Stock Return Predictability in Volatility Time*

Stefan Jacewitz†
and
Joon Y. Park‡

October 13, 2009

Abstract

Are stock returns easily predicted as far as five years into the future? Surprising to some, much of the existing empirical literature finds that stock returns are not only predictable, but highly so. We approach this old question from a completely new direction, proposing an innovative and effective test of predictability in stock returns. When we apply our new technique to the data, we find no evidence supporting stock return predictability. We show that certain data characteristics which are ubiquitous and widely accepted, such as the largely unaddressed presence of stochastic volatility in stock returns, have a substantial impact on traditional hypothesis testing. Our new technique for testing predictability is uniquely suited to each of these characteristics of predictive regression data. The technique consists of a simple time change to volatility time to accommodate a quite general form of stochastic volatility in stock returns, and instrumental variable estimation to allow for a wide range of endogenous nonstationary covariates. We provide Monte Carlo evidence to show that our methodology is superior to any of the existing tests in terms size and comparable in terms of power.

JEL Classification: C13, C22.

Key words and phrases: predictive regression, time change, Cauchy estimator, nonstationarity, stochastic volatility.

*The authors would like to thank Yoosoon Chang and Hagen Kim for numerous helpful discussions and suggestions. This paper is based in part on a dissertation submitted to Texas A&M University by Stefan Jacewitz. Joon Y. Park gratefully acknowledges the financial support from the NSF under NSF Grant No. SES-0518619.

†Texas A&M University and the Federal Deposit Insurance Corporation. The views and opinions expressed here are those of the authors and do not necessarily reflect the views of the Federal Deposit Insurance Corporation.

‡Texas A&M University and University of Indiana

1 Introduction

'Stock returns are predictable.' So states one of the new facts in finance from Cochrane (1999) and (2005). However, perhaps the question of stock return predictability may not be as settled as this would suggest. Non-constant volatility is universally accepted as part of the return process, a fact completely ignored in the existing econometric treatment of return predictability. We suggest that if this and other data characteristics are properly addressed, we are left with little to no evidence supporting stock return predictability using some of the most common predictors. It may be worth stating from the start that we do not conclude that the standard ratios are not useful predictors, nor that stock returns are all together unpredictable. Instead, we simply show that common features of the data can cause false predictors to appear real and, once this is accounted for, there is no evidence of predictability in some commonly used ratios.

This stands in contrast to much of the existing research in this area, which has consistently concluded that stock returns are not only predictable, but highly so. This conclusion has lead to a fruitful and still expanding literature on two fronts. Many theorists have developed models supporting such predictability. On the other hand, econometricians have long been troubled by the techniques often used to test for predictability. This paper takes the latter approach demonstrating some of the weaknesses in the most popular tests and offering an original and stronger alternative.

In the past, the literature has produced dozens of papers most of which rely on the same basic machinery to address predictability. We, instead, offer a completely new approach. We combine a simple nonlinear instrument and a volatility chronometer. This test is uniquely suited to address the problematic characteristics found in predictive regression data. Unlike most previous work, however, we are left with virtually no evidence in favor of predictability in stock returns. Further, we also apply two other recently proposed tests which similarly give no support to predictability. With this in mind, based on the empirical evidence, there seems little reason to conclude that stock returns are predictable using these common predictive ratios.

In terms of consensus, the predictability of stock returns stands as a particularly peculiar topic. There are, somehow, two simultaneous "consensuses". It is both widely assumed to be true, and widely presumed to be false. Generally speaking, among departments of finance, the predictability of stock returns could not be clearer. Indeed, there exists copious empirical evidence that this is the case. In many economics departments, on the other hand, the blatantly obvious truth is that stock returns cannot be predictable. It can be difficult to explain an equilibrium that supports returns which are highly predictable.

In some ways, it is so widely accepted that stock returns are predictable that it has become a stylized fact. (2005) shows thatLettau and Ludvigson (2001) reiterate the broad acceptance that excess returns are predictable by variables such as dividend-price ratios, and earnings-price ratios. This conclusion is hardly restricted to academia. Wilcox (2007) states that the clear consensus within the investment industry is that prediction based on these ratios is highly useful. According to Ferson, Sarkissian, and Simin (2003), predictive regressions are used in tactical asset allocation, active portfolio management, conditional performance evaluation, and market timing, among others.

As a byproduct of its firm establishment as a stylized fact, there have been many economic models which can support some degree of return predictability in a general equilibrium setting. Theoretical devices used to do so include consumption smoothing in Balvers, Cosimano, and McDonald (1990), habit formation in Campbell and Cochrane

(1999), heterogeneous preferences in Chan and Kogan (2002), and time varying risk preferences in Menzly, Santos, and Veronesi (2004).

Despite its general acceptance as fact, testing the hypothesis of predictability has remained a popular research topic. Many have been uneasy with the commonplace application of the standard OLS hypothesis testing. Financial data in general, and return data in particular, are widely known to have extensive econometric complications. There are several widely recognized characteristics of the covariates used to test return predictability which can cause standard hypothesis tests to over reject a true null. The most well documented and explored of these characteristics are a persistence in the regressor and correlation between innovations in the regressor and returns (see, for instance, the seminal work by Stambaugh (1999)). As a simple exercise, we generated 10,000 random unit-root sequences with innovations correlated with the actual stock returns at the empirical level. For more than 70%, the lagged values of these obviously useless predictors were found to be valid at the 10% level using the standard one-sided t-test.

Perhaps more convincing than these technical objections: predictive regressions don't actually predict very well. This has been one of the most trenchant arguments against the validity of predictive regressions. Bossaerts and Hillion (1999) find that even the best prediction models have no out-of-sample predictive power. Welch and Goyal (2008) find that the standard predictive variables perform poorly in- and out-of-sample and are outperformed by something as simple as the historical average. This is difficult to reconcile with the "new fact" that a substatial amount of stock return variation can be predicted by some standard ratios.

Predictive regression data has other problematic tendencies as well. The existence of a stochastic, time-varying volatility has been widely considered and explored in stock return related literature, though never seriously in this context. Many of considered time varying risk in the context of predictive regressions (i.e., Goyal and Clara (2003)), though never in terms of its econometric effects. As the conditional variance is clearly variable, returns have long been modeled as an ARCH or GARCH process. These models explicitly assume the existence of a constant unconditional variance. In that case, time varying volatility will have no effect on OLS estimation. However, Loretan and Phillips (1994) and Stărică and Granger (2005) find strong evidence against a constant unconditional variance. A time-varying stochastic variance process such as a volatility regime switching model or the model by Heston (1993) are natural alternatives. Cavaliere (2004) has shown that, in the presence of stochastic and non-stationary volatility, the standard unit root tests are highly distorted. This immediately implies that standard predictability tests, as well, will be heavily distorted by stochastic volatility in returns.

We provide a new technique for testing predictability that is uniquely suited to address all of these characteristics. We hope that this approach will be appealing not only for its effectiveness, but also for its simplicity. We combine a simple instrumental variable estimator with a simple time change. The instrument directly addresses the inherent endogeneity emerging from the correlation between regressor innovations and returns and persistence. We use the Cauchy estimator, an instrumental variable which requires no additional data. The time change corrects any "poor behavior" on the part of the return distributions. Essentially, since stock returns have a volatility that varies over different time periods, we wait for volatility to reach a certain amount and then add that observation to our sample. So, across all observations there is a constant level of volatility. After this procedure, the regression errors and returns are guaranteed to be independent and normally distributed with a variance of our choosing. This point is worth reemphasizing: regardless of the form of the volatility structure of returns, they will always have a normal distribution. Thus, we have ensured the validity of the standard hypothesis tests.

The remainder of the paper is organized as follows. Section 2 provides the background and summarizes the main issues relating to return predictability. Section 3 introduces a novel approach to effectively deal with the various problems affecting the conventional approach. In particular, we introduce a time change to volatility time in order to correct for time-varying stochastic volatilities nonparametrically. In Section 4, we subsequently present a new test based on the Cauchy t-ratio for the samples collected after the time change. We show that the time-changed Cauchy t-ratio has standard normal limit distribution under truly mild regularity conditions. We show that the use of Cauchy t-ratio allows the predictors to have various statistical anomalies such as near nonstationarity, structural breaks and jumps, among many others. Section 5 provides Monte Carlo evidence that demonstrates the effects of stochastic volatility on traditional predictability tests. We find that in the presence of these problems, the use of OLS and standard hypothesis testing is wholly inappropriate. These results demonstrate that, using our combination of a time change and the Cauchy estimator, the data characteristics described above are no longer a factor. Furthermore, if the researcher desires increased power, existing tests which are more powerful and were invalid in the presence of stochastic volatility may be applied to the time-changed data. Next, in Section 6 we apply the technique directly to actual stock return data. We examine the data set recently used in the paper by Campbell and Yogo (2006). The empirical results are clear: we find no evidence for predictability in stock returns. Section 7 concludes the paper, and all the proofs are in the Mathematical Appendix.

