determined with  $\hat{\theta}_n^R$ :

$$\hat{\theta}_n^R = \left( \sum_{t=1}^n (w_{t-1} w'_{t-1}) h_{ct} \right)^{-1} \sum_{t=1}^n (y_t w_{t-1}) h_{ct}. \quad (12)$$

By design, the Huber weight  $0 \leq h_{ct} \leq 1$  reduces the influence of potential anomalous observations on the estimation results. Weights below one indicate a potentially anomalous and influential data-point, while weights equal one indicate unproblematic observations for the pos-

---

<sup>20</sup>Other examples of robust estimators for regression also include, e.g., the trimmed least square estimators adopted in Knez and Ready (1997).

tulated predictive regression model. In Equation (11), the weight  $h_{ct}$  decreases for observations  $(y_t, w'_{t-1})$  that imply an excessively large OLS score value  $(y_t - w'_{t-1}\theta)w_{t-1}$ . This is intuitive, because those are single observations that can have an excessive influence on the OLS estimation and testing results. It follows that a low value of estimated weight (11) is also informative for highlighting potential anomalous observations that might be particularly influential for the OLS fit of the predictive regression model; see, e.g., Hampel, Ronchetti, Rousseeuw and Stahel (1986).

Constant  $c > 0$  is useful in order to tune the degree of resistance to anomalous data of estimator  $\hat{\theta}_n^R$  in relevant applications, and it can be determined in a fully data-driven way.<sup>21</sup> Note that, as required by our robust resampling approach, the norm of function  $\psi_{n,c}$  in Equation (10) is bounded (by constant  $c$ ), and the breakdown point of estimator  $\hat{\theta}_n^R$  is maximal ( $b = 0.5$ , see, e.g., Huber, 1981).

### 4.3.1 Robust Resampling Tests

By applying the robust fast approach in Definition 1 to the estimating function (10), we can estimate the sampling distribution of the nonstudentized statistic  $T_n^{NS} = \sqrt{n} (\hat{\theta}_n^R - \theta_0)$ , using the following robust fast resampling distribution:

$$L_{n,m}^{NS,K^*}(x) = \frac{1}{N} \sum_{s=1}^N \mathbb{I} \left( \sqrt{k} \left( - [\nabla_{\theta} \psi_{n,c}(z_{(n)}, \hat{\theta}_n)]^{-1} \psi_{k,c}(z_{(n,m),s}^{K^*}, \hat{\theta}_n) \right) \leq x \right), \quad K = B, S, \quad (13)$$

where  $\theta_0 = (\alpha_0, \beta_0)'$  and  $k = n$  for the block bootstrap ( $k = m$  for the subsampling).

A key property of resampling distribution (13) is that it implies a maximal quantile breakdown point, i.e., the largest achievable degree of resistance to anomalous observations, independent of the probability level  $t$  and the selected block size  $m$ , both for the bootstrap and the subsampling. This feature follows directly from the breakdown point  $b = 0.5$  of the robust Huber estimator (10) and the fact that  $\nabla_{\theta} \psi_{n,c}(z_{(n)}, \hat{\theta}_n)$  also has maximal breakdown properties, as formally established in Corollary 4 of Appendix A.

---

<sup>21</sup>Appendix C presents in detail the data-driven method for the selection of the tuning constant  $c$ .

Using nonstudentized statistics, robust resampling distribution (13) provides consistent estimators of the sampling distribution of  $T_n^{NS} = \sqrt{n} \left( \hat{\theta}_n^R - \theta_0 \right)$  in stationary time series settings. With slight modifications, our robust approach can also be used to approximate the sampling distribution of the studentized statistic  $T_n = \sqrt{n} [\hat{\Sigma}_n^R]^{-1/2} (\hat{\theta}_n^R - \theta_0)$ , where  $\hat{\Sigma}_n^R$  is an estimator of the asymptotic variance of  $\hat{\theta}_n^R$ . Appendix A discusses in detail robust resampling approximations for the sampling distribution of  $T_n$ .<sup>22</sup>

### 4.3.2 Monte Carlo Evidence

To quantify the robustness properties of our robust resampling approach, established in Corollary 4 of Appendix A, we can study the sensitivity of confidence intervals estimated by the robust block bootstrap and subsampling distributions (13), with respect to contaminations by anomalous observations of increasing size. To this end, we consider the same Monte Carlo setting of Section 3.2. We plot in Figure 6 the percentage increase of the length in the average estimated confidence interval, with respect to contaminations of the available data by a single anomalous observation of increasing size. In evident contrast to the findings for conventional bootstrap and subsampling tests, Figure 6 shows that the inference implied by our robust approach is largely insensitive to anomalous data, with a percentage increase in the average confidence interval length that is less than 1%, even for an outlier of size  $y_{max} + 5$ .

The robustness of resampling distribution (13) has favorable implications for the power of bootstrap and subsampling tests in presence of anomalous observations. For the same Monte Carlo setting of Sections 2.2 and 2.3, Figure 3 shows that in presence of non-contaminated samples (straight line) the frequencies of null hypothesis rejections of robust resampling approaches are again very close to those observed for nonrobust methods. For instance, the frequency of rejections when  $\beta_0 = 0$  is close to the nominal level 10% of the test, while the power is larger than 95% for  $\beta_0 = 0.15$ . In presence of anomalous observations (dashed line), robust approaches still provide an accurate empirical size close to the actual nominal level, as

---

<sup>22</sup>We find that while the quantile breakdown points of these latter robust resampling distributions are clearly larger than those of the conventional bootstrap and subsampling in Section 3.3, they are typically smaller than the maximal breakdown point quantiles of robust resampling distribution (13). Therefore, we perform our robust empirical analysis based of the resampling distribution (13).

well as a power curve that is close to the one obtained in the non-contaminated Monte Carlo simulation. In particular, in contrast to the standard tests, the power of both robust tests is above 90% for  $\beta_0 = 0.15$ .<sup>23</sup>

## 5 Empirical Evidence of Return Predictability

Using our robust bootstrap and subsampling tests, we revisit the recent empirical evidence on return predictability for US stock market data from a robustness perspective. We study single-predictor and multi-predictor settings, using several well-known predictive variables suggested in the literature, such as the lagged dividend yield, the difference between option-implied volatility and realized volatility (Bollerslev, Tauchen and Zhou, 2009), and the share of labor income to consumption (Santos and Veronesi, 2006). We compare the evidence produced by our robust bootstrap and subsampling tests of predictability with the results of recent testing methods proposed in the literature, including the bias-corrected method in Amihud, Hurvich and Wang (2008), the Bonferroni approach for local-to-unity asymptotics in Campbell and Yogo (2006), and conventional bootstrap and subsampling tests. From our previous findings, we can expect significant differences in the empirical conclusions of conventional and robust tests, should anomalous observations be present in the data.

The dividend yield is the most common predictor of future stock returns, as suggested by a simple present-value logic.<sup>24</sup> However, its forecasting ability has been called into question, e.g., by the ambiguous empirical evidence of studies not rejecting the null of no predictability for a number of forecasting horizons and sample periods; see, e.g., Goyal and Welch (2003), and Ang and Bekaert (2007), among others. Whether these ambiguous results are related to the weakness of conventional tests in detecting predictability structures masked by anomalous observations, is an empirical question that we can analyze using our robust testing method.

The empirical study is articulated in three parts. Section 5.1 studies the forecast ability

---

<sup>23</sup>Unreported Monte Carlo results for the robust block bootstrap and subsampling distributions of studentized tests, defined in equation (27) of the Appendix, confirm the improved resistance properties of our approach.

<sup>24</sup>See, e.g., Rozeff (1984), Campbell and Shiller (1988), Fama and French (1988), Stambaugh (1999), Lewellen (2004), Torous, Valkanov and Yan (2004), Lettau and Ludvigson (2005), and Campbell and Yogo (2006).



of the lagged dividend yield for explaining monthly S&P 500 index returns, in a predictive regression model with a single predictor. This study allows us to compare the results of our methodology with those of the Bonferroni approach for local-to-unity asymptotics, which is applicable to univariate regression settings. Instead, Section 5.2 considers models with several predictive variables. In Section 5.2.1, we test the predictive power of the dividend yield and the variance risk premium, for quarterly S&P 500 index returns sampled at a monthly frequency in periods marked by a financial bubble and a financial crisis. Finally, Section 5.2.2 tests the predictive power of the dividend yield and the ratio of labor income to consumption for predicting quarterly value-weighted CRSP index returns.<sup>25</sup>

## 5.1 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  $t$ . Consistent with the literature, the annualized dividend series  $D_t$  is defined as

$$D_t = d_t + (1 + r_t)d_{t-1} + (1 + r_t)(1 + r_{t-1})d_{t-2} + \cdots + (1 + r_t) \dots (1 + r_{t-10})d_{t-11}, \quad (14)$$

where  $r_t$  is the one-month maturity Treasury-bill rate. We estimate the predictive regression model

$$\ln(R_t) = \alpha + \beta \ln\left(\frac{D_{t-1}}{P_{t-1}}\right) + \epsilon_t ; t = 1, \dots, n , \quad (15)$$

and test the null of no predictability,  $\mathcal{H}_0 : \beta_0 = 0$ .

We collect monthly observations in the sample period 1980-2010 and estimate the predictive regression model using rolling windows of 180 observations. Figure 7 plots the 90%-confidence intervals for parameter  $\beta$  in the sample period 1980-2010, while Table 3 reports the detailed point estimates and test results for the different testing procedures in the four subperiods 1980-1995, 1985-2000, 1990-2005, 1995-2010.

We find that while the robust bootstrap and subsampling tests always clearly reject the

---

<sup>25</sup>We also consider regressions with three predictive variables that additionally incorporate interest rate proxies. We discuss below the results, but we do not report the details for brevity.

hypothesis of no predictability at the 5%-significance level, the conventional testing approaches produce a weaker and more ambiguous predictability evidence. For instance, the bootstrap and subsampling tests cannot reject  $\mathcal{H}_0$  at the 10% significance level in subperiod 1984-1999, while the bias-corrected method and the Bonferroni approach fail to reject  $\mathcal{H}_0$  at the 10% significance level in the subperiod 1995-2010.

It is interesting to study to which extent anomalous observations in sample periods 1984-1999 and 1995-2010 might have caused the diverging conclusions of robust and nonrobust testing methods. We exploit the properties of our robust testing method to identify such data points ex post, i.e., after having estimated the relevant predictive relation. Figure 12 plots the time series of Huber weights estimated by the robust estimator (10) for the predictive regression model (15).

We find that subperiod 1998-2002 is characterized by a cluster of infrequent anomalous observations with Huber-weights less than one, 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.

Overall, observations with a Huber weight less than one are less than 4.2% of the whole data sample. Among these, a small subset of very influential data-points explains the failure of conventional testing methods in uncovering hidden predictability structures in these sample periods.

We find that the most influential observation before 1995 is November 1987, following the Black Monday on October 19 1987. During the subperiod 1998-2002, the most influential observation is October 2001, reflecting the impact on financial markets of the terrorist attack on September 11 2001. Finally, the most anomalous observation in the whole sample period 1980-2010 is October 2008, following the Lehman Brothers default on September 15 2008. The impact on the test results of the Lehman Brothers default emerges also in Figure 7, where nonrobust resampling methods no longer reject  $\mathcal{H}_0$  in 2009. In contrast, robust tests still find significance evidence in favor of predictability.

## 5.2 Two-Predictor Model

We extend our empirical study to two-predictor regression models. This approach has several purposes. First, we can assess the incremental predictive ability of the dividend yield, in relation to other well-known competing predictive variables. Second, we can verify the power properties of robust bootstrap and subsampling tests in settings with several predictive variables.

Section 5.2.1 borrows from Bollerslev, Tauchen and Zhou (2009) and studies the joint predictive ability of the dividend yield and the variance risk premium. Section 5.2.2 follows the two-predictor model in Santos and Veronesi (2006) and considers the ratio of labor income to consumption as an additional predictive variable to the dividend yield.

### 5.2.1 Bollerslev, Tauchen and Zhou

We consider again monthly S&P 500 index and dividend data between January 1990 and December 2010, and test the predictive regression model:

$$\frac{1}{k} \ln(R_{t+k,t}) = \alpha + \beta_1 \ln\left(\frac{D_t}{P_t}\right) + \beta_2 VRP_t + \epsilon_{t+k,t}, \quad (16)$$

where  $\ln(R_{t+k,t}) := \ln(R_{t+1}) + \dots + \ln(R_{t+k})$  and the variance risk premium  $VRP_t := IV_t - RV_t$  is defined by the difference of the S&P 500 index option-implied volatility at time  $t$ , for one month maturity options, and the ex-post realized return variation over the period  $[t-1, t]$ . Bollerslev, Tauchen and Zhou (2009) show that the variance risk premium is the most significant predictive variable of market returns over a quarterly horizon. Therefore, we test the predictive regression model (16) for  $k = 4$ .

