Predictability hidden by Anomalous Observations in Financial Data*

Lorenzo Camponovo

University of Applied Sciences and Arts Southern Switzerland

Olivier Scaillet

University of Geneva and Swiss Finance Institute

Fabio Trojani

University of Geneva and Swiss Finance Institute

April 2024

*Correspondence Address: fabio.trojani@unige.ch.

Abstract

Testing procedures for predictive regressions involving lagged autoregressive variables produce a suboptimal inference in presence of minor violations of ideal assumptions. A novel testing framework based on resampling methods that exhibits resistance to such violations and is reliable also in models with nearly integrated regressors is introduced. To achieve this objective, the robustness of resampling procedures for time series are defined by deriving new formulas quantifying their quantile breakdown point. For both the block bootstrap and subsampling, these formulas show a very low quantile breakdown point. To overcome this problem, a robust and fast resampling scheme applicable to a broad class of time series settings is proposed. This framework is also suitable for multi-predictor settings, particularly when the data only approximately conform to a predictive regression model. Monte Carlo simulations provide substantial evidence for the significant improvements offered by this robust approach. Using the proposed resampling methods, empirical coverages and rejection frequencies are very close to the nominal levels, both in the presence and absence of small deviations from the ideal model assumptions. Empirical analysis reveals robust evidence of market return predictability, previously obscured by anomalous observations, both in- and out-of-sample.

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

JEL: C12, C13, G1.

1 Introduction

Extensive research has examined the predictive power of economic variables, such as the price-dividend ratio, proxies of labor income, or interest rate, for stock returns. The econometric approach to testing predictability relies on a predictive regression of stock returns onto a set of lagged financial variables, as exemplified by Stambaugh (1999). Significant disparities in testing methodologies within the literature emerge due to variations in test statistics and asymptotic theories employed to assess the null hypothesis of no predictability. These discrepancies result in divergent findings and conclusions in numerous cases.

A general approach to obtaining tests that are less susceptible to finite sample biases or assumptions on the form of their asymptotic distribution involves nonparametric resampling methods, such as the bootstrap or subsampling, see Wolf (2000) and Ang and Bekaert (2007). A common characteristic of these approaches is their dependence on procedures that are significantly influenced by a small fraction of anomalous observations in the data. This issue has long been recognized for standard OLS estimators, see Huber (1981). Recent research has also demonstrated that inference provided by bootstrap and subsampling tests can be easily inflated by a small fraction of anomalous observations, see Singh (1998), Salibian-Barrera and Zamar (2002), and Camponovo, Scaillet, and Trojani (2012).

This failing robustness is caused by the often excessively high fraction of anomalous observations generated by standard bootstrap and subsampling procedures, which tends to surpass the actual fraction of such observations in the original data. Addressing this issue is not straightforward, as it is not solved by merely applying conventional bootstrap or subsampling methods to more robust estimators or test statistics. Resampling trimmed or winsorized estimators fails to produce a robust resampling method, as demonstrated in detailed examples by Singh (1988) and Camponovo, Trojani, and Scaillet (2012). Therefore, we introduce new robust bootstrap and subsampling methodologies aimed at developing more stable and reliable tests for predictability hypotheses in predictive regression settings.

Our methodology naturally complements also more recent literature proposing predictive regression methods covering models with parameter instabilities or with predictors nearly featuring a unit root behaviour; see, e.g., Lin and Tu (2020), Boudoukh, Israel, and Richardson, (2022), Andersen and Varneskov (2022), Boucher, Jasinski, and Tokpavi (2023) and Coqueret and Tavin (2023) among others. The main objective of our study is different, as we introduce new robust resampling tests that outperform standard resampling procedures in reliability, especially when dealing with anomalous observations in the data. Such anomalies may arise because of outliers that deviate from the true data-generating process. Alternatively, they might also be originate from the true data-generating process.

We theoretically characterize the robustness properties of resampling methods in a time series context, through the concept of breakdown point, which measures the resistance of a testing procedure to outliers, see Hampel (1971), and Donoho and Huber (1983). Our theoretical results affirm the dramatic non-robustness of conventional resampling procedures. To overcome this problem, we introduce robust bootstrap and subsampling procedures. Our approach relies on a straightforward weighted least-squares procedure. The data-driven weights are applied to dampen, when necessary, the influence of a few data points identified as anomalous with respect to the assumed predictive link. Furthermore, our robust resampling approach is built on the fast resampling concept proposed in, among others, Davidson and McKinnon (1999), and Andrews (2002). The methodology is applicable to a broad range of bootstrap and subsampling simulation schemes in the literature. Monte Carlo simulations confirm that these tests effectively mitigate the adverse impact of outliers, preserving desirable finite sample properties in the presence of anomalous observations.

Finally, we conduct a robust analysis of recent empirical evidence on stock return predictability in US stock market data. Understanding potential predictability patterns in asset returns is key not only for the associated implications for dynamic portfolio choice, but also to understand more broadly the origins of time-varying asset risk premia, which can be explained theoretically, see Koijen and Van Nieuwerburgh (2011). We explore single-predictor and multi-predictor models, employing well-known predictive variables from the literature, such as dividend yield, difference between implied volatility and realized volatility, interest rate, and share of labor income to consumption. Empirical analysis reveals robust evidence of return predictability, previously obscured by anomalous observations, both in- and out-of-sample.

The remainder of the paper is organized as follows. In Section 2, we introduce the standard predictive regression model and illustrate, through simulation, the robustness challenges associated with some recent tests of predictability proposed in the literature. In Section 3, we theoretically examine the robustness properties of tests based on resampling procedures in general time series settings. In Section 4, we present our robust approach and develop bootstrap and subsampling tests of predictability. In Section 5, we apply our robust testing procedure to US equity data and reevaluate some recent empirical evidence on market return predictability. Section 6 concludes. Appendices gather the statistical theory underlying our approach to assess predictabilty hidden by anomalous observations in financial data.

2 Predictability and Anomalous Observations

In Sections 2.1 we introduce the predictive regression model. In Section 2.2 and Section 2.3, we study the robustness properties of bias-corrected methods, testing procedures relying on local-to-unity asymptotics, bootstrap and subsampling tests.

2.1 The Predictive Regression Model

We consider the predictive regression model:

$$y_t = \alpha + \beta x_{t-1} + u_t, \qquad (1)$$

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

where, for $t = 1, \dots, n$, y_t denotes the stock return at time t, and x_{t-1} is an economic variable observed at time $t-1$, predicting y_t. 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 context, it is well-known that inference based on standard asymptotic theory is prone to small sample biases, potentially leading to an overrejection of the hypothesis of no predictability, denoted as $\mathcal{H}_0 : \beta_0 = 0$, where β_0 represents the true value of the unknown parameter β. This issue has been highlighted by Mankiw and Shapiro (1986) and Stambaugh (1986), among others. Additionally, as emphasized by 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, it suggests a local-to-unity framework, specifically $\rho = 1 + c/n$ where $c < 0$, for the autoregressive coefficient of model (2). This framework may imply a nonstandard asymptotic distribution for the OLS estimator $\hat{\beta}_n$ of parameter β. Several recent testing procedures have been proposed to address these challenges. In Section 2.2, we focus on the of bias-corrected methods proposed in Amihud, Hurvich, and Wang (2008) and the Bonferroni approach for local-to-unity asymptotics introduced by Campbell and Yogo (2006).

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

A common characteristic of bias-corrected methods and inference based on local-to-unity asymptotics is their susceptibility to anomalous observations, potentially leading to conclusions driven by the specific features of a small subset of the data. Indeed, these approaches utilize statistical tools sensitive to small deviations from the predictive regression model (1)-(2). Consequently, despite their accuracy under strict model assumptions, these testing procedures may become less efficient or biased, even with a small fraction of anomalous observations in the data.

We conduct a Monte Carlo simulation analyzing the bias-corrected method proposed by Amihud, Hurvich, and Wang (2008) and the Bonferroni approach for the local-to-unity asymptotic theory introduced by Campbell and Yogo (2006). Initially, we generate 1,000 samples $z_{(n)} = \big(z_1, \ldots, z_n\big)$ of size $n = 180$ according to model (1)-(2), with parameters chosen as $v_t \sim N(0,1)$, $e_t \sim N(0,1)$, $\phi = -1$, $\alpha = \mu = 0$, $\rho = 0.9$, and $\beta_0 \in [0, 0.15]$. These parameter choices align with the Monte Carlo setting studied in, for example, Choi and Chue (2007). In the next step, to investigate the robustness of the methods under analysis, we consider replacement outliers in random samples $\tilde{z}_{(n)} = \big(\tilde{z}_1, \ldots, \tilde{z}_n\big)$, where $\tilde{z}_t = (\tilde{y}_t, x_{t-1})'$ is generated according to,

$$y_{3max} = 3 \cdot \max(y_1, \ldots, y_n), \tag{3}$$

and p_t is an independent and identically distributed (iid) $0-1$ random sequence, independent of the process (1)-(2), such that $\mathbb{P}[p_t = 1] = \eta$. The probability of contamination by outliers is set to $\eta = 4\%$, a small contamination level compatible with the characteristics of the real data set analyzed in the empirical study in Section 5.1.

We investigate the finite sample properties of tests for the null hypothesis $\mathcal{H}_0 : \beta_0 = 0$ in the predictive regression model. Figure 1 illustrates the empirical frequency of rejecting null hypothesis $\mathcal{H}_0$ for various testing methods across different values of the alternative hypothesis $\beta_0 \in [0, 0.15]$, with a nominal significance level of 10%.

In the Monte Carlo simulation with non-contaminated samples (solid line), the fraction of null hypothesis rejections for all procedures closely aligns with the nominal level of 10% when $\beta_0 = 0$. As anticipated, the power of the tests increases with higher values of β_0. For $\beta_0 = 0.1$, both methods exhibit a rejection frequency close to 70%, and for $\beta_0 = 0.15$, a frequency exceeding 95% is observed. In the simulation with contaminated samples (dashed line), the size of all tests remains proximate to the nominal significance level. Conversely, the presence of anomalous observations significantly diminishes the power of both procedures. Indeed, for $\beta_0 > 0$, the rejection frequency of the null hypothesis for both tests is considerably lower than in the non-contaminated case. The power of both tests remains relatively flat and below 55%, even for large values of β_0. Unreported Monte Carlo results for different parameter choices and significance levels yield similar findings.