Throughout the paper, we use $=_d$ and $\to_d$ to denote equality and convergence in distribution, respectively. To denote a stochastic process Z with time index $s \geq 0$, we use Z_s and $Z(s)$ interchangeably, based on saving space and accentuating the argument.

2 Background and Main Issues

In the most basic general equilibrium models, excess stock returns can not be predictable. The null hypothesis of the unpredictability of stock returns (y_i) has been routinely tested with simple regression

$$y_i = \alpha + \beta x_{i-1} + u_i,$$

for $i = 1, \ldots, N$, where (x_i) is some covariate which is believed to have some predictive power on the future values of (y_i). If stock returns are not predictable, clearly we will have that $\alpha = \beta = 0$. The most commonly used covariates are those which make the most economic sense, such as the dividend-price ratio and the earnings-price ratio.

To simplify the subsequent discussions, we let (x_i) be univariate as in most applications, unless specified otherwise. Moreover, we momentarily assume in this section that the constant term α is zero or its nonzero value is already incorporated into the definition of (y_i), so that we may concentrate on the slope coefficient β in regression (2)[1]. It will be explained later which part of the discussions in this section should be modified when there is a constant term in the regression. In most of the related literature, β

[1]The presence of the constant term α in regression (2) may be very important from the statistical point of view, as illustrated by Chen and Deo (2008).

is estimated by its ordinary least-squares estimate $\hat{\beta}_N$ and tested using the associated t-ratio which we will denote by $\tau(\hat{\beta}_N)$. Under the standard assumptions, we have

$$\tau(\hat{\beta}_N) \to_d \mathbb{N}(0,1)$$

as $N \to \infty$. It is, however, well documented that certain data characteristics may cause the distribution of the standard t-statistic to be far from standard normal, yielding a substantial bias to a test result relying on standard normal critical values. This will be explained in detail below.

For a sequence (ξ_i) of random vectors, let Z_N be a normalized partial sum process defined for $r \in [0,1]$ as $Z_N(r) = N^{-1/2} \sum_{i=1}^{[Nr]} \xi_i$, where $[z]$ is the integral part of any real number z. Throughout this section, we say that the invariance principle holds for (ξ_i) if the normalized partial sum process Z_N converges in distribution to a vector Brownian motion on $[0,1]$ as $N \to \infty$. The covariance matrix of the limit vector Brownian motion will simply be called the long-run variance of (ξ_i) following the usual convention.

2.1 Persistence and Endogeneity of Predictors

The covariates, (x_i), commonly used in the predictive regressions all show strong persistency. This is easily observable and has been well noted by many authors. For instance, see Goyal and Welch (2003) or Torous, Valkanov, and Yan (2004). Indeed, it is routinely modeled as an autoregressive process with the autoregressive parameter that is close to one, or more rigorously as a local-to-unity process. To fix the idea, we let

$$x_i = (1 - c/N)x_{i-1} + v_i \tag{1}$$

for some $c \geq 0$. Moreover, we let $\xi_i = (u_i, v_i)'$ and assume that the invariance principle holds for (ξ_i) with the bivariate limit Brownian motion $B = (U, V)'$, whose covariance matrix is given by

$$\Omega = \begin{pmatrix} \omega_u^2 & \omega_{uv} \\ \omega_{uv} & \omega_v^2 \end{pmatrix}.$$

Note that (u_i) is a sequence of martingale differences, and consequently, we may expect $\sum_{i=1}^N u_i^2/N \to_p \omega_u^2$ to hold under mild conditions.

The asymptotic null distribution of the usual OLS t-ratio $\tau(\hat{\beta}_N)$ can now be easily deduced and is given by

$$\tau(\hat{\beta}_N) \to_d \frac{1}{\omega_u} \left(\int_0^1 V_c(r)^2 dr \right)^{-1/2} \int_0^1 V_c(r) dU(r), \tag{2}$$

where V_c is an Ornstein-Uhlenbeck process defined as a solution to the stochastic differential equation $dV_c(r) = -cV_c(r)dr + dV(r)$ driven by the limit Brownian motion V. To further analyze the limit distribution in (2), we introduce

$$W = U - \frac{\omega_{uv}}{\omega_v^2} V,$$

a Brownian motion independent of V. Note that we have $U = (\omega_{uv}/\omega_v^2)V + W$ by construction.

The limit distribution in (2) may then be written as the sum of

$$P = \frac{\omega_{uv}}{\omega_u \omega_v^2} \left(\int_0^1 V_c(r)^2 dr \right)^{-1/2} \int_0^1 V_c(r) dV(r)$$
$$Q = \frac{1}{\omega_u} \left(\int_0^1 V_c(r)^2 dr \right)^{-1/2} \int_0^1 V_c(r) dW(r).$$

It is well known that the distribution of Q is normal, due to the independence of V and W. On the other hand, the distribution of P is essentially that of the t-ratio of the AR coefficient in the near-unit root model, obtained previously by Phillips (1987). The actual null distribution of $\tau(\hat{\beta}_N)$ is a mixture of normal and near-unit root distributions, with the mixing weight given by the long-run correlation coefficient $\rho_{uv} = \omega_{uv}/\omega_u\omega_v$ of (u_i) and (v_i). We may indeed easily deduce $\tau(\hat{\beta}_N) \to_d \mathbb{N}(0,1)$ if $\rho_{uv} = 0$. As $\rho_{uv}^2 \to 1$, the asymptotic null distribution of $\tau(\hat{\beta}_N)$ diverges from standard normal. See, e.g., Elliott and Stock (1994) for more discussions on the asymptotic null distribution of $\tau(\hat{\beta}_N)$. It is therefore clear that only when $\rho_{uv} = 0$ will the test based on the standard normal distribution be valid. Moreover, we may well expect that the size of the test becomes more distorted as $\rho_{uv}^2 \to 1$. Perhaps more importantly, it should be noticed that the distribution of the test statistic depends critically on the parameter c. This parameter cannot be estimated consistently since as the sample size N increases, c/N gets smaller at the same rate.

This problem has been clearly demonstrated recently by Campbell and Yogo (2006). In particular, they note that the size distortion is most severe when $c \approx 0$ and $\rho_{uv}^2 \approx 1$. For instance, if there is an exact unit root in the covariate and perfect long-run correlation between the innovations of the covariate and regression errors, they find that the asymptotic size of the one-sided t-test at 5% significance is as large as 46%. The reality of the data does not seem to be far from this worst case scenario. Upon examination, one can easily see that the predictors are highly persistent. More formally, Campbell and Yogo (2006) report that for the commonly used predictors such as dividend-price ratio and earnings-price ratio, unity lies outside the 95% confidence interval for only ten out of twenty eight data combinations. Even when a unit root may be rejected, the predictors are highly persistent. Of the ten for which unity lies outside the 95% confidence bounds, seven include an autoregressive parameter above 0.95. Moreover, innovations of the predictors seem to be highly correlated with stock returns in the long-run. For instance, the empirical long-run correlation between differences in the dividend-price ratio and stock returns is -0.98.