Let  $\beta_{01}$  and  $\beta_{02}$  denote the true values of parameters  $\beta_1$  and  $\beta_2$ , 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 the dividend yield,  $\mathcal{H}_{01} : \beta_{01} = 0$ .

Figure 8 plots the 90%-confidence intervals for parameter  $\beta_1$ , based on rolling windows of 180 monthly observations in sample period 1990-2010, while Table 4 collects the detailed point

estimates and testing results. We find again that the robust tests always clearly reject the null of no predictability at the 5%-significance level. In contrast, the conventional bootstrap and subsampling tests produce weaker and more ambiguous results, with uniformly lower  $p$ -values (larger confidence intervals) and a non-rejection of the null of no predictability at the 5%-level in period 1994-2009. Since the Bonferroni approach in Campbell and Yogo (2006) is defined for single-predictor models, we cannot apply this method in model (16). Unreported results for the multi-predictor testing method in Amihud, Hurvich and Wang (2008) show that for data windows following window 1993-2008 the bias-corrected method cannot reject the null hypothesis  $\mathcal{H}_{01}$  at the 10% significance level.

By inspecting the Huber weights (11), implied by the robust estimation of the predictive regression model (16), we find again a cluster of infrequent anomalous observations, both during the NASDAQ bubble and the recent financial crisis. In this setting, the most influential observation is still October 2008, reflecting the Lehman Brothers default on September 15 2008. The impact of these anomalous observations emerges also in Figure 8, explaining the large estimated confidence intervals of nonrobust tests and their non-rejection of  $\mathcal{H}_{01}$  in subperiod 1994-2009.

We also test the hypothesis of no predictability by the variance risk premium,  $\mathcal{H}_{02} : \beta_{02} = 0$ . Figure 9 plots the resulting confidence intervals for parameter  $\beta_{02}$ , while Table 4 reports the estimates and testing results. In contrast to the previous evidence, we find that all tests under investigation clearly reject  $\mathcal{H}_{02}$  at the 5%-significance level, thus confirming the remarkable return forecasting ability of the variance risk premium noticed in Bollerslev, Tauchen and Zhou (2009), as well as the international evidence reported in Bollerslev, Marrone, Xu and Zhou (2012).<sup>26</sup>

---

<sup>26</sup>Besides the two-predictor model (16), we also consider the three-predictor model

$$\frac{1}{k} \ln(R_{t+k,t}) = \alpha + \beta_1 \ln\left(\frac{D_t}{P_t}\right) + \beta_2 VRP_t + \beta_3 LTY_t + \epsilon_{t+k,t}, \quad (17)$$

where  $LTY_t$  is the detrended long-term yield, defined as the ten-year Treasury yield minus its trailing twelve-month moving averages. Again, using the nonrobust bootstrap, the nonrobust subsampling, the robust bootstrap and the robust subsampling, we find evidence in favor of predictability at 5% significance level for the variance risk premium for the sample period 1990-2010. In contrast, all tests do not reject the null hypothesis of no predictability at 10% significance level for the detrended long-term yield. Finally, both conventional and

### 5.2.2 Santos and Veronesi

We finally focus on the two-predictor regression model proposed in Santos and Veronesi (2006):

$$\ln(R_t) = \alpha + \beta_1 \ln\left(\frac{D_{t-1}}{P_{t-1}}\right) + \beta_2 s_{t-1} + \epsilon_t, \quad (18)$$

where  $s_{t-1} = w_{t-1}/C_{t-1}$  is the share of labor income to consumption. We make use of quarterly returns on the value weighted CRSP index, which includes NYSE, AMEX, and NASDAQ stocks, in the sample period Q1,1955-Q4,2010. The dividend time-series is also obtained from CRSP, while the risk free rate is the three-months Treasury bill rate. Labor income and consumption are obtained from the Bureau of Economic Analysis.<sup>27</sup>

Let  $\beta_{01}$  and  $\beta_{02}$  denote the true values of parameters  $\beta_1$  and  $\beta_2$ , respectively. Using bootstrap and subsampling tests, as well as our robust testing method, we first test the null hypothesis of no predictability by the dividend yield,  $\mathcal{H}_{01} : \beta_{01} = 0$ . Figure 10 plots the 90%-confidence intervals for parameter  $\beta_{01}$  based on rolling windows of 180 quarterly observations in sample period 1950-2010, while Table 5 collects detailed point estimates and test results for the four subperiods 1950-1995, 1955-2000, 1960-2005, 1965-2010. We find again that our robust tests always clearly reject  $\mathcal{H}_{01}$  at the 5%-significance level. In contrast, conventional tests produce more ambiguous results, and cannot reject at the 10%-significance level the null hypothesis  $\mathcal{H}_{01}$  for subperiod 1955-2000.

Figure 11 reports the 90%-confidence intervals estimated in tests of the null hypothesis of no predictability by the labor income proxy,  $\mathcal{H}_{02} : \beta_{02} = 0$ . Table 5 summarizes estimation and testing results. While the conventional tests produce a weak and mixed evidence of return predictability using labor income proxies, e.g., by not rejecting  $\mathcal{H}_{02}$  at the 10%-level in subperiod 1950-1995, the robust tests produce once more a clear and consistent predictability evidence

---

robust tests reject the null hypothesis of no predictability at the 5% significance level for the dividend yield. The comparison of these empirical results with those obtained in the two-predictor model (16) again confirms the reliability of our robust tests and the (possible) failure of nonrobust procedures in uncovering predictability structures in presence of anomalous observations.

<sup>27</sup>As in Lettau and Ludvigson (2001), labor income is defined as wages and salaries, plus transfer payments, plus other labor income, minus personal contributions for social insurance, minus taxes. Consumption is defined as non-durables plus services.

for all sample periods.

The clusters of anomalous observations (less than 4.6% of the data in the full sample), highlighted by the estimated weights in Figure 14, further indicate that conventional tests might fail to uncover hidden predictability structures using samples of data that include particularly influential observations from the NASDAQ bubble or the recent financial crisis. Such a feature was noted already in Santos and Veronesi (2006) and Lettau and Van Nieuwerburgh (2007) from a completely different angle. Here, the robust bootstrap and subsampling tests are again found to control well the potential damaging effects of anomalous observations. They provide an effective way to consistently uncover hidden predictability features when the data may only approximately follow the given predictive regression model.

## 6 Conclusion

Predictive relations motivated by financial arguments should hold at least for the majority of the data, but they can be realistically violated in a small subset of observations. A large literature studies from different angles such predictive relations. A major focus in this research is on the predictive ability of a variety of economic variables for future market returns.

Several useful approaches for testing the null of no predictability in predictive regressions with, e.g., correlated errors and nearly integrated regressors, have been proposed, including tests that rely on nonparametric Monte Carlo simulation methods, such as the bootstrap and the subsampling. All these methods improve on the conventional asymptotic tests under the ideal assumption that the predictive regression model is satisfied by the totality of the data. However, they often fail to detect predictive relations that hold for the vast majority of the data, because rare anomalous observations can have an excessive influence on the inference results. We propose a novel robust method for testing the null of no predictability, based on predictive regression models that might be violated by a minority of anomalous data points.

We start by documenting theoretically and by Monte Carlo simulation the lack of robustness of conventional tests of predictability to the effects of anomalous data. We find that even though the probability of rejecting a null by chance alone features some degree of resistance, the test

ability to detect predictive relations valid for the vast majority of the data can be surprisingly low.

We propose a new robust testing method for predictive regressions by introducing a novel general class of fast and robust bootstrap and subsampling procedures for times series. These procedures are applicable to linear and nonlinear predictive regression models at sustainable computational costs. Our bootstrap and subsampling tests are resistant to anomalous observations in the data and imply more robust confidence intervals and inference results. We demonstrate theoretically and by Monte Carlo simulations their improved accuracy and robustness properties.

In the empirical study for US stock market data, we study single-predictor and multi-predictor models, using well-known predictive variables in the literature, such as the market dividend yield, the difference between index option-implied volatility and realized volatility (Bollerslev, Tauchen and Zhou, 2009), and the share of labor income to consumption (Santos and Veronesi, 2006).

Using the robust tests, we obtain new unambiguous evidence of market return predictability. First, we find a clear-cut evidence that the dividend yield is a robust predictive variable for the vast majority of market returns, in each subperiod and for each sampling frequency and forecasting horizon considered. In contrast, conventional tests based on, e.g., bias-corrections, near-to-unity asymptotics, or standard resampling procedures provide ambiguous findings, by not rejecting the null of no predictability in a number of cases.

Second, we find that the difference between option-implied volatility and realized volatility is a robust predictive variable of future market returns at quarterly forecasting horizons, both using robust and nonrobust testing methods. This finding confirms the remarkable return forecasting ability of the variance risk premium, first noticed in Bollerslev, Tauchen and Zhou (2009).

Third, we find that conventional testing approaches deliver an ambiguous evidence of return predictability by proxies of labor income (Santos and Veronesi, 2006), which is either absent or weak in the sample periods 1955-2000 and 1965-2010, respectively. In contrast, the null of no predictability is always clearly rejected using the robust testing approach, indicating

that the weak findings of the conventional tests are likely deriving from their low ability to detect predictive relations valid for the vast majority of the data, in presence of rare anomalous observations.

Fourth, we exploit the properties of our robust tests to identify *ex post*, not *ex ante*, anomalous observations that explain the diverging conclusions of robust and nonrobust methods. We find a fraction of less than about 5% of anomalous observations in the data, which tend to cluster during the NASDAQ bubble and the more recent financial crisis. A small subset of these data points, including the Lehman Brothers default on September 2008, the terrorist attack of September 2001, the Black Monday on October 1987, and the Dot-Com bubble collapse in August 2002, is largely responsible for the failure of conventional testing methods in uncovering the hidden predictability structures.

In summary, this evidence shows that a number of key predictive relations implied by finance theory are well supported by the vast majority of the data. At the same time, infrequent anomalous deviations from the assumed predictive regression model can substantially distort the inference of conventional testing methods, thus leading to misleading conclusions about the existence of predictive relations overall. Anomalous data points also appear to be extraordinarily difficult to model, as they seem to be mostly related to sporadic and very turbulent market events. Such events are unlikely to be predictable in a reliable way using the conditioning information typically specified by most predictive regression models.

Finally, while our bootstrap and subsampling tests have been developed in the context of standard predictive systems with autocorrelated regressors, our approach is extendable also to more general settings, including potential nonlinear predictive relations or unobserved state variables. For instance, van Binsbergen and Koijen (2010) propose a latent-variable approach and a Kalman filter to estimate a present value model with hidden and persistent expected return and dividend growth, in order to formulate powerful tests for the joint predictability of stock returns and dividend growth. The application of our robust bootstrap and subsampling tests in the context of such present value models is an interesting avenue for future research.



## Appendix A: Quantile Breakdown Points

We first introduce formally the breakdown point  $b_{t,n,m}^K$  of the quantile  $Q_{t,n,m}^{K*}$  of the bootstrap ( $K = B$ ) and subsampling ( $K = S$ ) distributions of statistic  $T_{n,m}^{K*}$  defined in (4). Then, we derive upper and lower bounds for  $b_{t,n,m}^K$  as a function of the sample size  $n$ , the block size  $m$ , the quantile  $t$ , and the breakdown point  $b$  of statistic  $T_n$ . Finally, we also consider robust bootstrap and robust subsampling distributions, and compute their quantile breakdown points.

### Definition

The breakdown point of quantile (4) is the smallest fraction of outliers in the original sample such that  $Q_{t,n,m}^{K*}$  diverges to infinity. Borrowing the notation in Genton and Lucas (2003), we formally define the breakdown point of the  $t$ -quantile  $Q_{t,n,m}^{K*} := Q_{t,n,m}^{K*}(X_{(n)})$  as,

$$b_{t,n,m}^K := \frac{1}{n} \cdot \left[ \inf_{\{1 \leq p \leq \lceil n/2 \rceil\}} \{p \mid \text{there exists } Z_{n,p}^\zeta \in \mathcal{Z}_{n,p}^\zeta \text{ such that } Q_{t,n,m}^{K*}(X_{(n)} + Z_{n,p}^\zeta) = +\infty\} \right], \quad (19)$$

where  $\lceil x \rceil = \inf\{n \in \mathbb{N} \mid x \leq n\}$ , and  $\mathcal{Z}_{n,p}^\zeta$  denotes the set of all  $n$ -samples  $Z_{n,p}^\zeta$  with exactly  $p$  non-zero components that are  $d_x$ -dimensional outliers of size  $\zeta \in \bar{\mathbb{R}}^{d_x}$ .<sup>28</sup> Literally,  $b_{t,n,m}^K$  is the smallest fraction of anomalous observations of arbitrary size, in a generic outlier-contaminated sample  $X_{(n)} + Z_{n,p}^\zeta$ , such a quantile  $Q_{t,n,m}^{K*}$ , estimated by a bootstrap or a subsampling Monte Carlo simulation scheme, can become meaningless.