The outcomes in Figure 1 underscore the susceptibility of bias-corrected methods and inference based on local-to-unity asymptotics to anomalous data. Due to a small fraction of anomalous observations, these testing procedures become unreliable and fail to reject the null hypothesis of no predictability, even for large values of β_0. It is a critical consideration for applications where the statistical evidence of predictability is typically weak.

To address this robustness issue, a natural approach is to develop more resistant versions of the non-robust tests employed in our Monte Carlo exercise. However, achieving this task can be challenging in general. Robustifying the bias-corrected procedure in Amihud, Hurvich, and Wang (2008) would necessitate deriving an expression for the bias of robust estimators of regressions and subsequently establishing the asymptotic distribution of such bias-corrected robust

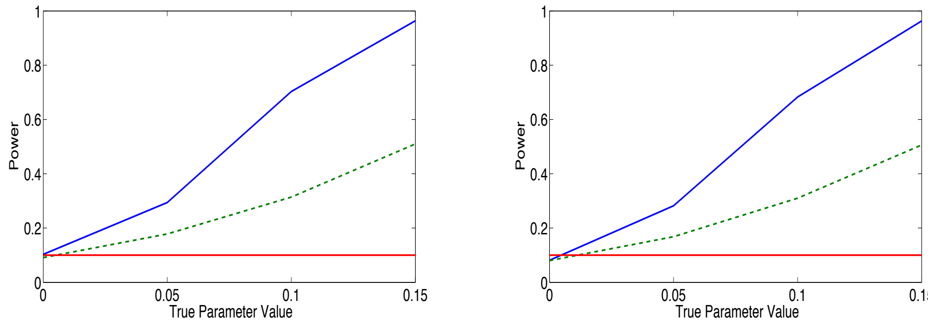

Figure 1: **Power curves of bias-corrected and local-to-unity asymptotics.** We plot the proportion of rejections of the null hypothesis $\mathcal{H}_0 : \beta_0 = 0$, when the true parameter value is $\beta_0 \in [0, 0.15]$. In the left panel, we consider the bias-corrected method proposed in Amihud, 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).

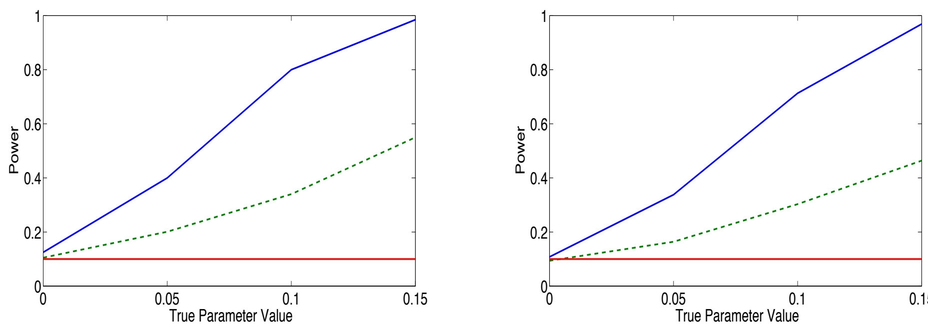

Figure 2: **Power curves of block bootstrap and subsampling.** We plot the proportion of rejections of the null hypothesis $\mathcal{H}_0 : \beta_0 = 0$, when the true parameter value is $\beta_0 \in [0, 0.15]$. In the left panel, we consider the block bootstrap, while in the right panel we consider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

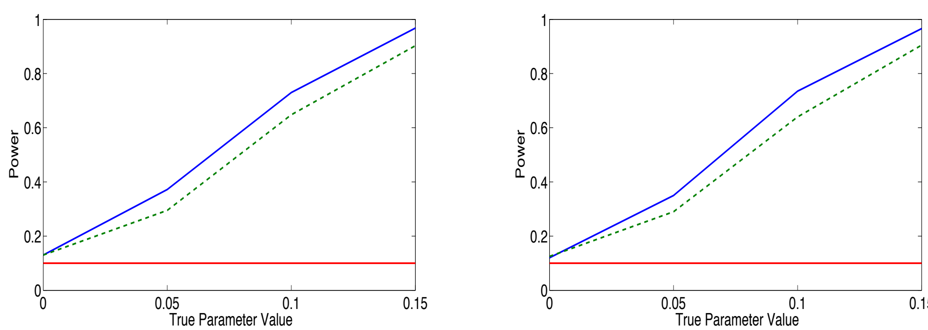

Figure 3: **Power curves of robust block bootstrap and robust subsampling.** We plot the proportion of rejections of the null hypothesis $\mathcal{H}_0 : \beta_0 = 0$, when the true parameter value is $\beta_0 \in [0, 0.15]$. In the left panel, we consider our robust block bootstrap, while in the right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

estimators. For nearly-integrated settings, robustifying the procedure proposed in Campbell and Yogo (2006) would require a nontrivial extension of the robust local-to-unity asymptotics

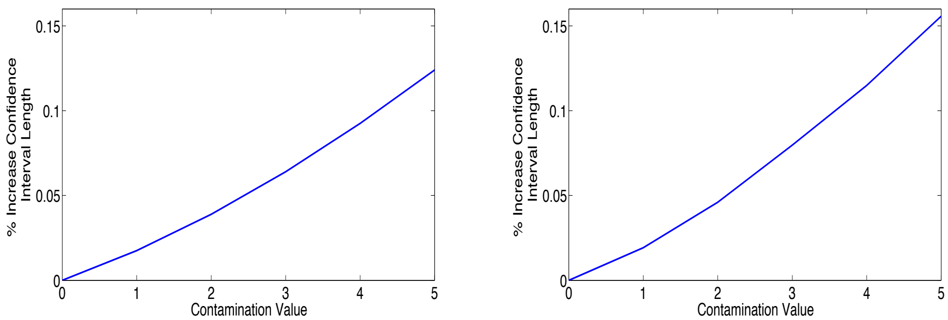

Figure 4: **Sensitivity analysis of bias-corrected and local-to-unity asymptotics.** We plot the percentage of increase of the confidence interval lengths with respect to variation of y_{max}, in each Monte Carlo sample, within the interval $[0, 5]$. In the left panel, we consider the bias-corrected 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).

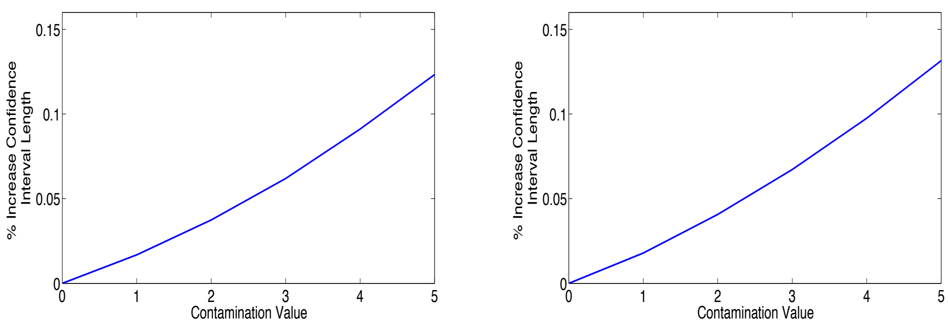

Figure 5: **Sensitivity analysis of block bootstrap and subsampling.** We plot the percentage of increase of the confidence interval lengths with respect to variation of y_{max}, in each Monte Carlo sample, within the interval $[0, 5]$. In the left panel, we consider the block bootstrap, while in the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

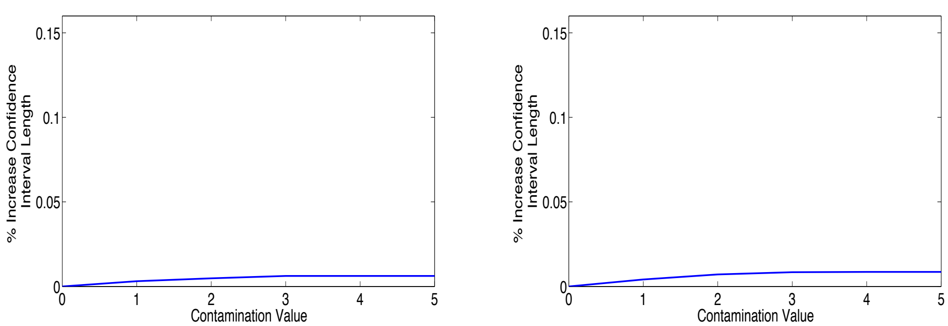

Figure 6: **Sensitivity analysis of robust block bootstrap and robust subsampling.** We plot the percentage of increase of the confidence interval lengths with respect to variation of y_{max}, in each Monte Carlo sample, within the interval $[0, 5]$. In the left panel, we consider our robust block bootstrap,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

developed in Lucas (1995, 1997) for the predictive regression model. A more general approach to obtaining tests less susceptible to finite sample biases or assumptions on their asymptotic distribution involves resampling methods, such as the bootstrap or subsampling.

2.3 Bootstrap and Subsampling Tests

Bootstrap and subsampling, offer potential improvements in inferences for time series models including predictive regression models, see Hall and Horowitz (1996) and Andrews (2002). Moreover, as demonstrated by Choi and Chue (2007) and Andrews and Guggenberger (2010), subsampling can produce accurate inferences in nearly integrated settings. To analyze these methods, we introduce block bootstrap and subsampling procedures, focusing on the predictive regression model (1)-(2). Subsequently, we employ Monte Carlo simulations to assess the degree of resistance to anomalous observations exhibited by bootstrap and subsampling tests.