2.2 Nonstationary Stochastic Volatility in Returns

Stock returns, (y_i), which would be identical to the regression errors (u_i) under the null of no predictability, are widely believed to have time-varying stochastic volatility. In this subsection, we introduce various nonstationary stochastic volatility models considered in the literature. Following the usual specification for volatility model, we let

$$u_i = \sigma_{i-1} \varepsilon_i, \tag{3}$$

where (ε_i) is a martingale difference sequence with respect to filtration $(\mathcal{F}_i)$ such that $\mathbb{E}(\varepsilon_i^2 | \mathcal{F}_{i-1}) = 1$ for all $i \geq 1$. Under this specification, we have $\mathbb{E}(u_i^2 | \mathcal{F}_{i-1}) = \sigma_{i-1}^2$, and therefore, σ_{i-1}^2 becomes the conditional variance of u_i given information at time

$i-1$, $i \geq 1$. The volatility process (σ_i) is known to be very persistent and at least nearly nonstationary. Indeed, many authors have found that the AR parameter for the volatility process is close to unity under some appropriate functional transformations. See, for instance, Jacquier, Polson, and Rossi (2004), who provide convincing evidence that the log of volatility process follows a near-unit root process for a very wide range of equity and exchange rate time series. We may then conclude that the true volatility process is highly nonstationary, since it will be the exponential of a near-unit root process.

To accommodate this nonstationarity in volatility, we let

$$\sigma_i = \varpi(z_i) \tag{4}$$

with some near-unit root process (z_i) and $\varpi : \mathbb{R} \to \mathbb{R}_+$. In what follows, we let

$$z_i = (1 - c/N)z_{i-1} + w_i \tag{5}$$

for some $c \geq 0$, [2] and assume that ϖ is asymptotically homogeneous in the sense of Park and Phillips (1999), i.e., $\varpi(\lambda x) \approx \pi(\lambda)\bar{\varpi}(x)$ in x uniformly over any compact subset of $\mathbb{R}$ for all large λ. We call π and $\bar{\varpi}$ respectively the asymptotic order and limit homogeneous function. Loosely, an asymptotic homogeneous function is a function that behaves like a homogeneous function in the limit.

In place of (4), we may set

$$\sigma_i = \varpi\left(\frac{i}{N}\right) \tag{6}$$

with ϖ being a fixed function or a random function independent of other stochastic components of the model as in Cavaliere (2004) or Cavaliere and Taylor (2007), who studied the unit root test in the presence of stochastic volatility in the innovations. We may also consider the volatility model given by

$$\sigma_i = \varpi\left(\frac{z_i}{\sqrt{N}}\right). \tag{7}$$

The asymptotic distribution of $\tau(\hat{\beta}_N)$ under the specification of volatility in (6) and (7) is largely comparable and can be easily obtained from our result based on (4).

The theory of regression with errors (u_i) specified as in (3) and (4) has recently been developed by Chung and Park (2007). Redefine $\xi_i = (\varepsilon_i, v_i, w_i)'$, where (v_i), (ε_i) and (w_i) are introduced respectively in (1), (3) and (5), and assume that the invariance principle holds for (ξ_i) with the limit Brownian motion denoted by $B = (U, V, W)'$. Also, we define W_c to be the Ornstein-Uhlenbeck process with the mean reversion parameter $c \geq 0$. Then under the additional condition that $\sup_{i\geq 1} \mathbb{E}(|\varepsilon_i|^{2+\epsilon}|\mathcal{F}_{i-1}) < \infty$ a.s. for some $\epsilon > 0$, we have

$$\tau(\hat{\beta}_N) \to_d \frac{\int_0^1 V_c(r)\bar{\varpi}(W_c(r))dU(r)}{\left(\int_0^1 \bar{\varpi}(W_c(r))^2 dr\right)^{1/2}\left(\int_0^1 V_c(r)^2 dr\right)^{1/2}}.$$

Under the specification of volatility in (6) and (7), we have the same result only with the replacement of $\bar{\varpi}(W_c(r))$ by $\varpi(r)$ and by $\varpi(W_c(r))$, respectively.

[2] We may of course allow the local-to-unity parameter c of (z_i) to be different from that of (v_i) in (1). The same $c \geq 0$ is used simply to avoid introducing an additional parameter.

The limit null distribution of the standard t-ratio $\tau(\hat{\beta}_N)$ is non-normal, even when the innovations of the covariate (v_i) and those of the volatility factor (w_i) are completely independent of the innovations of errors (ε_i). How far it is away from the standard normal depends on many factors including the volatility function, asymptotic covariances of the innovations and local-to-unity parameters. Given the previous simulation studies by Chung and Park (2007) and Cavaliere and Taylor (2007), we may expect substantial size distortion from using the standard normal critical values in the predictive regression setting.

2.3 Other Issues

The limiting null distribution of the OLS t-ratio $\tau(\hat{\beta}_N)$ which we obtained in (2) is not robust with respect to a wide range of other statistical problems in stock return data. In particular, the distributions are dependent upon the presence of deterministic trends, thick tails in the innovations, jumps, and structural breaks, among other things, in the predictive ratios. The existence of any of these problems generally affects the limiting null distributions of $\tau(\hat{\beta}_N)$, and is likely to introduce further size distortion to the test based on standard normal critical values.

The presence of deterministic trends in some predictive ratios, especially in the 1990's, has been widely discussed. The possibility of their having structural breaks in the mean and volatility has also drawn some attentions in the literature. Lettau and Van Nieuwerburgh (2008) focus on structural breaks in the mean of predictive ratios. They show that a small break in the mean can explain some of the common characteristics of predictive regressions. Moreover, we may also infer from Kim, Leybourne, and Newbold (2004), which studies the behavior of unit root tests, that a structural break in the variance of innovation can seriously distort the distribution of the OLS t-ratio in predictability tests. In this paper, we make no claims regarding the existence of time trends and structural breaks, or the lack there of. Instead, we simply point out that if there is a deterministic trend or a structural break and a researcher uses the critical values appropriate to the no-trend or no-break case (or vice versa), inference is rendered invalid.

Thick tails in the innovations of predictive ratios would also affect the distribution of the OLS t-ratio. In the context of the unit root test, Ahn, Fotopoulos, and He (2001) show that, for the value of stability index 1.5, the rejection probability of the 5% test is only 0.6% under the null hypothesis. The asymptotic distribution of $\tau(\hat{\beta}_N)$ for this case indeed follows straightforwardly from the invariance principle for stable innovations, for which the reader is referred to Borodin and Ibragimov (1995). Returns also have long been observed to have relatively thick tails. However, we do not use stable distributions to model thick tails in returns. The nonstationary stochastic volatility in returns also generates thick tails, as shown in Park (2002), and for the purposes of this paper it does not seem productive to introduce an additional source of thick tails in returns.

It is worth emphasizing that the size distortion problem of the usual OLS t-ratio we address here is not simply a finite sample phenomenon. Virtually all the previous studies on stock return predictability are based on relatively large samples, with the sample size large enough for the asymptotic theory to provide a good approximation. Further, many earlier studies try to bridge the remaining gap between the finite sample distribution and theoretical asymptotic distribution by introducing the local-to-unity formulation in specifying the data generating process for the covariate.

Our previous results for the OLS t-ratio do not change if we include the constant term as in regression (2). To get the exact limit null distributions in this case, we simply need to replace V_c by $\bar{V}_c$, where $\bar{V}_c(r) = V_c(r) - \int_0^1 V_c(s)ds$. In general, the limit null distribution of the OLS t-ratio is further away from the standard normal if the constant term is included. The test based on the standard normal critical values would yield more serious size distortions. See, e.g., Chen and Deo (2008) for some related discussions.

3 A Novel Approach

In this section, we develop a methodology to deal with these problems very effectively. It consists of two separate procedures: a time change and Cauchy estimation. When used together, these two procedures are incredibly robust and well suited to address the problems which we have developed thus far. The time change effectively and non-parametrically corrects for time-varying stochastic volatility in returns. The Cauchy estimator, with a relatively small loss of power, solves all of the potential statistical problems caused by the covariate nonstationarity with endogeneity, and many other problems that could possibly exist in the covariate such as the presence of deterministic trends, structural breaks, thick tails in the innovations, jumps and outliers. Regardless of the presence of any of these econometric problems, the Cauchy t-ratio of the time-changed data will have an asymptotically standard normal distribution.

3.1 A Change of Chronometer

It is an indubitable fact that volatility changes over time. Consider instead *changing time over volatility.* In the most simple terms, this is precisely how one can deal with time-varying and possibly stochastic volatility. In essence, during periods of high volatility, time is elongated so there is less volatility in any fixed interval. Likewise, when volatility is low, time is compressed to increase the volatility in a given interval. This is done until volatility remains constant throughout the span of the data. Of course, this explanation is vastly over-simplified, but it will be useful to keep this in mind when proceding through a formal and rigorous explanation.