### Quantile Breakdown Point Bounds

In Theorem 2, we compute explicit quantile breakdown point bounds as a function of the sample size  $n$ , the block size  $m$ , the quantile  $t$ , and the breakdown point  $b$  of statistic  $T_n$ .

---

<sup>28</sup>When  $p > 1$ , we do not necessarily assume outliers  $\zeta_1, \dots, \zeta_p$  to be all equal to  $\zeta$ , but we rather assume existence of constants  $c_1, \dots, c_p$ , such that  $\zeta_i = c_i \zeta$ . To better capture the presence of outliers in predictive regression models, our definitions for the breakdown point and the set  $\mathcal{Z}_{n,p}^\zeta$  of all  $n$ -components outlier samples are slightly different from those proposed in Genton and Lucas (2003) for general settings. However, we can modify our results to cover alternative definitions of breakdown point and outlier sets  $\mathcal{Z}_{n,p}^\zeta$ .

**Theorem 2** Let  $b$  be the breakdown point of  $T_n$  and  $t \in (0, 1)$ . The quantile breakdown point  $b_{t,n,m}^S$  and  $b_{t,n,m}^B$  of subsampling and block bootstrap procedures, respectively, satisfy following bounds,

$$\frac{\lceil mb \rceil}{n} \leq b_{t,n,m}^S \leq \frac{1}{n} \cdot \left[ \inf_{\{p \in \mathbb{N}, p \leq r-1\}} \left\{ p \cdot \lceil mb \rceil \left| p > \frac{(1-t)(n-m+1) + \lceil mb \rceil - 1}{m} \right. \right\} \right], \quad (20)$$

$$\frac{\lceil mb \rceil}{n} \leq b_{t,n,m}^B \leq \frac{1}{n} \cdot \left[ \inf_{\{p_1, p_2\}} \left\{ p = p_1 \cdot p_2 \left| P \left( \text{BIN} \left( r, \frac{mp_2 - p_1 + 1}{n - m + 1} \right) \geq \left\lceil \frac{nb}{p_1} \right\rceil \right) > 1 - t \right. \right\} \right], \quad (21)$$

where  $p_1, p_2 \in \mathbb{N}$ , with  $p_1 \leq m, p_2 \leq r - 1$ .

**Proof of Theorem 2.** We first consider the subsampling and focus on formula (20). The value  $\frac{\lceil mb \rceil}{n}$  is the smallest fraction of outliers, that causes the breakdown of statistic  $T$  in a block of size  $m$ . Therefore, the first inequality is satisfied.

For the second inequality of formula (20), we denote by  $X_{(m),i}^N = (X_{(i-1)m+1}, \dots, X_{im})$ ,  $i = 1, \dots, r$  and  $X_{(m),i}^O = (X_i, \dots, X_{i+m-1})$ ,  $i = 1, \dots, n-m+1$ , the nonoverlapping and overlapping blocks of size  $m$ , respectively. Given the original sample  $X_{(n)}$ , for the first nonoverlapping block  $X_{(m),1}^N$ , consider the following type of contamination:

$$X_{(m),1}^N = (X_1, \dots, X_{m-\lceil mb \rceil}, Z_{m-\lceil mb \rceil+1}, \dots, Z_m), \quad (22)$$

where  $X_i$ ,  $i = 1, \dots, m - \lceil mb \rceil$  and  $Z_j$ ,  $j = m - \lceil mb \rceil + 1, \dots, m$ , denote the noncontaminated and contaminated points, respectively. By construction, the first  $m - \lceil mb \rceil + 1$  overlapping blocks  $X_{(m),i}^O$ ,  $i = 1, \dots, m - \lceil mb \rceil + 1$ , contain  $\lceil mb \rceil$  outliers. Consequently,  $T(X_{(m),i}^O) = +\infty$ ,  $i = 1, \dots, m - \lceil mb \rceil + 1$ . Assume that the first  $p < r - 1$  nonoverlapping blocks  $X_{(m),i}^N$ ,  $i = 1, \dots, p$ , have the same contamination as in (22). Because of this contamination, the number of statistics  $T_{n,m}^{OS*}$  which diverge to infinity is  $mp - \lceil mb \rceil + 1$ .

$Q_{t,n,m}^{OS*} = +\infty$ , when the proportion of statistics  $T_{n,m}^{OS*}$  with  $T_{n,m}^{OS*} = +\infty$  is larger than  $(1-t)$ . Therefore,

$$b_{t,n,m}^{OS} \leq \inf_{\{p \in \mathbb{N}, p \leq r-1\}} \left\{ p \cdot \frac{\lceil mb \rceil}{n} \left| \frac{mp - \lceil mb \rceil + 1}{n - m + 1} > 1 - t \right. \right\}.$$

Finally, we consider formula (21). The proof of the first inequality in formula (21) follows the same lines as the proof of the first inequality in the formula (20). We focus on the second inequality.

Consider  $X_{(m),i}^N$ ,  $i = 1, \dots, r$ . Assume that  $p_2$  of these nonoverlapping blocks are contaminated with exactly  $p_1$  outliers for each block, while the remaining  $(r - p_2)$  are noncontaminated (0 outlier), where  $p_1, p_2 \in \mathbb{N}$  and  $p_1 \leq m$ ,  $p_2 \leq r - 1$ . Moreover, also assume that the contamination of the  $p_2$  contaminated blocks has the structure defined in (22). The block bootstrap constructs a  $n$ -sample randomly selecting with replacement  $r$  overlapping blocks of size  $m$ . Let  $X$  be the random variable which denotes the number of contaminated blocks in the random bootstrap sample. It follows that  $X \sim \text{BIN}(r, \frac{mp_2 - p_1 + 1}{n - m + 1})$ .

By Equation (19),  $Q_{t,n,m}^{OB*} = +\infty$ , when the proportion of statistics  $T_{n,m}^{OB*}$  with  $T_{n,m}^{OB*} = +\infty$  is larger than  $(1 - t)$ . The smallest number of outliers such that  $T_{n,m}^{OB*} = +\infty$  is by definition  $nb$ . Let  $p_1, p_2 \in \mathbb{N}$ ,  $p_1 \leq m$ ,  $p_2 \leq r - 1$ . Consequently,

$$b_{t,n,m}^{OB} \leq \frac{1}{n} \cdot \left[ \inf_{\{p_1, p_2\}} \left\{ p = p_1 \cdot p_2 \left| P\left( \text{BIN}\left( r, \frac{mp_2 - p_1 + 1}{n - m + 1} \right) \geq \left\lceil \frac{nb}{p_1} \right\rceil \right) > 1 - t \right\} \right].$$

This concludes the proof of Theorem 2. ■

## Robust Subsampling and Robust Bootstrap Distributions

After the definition of quantile breakdown point and the robustness analysis of the conventional bootstrap and subsampling, we consider robust subsampling and robust bootstrap distributions. In the next Corollary, we compute the quantile breakdown point of the robust fast resampling distribution  $L_{n,m}^{K*}$  introduced in Definition 1.

**Corollary 3** *Let  $b$  be the breakdown point of the robust  $M$ -estimator  $\hat{\theta}_n$  defined in (6). The  $t$ -quantile breakdown point of resampling distribution (8) equals  $b_{t,n,m}^K = \min(b, b_{\nabla\psi})$ , where  $b_{\nabla\psi}$  is the breakdown point of matrix  $\nabla_{\theta}\psi_n(X_{(n)}, \hat{\theta}_n)$ , defined by:*

$$b_{\nabla\psi} = \frac{1}{n} \cdot \inf_{1 \leq p \leq \lceil n/2 \rceil} \left\{ p \mid \text{there exists } Z_{n,p}^{\zeta} \in \mathcal{Z}_{n,p}^{\zeta} \text{ such that } \det(\nabla_{\theta}\psi_n(X_{(n)} + Z_{n,p}^{\zeta}, \hat{\theta}_n)) = 0 \right\}. \quad (23)$$

**Proof of Corollary 3.** Consider the robust fast approximation of  $\hat{\theta}_{n,m}^{K*} - \hat{\theta}_n$  given by:

$$-[\nabla_{\theta}\psi_n(X_{(n)}, \hat{\theta}_n)]^{-1}\psi_k(X_{(n,m),s}^{K*}, \hat{\theta}_n), \quad (24)$$

where  $k = n$  or  $k = m$ ,  $K = B, S$ . Assuming a bounded estimating function, Expression (24) may degenerate only when either (i)  $\hat{\theta}_n \notin \mathbb{R}$  or (ii) matrix  $[\nabla_{\theta}\psi_n(X_{(n)}, \hat{\theta}_n)]$  is singular, i.e.,  $\det([\nabla_{\theta}\psi_n(X_{(n)}, \hat{\theta}_n)]) = 0$ . If (i) and (ii) are not satisfied, then, quantile  $Q_{t,n,m}^{K*}$  is bounded, for all  $t \in (0, 1)$ . Let  $b$  be the breakdown point of  $\hat{\theta}_n$  and  $b_{\nabla\psi}$  the smallest fraction of outliers in the original sample such that condition (ii) is satisfied. Then, the breakdown point of  $Q_{t,n,m}^{K*}$  is  $b_{t,n,m}^K = \min(b, b_{\nabla\psi})$ . ■

Using the results in Corollary 3, we analyze the robustness properties of the robust subsampling and robust bootstrap for predictive regression model. In particular, in Corollary 4, we compute the quantile breakdown point of the robust subsampling and robust bootstrap for the sampling distribution of the nonstudentized statistic  $T_n^{NS} = \sqrt{n}(\hat{\theta}_n^R - \theta_0)$  defined in (13).

**Corollary 4** *Let  $t \in (0, 1)$ . The  $t$ -quantile breakdown point of the resampling distribution (13) is  $b_{t,n,m}^K = 0.5$ ,  $K = B, S$ .*

**Proof of Corollary 4.** First note that the breakdown point of the robust estimator  $\hat{\theta}_n^R$  defined in (10) is  $b = 0.5$ ; see, e.g., Huber (1984). Therefore, we have only to focus on the breakdown point of matrix  $\nabla_{\theta}(\psi_{n,c}(z_{(n)}, \hat{\theta}_n^R))$ . Without loss of generality, assume that  $\theta = (\alpha, \beta) \in \mathbb{R}^2$ . Consider the function

$$g_c(y_t, w_{t-1}, \theta) = (y_t - \theta'w_{t-1})w_{t-1} \cdot \min\left(1, \frac{c}{\|(y_t - \theta'w_{t-1})w_{t-1}\|}\right). \quad (25)$$

Using some algebra, we can show that

$$\nabla_{\theta}g_c(y_t, w_{t-1}, \theta) = \begin{cases} -(1, x_{t-1})'(1, x_{t-1}), & \text{if } \|(y_t - \theta'w_{t-1})w_{t-1}\| \leq c, \\ \mathbb{O}_{2 \times 2}, & \text{if } \|(y_t - \theta'w_{t-1})w_{t-1}\| > c, \end{cases} \quad (26)$$

where  $\mathbb{O}_{2 \times 2}$  denotes the  $2 \times 2$  null matrix. It turns out that by construction the matrix  $\nabla_{\theta}(\psi_{n,c}(z_{(n)}, \hat{\theta}_n^R))$  is semi-positive definite, and in particular  $\det(\nabla_{\theta}(\psi_{n,c}(z_{(n)}, \hat{\theta}_n^R))) = 0$ , only when  $\|(y_t - \hat{\theta}_n^R w_{t-1})w_{t-1}\| > c$ , for all the observations  $(y_t, w_{t-1})'$ , i.e.,  $b_{\nabla\psi_c} = 1$ . Therefore, using Corollary 3 we obtain that  $b_{t,n,m}^K = \min(0.5, 1) = 0.5$ ,  $K = B, S$ . ■

Finally, consider the studentized statistic  $T_n = \sqrt{n}[\hat{\Sigma}_n^R]^{-1/2}(\hat{\theta}_n^R - \theta_0)$ . Here, we can use the robust fast approach introduced in Definition 1 with minor modifications. In particular, we propose to estimate the sampling distribution of statistic  $T_n$  by the following robust fast resampling distribution:

$$L_{n,m}^{K*}(x) = \frac{1}{N} \sum_{s=1}^N \mathbb{I} \left( \sqrt{k} \left( - [\hat{\Sigma}_{k,s}^{R*}]^{-1/2} [\nabla_{\theta} \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n)]^{-1} \psi_k(z_{(n,m),s}^{K*}, \hat{\theta}_n) \right) \leq x \right), \quad (27)$$

where  $k = n$  for the block bootstrap ( $k = m$  for the subsampling) and  $\hat{\Sigma}_{k,s}^{R*} = \hat{\Sigma}_n^R(z_{(n,m),s}^{K*})$  is an estimate of the asymptotic variance of the robust M-estimator (10) based on the  $s$ -th block bootstrap (subsampling) random block.