Consider a random sample $X_{(n)} = (X_1, \ldots, X_n)$ 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 involve dividing the original sample $X_{(n)}$ into overlapping blocks of size $m < n$. Bootstrap samples $X^*_{(n)}$ of size n are then randomly generated from these blocks. Alternatively, nonoverlapping blocks can be utilized. The empirical distribution of the statistic $T(X^{(n)})$ is employed to estimate the sampling distribution of $T(X(n))$. Similarly, the subsampling method applies the statistic T directly to overlapping random blocks $X^*_{(m)}$ of size m strictly less than n.

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

Given the sensitivity of OLS estimators and empirical averages to even small fractions of anomalous observations, and since block bootstrap and subsampling tests rely on such statistics, inference based on resampling methods may inherit this lack of robustness. To validate this intuition, we study the finite-sample properties of block bootstrap and subsampling tests of predictability in the presence of anomalous observations through Monte Carlo simulations. For comparison, we consider the same simulation setting as in the previous section, studying tests of the null hypothesis $\mathcal{H}_0 : \beta_0 = 0$, using symmetric (bootstrap and subsampling) confidence intervals for parameter β. Figure 2 displays the empirical frequencies of rejecting the null hypothesis $\mathcal{H}_0$ for different values of the alternative hypothesis $\beta_0 \in [0, 0.15]$ with a nominal significance level of 10%. In non-contaminated samples (solid line), the rejection frequency of block bootstrap and subsampling tests closely mirrors that of the bias-corrected method and the Bonferroni approach observed in the previous section, showing a size close to the nominal level 10% for $\beta_0 = 0$ and a power exceeding 95% for $\beta_0 = 0.15$. However, contamination with anomalous observations drastically reduces the power of the tests (dashed line), with a rejection frequency consistently below 55% even for large values of β_0. Particularly, when $\beta_0 = 0.15$, the difference in power for the subsampling applied to non-contaminated and contaminated samples is larger than 50%.

The outcomes depicted in Figure 2 reveal that bootstrap and subsampling tests not only inherit but also, to some extent, exacerbate the lack of robustness observed in OLS estimators for predictive regressions. To enhance the robustness of the inferences derived from resampling methods, a logical approach is to apply standard bootstrap and subsampling simulation procedures to a more robust statistic, such as a robust linear regression estimator. Regrettably, as demonstrated in Singh (1998), Salibian-Barrera and Zamar (2002), and Camponovo, Scaillet, and Trojani (2012) for independent and identically distributed settings, resampling a robust statistic does not guarantee robust inference due to the intrinsic non-resistance to outliers in standard block bootstrap and subsampling procedures. This issue arises because the fraction of anomalous observations generated in bootstrap and subsampling blocks often exceeds the fraction of outliers in the data. Addressing this problem requires a more systematic analysis of the robustness of bootstrap and subsampling methods specifically tailored for time series data.

3 Robust Resampling 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 degree of resistance to outliers of bootstrap and subsampling tests for predictability, before applying them to the data.

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 $X^{B*}_{(n,m)} = (X^*_1, \ldots, X^*_n)$ denote a block bootstrap sample, constructed using overlapping blocks of size m. Similarly, let $X^{S*}_{(n,m)} = (X^*_1, \ldots, X^*_m)$ denote an overlapping subsampling block. We denote by $T^{K*}_{n,m} := T(X^{K*}_{(n,m)})$, $K = S, B$, the corresponding block bootstrap and subsampling statistics, respectively. 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.

For $t \in (0,1)$, the quantile $Q^{K*}_{t,n,m}$ of $T^{K*}_{n,m}$ is defined by

$$Q^{K*}_{t,n,m} = \inf\{x | \mathbb{P}^*(T^{K*}_{n,m} \leq x) \geq t\}, \tag{4}$$

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

Quantile $Q^{K*}_{t,n,m}$ is effectively a useful nonparametric estimator of the corresponding finite-sample quantile of statistic $T(X_1, \ldots, X_n)$. We characterize the robustness properties of block bootstrap and subsampling by the breakdown point $b^K_{t,n,m}$ of the quantile (4), which is defined as the smallest fraction of outliers in the original sample such that $Q^{K*}_{t,n,m}$ diverges to infinity. In Appendix A, we provide the formal definition of the breakdown point $b^K_{t,n,m}$.

Intuitively, when a breakdown occurs, inference about the distribution of $T(X_1, \ldots, X_n)$ 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 $b^K_{t,n,m}$ 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, 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 $z_{(n)} = \big(z_1, \ldots, z_n\big)$ 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

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

and we modify $y_{\max}$ over the interval $[y_{\max}, y_{\max} + 5]$. It 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 β, estimated by the standard block bootstrap and the subsampling. It is a natural exercise, as the length of the confidence interval for parameter β 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 achieve bootstrap and subsampling tests with improved breakdown properties, it is essential to employ resampling procedures on a robust statistic with a nontrivial breakdown point ($b > 1/n$), such as a robust estimator of linear regression. Without loss of generality, let $T_n = T(X_{(n)})$ be a statistic with a breakdown point of $1/n < b \leq 0.5$.

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. 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).

In our theorems, we show that $b^S_{t,n,m}$ and $b^B_{t,n,m}$ satisfy the following bounds,

$$\frac{\lceil mb \rceil}{n} \leq b^S_{t,n,m} \leq \frac{1}{n} \cdot \left[\inf_{\{p \in \mathbb{N}, p \leq r-1\}} \left\{ p \cdot \lceil mb \rceil \middle| p > \frac{(1-t)(n-m+1) + \lceil mb \rceil - 1}{m} \right\} \right],$$
$$\frac{\lceil mb \rceil}{n} \leq b^B_{t,n,m} \leq \frac{1}{n} \cdot \left[\inf_{\{p_1, p_2\}} \left\{ p = p_1 \cdot p_2 \middle| P\left(Z \geq \left\lceil \frac{nb}{p_1} \right\rceil \right) > 1-t \right\} \right],$$

where $Z \sim BIN\left(r, \frac{mp_2-p_1+1}{n-m+1}\right)$, $p_1, p_2 \in \mathbb{N}$, with $p_1 \leq m, p_2 \leq r-1$. The term $\frac{(1-t)(n-m+1)}{m}$ represents the number of degenerated subsampling statistics necessary in order to cause the breakdown of $Q^{S*}_{t,n,m}$, while $\frac{[mb]}{n}$ 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 (22), for $m = 1$.

$n = 120$, $b = 0.5$	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.

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 ($b = 0.5$). 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 outliers, the subsampling quantile can collapse with just 5 outliers out of 100 observations.

This breakdown point is also clearly lower than in the iid case; see Camponovo, Scaillet and Trojani (2012). 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. 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. 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). Overall, the results in Theorem 2 imply that subsampling and bootstrap tests for time series feature an intrinsic non-resistance to anomalous observations, which 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, which is consistent with our Monte Carlo results in the predictive regression model. To address this issue, it becomes imperative to introduce a novel class of robust bootstrap and subsampling tests within the context of time series. We have developed such robust methods by drawing inspiration from the fast resampling approaches discussed in works such as Davidson and McKinnon (1999), Andrews (2002), Hong and Scaillet (2006), and Camponovo, Scaillet, and Trojani (2012). Section 4.1 introduces our robust bootstrap and subsampling approach, while Section 4.2 demonstrates its favorable breakdown properties. Additionally, Section 4.3 introduces new robust bootstrap and subsampling tests specifically tailored for 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

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

where $\psi_n(X_{(n)}, \cdot) : \mathbb{R}^d \to \mathbb{R}^d$ depends on parameter θ and a bounded estimating function g. Boundedness of function g is a characterizing feature of robust M-estimators.

Conventional bootstrap (subsampling) methods solve equation $\psi_k(X^{K*}_{(n,m)}, \hat{\theta}^{K*}_{n,m}) = 0$, for each bootstrap (subsampling) random sample $X^{K*}_{(n,m)}$, which can be a computationally demanding task. Instead, we consider a standard Taylor expansion of (6) around the true parameter θ_0,

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

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

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

$$L^{K*}_{n,m}(x) = \frac{1}{N} \sum_{s=1}^{N} \mathbb{I}\Bigg(\tau_k \Bigg(-[\nabla_\theta \psi_n(X_{(n)}, \hat{\theta}_n)]^{-1} \psi_k(X^{K*}_{(n,m),s}, \hat{\theta}_n) \Bigg) \leq x \Bigg), \tag{8}$$

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).

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 θ_0 and matrix $-[\nabla_\theta \psi_n(X_{(n)}, \theta_0)]^{-1}$, based on the original sample $X_{(n)}$. A closer inspection of quantity $-[\nabla_\theta \psi_n(X_{(n)}, \hat{\theta}_n)]^{-1}\ \psi_k(X^{K*}_{(n,m),s}, \hat{\theta}_n)$ 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 bounded estimating function g, situation (ii) cannot arise. Intuitively, we expect the breakdown of the quantiles of robust fast resampling distribution (8) to arise only when condition (i) is met. In Corollary 3 in Appendix A, we compute the quantile breakdown point $b^K_{t,n,m}$ of the robust fast resampling distribution (8), which depends only on the breakdown properties of $\hat{\theta}_n$ and $-[\nabla_\theta \psi_n(X_{(n)}, \hat{\theta}_n)]^{-1}$. 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 predictive regression model

$$y_t = \theta' w_{t-1} + u_t\ ,\ \ t = 1, \ldots, n\ , \tag{9}$$

with $\theta = (\alpha, \beta)'$ and $w_{t-1} = (1, x_{t-1})'$, and denote by $z_{(n)} = (z_1, \ldots, z_n)$ 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. Among those estimators, a convenient choice is the Huber estimator of regression, which 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

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

where the function h_c is defined as

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

In Equation (10), we can write the Huber estimator $\hat{\theta}_n^R$ as a weighted least square estimator with data-driven weights h_c defined by (11). By design, the Huber weight $0 \leq h(z_t, \theta) \leq 1$ reduces the influence of potential anomalous observations on the estimation results. Even if the weights in (11) are nonlinear, the underlying model is still the linear model (9). Equation (10) is an estimating function and not the way we define the predictive relationship. Weights below one indicate a potentially anomalous data-point, while weights equal to one indicate unproblematic observations for the postulated model. Therefore, the value of weight (11) provides a useful way for highlighting potential anomalous observations that might be influential for the 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 can be determined in a fully data-driven way. Appendix C presents in detail the data-driven method for the selection of the tuning constant c. 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 $T_n^{NS} = \sqrt{n}\left(\hat{\theta}_n^R - \theta_0\right)$, using the following robust fast resampling distribution:

$$L_{n,m}^{NS,K*}(x) = \frac{1}{N}\sum_{s=1}^{N}\mathbb{I}\left(\sqrt{k}\left(- [\nabla_\theta \psi_{n,c}(z_{(n)}, \hat{\theta}_n)]^{-1}\psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n)\right) \leq x\right) \ , \ K = B, S, \quad (12)$$

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 (12) 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 of the robust Huber estimator (10) $b = 0.5$, and $\nabla_\theta \psi_{n,c}(z_{(n)}, \hat{\theta}_n)$ possessing maximal breakdown properties, as established in Corollary 4 of Appendix A.