The approach here relies heavily on the theory of continuous time stochastic processes. Therefore, throughout the paper let the log of stock price, (Y_t), be a stochastic process in continuous time, which is adapted to a filtration $(\mathcal{F}_t)$ representing the information accumulated up to, and available at, time t. We are interested in testing for the null hypothesis of no predictability of (Y_t) given by

$$\mathbb{E}\left(dY_t|\mathcal{F}_t\right) = 0,$$

which implies, in particular, that

$$\mathbb{E}\left(Y_{t+h} - Y_t|\mathcal{F}_t\right) = 0$$

for all t and $h > 0$.

Under the null hypothesis, (Y_t) becomes a martingale with respect to the filtration $(\mathcal{F}_t)$. That is, given all available information up to the current date, t, the best possible prediction in the mean squared sense for any future value of a stock is the current value of that stock. Note that $(\mathcal{F}_t)$ is generally larger than the natural filtration of (Y_t), i.e., $(\mathcal{F}_t)$ contains more information than is provided only by the realized values of (Y_s) up to time $t > 0$. In our framework, (y_i) in regression (2) may be obtained by

$$y_i = Y_{t_i} - Y_{t_{i-1}} \tag{8}$$

for some choice of discrete set of time indexes $0 = t_0 < \cdots < t_N = T$ over the time interval $[0, T]$. In the subsequent development of our theory, we assume that $x_i = X_{t_i}$ with $t_i \in [0, T]$ for some continuous time stochastic process (X_t), which is adapted to the filtration $(\mathcal{F}_t)$.

Now, we let $\langle Y \rangle$ be the quadratic variation of Y, which is defined as

$$\langle Y \rangle_t = \operatorname*{plim}_{\pi_t \to 0} \sum_{k=1}^{m} (Y_{s_k} - Y_{s_{k-1}})^2,$$

where π_t is the mesh of partition $0 = s_0 < \cdots < s_m = t$ of the interval $[0, t]$, i.e., $\pi_t = \max_{1 \leq k \leq m} |s_k - s_{k-1}|$. Our methodology heavily depends upon the following celebrated theorem by Dambis, Dubins and Schwarz.

Lemma 3.2 Let Y be a continuous martingale. Then there exists a standard Brownian motion W such that

$$Y_t = W_{\langle Y \rangle_t},$$

or equivalently,

$$Y_{T_t} = W_t,$$

where T is a time change defined by

$$T_t = \inf_{s \geq 0} \{\langle Y \rangle_s > t\}.$$

The Brownian motion W is called the DDS Brownian motion of Y.

See, e.g., Revus and Yor (2005) for the proof and more discussions about this result, which is often referred to as the DDS theorem. Note that if $\langle Y \rangle$ is strictly increasing as in most potential applications, T is nothing but the time inverse of $\langle Y \rangle$. See Figure 1 for a demonstration of the change to volatility time. The quadratic variation increases at a regular interval of our choosing, while the calender time between observations varies greatly. Notice that when the returns are more volatile, which implies that quadratic variation increases quickly, the random time observations are very close to one another. In less volatile periods, when quadratic variation is increasing more slowly, the random time observations are much further apart.

The DDS theorem implies that all continuous martingales are essentially a Brownian motion read according to a different clock. They are all merely a time deformation of a Brownian motion. They are different from a Brownian motion only in that a general continuous martingale's quadratic variation is not given by the chronological time clock. On the other hand, this implies that for any continuous martingale, if we use the time clock inversely proportional to its quadratic variation, the martingale reduces to the standard Brownian motion. We use this important fact to develop a methodology to deal with a general time-varying stochastic volatility in the regression errors (u_i) in regression (2).

Under the null hypothesis of no predictability, we have $u_i = y_i$, which is given by (8) in our continuous time setup. The regressions errors (u_i) therefore become a sequence of martingale differences, i.e., $\mathbb{E}(u_i | \mathcal{F}_{t_{i-1}}) = 0$ for $i = 1, \ldots, N$. However, this provides

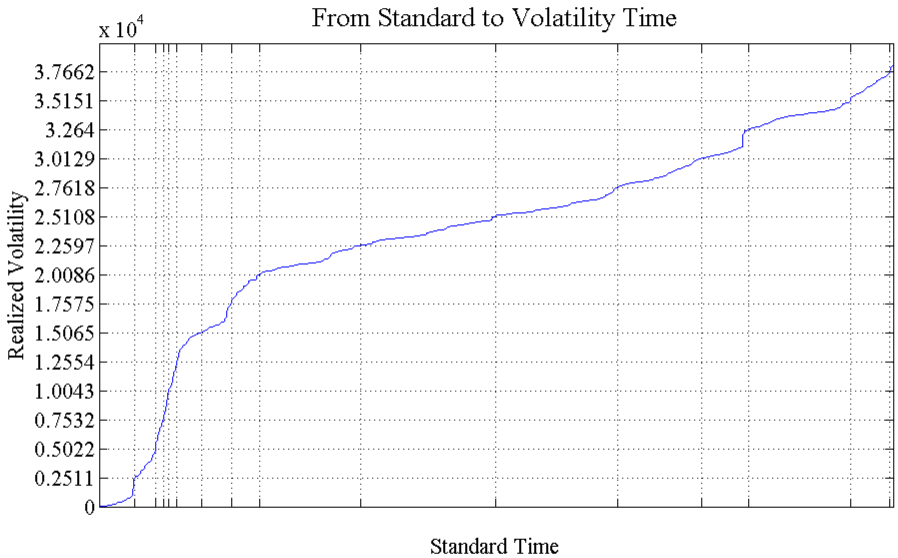

Figure 1: Calender versus volatility time. This figure demonstrates the conversion between calender time and volatility time. Instead of proceeding by one interval of time (i.e., one month), we proceed by one interval of volatility between observations. For illustrative purposes, not all of the random time observations are shown here.

no information on the conditional and unconditional higher moments. As we explained in the previous section, the presence of heterogeneous volatility makes the use of the standard testing procedure invalid if it is nonstationary.

We may use the DDS theorem to nonparametrically correct for a wide spectrum of nonstationary volatilities in stock returns. To see this, consider a sequence of stopping times

$$0 = T_0 \leq T_\Delta \leq \cdots \leq T_{N\Delta} = T$$

for fixed Δ, and let $N = \langle Y \rangle_T / \Delta$. Then we define

$$y_i^* = Y_{T_{i\Delta}} - Y_{T_{(i-1)\Delta}}$$

for $i = 1, \ldots, N$. Note that we have

$$Y_{T_{i\Delta}} - Y_{T_{(i-1)\Delta}} = W_{i\Delta} - W_{(i-1)\Delta} =_d \mathbb{N}(0, \Delta),$$

where W is the DDS Brownian motion of Y. Finally, we set $x_{i-1}^* = \sum_{s=i-1}^{i} X_{T_{s\Delta}}$ and consider the regression

$$y_i^* = \alpha c_i^* + \beta x_{i-1}^* + u_i^* \tag{9}$$

in place of regression (2), where $c_i^* = T_{i\Delta} - T_{(i-1)\Delta}$. The variable regressor (c_i^*) is necessary since we have a variable interval between observations rather than the constant interval of regression (2).

In sharp contrast to the original predictive regression (2) having errors potentially contaminated with various kinds of nonstationary volatilities, the newly proposed regression (9) has errors that are independent normals under the null hypothesis. This extremely useful property is easily attainable. To set up the new regression, instead of obtaining a sample at a fixed time interval, such as monthly, quarterly, or yearly, simply collect the data via a random sampling time based on volatility. The random sampling scheme used here is quite intuitive. We collect the data more often when the market is more volatile, and less often when the market is more stable. The resulting stock returns would then have a constant volatility. Due to the celebrated DDS theorem, if the log of stock price follows a continuous martingale as we assume under the null hypothesis, the stock returns collected in such a manner will always be independent normals.

Of course, the log of stock process Y is not continuously observable. Moreover, its quadratic variation $\langle Y \rangle$ has to be estimated, since it cannot be directly observed. To implement the idea of our time change in practical applications, we should therefore use observations of Y collected at sampling intervals substantially smaller than Δ. For the purposes of this paper, we use the daily stock price data to estimate the quadratic variation of $\langle Y \rangle$ and consider regression (9) with Δ, which is the average estimated increase in quadratic variation of $\langle Y \rangle$ over a month, quarter or year. This will be explained in more detail later.