The quantile breakdown point properties of resampling distribution (27) are more complex than those obtained in the unstudentized case, and are summarized in the next Theorem.

**Theorem 5** *For simplicity, let  $r = n/m \in \mathbb{N}$ . The  $t$ -quantile breakdown points  $b_{t,n,m}^B$  and  $b_{t,n,m}^S$  of the robust block bootstrap and robust subsampling distributions (27), respectively, are given by*

$$b_{t,n,m}^S = \frac{1}{n} \left[ \inf_{\{p \in \mathbb{N}, p \leq n-m+1\}} \left\{ m + p \mid p > (1-t)(n-m+1) - 1 \right\} \right], \quad (28)$$

$$b_{t,n,m}^B = \frac{1}{n} \left[ \inf_{\{p \in \mathbb{N}, p \leq n-m+1\}} \left\{ m + p \mid P \left( \text{BIN} \left( r, \frac{p+1}{n-m+1} \right) = r \right) > 1-t \right\} \right], \quad (29)$$

where  $\text{BIN}(N, q)$  is a Binomial random variable with parameters  $N$  and  $q \in (0, 1)$ .

**Proof of Theorem 5.** Consider the resampling distribution (27). Let  $z_{(n,m),s}^{K*} = (z_1^*, \dots, z_k^*)$  denote a random bootstrap ( $K = B$  and  $k = n$ ) or subsampling ( $K = S$  and  $k = m$ ) sample.

Since the estimating function  $\psi_{n,c}$  is bounded, it turns out that

$$T_{n,m,s}^{K*} := [\hat{\Sigma}_{k,s}^{R*}]^{-1/2} [\nabla_{\theta} \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n^R)]^{-1} \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n^R), \quad (30)$$

may degenerate when (i)  $\det(\hat{\Sigma}_{k,s}^{R*}) = 0$  or (ii)  $\det(\nabla_{\theta} \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n^R)) = 0$ . Moreover, also note that  $\hat{\Sigma}_{k,s}^{R*} = \hat{J}'_{k,s} \hat{V}_{k,s}^{R*} \hat{J}_{k,s}^{R*}$ , where  $\hat{J}_{k,s}^{R*} = [\nabla_{\theta} \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n^R)]^{-1}$ . Because of Equations (37), (38), (26), it turns out that cases (i) and (ii) can be satisfied only when  $\|(y_t^* - \hat{\theta}_n^R w_{t-1}^*) w_{t-1}^*\| > c$  for all random observations  $z_i^* = (y_i^*, w_{i-1}^*)'$ ,  $i = 1, \dots, k$ , where  $k = n, m$ , for the bootstrap and subsampling, respectively.

For the original sample, consider following type of contamination

$$z_{(n)} = (z_1, \dots, z_j, C_{j+1}, \dots, C_{j+p}, z_{j+p+1}, \dots, z_n), \quad (31)$$

where  $z_i$ ,  $i = 1, \dots, j$  and  $i = j + p + 1, \dots, n$  and  $C_i$ ,  $i = j + 1, \dots, j + p$ , denote the noncontaminated and contaminated points, respectively, where  $p \geq m$ . It turns out that all the  $p - m + 1$  overlapping blocks of size  $m$

$$(C_{j+i}, \dots, C_{j+i+m-1}), \quad (32)$$

$i = 1, \dots, p - m + 1$  contain only outliers. Therefore, for these  $p - m + 1$  blocks we have that  $\det(\nabla_{\theta} \psi_{m,c}(C_{j+i}, \dots, C_{j+i+m-1}, \hat{\theta}_n^R)) = 0$ , i.e., some components of vector (30) may degenerate to infinity. Moreover,  $Q_{t,n,m}^{S*} = +\infty$  when the proportion of statistics  $T_{n,m}^{S*}$  with  $T_{n,m}^{S*} = +\infty$  is larger than  $(1 - t)$ . Therefore,  $b_{t,n,m}^S = \inf_{\{p \in \mathbb{N}, m \leq p \leq n - m + 1\}} \left\{ \frac{p}{n} \left| \frac{p - m + 1}{n - m + 1} > 1 - t \right. \right\}$ , which proves the result in Equation (28).

For the result in Equation (29), note that because of the contamination defined in (31), by construction we have  $p - m + 1$  overlapping blocks of size  $m$  with exactly  $m$  outliers, and  $n - (p - m + 1)$  blocks with less than  $m$  outliers. Let  $X$  be the random variable which denotes the number of full contaminated blocks in the random bootstrap sample. It follows that  $X \sim \text{BIN}(r, \frac{p - m + 1}{n - m + 1})$ . To imply (i) or (ii), all the random observations  $(z_1^*, \dots, z_k^*)$  have

to be outliers, i.e.,  $X = r$ . By Equation (19),  $Q_{t,n,m}^{B*} = +\infty$ , when the proportion of statistics  $T_{n,m}^{B*}$  with  $T_{n,m}^{B*} = +\infty$  is larger than  $(1 - t)$ . Consequently,

$$b_{t,n,m}^B = \frac{1}{n} \cdot \left[ \inf_{\{p \in \mathbb{N}, p \leq n-m+1\}} \left\{ p \left| P \left( \text{BIN} \left( r, \frac{p-m+1}{n-m+1} \right) = r \right) > 1-t \right. \right\} \right].$$

This concludes the proof. ■

Formulas (28) and (29) improve on the results in Equations (20) and (21) for the conventional subsampling and bootstrap, respectively. The quantile breakdown point of the robust block bootstrap and subsampling approach is often much higher than the one of conventional resampling methods. Table 2 quantifies these differences. For instance, for  $m = 10, 20$ , the 0.95-quantile of the robust block bootstrap is maximal. Similarly, the robust subsampling quantile breakdown points in Table 2 are considerably larger than those in Table 1 for conventional subsampling methods, even if they do not always attain the upper bound of 0.5.

In contrast to the unstudentized case, the quantile breakdown point of robust resampling distribution (27) is not always maximal, because we need to compute matrices  $\hat{\Sigma}_k^{R*}$  and  $\nabla_{\theta} \psi_{k,c}(z_{(n,m)}^{K*}, \hat{\theta}_n^R)$  in each bootstrap or subsampling block. While this approach can yield consistency also in nonstationary settings and a potentially improved convergence, the additional estimation step can imply a loss in the resistance of the whole procedure to anomalous observations. Thus, a tradeoff arises between resistance to anomalous observations and improved finite-sample inference, which has to be considered and evaluated case-by-case in applications.

## Appendix B: Robust Bootstrap and Subsampling Tests of Predictability

We present in more details how to compute the robust resampling distribution (27) and construct robust confidence intervals for the components of a general  $d$ -dimensional parameter  $\beta$ , where  $d \geq 1$ . The extension to the nonstudentized distributions (13) is straightforward. We construct symmetric resampling confidence intervals for the parameter of interest.<sup>29</sup>

Let  $\theta = (\alpha, \beta)'$  and let  $z_{(n)} = (z_1, \dots, z_n)$  be an observation sample generated according to the multi-predictor regression model

$$\begin{aligned} y_t &= \alpha + \beta'x_{t-1} + u_t, \\ x_t &= \Phi + Rx_{t-1} + V_t, \end{aligned}$$

where  $z_t = (y_t, w'_{t-1})'$ ,  $w_{t-1} = (1, x'_{t-1})'$ ,  $\Phi$  is a  $d$ -dimensional parameter vector and  $R$  is a  $d \times d$  parameter matrix. First, we compute the robust Huber estimator  $\hat{\theta}_n^R = (\hat{\alpha}_n^R, \hat{\beta}_n^R)'$  as the solution of  $\psi_{n,c}(z_{(n)}, \hat{\theta}_n^R) = 0$ , where the estimating function  $\psi_{n,c}$  is defined in (10) and  $c > 0$  is a tuning constant selected through the data-driven method introduced in Appendix C below. Note that the asymptotic variance of the robust M-estimator  $\hat{\theta}_n^R$  is given by

$$\Sigma^R = J'^R V^R J^R, \tag{33}$$

where  $J^R = \left( \lim_{n \rightarrow \infty} E [\nabla_{\theta} \psi_{n,c}(z_{(n)}, \theta_0)] \right)^{-1}$ ,  $V^R = \lim_{n \rightarrow \infty} Var \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n g_c(z_{(i)}, \theta_0) \right]$ , and  $\nabla_{\theta} \psi_{n,c}(z_{(n)}, \theta_0)$  denotes the derivative of function  $\psi_{n,c}$  with respect to parameter  $\theta$  with  $\theta_0 =$

---

<sup>29</sup>Hall (1988) and more recent contributions, as for instance Politis, Romano and Wolf (1999), highlight a better accuracy of symmetric confidence intervals, which even in asymmetric settings can be shorter than asymmetric confidence intervals. Mikusheva (2007) and Andrews and Guggenberger (2010) also show that because of a lack of uniformity in pointwise asymptotics, nonsymmetric subsampling confidence intervals for autoregressive models can imply a distorted asymptotic size, which is instead correct for symmetric confidence intervals.



$(\alpha_0, \beta_0)'$ . We compute the estimator  $\hat{\Sigma}_n^R = \hat{J}_n^{R'} \hat{V}_n^R \hat{J}_n^R$  of the asymptotic variance  $\Sigma^R$  where

$$\hat{J}_n^R = \left[ \nabla_{\theta} \psi_{n,c}(z_{(n)}, \hat{\theta}_n^R) \right]^{-1}, \quad (34)$$

$$\hat{V}_n^R = \frac{1}{n} \sum_{i=1}^n g_c(z_i, \hat{\theta}_n^R) g_c(z_i, \hat{\theta}_n^R)', \quad (35)$$

respectively.<sup>30</sup> Let  $(p)^{(j)}$  denotes the  $j$ -th component of a  $d$ -dimensional vector  $p$ ,  $1 \leq j \leq d$ . To construct symmetric confidence intervals for each  $j$ -th component of the parameter  $\beta$ , we compute an estimator of the sampling distribution of  $(T_{n,|\cdot|})^{(j+1)} := \frac{\sqrt{n}}{(\hat{\sigma}_n^R)^{(j+1)}} \left| \left( \hat{\theta}_n^R - \theta_0 \right)^{(j+1)} \right|$ , by applying our robust resampling approach, where the subscript  $|\cdot|$  indicates that we consider the absolute value of each component of the statistic  $T_n$ , and  $(\hat{\sigma}_n^R)^{(j+1)}$  is the square root of the  $(j+1)$ -th diagonal component of matrix  $\hat{\Sigma}_n^R$ .