Using nonstudentized statistics, robust resampling distribution (12) 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 we can also apply our robust approach 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 $\hat{\theta}_n^R$. In Appendix A, we discuss robust resampling approximations for the sampling distribution of T_n. Moreover, we analyze their robustness properties by deriving quantile breakdown point formulas.

While the quantile breakdown points of these latter robust resampling distributions are clearly larger than those of the conventional bootstrap and subsampling derived in Section 3.3, they are instead typically smaller than the maximal breakdown point quantiles implied by the robust resampling distribution (12). Therefore, we perform our robust empirical analysis based of the distribution (12).

4.3.2 Monte Carlo Evidence

To quantify the implications of Corollary 4, we can study the sensitivity of confidence intervals estimated by the robust block bootstrap and the robust subsampling distributions (12), 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 outliers, with a percentage increase in the average confidence interval length that is less than 1%, even for an outliers of size $y_{max} + 5$.

The robustness exhibited by the resampling distribution (12) holds promising implications for the power of bootstrap and subsampling tests in the presence of anomalous observations. In the same Monte Carlo setting as discussed in Sections 2.2 and 2.3, Figure 3 illustrates that under non-contaminated samples (depicted by the straight line), the rates of null hypothesis rejections for robust resampling approaches closely align with those observed for non-robust methods. Specifically, for instances like $\beta_0 = 0$, the rejection frequency approximates the nominal level of 10%, while the power exceeds 95% for $\beta_0 = 0.15$. It suggests that the asymptotic efficiency loss of robust estimators, in the absence of anomalous observations, does not seem to diminish the performance of robust resampling methods compared to non-robust procedures. In the presence of anomalous observations (indicated by the dashed line), robust approaches maintain an accurate empirical size close to the actual nominal level, along with a power curve resembling that obtained in the non-contaminated Monte Carlo simulation. Notably, in contrast to standard tests, both robust tests exhibit a power exceeding 90% for $\beta_0 = 0.15$. Additionally, unreported Monte Carlo results for the robust block bootstrap and subsampling distributions of studentized tests (28) confirm the enhanced resistance properties of our approach.

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 and the share of labor income to consumption. 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.

The empirical study is articulated in three parts. Section 5.1 studies the forecast ability 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. 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. 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. We also consider regressions with three predictive variables that additionally incorporate interest rate proxies. Below, we discuss the main results.

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

$$D_t = d_t + (1 + r_t)d_{t-1} + (1 + r_t)(1 + r_{t-1})d_{t-2} + \cdots + (1 + r_t) \dots (1 + r_{t-10})d_{t-11}, \qquad (13)$$

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

$$\ln(R_t) = \alpha + \beta \ln\left(\frac{D_{t-1}}{P_{t-1}}\right) + \epsilon_t \ ; \ t = 1, \ldots, n, \tag{14}$$

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 β in the sample period 1980-2010.

We find that while the robust bootstrap and subsampling tests always clearly reject the 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. Figure 8 plots the time series of Huber weights estimated by the robust estimator (10) of the predictive regression model (14).

We find that subperiod 1998-2002 is characterized by a cluster of infrequent anomalous observations, which are likely related to the abnormal stock market performance during the NASDAQ bubble in the second half of the 1990s. Similarly, we find a second cluster of anomalous observations in subperiod 2008-2010, which is linked to the extraordinary events of the recent financial crisis. Overall, anomalous observations are less than 4.2% of the whole data sample, and they explain 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

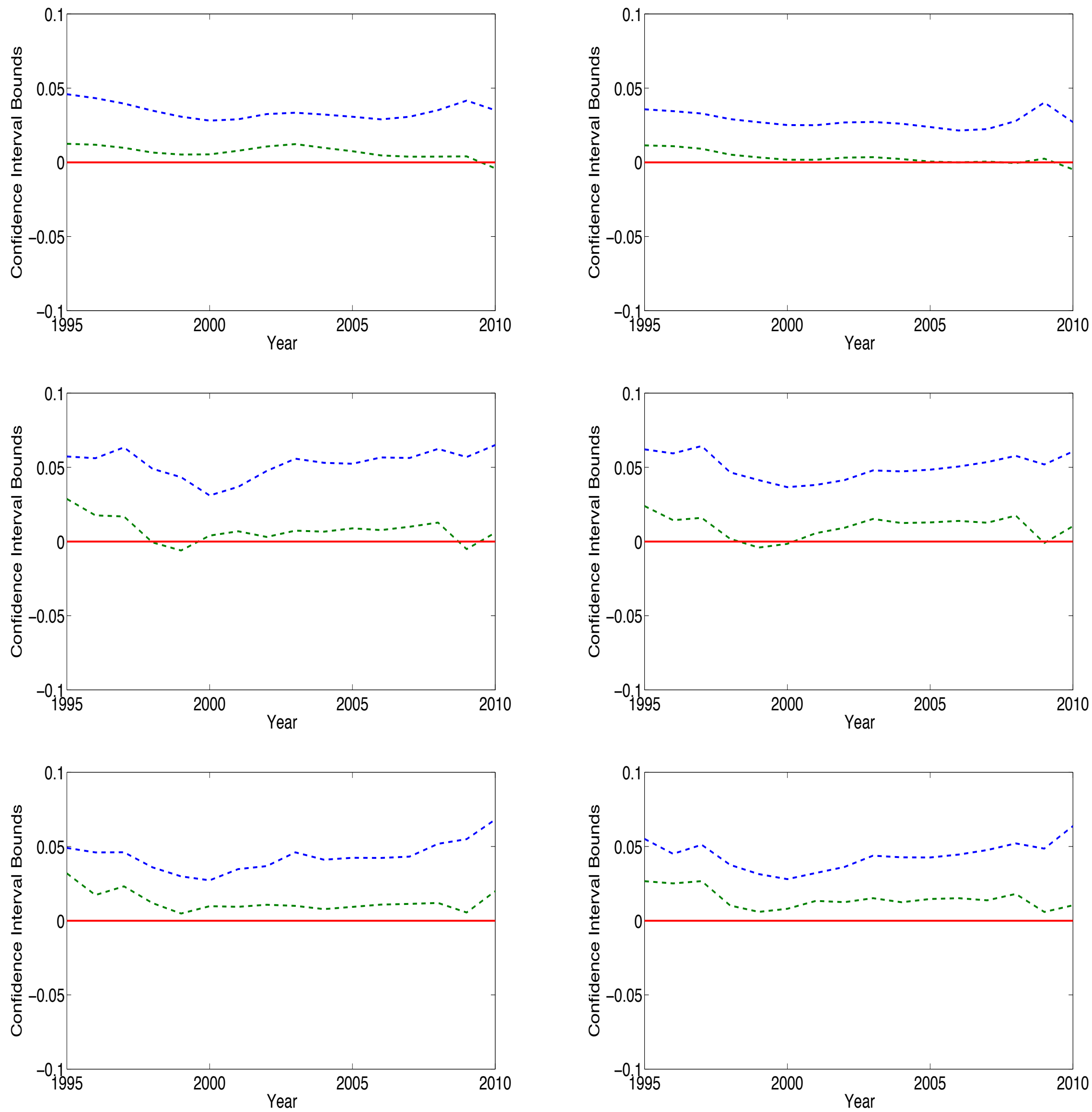

Figure 7: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β in the predictive regression model (14). We consider rolling windows of 180 observations for the period 1980-2010. In the first line, we present the 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).

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

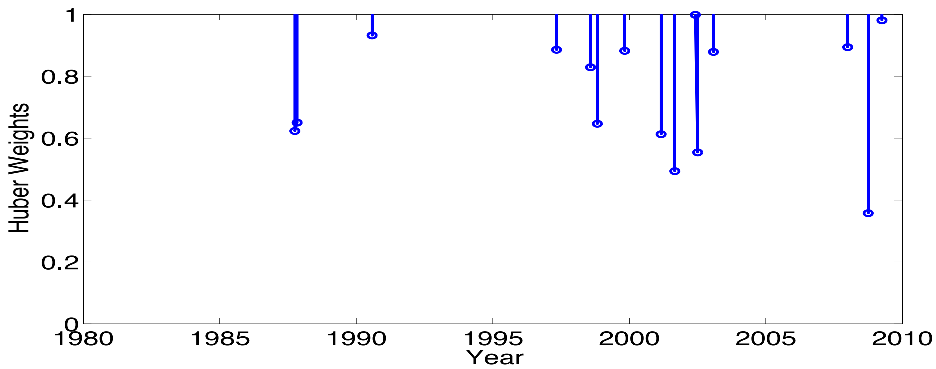

Figure 8: **Huber Weights under the Predictive Regression Model (14).** We plot the Huber weights for the predictive regression model (14) in the period 1980-2010.

significance evidence in favor of predictability.