The DDS theorem requires continuity and thus does not apply if the process contains jumps. Our previous discussions based on the DDS theorem are therefore no longer true if the log of the stock price process has jumps, which are likely especially when we use high frequency data. However, our methodology can be easily modified to accommodate for the presence of jumps. One of the easiest remedies is to test for the existence of jumps in the intervals $[T_{(i-1)\Delta}, T_{i\Delta}]$, $i = 1, \ldots, N$. If any of the intervals is suspected of containing jumps, then we may just delete the corresponding time changed observation from the regression. We may of course also try to find the exact timings of jumps and remove them before we apply our methodology. If we can identify the precise location and magnitude of the jump, we can simply remove the jump from the data. For the detection of jumps, the reader is referred to Aït-Sahalia (2004) and Barndorff-Nielsen and Shephard (2006). Clearly, as long as we have finite number of jumps and if the number of jumps is small relatively to N, the effect of discarding the stock returns over the intervals with jumps would be in no way detrimental. The same remedy may also be applicable for other types of discontinuities, such as structural breaks, in the sample path of the price process.

3.2 Efficient Tests

Once errors in regression (2) have been made into independent standard normally distributed observations, the researcher may apply some appropriate efficient tests. The tests we examine here are the Bonferroni based Q-test of Campbell and Yogo (2006) and the Restricted Maximum Likelihood (REML) based test of Chen and Deo (2008). These two tests have been shown to be highly efficient in the case in which returns and regressor innovations are normally distributed and independent. This convenient form will always be guaranteed by the use of volatility time.

Campbell and Yogo's (2006) test is efficient and asymptotically has the correct size. Since c cannot be consistently estimated, the test relies on placing confidence intervals on c. Then, using the endpoints of this set, confidence bounds may be placed on β. Suppose that a researcher were interested in a 5% test. The researcher could construct a 2.5% confidence interval on c. For each value in this interval, the researcher could then construct a 2.5% interval on β given the value of c. From the Bonferroni inequality, we know that such a confidence interval will have a coverage of at least 5%. The basic

Table 1: Estimates based on Campbell and Yogo's (2006) efficient test of predictability performed on the time-changed data, with 90% confidence intervals on the value of β.

Frequencies	Covariate	$\hat{\beta}_{CY}$	t-test	Q-test
Monthly	DP	-0.003	[-0.022, 0.013]	[-0.025 , 0.008]
	EP	0.005	[-0.012, 0.018]	[-0.014 , 0.014]
Quarterly	DP	-0.011	[-0.069, 0.033]	[-0.065 , 0.032]
	EP	0.011	[-0.041, 0.047]	[-0.036 , 0.048]
Annual	DP	-0.036	[-0.296, 0.140]	[-0.273 , 0.136]
	EP	0.061	[-0.169, 0.207]	[-0.145 , 0.211]

Table 2: Estimates for Chen and Deo's (2008) RLRT performed on the time-changed data.

Frequencies	Covariate	$\hat{\beta}_{REML}$	RLRT Statistic	p-value
Monthly	DP	-0.1114	0.0943	0.7588
	EP	0.1858	0.2623	0.6085
Quarterly	DP	-0.4852	0.1849	0.6672
	EP	0.4169	0.1365	0.7118
Annual	DP	-1.7616	0.1452	0.7032
	EP	2.1746	0.2275	0.6334

asymptotics developed by Campbell and Yogo (2006) are based on a simple standard normal error setup. Thus, the time-changed data should be ideal for optimal use of this estimator.

Chen and Deo (2008) show that a major source of distortion of test statistics in predictive regression is actually from the intercept parameter. Their REML likelihood ratio test (RLRT) achieves roughly half as much bias as OLS estimates for nearly unit root processes while suffering no loss in efficiency. For now, let η be the autoregressive parameter and let X_c and Y_c be the mean corrected independent and dependent variables, respectively. From Chen and Deo (2008), we have that the REML estimates are given by

$$\hat{\eta}_{\text{REML}} = \arg\min_{\eta} \left\{ n \log \left(\frac{1+\eta}{(n-1)(1-\eta)+2} \right) \right\},$$

$$(\hat{\rho}_{uv,\text{REML}}, \hat{\beta}_{\text{REML}}) = \begin{pmatrix} 1 & 0 \\ \hat{\eta} & 1 \end{pmatrix} (X_c' X_c)^{-1} X_c' Y_c).$$

Similarly to the OLS estimate, the REML estimate of β is critically dependent on the bias in η. Of course, the structure of the likelihood function that forms the basis of this estimation technique is also critically dependent on the distribution of errors. This makes a pre-test time change especially convenient for this estimator so that we will then know precisely the distribution of the structural error.

From this subsection, the benefits of the time change may easily be seen. Likewise, the results in Tables 1 and 2 show results which are much more in-line with what we economically would expect. The data used here is described more fully in the empirics in Section 6. At no frequency and for no covariate is the so called predictive ratio a significant predictor of future returns. For the Bonferroni type test of Campbell and Yogo (2006), the right most columns display the 90% confidence interval for each estimate for the t- and Q-tests. In all cases, the null hypothesis of $\beta = 0$ lies within the confidence intervals. For bias reduced likelihood based test of Chen and Deo (2008), the estimates are not significant at any of the conventional levels.

3.3 A Robust Test - The Cauchy Estimator

Now, using regression (9), the errors are guaranteed to be well behaved. As previously stated, it may then be acceptable to apply predictability tests based on Campbell and Yogo (2006) or Chen and Deo (2008). However, we also provide an effective way to deal with a wide variety of problems in the covariate of our predictive regression which may remain despite the time change, including persistence and endogeneity and the other data anomalies we discussed in the previous section. In fact, this section demonstrates how to properly address the issues that have been discussed in virtually every previous paper on predictive regressions.

To convey the main idea, we consider regression (2) with no constant term, i.e., $\alpha = 0$, and introduce the Cauchy estimator $\tilde{\beta}_N$ for β, which is given by

$$\tilde{\beta}_N = \left(\sum_{i=1}^{N} |x_{i-1}| \right)^{-1} \sum_{i=1}^{N} \mathrm{sgn}(x_{i-1}) y_i,$$

where $\mathrm{sgn}(\cdot)$ is the sign function defined as $\mathrm{sgn}(x) = 1$ if $x \geq 0$ and $\mathrm{sgn}(x) = -1$ if $x < 0$. Clearly, $\tilde{\beta}_N$ is nothing but the IV estimator using $\mathrm{sgn}(x_{i-1})$ as an instrument. This estimator was first proposed by Cauchy in 1836 and hence is referred to as the "Cauchy estimator". It has more recently brought some attention in the econometric literature. See So and Shin (1999) and Chang (2002) for the use of Cauchy estimator to test for a unit root. We will show that this estimator will be robust to a truly wide variety of data characteristics.

The associated Cauchy t-ratio $\tau(\tilde{\beta}_N)$ for β is given by

$$\tau(\tilde{\beta}_N) = \frac{\tilde{\beta}_N}{s(\tilde{\beta}_N)},$$

where $s(\tilde{\beta}_N)$ is the standard error of the Cauchy estimator $\tilde{\beta}_N$, which is given by

$$s(\tilde{\beta}_N) = \hat{\sigma}_N \sqrt{N} \left(\sum_{i=1}^{N} |x_{i-1}| \right)^{-1}$$

with any consistent estimator $\hat{\sigma}_N^2$ for the asymptotic variance σ^2 of the regression errors (u_i). [3]

The Cauchy estimator has many qualities well suited to the problems at hand. As has been well established by previous authors, the distortion in the distribution of the OLS t-statistic depends on values of the local-to-unity parameter c and the covariance ω_{uv}. The distortion is worst when $c \approx 0$ and $\omega_{uv} \approx 1$. Whereas the Cuachy t-statistic

[3] Under very mild conditions, the OLS residuals $(\hat{u}_i)$, $u_i = y_i - x_i\hat{\beta}_N$, as well as the Cauchy residuals $(\tilde{u}_i)$, $u_i = y_i - x_i\tilde{\beta}_N$ are consistent. This will be made clear in our subsequent development of our theory.