More precisely, let  $z_{(n,m)}^{K*}$  be a block bootstrap ( $K = B$ ) or subsampling ( $K = S$ ) random sample based on blocks of size  $m$ , where  $m$  is selected through the data-driven method introduced in Appendix C below. For each component  $1 \leq j \leq d$ , we estimate the sampling distribution of  $(T_{n,|\cdot|})^{(j+1)}$  through the robust resampling distribution

$$(L_{n,m,|\cdot|}^{K*}(x))^{(j+1)} = \frac{1}{N} \sum_{s=1}^N \mathbb{I} \left( \frac{\sqrt{k}}{(\hat{\sigma}_{k,s}^{R*})^{(j+1)}} \left| \left( [\nabla_{\theta} \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n^R)]^{-1} \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n^R) \right)^{(j+1)} \right| \leq x \right), \quad (36)$$

where  $k = n$  for the block bootstrap ( $K = B$ ), while  $k = m$  for the subsampling ( $K = S$ ),  $(\hat{\sigma}_{k,s}^{R*})^{(j+1)}$  is the square root of the  $(j+1)$ -th diagonal component of matrix  $\hat{\Sigma}_{k,s}^{R*}$ , and  $\hat{\Sigma}_{k,s}^{R*} = \hat{J}_{k,s}^{R*'} \hat{V}_{k,s}^{R*} \hat{J}_{k,s}^{R*}$  with

$$\hat{J}_{k,s}^{R*} = \left[ \nabla_{\theta} \psi_{c,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n^R) \right]^{-1}, \quad (37)$$

$$\hat{V}_{k,s}^{R*} = \frac{1}{k} \sum_{i=1}^k g_c(z_{i,s}^{K*}, \hat{\theta}_n^R) g_c(z_{i,s}^{K*}, \hat{\theta}_n^R)'. \quad (38)$$

---

<sup>30</sup>For the sake of brevity, we assume that  $E[g_c(z_i, \theta_0) g_c(z_j, \theta_0)'] = 0$ ,  $i \neq j$ , and consequently we consider  $\hat{V}_n^R$  as estimator of  $V^R$ . However, if this assumption is not satisfied, then  $\hat{V}_n^R$  has to be replaced with the Newey-West covariance estimator.

Finally, let  $t \in (0, 1)$ , and let  $(Q_{t,n,m}^{K*})^{(j+1)}$  be the  $t$ -quantile of the block bootstrap or subsampling empirical distribution (36),  $K = B, S$ , respectively. Then, the symmetric  $t$ -confidence interval for the  $j$ -th component  $(\beta_0)^{(j)}$  of  $\beta_0$  is given by

$$(CI_t)^{(j)} = \left[ \left( \hat{\theta}_n^R \right)^{(j+1)} - \left( \hat{\sigma}_n^R \right)^{(j+1)} \left( Q_{t,n,m}^{K*} \right)^{(j+1)}, \left( \hat{\theta}_n^R \right)^{(j+1)} + \left( \hat{\sigma}_n^R \right)^{(j+1)} \left( Q_{t,n,m}^{K*} \right)^{(j+1)} \right], \quad (39)$$

We summarize our robust approach in the following algorithm.

- (1) Compute  $\hat{\theta}_n^R = (\hat{\alpha}_{ROB}, \hat{\beta}'_{ROB})'$ , as the solution of (10), where  $c$  is selected using the data-driven method introduced in Appendix C below.
- (2) Compute  $\hat{\Sigma}_n^R = \hat{J}_n^R \hat{V}_n^R \hat{J}_n^R$ , where  $\hat{J}_n^R$  and  $\hat{V}_n^R$  are defined in (34) and (35), respectively.
- (3B) For the robust block bootstrap, generate  $B_B = 999$  random bootstrap samples based on the overlapping blocks  $(z_i, \dots, z_{i+m-1})$ ,  $i = 1, \dots, n-m+1$ , where  $m$  is selected according to the data-driven method introduced in Appendix C below.
- (3S) For the robust subsampling, consider the  $B_S = n-m+1$  overlapping blocks  $(z_i, \dots, z_{i+m-1})$ ,  $i = 1, \dots, n-m+1$ , where  $m$  is selected according to the data-driven method introduced in Appendix C below.
- (4) Compute the robust resampling distribution (36).
- (5) The robust symmetric  $t$ -confidence interval for the  $j$ -th component  $(\beta_0)^{(j)}$  of  $\beta_0$  is given by (39).

## Appendix C: Data-Driven Selection of the Block Size and the Robust Estimating Function Bound

The implementation of our robust resampling methods requires the selection of the block size  $m$  and the degree of robustness  $c$  of the estimating function (10). To this end, we propose a data-driven procedure for the choice of the block size  $m$  and the estimating function bound  $c$ , by extending the calibration method (CM) discussed in Romano and Wolf (2001) in relation to subsampling procedures.

Let  $\mathcal{MC} := \{(m, c) | m \in \mathcal{M}, c \in \mathcal{C}\}$ , where  $\mathcal{M} := \{m_{min} < \dots < m_{max}\}$  and  $\mathcal{C} := \{c_{min} < \dots < c_{max}\}$  are the sets of admissible block sizes and estimating functions bounds, respectively. Let  $T_n^{NS} = \sqrt{n} (\hat{\theta}_n^R - \theta_0)$  be the nonstudentized statistic of interest, where  $\hat{\theta}_n^R$  is the robust Huber estimator solution of Equation (10) with  $c = c_1$  fixed, as preliminary value of the estimating function bound. Furthermore, let  $(X_1^*, \dots, X_n^*)$  be a block bootstrap sample generated from  $(X_1, \dots, X_n)$ , with the block size  $m \in \mathcal{M}$ . For each bootstrap sample, compute a  $t$ -subsampling (or bootstrap) confidence interval  $CI_{t,(m,c)}$  as described in Appendix B according to block size  $m \in \mathcal{M}$  and bound  $c \in \mathcal{C}$ . The data-driven block size and estimating function bound according to the calibration method are defined as

$$(m, c)_{CM} := \arg \inf_{(m,c) \in \mathcal{MC}} \left\{ \left| t - P^* \left[ \hat{\theta}_n^R \in CI_{t,(m,c)} \right] \right| \right\}, \quad (40)$$

where, by definition,  $\arg \inf(\emptyset) := \infty$ , and  $P^*$  denotes the probability with respect to the bootstrap distribution. By definition,  $(m, c)_{CM}$  is the pair for which the bootstrap probability of the event  $\{\hat{\theta}_n^R \in CI_{t,(m,c)}\}$  is as near as possible to the nominal level  $t$  of the confidence interval. The extension to the studentized statistic  $T_n = \sqrt{n} \left( [\hat{\Sigma}_n^R]^{-1/2} (\hat{\theta}_n^R - \theta_0) \right)$  is straightforward.

We summarize the calibration method for the selection of the block size  $m$  and estimating function bound  $c$  in the following steps.

- (1) Compute  $\hat{\theta}_n^R = (\hat{\alpha}_n^R, \hat{\beta}_n^R)'$ , as the solution of (10), with  $c = c_1$  as preliminary value of the estimating function bound.

- (2) For each  $m \in \mathcal{M}$ , generate  $K$  random bootstrap samples  $(z_1^*, \dots, z_n^*)$  based on overlapping blocks of size  $m$ .
- (3) For each random bootstrap sample  $(z_1^*, \dots, z_n^*)$  and  $c \in \mathcal{C}$ , compute confidence intervals  $CI_{t,(m,c)}$  for the parameter  $\hat{\beta}_n^R$ , by applying steps (1)-(5) of the algorithm in the previous Appendix A.
- (4) For each pair  $(m, c) \in \mathcal{MC}$  compute  $h(m, c) = \#\{\hat{\beta}_n^R \in CI_{t,(m,c)}\}/K$ .
- (5) The data-driven block size and estimating function bound according to the calibration method are defined as  $(m, c)_{CM} := \arg \inf_{(m,c) \in \mathcal{MC}} \{|t - h(m, c)|\}$ .

## References

- [1] Amihud, Y., and C.M. Hurvich, 2004. Predictive regressions: a reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39, 813–841.
- [2] Amihud, Y., Hurvich C.M., and Y. Wang, 2008. Multiple-predictor regressions: hypothesis testing. *The Review of Financial Studies*, 22, 413–434.
- [3] Amihud, Y., Hurvich C.M., and Y. Wang, 2010. Predictive regression with order-p autoregressive predictors. *Journal of Empirical Finance*, 17, 513–525.
- [4] Andrews, D., 2002. Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica*, 70, 119–162.
- [5] Andrews, D., 2004. The block-block bootstrap: improved asymptotic refinements. *Econometrica*, 72, 673–700.
- [6] Andrews, D., and P. Guggenberger, 2010. Application of subsampling, hybrid and size-correction methods. *Journal of Econometrics*, 158, 285–305.
- [7] Ang, A., and G. Bekaert, 2007. Stock return predictability: is it there? *The Review of Financial Studies*, 20, 651–707.
- [8] Bollerslev, T., Marrone, J., Xu, L., and H. Zhou, 2012. Stock return predictability and variance risk premia: statistical inference and international evidence. Working paper.
- [9] Bollerslev, T., Tauchen, G., and H. Zhou, 2009. Expected stock returns and variance risk premia. *Review of Financial Studies*, 22, 4463–4492.
- [10] Campbell, J.Y., and R.J. Shiller, 1988. The dividend ratio model and small sample bias: a Monte Carlo study. *Economics Letters*, 29, 325–331.
- [11] Campbell, J.Y., and M. Yogo, 2006. Efficient tests of stock return predictability. *Journal of Financial Economics*, 81, 27–60.

- [12] Camponovo, L., Scaillet, O., and F. Trojani, 2012. Robust subsampling. *Journal of Econometrics*, 167, 197–210.
- [13] Carlstein, E., 1986. The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Annals of Statistics*, 14, 1171–1179.
- [14] Cavanagh, C., Elliott, G., and J.H. Stock, (1995). Inference in models with nearly integrated regressors. *Econometric Theory* 11, 1131–1147.
- [15] Choi, I., and T.K. Chue, 2007. Subsampling-based tests of stock-return predictability, Working paper.
- [16] Davidson, R., and J. McKinnon, 1999. Bootstrap testing in nonlinear models. *International Economic Review*, 40, 487–508.
- [17] Donoho, D. L. and P. J. Huber (1983) The notion of breakdown point, in *A Festschrift for Erich L. Lehmann* by Bickel, P. J., Doksum, K. A. and J. L. Hodges Jr. (eds.), 157-184, Wadsworth, Belmont, California.
- [18] Fama, E., and K. French, 1988. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22, 3-25.
- [19] Gagliardini P., Trojani F., and G. Urga, (2005). Robust GMM tests for structural breaks. *Journal of Econometrics*, 129, 139-182.
- [20] Goetzmann, W.N., and P. Jorion, 1995. A longer look at dividend yields. *Journal of Business*, 68, No.4.
- [21] Genton, M., and A. Lucas, 2003. Comprehensive definitions of breakdown point for independent and dependent observations. *Journal of the Royal Statistical Society, Series B*, 65, 81–94.
- [22] Goncalves, S., and H. White, 2004. Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*, 119, 199–220.

- [23] Goyal, A., and I. Welch, 2003. Predicting the equity premium with dividend ratios. *Management Science*, 49, 639–654.
- [24] Hall, P., 1985. Resampling a coverage process. *Stochastic Processes and their Applications*, 19, 259–269.
- [25] Hall, P., 1988. On symmetric bootstrap confidence intervals. *Journal of the Royal Statistical Society, Ser.B*, 50, 35–45.
- [26] Hall, P., and J. Horowitz, 1996. Bootstrap critical values for tests based on Generalized-Method-of-Moment estimators. *Econometrica*, 64, 891–916.
- [27] Hall, P., Horowitz J., and B.-Y. Jing, 1995. On blocking rules for the bootstrap with dependent data. *Biometrika*, 82, 561–574.
- [28] Hampel, F. R., (1971) A General Qualitative Definition of Robustness, *Annals of Mathematical Statistics*, 42, 1887-1896.
- [29] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P . J., and W. A. Stahel, 1986. *Robust statistics. The approach based on influence functions*. Wiley, New York.
- [30] Heritier, S., and E. Ronchetti, 1994. Robust bounded-influence tests in general parametric model. *Journal of the American Statistical Association*, 89, 897-904.
- [31] Hong, H., and O. Scaillet, 2006. A fast subsampling method for nonlinear dynamic models. *Journal of Econometrics*, 133, 557-578.
- [32] Hu, F., and J. Kalbfleisch, 2000. The estimating function bootstrap. *Canadian Journal of Statistics*, 28, 449–499
- [33] Huber, P. J., 1981. *Robust Statistics*. Wiley, New York.
- [34] Huber, P.J., 1984. Finite sample breakdown point of M- and P-estimators. *Annals of Statistics*, 12, 119–126.