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

$$R^2_{OS} = 1 - \frac{\sum_{t=t_1+1}^{t_2}(y_t - \hat{y}_{t,ROB})^2}{\sum_{t=t_1+1}^{t_2}(y_t - \hat{y}_{t,OLS})^2}, \tag{15}$$

where $\hat{y}_{t,ROB}$ and $\hat{y}_{t,OLS}$ are the fitted values from a predictive regression estimated up to period t_1 for the out-of-sample forecast periods $t_1+1,\dots,t_2$, using the robust Huber estimator and the OLS estimator, respectively. Whenever statistic R^2_{OS} is positive, the robust approach yields a lower average mean squared prediction error than the nonrobust method, providing more accurate out-of-sample forecasts. As reported in Table 3, we obtain $R^2_{OS} = 0.51\%$. Therefore, besides the more robust in-sample results, our robust approach also yields better out-of-sample predictions. To compare the out-of-sample accuracy of the nonrobust and robust approaches with respect to the simple forecast based on the sample mean of market returns, we consider also the out-of-sample $R^2_{OS,K}$ statistic, defined as

$$R^2_{OS,K} = 1 - \frac{\sum_{t=t_1+1}^{t_2}(y_t - \hat{y}_{t,K})^2}{\sum_{t=t_1+1}^{t_2}(y_t - \bar{y}_t)^2}, \tag{16}$$

where $\bar{y}_t$ is the historical average return estimated through period t_1, and $K = ROB, OLS$.

As reported in Table 3, we obtain $R^2_{OS,ROB} = 4.04\%$, and $R^2_{OS,OLS} = 3.51\%$. Therefore, both nonrobust and robust methods provide more accurate out-of-sample predictions than simple forecast based on the sample mean of market returns.

	R^2_{OS}	$R^2_{OS,OLS}$	$R^2_{OS,ROB}$
Shiller	0.0051	0.0351	0.0404
Bollerslev et al.	0.0140	0.0437	0.0570
Santos and Veronesi	0.0113	−0.0389	−0.0273

Table 3: **Out-of-Sample R^2 Statistics.** We report the out-of-sample R^2 statistics for the single predictor model introduced in Section 5.1 (Shiller), and the two-predictor models analyzed in Sections 5.2.1 and 5.2.2 (Bollerslev et al., and Santos and Veronesi), respectively.

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), which 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:

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

where $\ln(R_{t+k,t}) := \ln(R_{t+1}) + \cdots + \ln(R_{t+k})$ 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 set $k = 4$.

Let β_{01} and β_{02} denote the true values of parameters β_1 and β_2, respectively. Using the conventional bootstrap and subsampling tests, as well as our robust tests, we first test the null hypothesis of no return predictability by the dividend yield, $\mathcal{H}_{01} : \beta_{01} = 0$.

Figure 9 plots the 90%-confidence intervals for parameter β_1, based on rolling windows of 180 monthly observations in sample period 1990-2010. 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 (17). 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 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 (17), 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 9, explaining the large estimated confidence intervals of nonrobust tests in subperiod 1994-2009.

We also test the hypothesis of no predictability by the variance risk premium, $\mathcal{H}_{02} : \beta_{02} = 0$. Figure 10 plots the resulting confidence intervals for parameter β_{02}. 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 (2014). Finally, for this predictive regression model, we obtain out-of-sample statistics $R^2_{OS} = 1.40\%$ and $R^2_{OS,ROB} = 5.70\%$, indicating again an

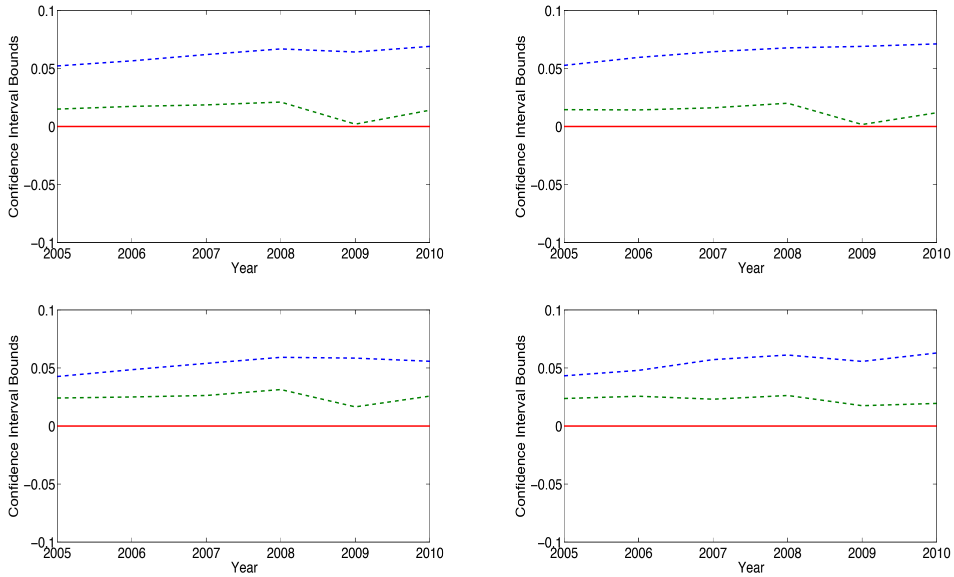

Figure 9: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_1 in the predictive regression model (17). We consider rolling windows of 180 observations for the period 1990-2010. In the top line, we present the 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).

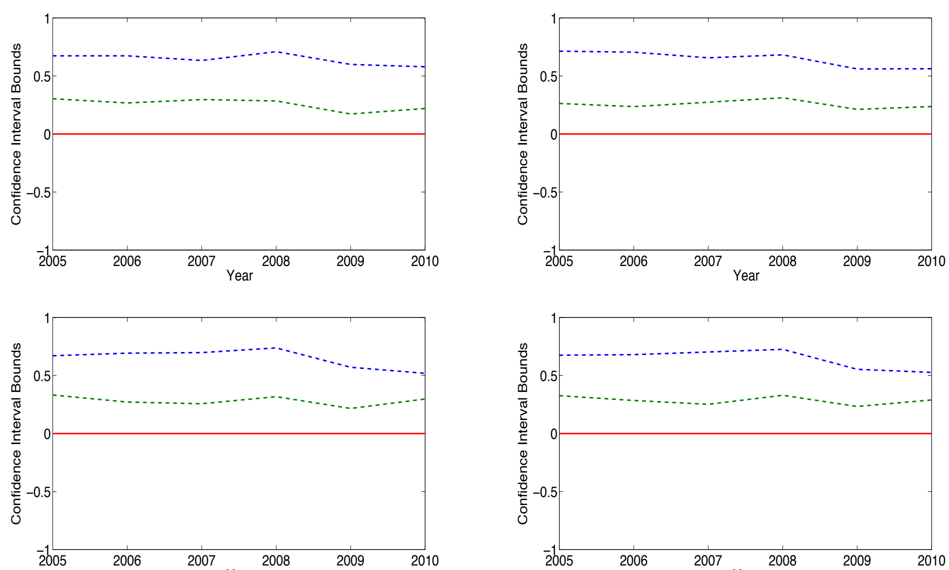

Figure 10: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_2 in the predictive regression model (17). We consider rolling windows of 180 observations for the period 1990-2010. In the top line, we present the 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).

improved out-of-sample predictive power for our robust approach.

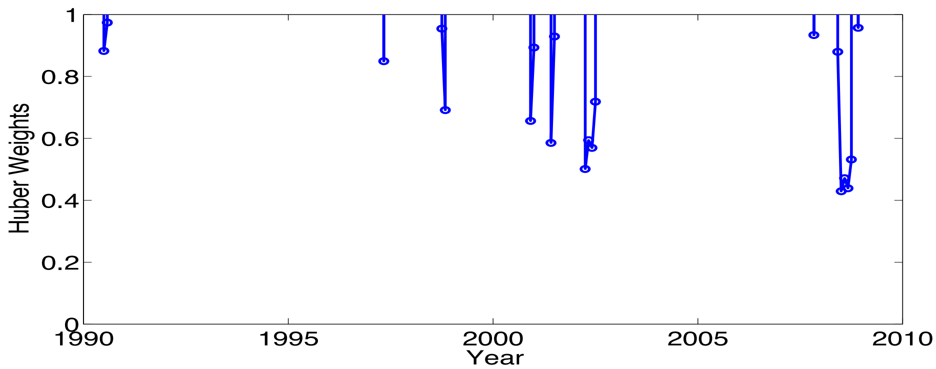

Figure 11: **Huber Weights under the Predictive Regression Model (17).** We plot the Huber weights for the predictive regression model (17) in the period 1990-2010.

Besides the two-predictor model (17), we also consider the three-predictor model

$$\frac{1}{k}\ln(R_{t+k,t}) = \alpha + \beta_1 \ln\left(\frac{D_t}{P_t}\right) + \beta_2 VRP_t + \beta_3 LTY_t + \epsilon_{t+k,t}, \tag{18}$$

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 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 (17) again confirms the reliability of our robust tests and the (possible) failure of nonrobust procedures in uncovering predictability structures in presence of anomalous observations.

5.2.2 Santos and Veronesi

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

$$\ln(R_t) = \alpha + \beta_1 \ln\left(\frac{D_{t-1}}{P_{t-1}}\right) + \beta_2 s_{t-1} + \epsilon_t, \tag{19}$$

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. 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.