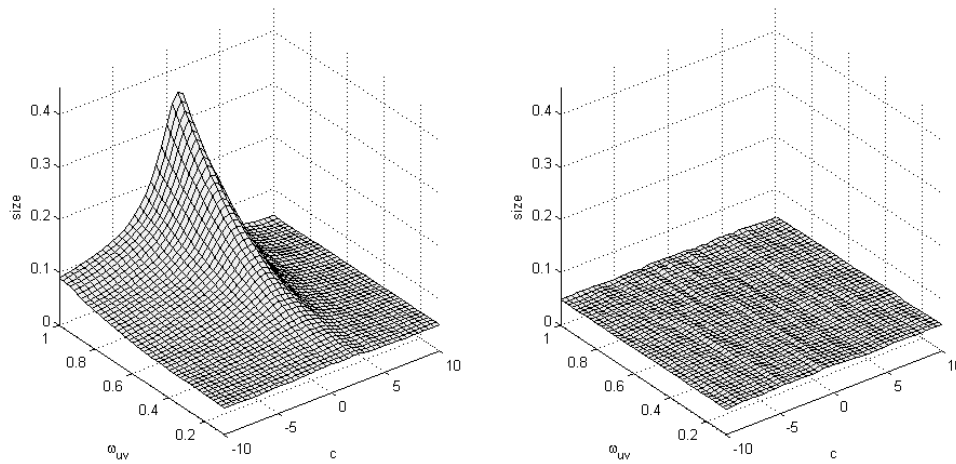

Figure 2: Size comparison. A comparison of size at a standard normal 5% critical value for t-tests performed on the OLS t-statistic (on the left) and for the Cauchy t-test (on the right). This figure is generated using 10,000 replications of a 500 observation sample.

will be normally distributed regardless of persistence in the regressor and correlation between innovations in the regressor and regressand. Figure 2 shows a comparison of t-tests based on the usual OLS t-statistic and the Cuachy t-statistic.

To develop the asymptotic theory for the Cauchy estimator $\tilde{\beta}_N$ and its t-ratio $\tau(\tilde{\beta}_N)$, we introduce

Assumption 3.1 Let $(u_i, \mathcal{F}_i)$ be a martingale difference sequence such that

(a) $\displaystyle\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(u_i^2|\mathcal{F}_{i-1}\right) \to_p \sigma^2$, and

(b) $\displaystyle\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(u_i^2 1\{|u_i| \geq \epsilon\sqrt{N}\}\Big|\mathcal{F}_{i-1}\right) \to_p 0$ for any $\epsilon > 0$,

as $N \to \infty$. The conditions in Assumption 3.1 are not stringent, and are required for the central limit theory to be applicable for (u_i). See, e.g., Hall and Heyde (1980). The condition in (a) is satisfied for a wide class of martingale sequences. It actually allows for stationary stochastic volatility to be present in (u_i). If we write $u_i = \sigma_{i-1}\varepsilon_i$ as before, where (σ_i^2) is ergodic and stationary with $\mathbb{E}\sigma_i^2 = \sigma^2$, then the condition obviously holds. The condition in (b) is the conditional version of the usual Lindeberg condition. It is met if we set $\sup_{i\geq 1}\mathbb{E}(|u_i|^{2+\epsilon}|\mathcal{F}_{i-1}) < \infty$ a.s. with some $\epsilon > 0$, as is often assumed in the literature. Consequently, Assumption 3.1 is expected to hold for a general class of martingale difference sequences. It should, however, be emphasized that it does not hold if the nonstationary volatility introduced in Section 2.2 is present, and our subsequent results in this section do not apply. This case will be fully addressed in the next subsection.

Assumption 3.2 There exists a sequence κ_N of numbers such that

$$\left(\kappa_N^{-1}\sum_{i=1}^N |x_i|\right)^{-1} = O_p(1)$$

for all large N.

The required condition in Assumption 3.2 is extremely mild, and should hold for a truly wide variety of the predictor (x_i). If (x_i) is a nonconstant, stationary and ergodic time series, then the condition is satisfied with $\kappa_N = N$. If, on the other hand, (x_i) has a near-to-unit root and if its innovations satisfy the invariance principle, then we have

$$N^{-3/2}\sum_{i=1}^N |x_{i-1}| \to_d \int_0^1 |V_c(r)|dr,$$

where V_c is the Ornstein-Uhlenbeck process introduced in the previous section. Therefore, the condition holds for $\kappa_N = N^{3/2}$. We may also allow for fat-tailed innovations of (x_i). In this case, the required condition is satisfied under general regularity conditions with $\kappa_N = N^{1+1/\alpha}\ell(N)$ for $0 < \alpha < 2$, where α is the stability index and ℓ is a function that is slowly varying at infinity. See, e.g., Borodin and Ibragimov (1995) for more details on the invariance principle for stable innovations. More importantly, the condition in Assumption 3.2 allows for various kinds of data anomalies in (x_i). The jumps and structural breaks in (x_i) are generally permitted, as long as their numbers are finite.

Now we may readily deduce the asymptotics of the Cauchy estimator $\tilde{\beta}_N$ and the Cauchy t-ratio $\tau(\tilde{\beta}_N)$.

Lemma 3.1

(a) If Assumptions 3.1 and 3.2 hold, then $\tilde{\beta}_N = \beta + O_p(N^{1/2}/\kappa_N)$ for all large N.

(b) If Assumption 3.1 holds and $\beta = 0$, then $\tau(\tilde{\beta}_N) \to_d \mathbb{N}(0,1)$ as $N \to \infty$.

The Cauchy estimator $\tilde{\beta}_N$ is therefore generally consistent. For each of the cases where (x_i) is stationary, nearly-nonstationary and nonstationary with fat-tailed innovations, its convergence rate is given by $N^{1/2}, N$ and $N^{(2+\alpha)/2\alpha}\ell(N)$. This gives us two highly useful properties of the Cauchy estimator $\tilde{\beta}_N$. First, its convergence rate is generally the same as the OLS estimator $\hat{\beta}_N$. Second, the Cauchy t-ratio has the standard normal limit distribution. Note that we only impose the conditions necessary for the central limit theory to hold for (u_i). It is important to notice in particular that we do not require any regularity conditions on (x_i). This is truly remarkable. In light of this fact, the Cauchy t-ratio is well suited to test for return predictability.

In the classical regression setting, the Cauchy estimator is less efficient than the OLS estimator. If (x_i) is stationary and ergodic, then the asymptotic variance of the former is $\sigma^2(\mathbb{E}|x_i|)^{-2}$, while that of the latter is $\sigma^2(\mathbb{E}x_i^2)^{-1}$, if we assume that they exist. We may therefore easily see that the OLS estimator has a smaller variance than the Cauchy estimator due to the Jensen's inequality. If, on the other hand, (x_i) is a near-unit root process having innovations (v_i) that are asymptotically independent of (u_i), then the limit distributions of the Cauchy and the OLS estimators are normal mixtures with mixing variates given respectively by $\sigma(\int_0^1 |V_c(r)|dr)^{-1}$ and $\sigma(\int_0^1 V_c(r)^2dr)^{-1/2}$. It can therefore easily deduced from the Cauchy-Schwarz inequality $\int_0^1 |V_c(r)|dr \leq (\int_0^1 V_c(r)^2dr)^{1/2}$

a.s. that the OLS estimator is more efficient than the Cauchy estimator. So, in these highly counterfactual circumstances, the OLS estimator is preferable.

However, in the context of predictive regressions, none of the above standard comparisons between the OLS and Cauchy estimators is applicable. In particular, the relative efficiency of the OLS estimator does not apply to the case where we have persistence and endogeneity considered in Section 2.1. The asymptotic distribution of the OLS estimator is generally biased and skewed, as well as non-normal, whereas that of the Cauchy estimator is normal even in this case. Therefore, the strict comparison based on their asymptotic variances cannot be made. Their relative efficiency will be dependent on the asymptotic correlation between the regression errors (u_i) and the innovations of predictive ratios (v_i), and also on the realization of the predictive ratios (x_i) themselves. Moreover, as discussed in Section 2.3, the OLS estimator is non-robust to other aberrant data characteristics such as deterministic trends, structural breaks and thick tails. In contrast, the Cauchy estimator is robust against these and many other potential problems in the data.