- [35] Knez, P.J., and M.J. Ready, 1997. On the robustness of size and book-to-market in cross-sectional regressions. *Journal of Finance*, 52, 1355-1382.
- [36] Kojien, R.S.J., and S. Van Nieuwerburgh, 2001. Predictability of returns and cash flows. *Annual Review of Financial Economics*, 3, 467-491.
- [37] Kothari, S.P., and J. Shanken, 1997. Book-to-market, dividend yield, and expected market returns: a time series analysis. *Journal of Financial Economics*, 2, 169-203.
- [38] Künsch, H., 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17, 1217–1241.
- [39] Jansson, M., and M.J. Moreira, 2006. Optimal inference in regression models with nearly integrated regressors. *Econometrica*, 74, 681–714.
- [40] Lettau, M., and S. Ludvigson, 2001. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 3, 815–849.
- [41] Lettau, M., and S. Van Nieuwerburgh, 2007. Reconciling the return predictability evidence: in-sample forecasts, out-of-sample forecasts, and parameter instability. *Review of Financial Studies*, 21, 1607–1652.
- [42] Lewellen, J., 2004. Predicting returns with financial ratios. *Journal of Financial Economics*, 74, 209–235.
- [43] Lucas, A., 1995. Unit root tests based on M-estimators. *Econometric Theory*, 11, 331–346.
- [44] Lucas, A., 1997. Cointegration testing using pseudolikelihood ratio tests. *Econometric Theory*, 13, 149–169.
- [45] Mancini, L., Ronchetti, E., and F. Trojani, 2005. Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models. *Journal of the American Statistical Association*, 100, 628-641.



- [46] Mankiv, N.G., and M.D. Shapiro, 1986. Do we reject too often? Small sample properties of tests of rational expectation models. *Economics Letters* 20, 139-145.
- [47] Mikusheva, A., 2007. Uniform inference in autoregressive models. *Econometrica*, 75, 1411–1452.
- [48] Nelson, C.R., and M.J. Kim, 1993. Predictable stock returns: the role of small sample bias. *Journal of Finance*, 48, No.2, 641-661.
- [49] Politis, D. N., and J. P. Romano, 1994. The stationary bootstrap. *Journal of the American Statistical Association*, 89, 1303–1313.
- [50] Politis, D.N., Romano J.P., and M. Wolf, 1999. *Subsampling*. Springer, New York.
- [51] Polk, C., S. Thompson, and T. Vuolteenaho, 2006. Cross-sectional forecast of the equity premium. *Journal of Financial Economics*, 81, 101-141.
- [52] Romano, J. P., and M. Wolf, 2001. Subsampling intervals in autoregressive models with linear time trend. *Econometrica*, 69, 1283-1314.
- [53] Ronchetti, E., and F. Trojani, 2001. Robust inference with GMM estimators, *Journal of Econometrics*, 101, 37–69.
- [54] Rozeff, M., 1984. Dividend yields are equity risk premium. *Journal of Portfolio Management*, 11, 68-75.
- [55] Salibian-Barrera, M., and R. Zamar, 2002. Bootstrapping robust estimates of regression. *Annals of Statistics*, 30, No. 2, 556–582.
- [56] Salibian-Barrera, M., Van Aelst, S., and G. Willems, 2006. Principal components analysis based on multivariate MM estimators with fast and robust bootstrap. *Journal of the American Statistical Association*, 101, 1198–1211.
- [57] Salibian-Barrera, M., Van Aelst, S., and G. Willems, 2007. Fast and robust bootstrap. *Statistical Methods and Applications*, 17, 41–71

- [58] Santos, T., and P. Veronesi, 2006. Labor income and predictable stock returns. *Review of Financial Studies*, 19, 1–44.
- [59] Shao, J., and D. Tu, 1995. *The jackknife and bootstrap*. Springer, New York.
- [60] Shiller, R.J., 2000. *Irrational Exuberance*. Princeton University Press, Princeton, NJ.
- [61] Singh, K., 1998. Breakdown theory for bootstrap quantiles. *Annals of Statistics*, 26, 1719–1732.
- [62] Stambaugh, R.F., 1986. Bias in regressions with lagged stochastic regressors. Graduate School of Business, University of Chicago, Working Paper No. 156.
- [63] Stambaugh, R.F., 1999. Predictive regressions. *Journal of Financial Economics*, 54, 375–421.
- [64] Torous, W., R. Valkanov, and S. Yan, 2004. On predicting stock returns with nearly integrated explanatory variables. *Journal of Business*, 77, 937–966.
- [65] van Binsbergen, J.H., and R.S. Koijen, 2010. Predictive regressions: A present-value approach. *Journal of Finance*, 65, 1439–1471.
- [66] Wolf, M., 2000. Stock returns and dividend yields revisited: a new way to look at an old problem. *Journal of Business and Economic Statistics*, 18, 18–30.

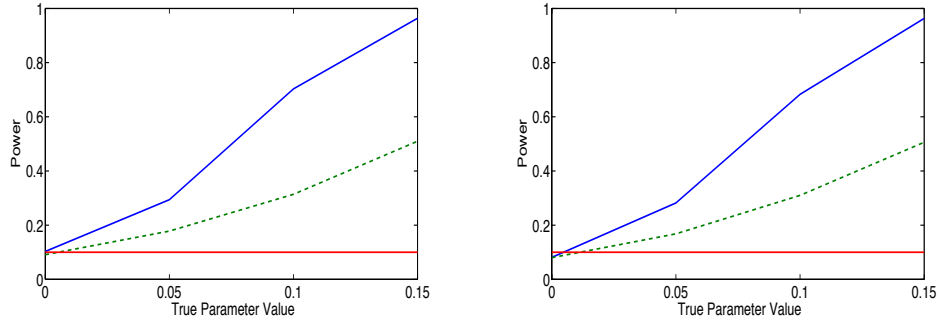


Figure 1: **Power curves of bias-corrected and local-to-unity asymptotics.** We plot the proportion of rejections of the null hypothesis  $\mathcal{H}_0 : \beta_0 = 0$ , when the true parameter value is  $\beta_0 \in [0, 0.15]$ . In the left panel, we consider the bias-corrected method proposed in Amihud, Hurvich and Wang (2008), while in the right panel we consider the Bonferroni approach for the local-to-unity asymptotic theory introduced in Campbell and Yogo (2006). We consider non-contaminated samples (straight line) and contaminated samples (dashed line).

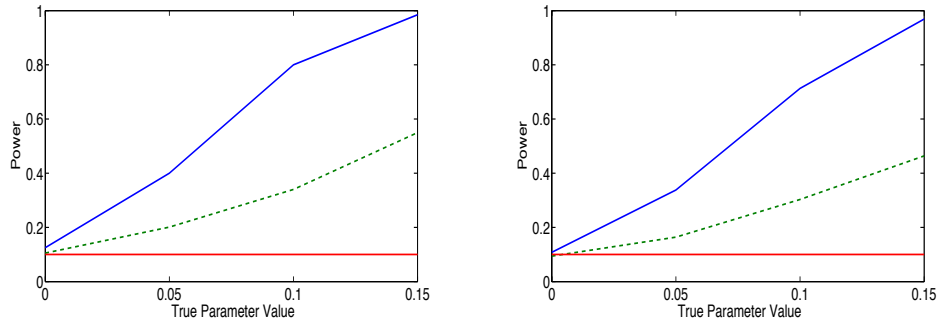


Figure 2: **Power curves of block bootstrap and subsampling.** We plot the proportion of rejections of the null hypothesis  $\mathcal{H}_0 : \beta_0 = 0$ , when the true parameter value is  $\beta_0 \in [0, 0.15]$ . In the left panel, we consider the block bootstrap, while in the right panel we consider the subsampling. We consider non-contaminated samples (straight line) and contaminated samples (dashed line).

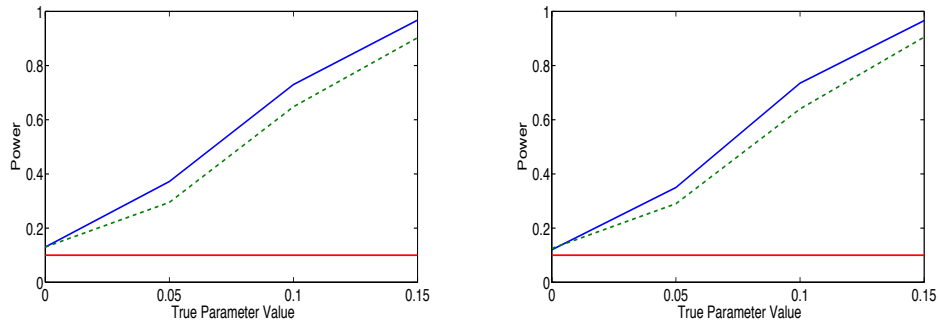


Figure 3: **Power curves of robust block bootstrap and robust subsampling.** We plot the proportion of rejections of the null hypothesis  $\mathcal{H}_0 : \beta_0 = 0$ , when the true parameter value is  $\beta_0 \in [0, 0.15]$ . In the left panel, we consider our robust block bootstrap, while in the right panel we consider our robust subsampling. We consider non-contaminated samples (straight line) and contaminated samples (dashed line).

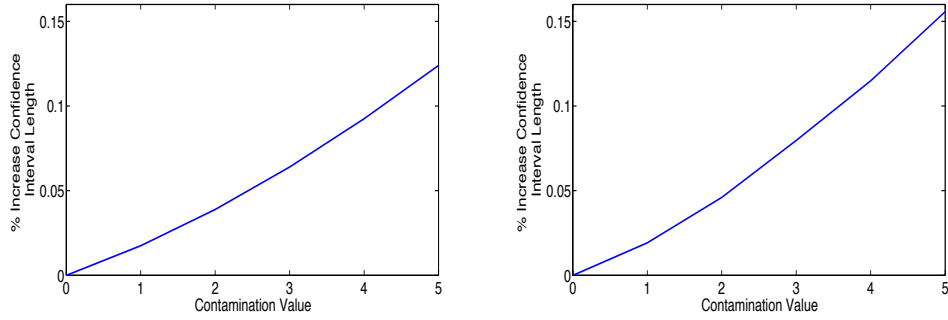


Figure 4: **Sensitivity analysis of bias-corrected and local-to-unity asymptotics.** We plot the percentage of increase of the confidence interval lengths with respect to variation of  $y_{max}$ , in each Monte Carlo sample, within the interval  $[0, 5]$ . In the left panel, we consider the bias-corrected method proposed in Amihud, Hurvich and Wang (2008), while in the right panel we consider the Bonferroni approach for the local-to-unity asymptotic theory introduced in Campbell and Yogo (2006).

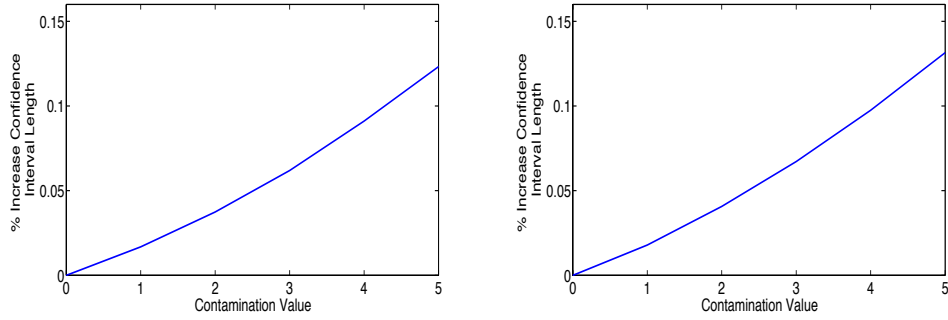


Figure 5: **Sensitivity analysis of block bootstrap and subsampling.** We plot the percentage of increase of the confidence interval lengths with respect to variation of  $y_{max}$ , in each Monte Carlo sample, within the interval  $[0, 5]$ . In the left panel, we consider the block bootstrap, while in the right panel we consider the subsampling.

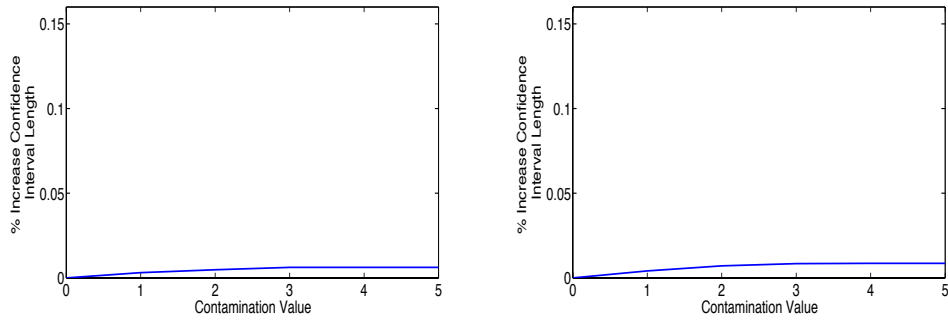


Figure 6: **Sensitivity analysis of robust block bootstrap and robust subsampling.** We plot the percentage of increase of the confidence interval lengths with respect to variation of  $y_{max}$ , in each Monte Carlo sample, within the interval  $[0, 5]$ . In the left panel, we consider our robust block bootstrap, while in the right panel we consider our robust subsampling.