Let β_{01} and β_{02} denote the true values of parameters β_1 and β_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 12 plots the 90%-confidence intervals for parameter β_{01} based on rolling windows of 180 quarterly observations in sample period 1950-2010. 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 13 reports the 90%-confidence intervals estimated in tests of the null hypothesis of no predictability by the labor income proxy, $\mathcal{H}_{02} : \beta_{02} = 0$. 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 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 observations from the NASDAQ bubble or the recent financial crisis, a feature that have already noted also in Santos and Veronesi (2006) and Lettau and Van Nieuwerburgh (2007) from a different angle. In such contexts, the robust bootstrap and subsampling tests are again found to control well the potential damaging effects of anomalous observations, by providing a way to consistently uncover hidden predictability features also when the data may only approximately follow the given predictive regression model. We do not find evidence of structural breaks at

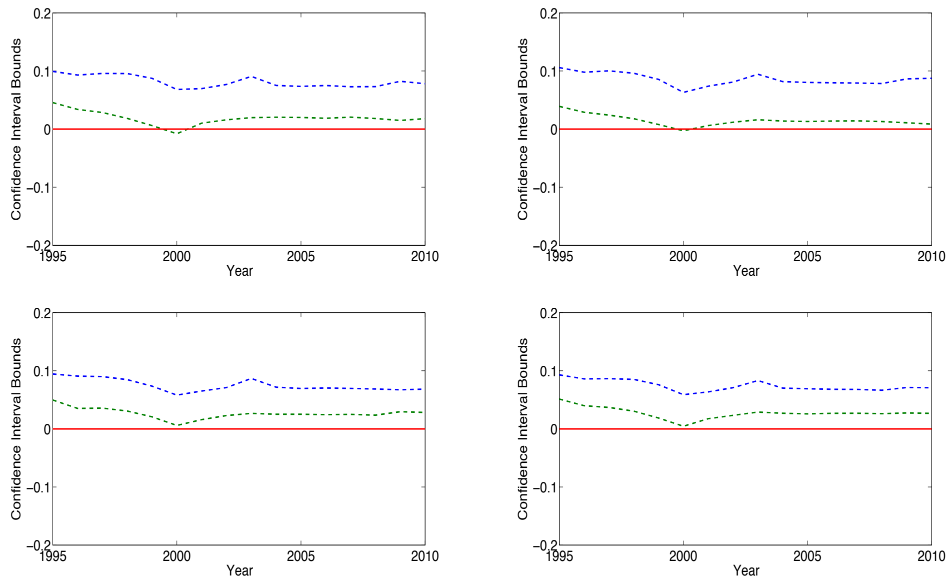

Figure 12: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_1 in the predictive regression model (19). We consider rolling windows of 180 observations for the period 1950-2010. In the top line, we present the 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).

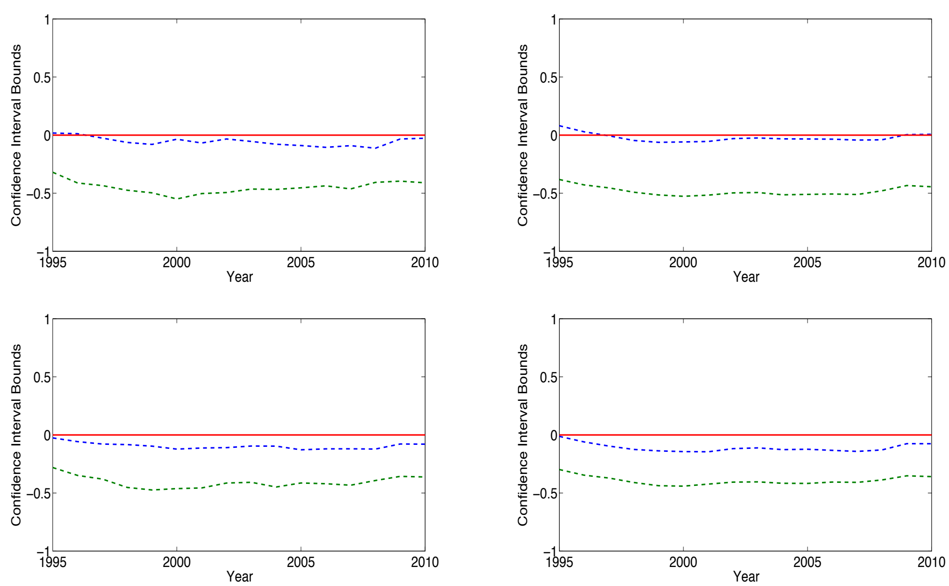

Figure 13: **Upper and lower bounds of the confidence intervals.** We plot the upper and lower bound of the 90% confidence intervals for the parameter β_2 in the predictive regression model (19). We consider rolling windows of 180 observations for the period 1950-2010. In the top line, we present the 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).

the 10% significance level, while we obtain an out-of-sample statistic $R^2_{OS} = 1.13\%$, indicating that our robust approach improves the out-of-sample predictions of classical predictive regression methods. However, in this case the out-of-sample statistic $R^2_{OS,ROB} = -2.73\%$ shows no improvement over quarterly forecasts provided by standard sample mean of market returns.

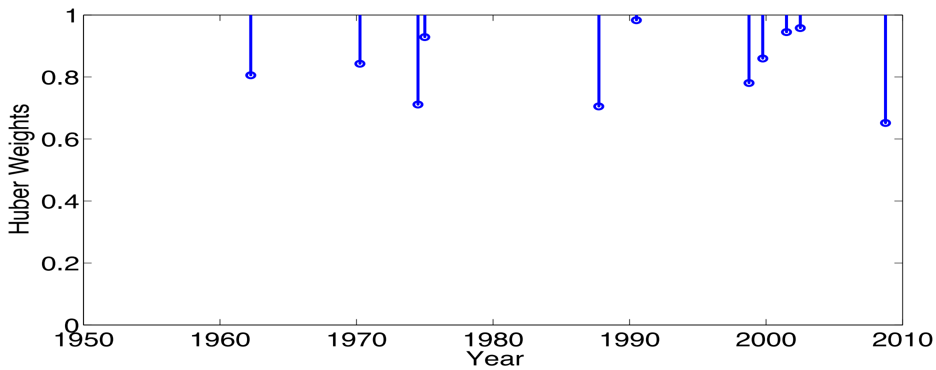

Figure 14: **Huber Weights under the Predictive Regression Model (19).** We plot the Huber weights for the predictive regression model (19) in the period 1950-2010.

6 Conclusion

Extensive research examines the predictive capabilities of various economic variables concerning future market returns. Several testing approaches have been proposed to assess the null hypothesis of no predictability in predictive regressions featuring correlated errors and nearly integrated regressors. These approaches include resampling methods, such as bootstrap and subsampling, which demonstrate improvements over conventional asymptotic tests. However, Monte Carlo analysis reveals that even minor violations of strict model assumptions can significantly compromise the reliability of these tests.

To comprehensively understand the issue, we theoretically characterize the robustness properties of bootstrap and subsampling tests in a time series context, using the concept of quantile breakdown point as a measure of the global resistance of a testing procedure to outliers. We derive general quantile breakdown point formulas, highlighting the limited resistance of these tests to anomalous observations that may sporadically contaminate the predictive regression model. It confirms the fragility identified in our Monte Carlo study.

In response, we propose a more robust testing method for predictive regressions with correlated errors and nearly integrated regressors. It involves introducing a novel class of fast and robust bootstrap and subsampling procedures for time series, applicable to both linear and nonlinear predictive regression models at sustainable computational costs. The new tests exhibit resistance to anomalous observations in the data, leading to more robust confidence intervals and inference results. Monte Carlo simulations demonstrate their strong resistance to outliers and improved finite-sample properties in the presence of anomalous observations.

In an empirical study using US stock market data, we investigate single-predictor and multi-predictor models employing well-known predictive variables from the literature, such as dividend yield, difference between implied volatility and realized volatility, interest rate, and share of labor income to consumption. Empirical analysis reveals robust evidence of return predictability, previously obscured by anomalous observations, both in- and out-of-sample.

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,

$$b_{t,n,m}^{K} := \frac{1}{n} \cdot \left[\inf_{\{1 \leq p \leq \lceil n/2 \rceil\}} \left\{ p \middle| \text{there exists } Z_{n,p}^{\zeta} \in \mathcal{Z}_{n,p}^{\zeta} \text{ such that } Q_{t,n,m}^{K*}(X_{(n)} + Z_{n,p}^{\zeta}) = +\infty \right\} \right], \tag{20}$$

where $\lceil x \rceil = \inf\{n \in \mathbb{N} | x \leq n\}$, 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 $\zeta \in \bar{\mathbb{R}}^{d_x}$. When $p > 1$, we do not necessarily assume outliers $\zeta_1, \ldots, \zeta_p$ to be all equal to ζ, but we rather assume existence of constants $c_1, \ldots, c_p$, 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 $\mathcal{Z}_{n,p}^{\zeta}$. 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.

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

$$\frac{\lceil mb \rceil}{n} \leq b^S_{t,n,m} \leq \frac{1}{n} \cdot \left[\inf_{\{p \in \mathbb{N}, p \leq r-1\}} \left\{ p \cdot \lceil mb \rceil \middle| p > \frac{(1-t)(n-m+1) + \lceil mb \rceil - 1}{m} \right\} \right], \tag{21}$$

$$\frac{\lceil mb \rceil}{n} \leq b^B_{t,n,m} \leq \frac{1}{n} \cdot \left[\inf_{\{p_1, p_2\}} \left\{ p = p_1 \cdot p_2 \middle| P\left(Z \geq \left\lceil \frac{nb}{p_1} \right\rceil \right) > 1-t \right\} \right], \tag{22}$$

where $Z \sim BIN\left(r, \frac{mp_2 - p_1 + 1}{n-m+1}\right)$, $p_1, p_2 \in \mathbb{N}$, with $p_1 \leq m, p_2 \leq r-1$.