There are more compelling reasons why the Cauchy t-ratio $\tau(\tilde{\beta}_N)$ should be preferred to the OLS t-ratio $\tau(\hat{\beta}_N)$ in testing for return predictability. First, the asymptotic null distribution of $\tau(\hat{\beta}_N)$ is generally biased, skewed and has a tail thicker than that of normal distribution. This is in striking contrast to $\tau(\tilde{\beta}_N)$, whose asymptotic null distribution is exactly standard normal. Therefore, against the general alternative of predictability, the Cauchy t-ratio is expected to be more powerful than the OLS t-ratio. Second, more importantly, the asymptotic null distribution of the OLS t-ratio is dependent upon the local-to-unity parameter c of the predictive ratio, which is often assumed to be a near-unit root process. Strictly speaking, this nuisance parameter dependency makes the OLS t-ratio unusable, since it cannot be consistently estimated. Campbell and Yogo (2006) circumvent this problem by constructing a confidence interval for c and using a Bonferroni type inequality.

The robustness of the Cauchy estimator, as with all instrumental variable estimation, comes with a loss of some power. In Figure 3, we compare the powers of the Bonferroni-type Q-test of Campbell and Yogo (2006) and the RLRT of Chen and Deo (2008) with the Cauchy t-test for some selected values of c and ρ_{uv} when regressors and structural errors are well behaved, in that they follow the requirements of Campbell and Yogo (2006). The Q-test generally out performs the Cauchy t-test in terms of power. The RLRT outperforms the Cauchy t- and the Bonferroni Q-tests in most instances. The difference is most striking Though, the gap in the powers of the Cauchy t-test and the other tests diminishes as $\rho_{uv} \to 0$. This is well expected, since the asymptotic null distribution of the Bonferroni-type tests approaches the standard normal distribution as $\rho_{uv} \to 0$. In fact, the power curves of all the tests considered here become roughly comparable when $\rho_{uv} = 0$.

It is important to note that in this subsection, we maintain that the regression errors (u_i) are martingale differences whose volatilities are asymptotically constant. Under this circumstance, we clearly demonstrate that the simple Cauchy t-test alone may effectively deal with the persistence and endogeneity of predictive ratios, the issue specifically considered in the vast majority of previous literature in the predictability of stock returns. However, the Cauchy t-test, as well as all the other existing tests, is not applicable if the stock returns have nonstationary volatilities considered in Section 2.2. In this case, the results in Lemma 3.1 are not valid, and in particular, the limit distribution of the Cauchy t-ratio $\tau(\tilde{\beta}_N)$ is not normal. Therefore, we may no longer use the normal critical values. Perhaps the most valuable contribution of this paper is

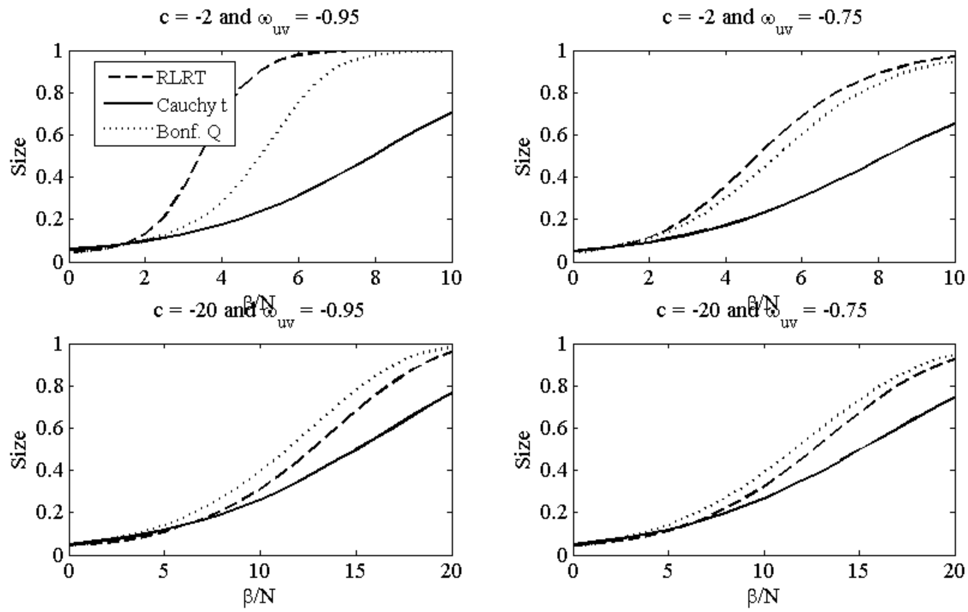

Figure 3: Power comparison. A comparison of local asymptotic power for the Cauchy t-test, the infeasible t-test, and the Bonferroni Q-test. Here, let $b = T\beta$ to account for the sample size. This figure is generated using 20,000 replications of a 500 observation sample.

that, given that the data has already been observed in volatility time, we may maintain these ordinarily strenuous restrictions without loss of generality.

4 A New Test and Its Statistical Theory

Our test is based on regression (9) consisting of the samples collected from the process Y using the chronometer running at the speed inversely proportional to the quadratic variation $\langle Y \rangle$ of Y. To implement our methodology for practical applications, we therefore need to observe Y and its quadratic variation $\langle Y \rangle$. However, the observations of Y are made only discretely in time. As a result, $\langle Y \rangle$ is not directly observable. We need to estimate the quadratic variation using a discrete set of observations. In this section, we assume that the samples $(Y_{i\delta})$, $i = 1, \ldots, n$ are observed over the time interval $[0, T]$ with $T = n\delta$. Note that the size n of the available sample is in general different from the number N of observations used to run regression (9). In our applications reported in the paper, we use n to denote the number of daily observations, while N refers to the sample size number of monthly, quarterly or yearly observations.

4.1 A New Test

To employ our method, we first need to estimate the quadratic variation $\langle Y \rangle$ of Y. The natural estimate for $\langle Y \rangle$ is

$$\langle Y \rangle_t^\delta = \sum_{i=1}^{m} \left(Y_{i\delta} - Y_{(i-1)\delta} \right)^2$$

for $(m-1)\delta \leq t < m\delta$, $m = 1, \ldots, n$, which is often referred to as the realized variance. Subsequently, we let

$$T_t^\delta = \inf_{s \geq 0} \left\{ \langle Y \rangle_s^\delta > t \right\},$$

i.e., the required time change defined from the realized variance process $\langle Y \rangle^\delta$. For a fixed $\Delta > 0$, we let

$$\begin{aligned} y_i^{*\delta} &= Y_{T_{i\Delta}^\delta} - Y_{T_{(i-1)\Delta}^\delta}, \\ c_i^{*\delta} &= T_{i\Delta}^\delta - T_{(i-1)\Delta}^\delta, \end{aligned}$$

and $x_{i-1}^{*\delta} = X_{T_{(i-1)\Delta}^\delta}$. Then we consider the regression

$$y_i^{*\delta} = \alpha c_i^{*\delta} + \beta x_{i-1}^{*\delta} + u_i^{*\delta}, \tag{10}$$

which corresponds to regression (9) in the previous section.

For the actual test of no predictability in stock returns, we use the excess returns to adjust for the nonzero constant term in regression. This is in accordance with virtually all the previous studies reported in the literature. All those studies found that the constant term, after the adjustment, is insignificant. Therefore, assuming $\alpha = 0$ in regression (10), we first consider the test of hypothesis $\beta = 0$. The test of whether the mean of the excessive returns is zero will be introduced later in this section. As we will explain in more detail later, our test results are consistent with earlier findings by others, and strongly support that the excess returns have a mean of zero.

To test for the hypothesis $\beta = 0$, we consider the regression

$$y_i^{*\delta} = \beta x_{i-1}^{*\delta} + u_i^{*\delta} \tag{11}$$

and use the Cauchy t-ratio $\tau(\tilde{\beta}_N^{*\delta})$ for β, which is given by

$$\tau(\tilde{\beta}_N^{*\delta}) = \frac{\tilde{\beta}_N^{*\delta}}{s(\tilde{\beta}_N^{*\delta})}.$$

As before, $\tilde{\beta}_N^{*\delta}$ is the Cauchy estimator for β, i.e.,

$$\tilde{\beta}_N^{*\delta} = \left(\sum_{i=1}^N |x_{i-1}^{*\delta}| \right)^{-1} \sum_{i=1}^N \mathrm{sgn}(x_{i-1}^{*\delta}) y_i^{*\delta},$$

and $s(\tilde{\beta}_N^{*\delta})$ is the standard error of the Cauchy estimator $\tilde{\beta}_N^{*\delta}$ that is given as

$$s(\tilde{\beta}_N^{*\delta}) = \hat{\sigma}_N^* \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N |x_{i-1}^{*\delta}| \right)^{-1}$$

with any consistent estimator $\hat{\sigma}_N^{*2}$ for the variance σ^{*2} of (u_i^*). Note that, by construction, we know that $\sigma^{*2} = \Delta$ if $u_i^* = y_i^*$ for all $i = 1, \ldots, N$. We will show that the Cauchy t-ratio $\tau(\tilde{\beta}_N^{*\delta})$ has the standard normal null limiting distribution under very general regularity conditions. The usual normal critical values can therefore be used for the test.