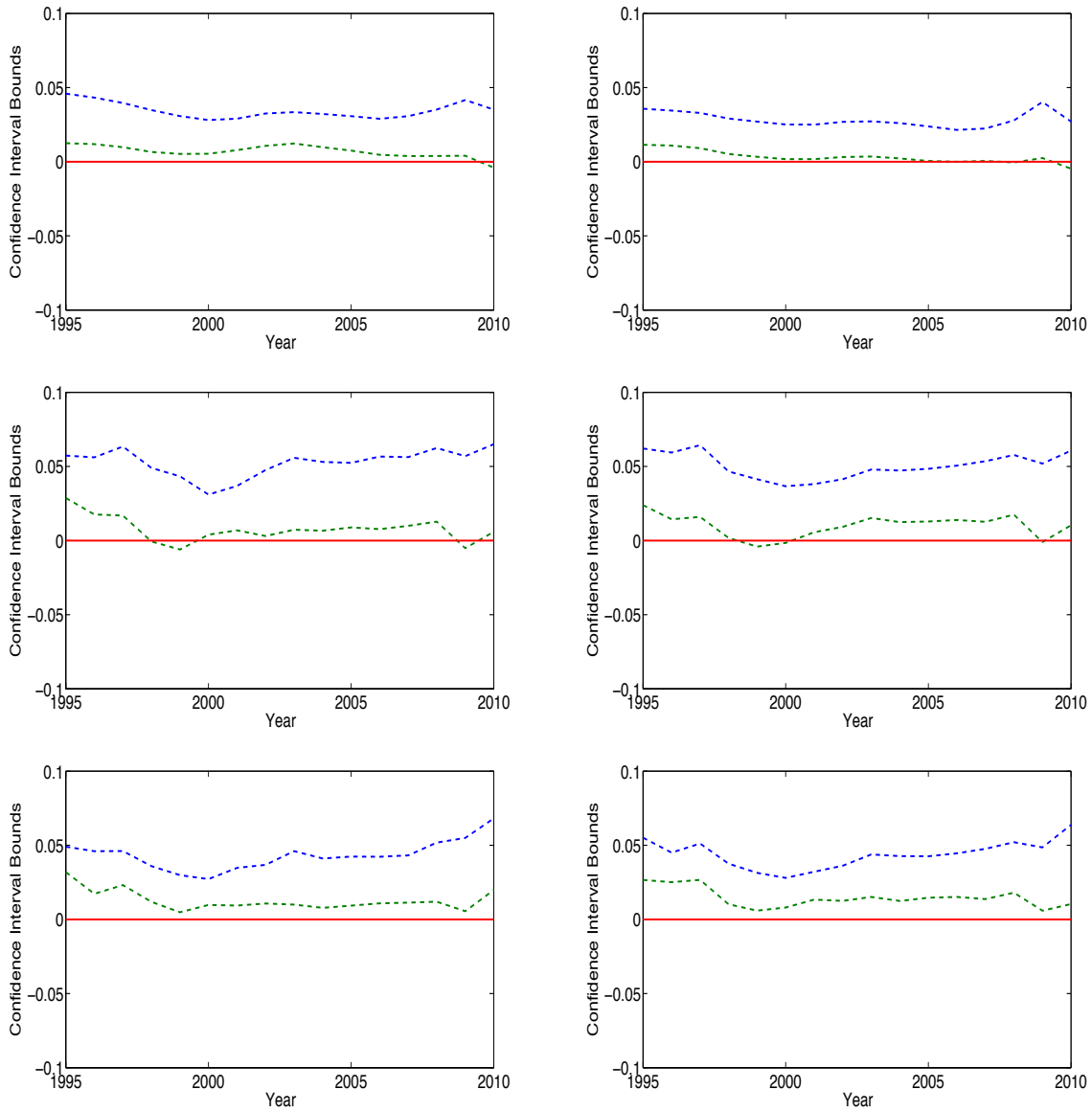


Figure 7: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter  $\beta$  in the predictive regression model (15). We consider rolling windows of 180 observations for the period 1980-2010. In the first line, we present the bias-corrected method (left panel) and the Bonferroni approach (right panel). In the second line, we consider the classic bootstrap (left panel) and the classic subsampling (right panel), while in the third line we consider our robust bootstrap (left panel) and our robust subsampling (right panel).

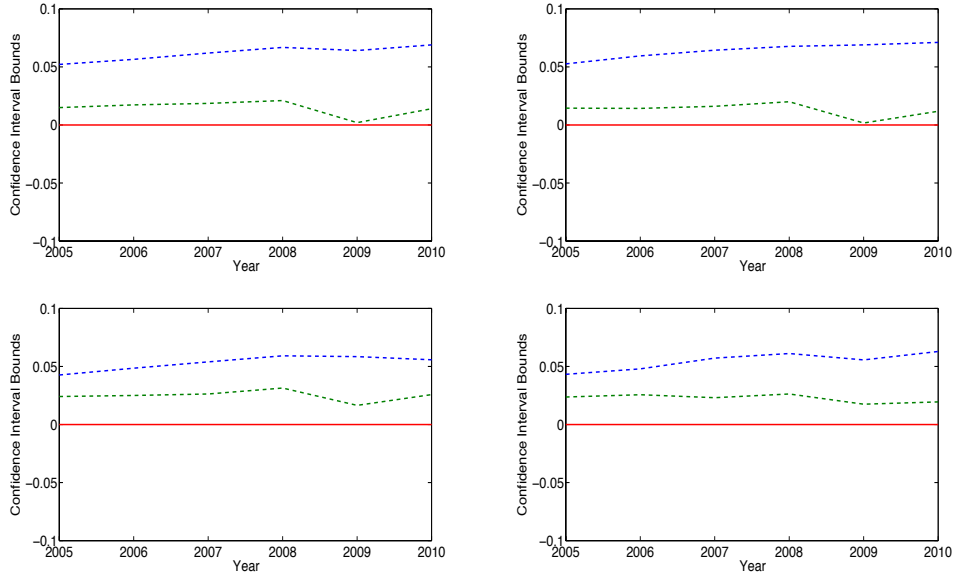


Figure 8: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter  $\beta_1$  in the predictive regression model (16). We consider rolling windows of 180 observations for the period 1990-2010. In the top line, we present the classic bootstrap (left panel) and the classic subsampling (right panel), while in the bottom line we consider our robust bootstrap (left panel) and our robust subsampling (right panel).

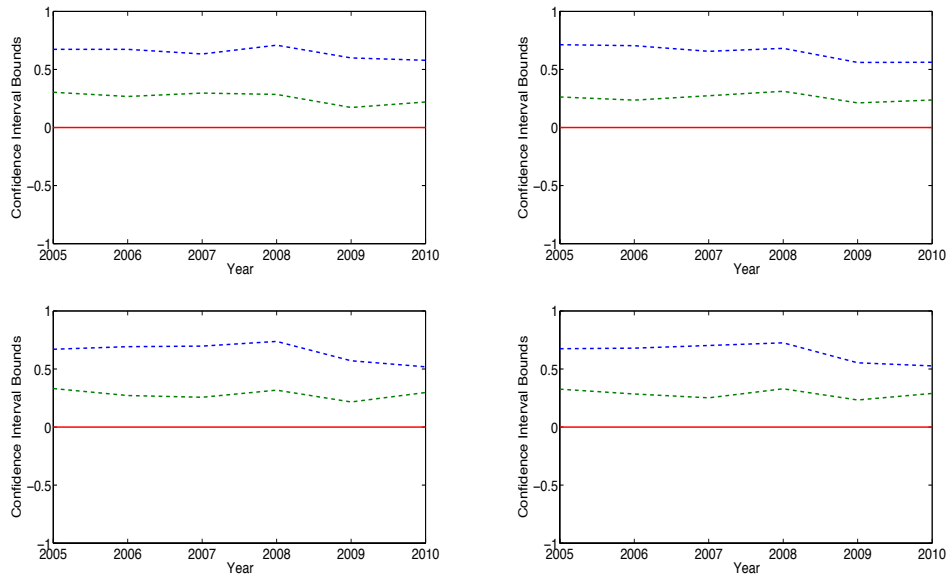


Figure 9: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter  $\beta_2$  in the predictive regression model (16). We consider rolling windows of 180 observations for the period 1990-2010. In the top line, we present the classic bootstrap (left panel) and the classic subsampling (right panel), while in the bottom line we consider our robust bootstrap (left panel) and our robust subsampling (right panel).

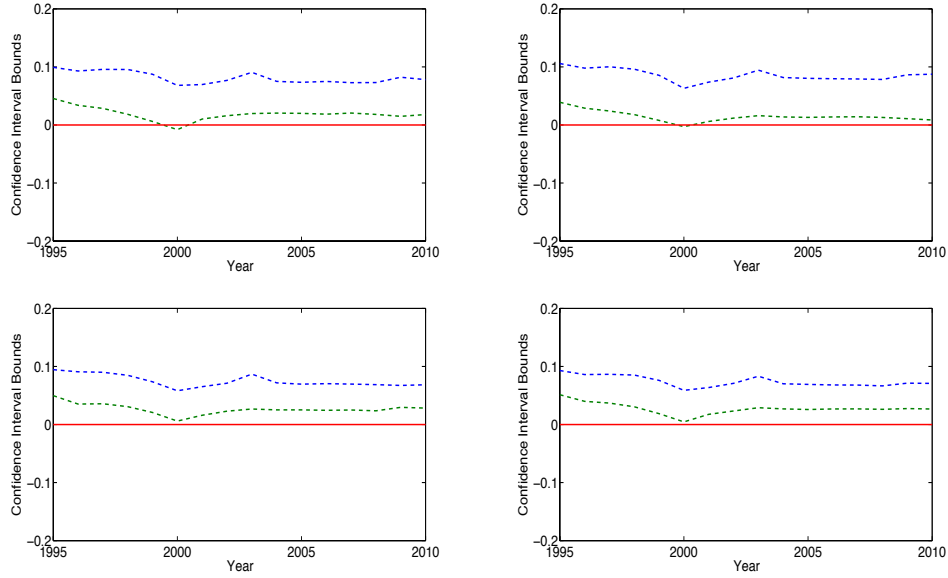


Figure 10: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter  $\beta_1$  in the predictive regression model (18). We consider rolling windows of 180 observations for the period 1950-2010. In the top line, we present the classic bootstrap (left panel) and the classic subsampling (right panel), while in the bottom line we consider our robust bootstrap (left panel) and our robust subsampling (right panel).

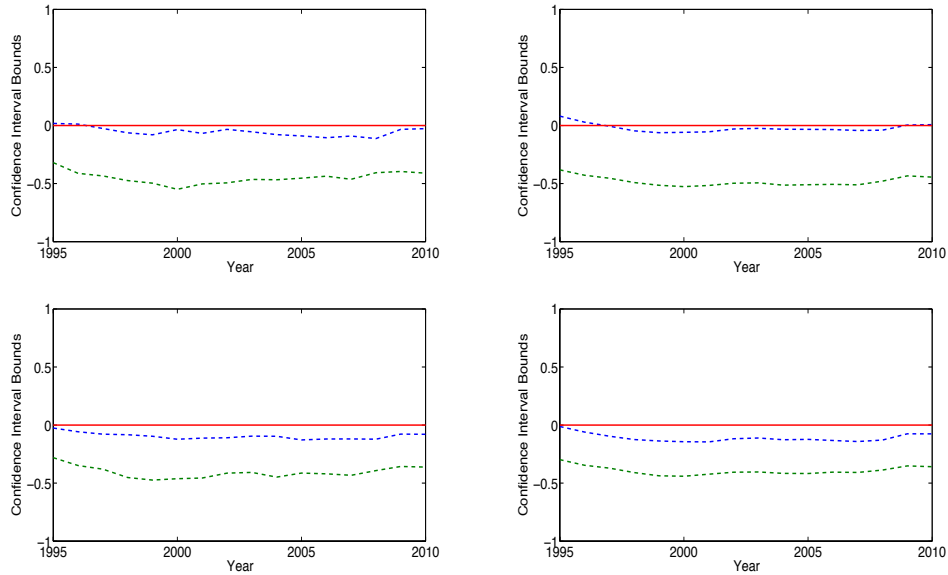


Figure 11: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter  $\beta_2$  in the predictive regression model (18). We consider rolling windows of 180 observations for the period 1950-2010. In the top line, we present the classic bootstrap (left panel) and the classic subsampling (right panel), while in the bottom line we consider our robust bootstrap (left panel) and our robust subsampling (right panel).