Proof of Theorem 2. We first consider the subsampling and focus on formula (21). 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 (21), we denote by $X^N_{(m),i} = (X_{(i-1)m+1}, \ldots, X_{im})$, $i = 1, \ldots, r$ and $X^O_{(m),i} = (X_i, \ldots, X_{i+m-1})$, $i = 1, \ldots, n-m+1$, the nonoverlapping and overlapping blocks of size m, respectively. Given the original sample $X_{(n)}$, for the first nonoverlapping block $X^N_{(m),1}$, consider the following type of contamination:

$$X^N_{(m),1} = (X_1, \ldots, X_{m-\lceil mb \rceil}, Z_{m-\lceil mb \rceil+1}, \ldots, Z_m), \tag{23}$$

where X_i, $i = 1, \ldots, m - \lceil mb \rceil$ and Z_j, $j = m - \lceil mb \rceil + 1, \ldots, m$, denote the noncontaminated and contaminated points, respectively. By construction, the first $m - \lceil mb \rceil + 1$ overlapping blocks $X^O_{(m),i}$, $i = 1, \ldots, m - \lceil mb \rceil + 1$, contain $\lceil mb \rceil$ outliers. Consequently, $T(X^O_{(m),i}) = +\infty$, $i = 1, \ldots, m - \lceil mb \rceil + 1$. Assume that the first $p < r-1$ nonoverlapping blocks $X^N_{(m),i}$, $i = 1, \ldots, p$, have the same contamination as in (23). Because of this contamination, the number of statistics $T^{OS*}_{n,m}$ which diverge to infinity is $mp - \lceil mb \rceil + 1$.

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

Therefore,

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

Finally, we consider formula (22). The proof of the first inequality in formula (22) follows the same lines as the proof of the first inequality in the formula (21). We focus on the second inequality.

Consider $X^N_{(m),i}$, $i = 1, \ldots, r$. 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 (23). 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 $X \sim BIN(r, \frac{mp_2 - p_1 + 1}{n-m+1})$.

By Equation (20), $Q^{OB*}_{t,n,m} = +\infty$, when the proportion of statistics $T^{OB*}_{n,m}$ with $T^{OB*}_{n,m} = +\infty$ is larger than $(1-t)$. The smallest number of outliers such that $T^{OB*}_{n,m} = +\infty$ is by definition nb. Let $p_1, p_2 \in \mathbb{N}, p_1 \leq m, p_2 \leq r-1$. Consequently,

$$b^{OB}_{t,n,m} \leq \frac{1}{n} \cdot \left[\inf_{\{p_1,p_2\}} \left\{ p = p_1 \cdot p_2 \middle| P\left(Z \geq \left\lceil \frac{nb}{p_1} \right\rceil \right) > 1-t \right\} \right],$$

where $Z \sim BIN\left(r, \frac{mp_2 - p_1 + 1}{n-m+1}\right)$. It 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 $L^{K*}_{n,m}$ 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* $b^K_{t,n,m} = \min(b, b_{\nabla\psi})$*, where* $b_{\nabla\psi}$ *is the breakdown point of matrix* $\nabla_\theta \psi_n(X_{(n)}, \hat{\theta}_n)$*, defined by:*

$$b_{\nabla\psi} = \frac{1}{n} \cdot \inf_{1 \le p \le \lceil n/2 \rceil} \left\{ p \middle| \text{there exists } Z^\zeta_{n,p} \in \mathcal{Z}^\zeta_{n,p} \text{ such that } \det(\nabla_\theta \psi_n(X_{(n)} + Z^\zeta_{n,p}, \hat{\theta}_n)) = 0 \right\}. \tag{24}$$

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

$$-[\nabla_\theta \psi_n(X_{(n)}, \hat{\theta}_n)]^{-1} \psi_k(X^{K*}_{(n,m),s}, \hat{\theta}_n), \tag{25}$$

where $k = n$ or $k = m$, $K = B, S$. Assuming a bounded estimating function, Expression (25) 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 $Q^{K*}_{t,n,m}$ 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 $Q^{K*}_{t,n,m}$ is $b^K_{t,n,m} = \min(b, b_{\nabla\psi})$. ■

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 $T^{NS}_n = \sqrt{n}\left(\hat{\theta}^R_n - \theta_0\right)$ defined in (12).

Corollary 4 *Let* $t \in (0, 1)$*. The t-quantile breakdown point of the resampling distribution (12) is* $b^K_{t,n,m} = 0.5$*,* $K = B, S$*.*

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

$$g_c(y_t, w_{t-1}, \theta) = (y_t - \theta' w_{t-1}) w_{t-1} \cdot \min\left(1, \frac{c}{||(y_t - \theta' w_{t-1}) w_{t-1}||}\right). \tag{26}$$

Using some algebra, we can show that

$$\nabla_\theta g_c(y_t, w_{t-1}, \theta) = \begin{cases} -(1, x_{t-1})'(1, x_{t-1}), & \text{if } ||(y_t - \theta' w_{t-1}) w_{t-1}|| \leq c, \\ \mathbb{O}_{2\times 2}, & \text{if } ||(y_t - \theta' w_{t-1}) w_{t-1}|| > c, \end{cases} \tag{27}$$

where $\mathbb{O}_{2\times 2}$ denotes the 2×2 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:

$$L_{n,m}^{K*}(x) = \frac{1}{N} \sum_{s=1}^{N} \mathbb{I}\left(\sqrt{k}\left(- [\hat{\Sigma}_{k,s}^{R*}]^{-1/2} [\nabla_\theta \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n)]^{-1} \psi_k(z_{(n,m),s}^{K*}, \hat{\theta}_n)\right) \leq x\right), \tag{28}$$

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 (28) 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 (28), respectively, are given by*

$$b_{t,n,m}^S = \frac{1}{n}\left[\inf_{\{p\in\mathbb{N}, p\leq n-m+1\}} \left\{ m + p \,\middle|\, p > (1-t)(n-m+1) - 1 \right\}\right], \tag{29}$$

$$b_{t,n,m}^B = \frac{1}{n}\left[\inf_{\{p\in\mathbb{N}, p\leq n-m+1\}} \left\{ m + p \,\middle|\, P\Big(Z = r\Big) > 1 - t \right\}\right], \tag{30}$$

where $Z \sim BIN\left(r, \frac{p+1}{n-m+1}\right)$ *and* $q \in (0,1)$.

Proof of Theorem 5. Consider the resampling distribution (28). Let $z^{K*}_{(n,m),s} = (z^*_1, \ldots, z^*_k)$ 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

$$T^{K*}_{n,m,s} := [\hat{\Sigma}^{R*}_{k,s}]^{-1/2}[\nabla_\theta \psi_{k,c}(z^{K*}_{(n,m),s}, \hat{\theta}^R_n)]^{-1} \psi_{k,c}(z^{K*}_{(n,m),s}, \hat{\theta}^R_n), \tag{31}$$

may degenerate when (i) $det\left(\hat{\Sigma}^{R*}_{k,s}\right) = 0$ or (ii) $det\left(\nabla_\theta \psi_{k,c}(z^{K*}_{(n,m),s}, \hat{\theta}^R_n)\right) = 0$. Moreover, also note that $\hat{\Sigma}^{R*}_{k,s} = \hat{J}'^{R*}_{k,s} \hat{V}^{R*}_{k,s} \hat{J}^{R*}_{k,s}$, where $\hat{J}^{R*}_{k,s} = \left[\nabla_\theta \psi_{k,c}(z^{K*}_{(n,m),s}, \hat{\theta}^R_n)\right]^{-1}$. Because of Equations (38), (39), (27), it turns out that cases (i) and (ii) can be satisfied only when $||(y^*_t - \hat{\theta}'^R_n w^*_{t-1}) w^*_{t-1}|| > c$ for all random observations $z^*_i = (y^*_i, w^*_{i-1})'$, $i = 1, \ldots, k$, where $k = n, m$, for the bootstrap and subsampling, respectively.

For the original sample, consider following type of contamination

$$z_{(n)} = (z_1, \ldots, z_j, C_{j+1}, \ldots, C_{j+p}, z_{j+p+1}, \ldots, z_n), \tag{32}$$

where z_i, $i = 1, \ldots, j$ and $i = j+p+1, \ldots, n$ and C_i, $i = j+1, \ldots, j+p$, 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

$$(C_{j+i}, \ldots, C_{j+i+m-1}), \tag{33}$$

$i = 1, \ldots, p - m + 1$ contain only outliers. Therefore, for these $p - m + 1$ blocks we have that $det\left(\nabla_\theta \psi_{m,c}(C_{j+i}, \ldots, C_{j+i+m-1}, \hat{\theta}^R_n)\right) = 0$, i.e., some components of vector (31) may degenerate to infinity. Moreover, $Q^{S*}_{t,n,m} = +\infty$ when the proportion of statistics $T^{S*}_{n,m}$ with $T^{S*}_{n,m} = +\infty$ is larger than $(1-t)$. Therefore, $b^S_{t,n,m} = \inf_{\{p \in \mathbb{N}, m \leq p \leq n-m+1\}} \left\{ \frac{p}{n} \middle| \frac{p-m+1}{n-m+1} > 1 - t \right\}$, which proves the result in Equation (29).

For the result in Equation (30), note that because of the contamination defined in (32), 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 $X \sim BIN\left(r, \frac{p-m+1}{n-m+1}\right)$. To imply (i) or (ii), all the random observations $(z_1^*, \ldots, z_k^*)$ have to be outliers, i.e., $X = r$. By Equation (20), $Q_{t,n,m}^{B*} = +\infty$, when the proportion of statistics $T_{n,m}^{B*}$ with $T_{n,m}^{B*} = +\infty$ is larger than $(1-t)$. Consequently,

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

where $Z \sim \left(r, \frac{p-m+1}{n-m+1}\right)$. It concludes the proof. ■

Formulas (29) and (30) improve on the results in Equations (21) and (22) 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 (28) 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 (28) and construct robust confidence intervals for the components of a general d-dimensional parameter β, where $d \geq 1$. The extension to the nonstudentized distributions (12) is straightforward. We construct symmetric resampling confidence intervals for the parameter of interest. Indeed, 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.

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

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

where $z_t = (y_t, w_{t-1}')'$, $w_{t-1} = (1, x_{t-1}')'$, Φ is a d-dimensional parameter vector and R is a $d \times d$ parameter matrix. First, we compute the robust Huber estimator $\hat{\theta}_n^R = (\hat{\alpha}_n^R, \hat{\beta}_n^{R\prime})'$ 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

$$
\Sigma^R = J'^R V^R J^R, \tag{34}
$$

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

$\nabla_\theta \psi_{n,c}(z_{(n)}, \theta_0)$ denotes the derivative of function $\psi_{n,c}$ with respect to parameter θ with $\theta_0 = (\alpha_0, \beta_0')'$. We compute the estimator $\hat{\Sigma}_n^R = \hat{J}_n'^R \hat{V}_n^R \hat{J}_n^R$ of the asymptotic variance Σ^R where

$$\hat{J}_n^R = \left[\nabla_\theta \psi_{n,c}(z_{(n)}, \hat{\theta}_n^R)\right]^{-1}, \tag{35}$$

$$\hat{V}_n^R = \frac{1}{n}\sum_{i=1}^{n} g_c(z_i, \hat{\theta}_n^R) g_c(z_i, \hat{\theta}_n^R)', \tag{36}$$

respectively. 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. 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 β, we compute an estimator of the sampling distribution of $\left(T_{n,|\cdot|}\right)^{(j+1)} := \frac{\sqrt{n}}{(\hat{\sigma}_n^R)^{(j+1)}} \left| \left(\hat{\theta}_n^R - \theta_0\right)^{(j+1)} \right|$, 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 $\left(\hat{\sigma}_n^R\right)^{(j+1)}$ 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 $\left(T_{n,|\cdot|}\right)^{(j+1)}$ through the robust resampling distribution

$$\left(L_{n,m,|\cdot|}^{K*}(x)\right)^{(j+1)} = \frac{1}{N}\sum_{s=1}^{N} \mathbb{I}\left(\frac{\sqrt{k}}{\left(\hat{\sigma}_{k,s}^{R*}\right)^{(j+1)}} \left| \left([\nabla_\theta \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n^R)]^{-1} \psi_{k,c}(z_{(n,m),s}^{K*}, \hat{\theta}_n^R) \right)^{(j+1)} \right| \leq x \right), \tag{37}$$

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

$\hat{J}'^{R*}_{k,s}\hat{V}^{R*}_{k,s}\hat{J}^{R*}_{k,s}$ with

$$\hat{J}^{R*}_{k,s} = \left[\nabla_\theta \psi_{c,c}(z^{K*}_{(n,m),s}, \hat{\theta}^R_n)\right]^{-1}, \tag{38}$$

$$\hat{V}^{R*}_{k,s} = \frac{1}{k}\sum_{i=1}^{k} g_c(z^{K*}_{i,s}, \hat{\theta}^R_n) g_c(z^{K*}_{i,s}, \hat{\theta}^R_n)'. \tag{39}$$

Finally, let $t \in (0,1)$, and let $\left(Q^{K*}_{t,n,m}\right)^{(j+1)}$ be the t-quantile of the block bootstrap or subsampling empirical distribution (37), $K = B, S$, respectively. Then, the symmetric t-confidence interval for the j-th component $(\beta_0)^{(j)}$ of β_0 is given by

$$(CI_t)^{(j)} = \left[\left(\hat{\theta}^R_n\right)^{(j+1)} - \left(\hat{\sigma}^R_n\right)^{(j+1)}\left(Q^{K*}_{t,n,m}\right)^{(j+1)}, \left(\hat{\theta}^R_n\right)^{(j+1)} + \left(\hat{\sigma}^R_n\right)^{(j+1)}\left(Q^{K*}_{t,n,m}\right)^{(j+1)}\right], \tag{40}$$

We summarize our robust approach in the following algorithm.

(1) Compute $\hat{\theta}^R_n = (\hat{\alpha}_{ROB}, \hat{\beta}'_{ROB})'$, as the solution of (10), where c is selected using the data-driven method introduced in Appendix C below.

(2) Compute $\hat{\Sigma}^R_n = \hat{J}'^R_n \hat{V}^R_n \hat{J}^R_n$, where $\hat{J}'^R_n$ and $\hat{V}^R_n$ are defined in (35) and (36), respectively.

(3B) For the robust block bootstrap, generate $B_B = 999$ random bootstrap samples based on the overlapping blocks $(z_i, \ldots, z_{i+m-1})$, $i = 1, \ldots, n-m+1$, 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 $(z_i, \ldots, z_{i+m-1})$, $i = 1, \ldots, n-m+1$, where m is selected according to the data-driven method introduced in Appendix C below.

(4) Compute the robust resampling distribution (37).

(5) The robust symmetric t-confidence interval for the j-th component $(\beta_0)^{(j)}$ of β_0 is is given by (40).

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 $\mathcal{M} := \{m_{min} < \cdots < m_{max}\}$ and $\mathcal{C} := \{c_{min} < \cdots < c_{max}\}$ are the sets of admissible block sizes and estimating functions bounds, respectively. Let $T_n^{NS} = \sqrt{n}\left(\hat{\theta}_n^R - \theta_0\right)$ 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

$$(m, c)_{CM} := \arg \inf_{(m,c) \in \mathcal{MC}} \left\{ \left| t - \mathbb{P}^* \left[\hat{\theta}_n^R \in CI_{t,(m,c)} \right] \right| \right\}, \qquad (41)$$

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 $h(m, c) = \sharp\{\hat{\beta}_n^R \in CI_{t,(m,c)}\}/K$.

(5) The data-driven block size and estimating function bound according to the calibration method are defined as $(m, c)_{CM} := \arg\inf_{(m,c)\in\mathcal{MC}} \{|t - h(m, c)|\}$.

Acknowledgements

We thank the Editor, the Associate Editor and two anonymous Referees for constructive criticism and numerous suggestions which have led to substantial improvements over the previous version. We also acknowledge the financial support of the Swiss National Science Foundation (NCCR FINRISK, grants 101312-103781/1, 100012-105745/1, and PDFM1-114533).

References

[1] Amihud, Y., Hurvich C.M., and Y. Wang, 2008. Multiple-predictor regressions: hypothesis testing. The Review of Financial Studies, 22, 413–434.

[2] Andersen, T.G., and R.T. Varneskov, 2022. Testing for parameter instability and structural change in persistent predictive regressions. Journal of Econometrics, 231, 361–386

[3] Andrews, D., 2002. Higher-order improvements of a computationally attractive k-step bootstrap- for extremum estimators. Econometrica, 70, 119–162.

[4] Andrews, D., and P. Guggenberger, 2010. Application of subsampling, hybrid and size-correction methods. Journal of Econometrics, 158, 285–305.

[5] Ang, A., and G. Bekaert, 2007. Stock return predictability: is it there? The Review of Financial Studies, 20, 651–707.

[6] Bollerslev, T., Marrone, J., Xu, L., and H. Zhou, 2014. Stock return predictability and variance risk premia: statistical inference and international evidence. Journal of Financial and Quantitative Analysis, 49, 633–611.

[7] Bollerslev, T., Tauchen, G., and H. Zhou, 2009. Expected stock returns and variance risk premia. Review of Financial Studies, 22, 4463–4492.

[8] Boucher,C., Jasinski, A.,and S. Tokpavi, 2023. Conditional mean reversion of financial ratios and the predictability of returns. Journal of International Money and Finance, forthcoming.

[9] Boudoukh, J., Israel, R., and M. Richardson, 2022. Biases in long-horizon predictive regressions Journal of Financial Economics, 145, 937–969 .

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

[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] Choi, I., and T.K. Chue, 2007. Subsampling-based tests of stock-return predictability, Working paper.

[14] Coqueret, G., and B. Tavin, 2023. Dynamic decision making with predictive panels. Journal of the Operational Research Society, forthcoming.

[15] Davidson, R., and J. McKinnon, 1999. Bootstrap testing in nonlinear models. International Economic Review, 40, 487–508.

[16] 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.

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

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

[19] Hall, P., 1988. On symmetric bootstrap confidence intervals. Journal of the Royal Statistical Society, Ser.B, 50, 35–45.

[20] Hall, P., and J. Horowitz, 1996. Bootstrap critical values for tests based on Generalized-Method-of-Moment estimators. Econometrica, 64, 891–916.

[21] Hampel, F. R., (1971) A General Qualitative Definition of Robustness, *Annals of Mathematical Statistics*, 42, 1887–1896.

[22] 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.

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

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

[25] Koijen, R.S., and S. Van Nieuwerburgh, 2011. Predictability of returns and cash-flows. Annual Review of Financial Economics, 3, 467–491.

[26] Lettau, M., and S. Ludvigson, 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance, 3, 815–849.

[27] 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.

[28] Lin, Y., and Y. Tu, 2020. Robust inference for spurious regressions and cointegrations involving processes moderately deviated from a unit root. Journal of Econometrics, 219, 52–65.

[29] Lucas, A., 1995. Unit root tests based on M-estimators. Econometric Theory, 11, 331–346.

[30] Lucas, A., 1997. Cointegration testing using pseudolikelihood ratio tests. Econometric Theory, 13, 149–169.

[31] 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.

[32] Mikusheva, A., 2007. Uniform inference in autoregressive models. Econometrica, 75, 1411–1452.

[33] Politis, D. N., and J. P. Romano, 1994. The stationary bootstrap. Journal of the American Statistical Association, 89, 1303–1313.

[34] Politis, D.N., Romano J.P., and M. Wolf, 1999. *Subsampling*. Springer, New York.

[35] Romano, J. P., and M. Wolf, 2001. Subsampling intervals in autoregressive models with linear time trend. Econometrica, 69, 1283–1314.

[36] Salibian-Barrera, M., and R. Zamar, 2002. Boostrapping robust estimates of regression. Annals of Statistics, 30, 556–582.

[37] Santos, T., and P. Veronesi, 2006. Labor income and predictable stock returns. Review of Financial Studies, 19, 1–44.

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

[39] Singh, K., 1998. Breakdown theory for bootstrap quantiles. Annals of Statistics, 26, 1719–1732.

[40] Stambaugh, R.F., 1986. Bias in regressions with lagged stochastic regressors. Graduate School of Business, University of Chicago, Working Paper No. 156.

[41] Stambaugh, R.F., 1999. Predictive regressions. Jornal of Financial Economics, 54, 375-421.

[42] Torous, W., R. Valkanov, and S. Yan, 2004. On predicting stock returns with nearly integrated explanatory variables. Journal of Business, 77, 937–966.

[43] 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.