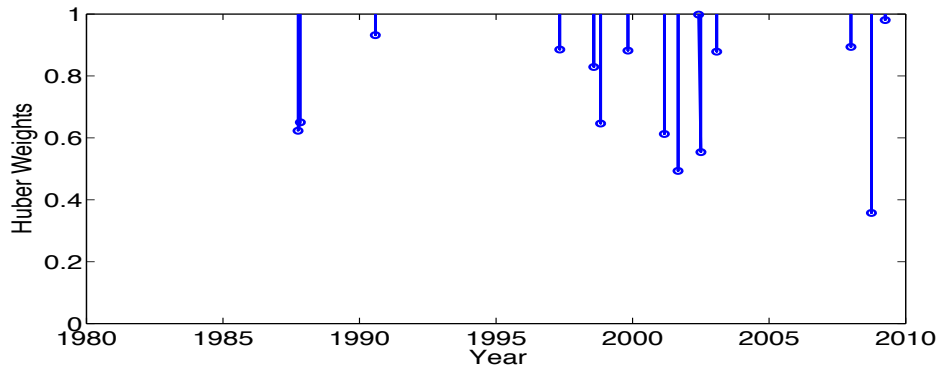


Figure 12: **Huber Weights under the Predictive Regression Model (15)**. We plot the estimated Huber weights for the single-predictor predictive regression model (15), with the dividend yield as predictive variable, in the period 1980-2010.

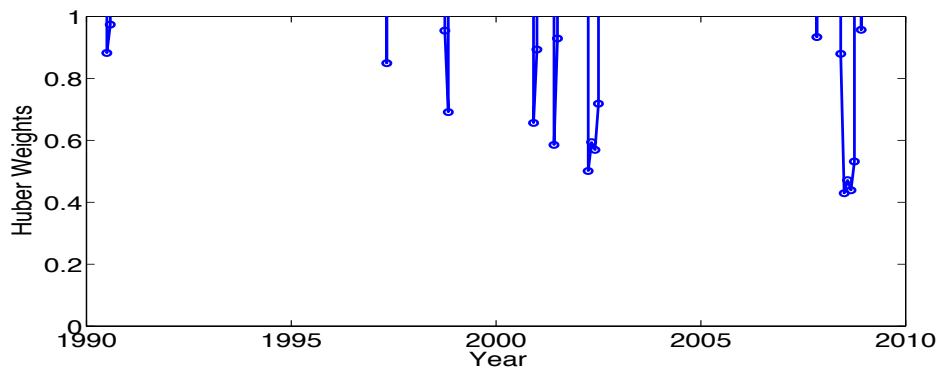


Figure 13: **Huber Weights under the Predictive Regression Model (16)**. We plot the estimated Huber weights for the two-predictor predictive regression model (16), with the dividend yield and the variance risk premium as predictive variables, in the period 1990-2010.



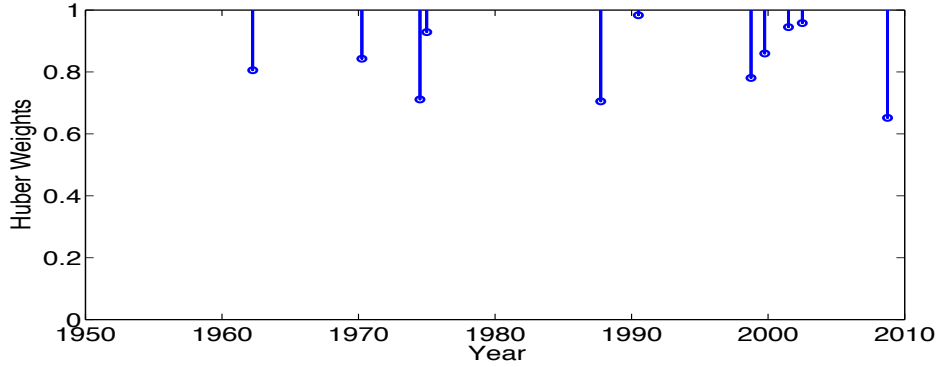


Figure 14: **Huber Weights under the Predictive Regression Model (18).** We plot the estimated Huber weights for the two-predictor predictive regression model (18), with the dividend yield and the ratio of labor income to consumption as predictive variables, in the period 1950-2010.

$n = 120, b = 0.5$	0.9	0.95
Subsampling ( $m = 10$ )	[0.0417; 0.0833]	[0.0417; 0.0417]
Subsampling ( $m = 20$ )	[0.0833; 0.0833]	[0.0833; 0.0833]
Subsampling ( $m = 30$ )	[0.1250; 0.1250]	[0.1250; 0.1250]
Bootstrap ( $m = 10$ )	[0.0417; 0.3750]	[0.0417; 0.3333]
Bootstrap ( $m = 20$ )	[0.0833; 0.3333]	[0.0833; 0.3333]
Bootstrap ( $m = 30$ )	[0.1250; 0.3333]	[0.1250; 0.2500]

Table 1: **Subsampling and Block Bootstrap Lower and Upper Bounds for the Quantile Breakdown Point.** Breakdown point of the subsampling and the block bootstrap quantiles. The sample size is  $n = 120$ , and the block size is  $m = 10, 20, 30$ . We assume a statistic with breakdown point  $b = 0.5$  and confidence levels  $t = 0.9, 0.95$ . Lower and upper bounds for quantile breakdown points are computed using Theorem 2.

$n = 120$	0.9	0.95
Subsampling ( $m = 10$ )	0.1750	0.1250
Subsampling ( $m = 20$ )	0.2500	0.2083
Subsampling ( $m = 30$ )	0.3250	0.2833
Bootstrap ( $m = 10$ )	0.5000	0.5000
Bootstrap ( $m = 20$ )	0.5000	0.5000
Bootstrap ( $m = 30$ )	0.4250	0.3583

Table 2: **Robust Subsampling and Robust Block Bootstrap for the studentized Statistic  $T_n$ .** Breakdown point of the robust subsampling and the robust block bootstrap quantiles for the studentized statistic  $T_n$ , in the predictive regression model (1)-(2). The sample size is  $n = 120$ , and the block size is  $m = 10, 20, 30$ . The quantile breakdown points are computed using Theorem 5.

	1980 – 1995	1985 – 2000	1990 – 2005	1995 – 2010
Bias-Corrected	0.0292 <sup>(**)</sup>	0.0167 <sup>(**)</sup>	0.0191 <sup>(**)</sup>	0.0156
Bonferroni	0.0236 <sup>(**)</sup>	0.0134 <sup>(**)</sup>	0.0117 <sup>(*)</sup>	0.0112
Bootstrap	0.0430 <sup>(**)</sup>	0.0175 <sup>(*)</sup>	0.0306 <sup>(**)</sup>	0.0355 <sup>(**)</sup>
Subsampling	0.0430 <sup>(**)</sup>	0.0175	0.0306 <sup>(**)</sup>	0.0355 <sup>(**)</sup>
R.Bootstrap	0.0405 <sup>(**)</sup>	0.0174 <sup>(**)</sup>	0.0245 <sup>(**)</sup>	0.0378 <sup>(**)</sup>
R.Subsampling	0.0405 <sup>(**)</sup>	0.0174 <sup>(**)</sup>	0.0245 <sup>(**)</sup>	0.0378 <sup>(**)</sup>

Table 3: **Point Estimates of Parameter  $\beta$** . We report the point estimates of the parameter  $\beta$  in the predictive regression model (15) for the subperiods 1994-2009, 1980-1995, 1985-2000, 1990-2005 and 1995-2010, all consisting of 180 observations. In the second and third line we consider the bias-corrected method and the Bonferroni approach, respectively. In the fourth and fifth line we consider the standard bootstrap and standard subsampling, respectively. Finally, in the sixth and seventh line we consider our robust bootstrap and robust subsampling, respectively. (\*) and (\*\*) mean rejection at 10% and 5% significance level, respectively.

	1991 – 2006	1992 – 2007	1993 – 2008	1994 – 2009	1995 – 2010
Bootstrap	0.0369 <sup>(**)</sup>	0.0402 <sup>(**)</sup>	0.0454 <sup>(**)</sup>	0.0368 <sup>(*)</sup>	0.0415 <sup>(**)</sup>
Subsampling	0.0369 <sup>(**)</sup>	0.0402 <sup>(**)</sup>	0.0454 <sup>(**)</sup>	0.0368 <sup>(*)</sup>	0.0415 <sup>(**)</sup>
R.Bootstrap	0.0368 <sup>(**)</sup>	0.0402 <sup>(**)</sup>	0.0437 <sup>(**)</sup>	0.0366 <sup>(**)</sup>	0.0412 <sup>(**)</sup>
R.Subsampling	0.0368 <sup>(**)</sup>	0.0402 <sup>(**)</sup>	0.0437 <sup>(**)</sup>	0.0366 <sup>(**)</sup>	0.0412 <sup>(**)</sup>

	1991 – 2006	1992 – 2007	1993 – 2008	1994 – 2009	1995 – 2010
Bootstrap	0.4700 <sup>(**)</sup>	0.4648 <sup>(**)</sup>	0.4968 <sup>(**)</sup>	0.3859 <sup>(**)</sup>	0.3993 <sup>(**)</sup>
Subsampling	0.4700 <sup>(**)</sup>	0.4648 <sup>(**)</sup>	0.4968 <sup>(**)</sup>	0.3859 <sup>(**)</sup>	0.3993 <sup>(**)</sup>
R.Bootstrap	0.4821 <sup>(**)</sup>	0.4771 <sup>(**)</sup>	0.5276 <sup>(**)</sup>	0.3932 <sup>(**)</sup>	0.4083 <sup>(**)</sup>
R.Subsampling	0.4821 <sup>(**)</sup>	0.4771 <sup>(**)</sup>	0.5276 <sup>(**)</sup>	0.3932 <sup>(**)</sup>	0.4083 <sup>(**)</sup>

Table 4: **Point Estimates of Parameters  $\beta_1$  and  $\beta_2$** . We report the point estimates of parameters  $\beta_1$  (first table) and  $\beta_2$  (second table) in the predictive regression model (16) for the subperiods 1991-2006, 1992-2007, 1993-2008, 1994-2009 and 1995-2010, all consisting of 180 observations. In the second and third line we consider the standard bootstrap and standard subsampling, respectively. In the fourth and fifth line we consider our robust bootstrap and robust subsampling, respectively. (\*) and (\*\*) mean rejection at 10% and 5% significance level, respectively.

	1950 – 1995	1955 – 2000	1960 – 2005	1965 – 2010
Bootstrap	0.0724 <sup>(**)</sup>	0.0301	0.0467 <sup>(**)</sup>	0.0480 <sup>(**)</sup>
Subsampling	0.0724 <sup>(**)</sup>	0.0301	0.0467 <sup>(**)</sup>	0.0480 <sup>(*)</sup>
R.Bootstrap	0.0721 <sup>(**)</sup>	0.0305 <sup>(**)</sup>	0.0474 <sup>(**)</sup>	0.0488 <sup>(**)</sup>
R.Subsampling	0.0721 <sup>(**)</sup>	0.0305 <sup>(**)</sup>	0.0474 <sup>(**)</sup>	0.0488 <sup>(**)</sup>

	1950 – 1995	1955 – 2000	1960 – 2005	1965 – 2010
Bootstrap	-0.1509	-0.2926 <sup>(**)</sup>	-0.2718 <sup>(**)</sup>	-0.2187 <sup>(*)</sup>
Subsampling	-0.1509	-0.2926 <sup>(**)</sup>	-0.2718 <sup>(**)</sup>	-0.2187
R.Bootstrap	-0.1532 <sup>(**)</sup>	-0.2920 <sup>(**)</sup>	-0.2701 <sup>(**)</sup>	-0.2173 <sup>(**)</sup>
R.Subsampling	-0.1532 <sup>(**)</sup>	-0.2920 <sup>(**)</sup>	-0.2701 <sup>(**)</sup>	-0.2173 <sup>(**)</sup>

Table 5: **Point Estimates of Parameters  $\beta_1$  and  $\beta_2$ .** We report the point estimates of parameters  $\beta_1$  (first table) and  $\beta_2$  (second table) in the predictive regression model (18) for the subperiods 1950-1995, 1955-2000, 1960-2005 and 1965-2010, all consisting of 180 observations. In the second and third line we consider the standard bootstrap and standard subsampling, respectively. In the fourth and fifth line we consider our robust bootstrap and robust subsampling, respectively. (\*) and (\*\*) mean rejection at 10% and 5% significance level, respectively.