In implementing our test, we pay a particular attention to maximizing the finite sample power. As we mentioned earlier, much of the existing literature finds at least some and usually strong evidence of return predictability in stocks. This is in sharp contrast with our results, which unambiguously finds no support for predictability at all horizons. To strengthen our conclusion, it is therefore particularly important to increase the finite sample power of our test. First, if the mean of excess returns is zero, then the nonzero mean (in the stationary case) or large starting value (in the nonstationary case) of the predictor may decrease the finite sample power of our test. To get rid of the dependency of our test on the nonzero mean or the initial value of the predictor, we may use

$$x^{*\delta}_{i-1} - x_0 \qquad \text{or} \qquad x^{*\delta}_{i-1} - \frac{1}{i-1}\sum_{j=1}^{i-1} x^{*\delta}_{j-1}$$

or any other transforms relying only on the past values of the predictor, instead of $(x^{*\delta}_{i-1})$ in regression (11). Second, the consistent estimator $\hat{\sigma}^{*2}_N$ for the variance of (u^*_i) is obtained by the OLS regression (11) with $(y^{*\delta}_i)$ replaced by

$$y^{*\delta}_i - \tilde{\alpha}^{*\delta}_N c^{*\delta}_i$$

with $\tilde{\alpha}^{*\delta}_N$ introduced subsequently below, in addition to the transformation of the predictor mentioned above. We use the demeaned regressand since the nonzero mean in the excess returns, though it is not significantly different from zero, may inflate the estimate of the error variance and falsely lead to nonrejection of no predictability. The OLS method, rather than the IV method, is used to minimize the variance estimate. This also makes it easier to reject the null of no predictability.

To see whether the excess returns have mean zero, we test $\alpha = 0$ in the regression

$$y^{*\delta}_i = \alpha c^{*\delta}_i + u^{*\delta}_i, \tag{12}$$

where $c^{*\delta}_i = T^{\delta}_{i\Delta} - T^{\delta}_{(i-1)\Delta}$. For regression (12), we use the IV approach using the lagged regressor as an instrument, which yields the estimator

$$\tilde{\alpha}^{*\delta}_N = \left(\sum_{i=1}^N c^{*\delta}_i c^{*\delta}_{i-1}\right)^{-1} \sum_{i=1}^N c^{*\delta}_{i-1} y^{*\delta}_i.$$

Our test for the hypothesis $\alpha = 0$ is based on the associated t-ratio, denoted similarly as before by $\tau(\tilde{\alpha}^{*\delta}_N)$. To be consistent with our test for β, we use the OLS method to obtain the error variance estimate. It is quite clear that the asymptotic distribution of the IV t-ratio $\tau(\tilde{\alpha}^{*\delta}_N)$ also has the standard normal distribution under general regularity conditions. However, we will not formally develop the asymptotics for $\tau(\tilde{\alpha}^{*\delta}_N)$ here. They are relatively straightforward, given the asymptotics for $\tau(\tilde{\beta}^{*\delta}_N)$ that are presented in this paper.

4.2 Asymptotic Theory

Our asymptotics require $\delta \to 0$ and $n \to \infty$, as well as $N \to \infty$. Not only must the number of samples increase, but also the maximal frequency of available observations must increase as well. We develop the null asymptotics in this section, and therefore, the null hypothesis of no predictability will be maintained throughout the section. Furthermore, we assume that the jumps and other discontinuities in the path of Y have already been addressed, as explained earlier in Section 3.2. Under the convention, Y becomes a continuous martingale.

Intuitively, it is clear that realized variance $\langle Y \rangle^\delta$ gets close to the actual quadratic variation $\langle Y \rangle$, and likewise, a time change based on realized variance is close to the ideal time change, as $\delta \to 0$. Therefore, if δ is sufficiently small, then we can use random time observations based on realized variance as a valid proxy for random time observations based on quadratic variation. They will be close enough, in fact, that the same asymptotic theory will hold, under very general regularity conditions. To proceed formally, we first introduce some technical conditions. For brevity, we denote $\langle Y \rangle_s^t = \langle Y \rangle_t - \langle Y \rangle_s$ in what follows.

Assumption 4.1 For any $0 < s < t$, there exists some $\kappa > 0$ such that

$$\mathbb{E}\left[(Y_t - Y_s)^2 - \langle Y \rangle_s^t\right]^2 \leq c\,|t-s|^{1+\kappa}$$

where $c > 0$ is some constant independent of $s > 0$ and, $t > 0$.

A large class of continuous martingales satisfy this assumption. The reader is referred to Park (2008) for a more thorough discussion on this assumption. Roughly, this assumption bounds the variance of the error when estimating quadratic variation with realized variance.

Assumption 4.2 Let $\delta \to 0$ and $n \to \infty$ such that $n\delta^{1+\kappa} \to 0$ for the κ introduced above, and such that

$$N \leq \frac{c\Delta}{(n\delta^{1+\kappa})^{1/2}\left[\log(1/n\delta^{1+\kappa}) + \log(N\Delta)\right]^2},$$

for some constant $c > 0$.

Essentially, the conditions in Assumption 4.2 require that the frequency of the data increases quickly and that the number of observations increases relatively slowly. They also provide an upper bound for the number of time-changed observations.

It is shown in Park (2008) that

Lemma 4.1 Under Assumptions 4.1 and 4.2, we have

$$\sup_{t \leq n\delta} \left|\langle Y \rangle_t^\delta - \langle Y \rangle_t\right| = O_p((n\delta^{1+\kappa})^{1/2}),$$

and

$$\max_{1 \leq i \leq N} \left|y_i^{*\delta} - y_i^*\right| = o_p(N^{-1/2})$$

for all large N.

The conditions in Assumptions 4.1 and 4.2 are therefore sufficient to ensure that realized variance uniformly and consistently estimates quadratic variation. As a result, the samples collected using a time change based on realized variance get close to those by the ideal time change, and the errors incurred by using realized variance rather than quadratic variation become negligible asymptotically. Below we show that the t-ratios derived from the estimated time change are asymptotically distributionally equivalent to the case in which the true time change is known.

We may readily deduce from Lemma 4.2 that

Theorem 4.2 Under Assumptions 4.1 and 4.2, we have

$$\tau(\tilde{\beta}_N^{*\delta}) \to_d \mathbb{N}(0,1)$$

as $N \to \infty$.

In Theorem 4.2, we formally establish that the limit distribution of the Cauchy t-ratio is standard normal. Note that we require only very minimal conditions here. We allow the underlying process Y to be a very general martingale having a variety of possibly nonstationary stochastic volatilities. Furthermore, we impose no conditions on the predictor. The asymptotic normality of the Cauchy t-ratio holds regardless of any statistical anomalies in covariates including nonstationarity, fat-tailed innovations, structural breaks and jumps.

Now consider the case where $\beta \neq 0$. In this case, we assume

Assumption 4.3 Let

(a) $\displaystyle\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(u_i^{*\delta 2}) < \infty$, and

(b) $\displaystyle\sum_{i=1}^{N}|x_i^{*\delta}|$ is of order κ_N with $\kappa_N/\sqrt{N} \to \infty$.

Then we have

Proposition 4.3 Let $\beta \neq 0$. Under Assumption 4.3, $\tau(\tilde{\beta}_N^{*\delta})$ diverges at the rate of $\kappa_N/\sqrt{N}$ as $N \to \infty$.

The conditions in Assumption 4.3 are very mild. Therefore, Proposition 4.3 shows that the test based on the Cauchy t-ratio $\tau(\tilde{\beta}_N^{*\delta})$ is generally consistent, and have unit power against the alternative $\beta \neq 0$.

Though condition (a) is weak and satisfied widely, it is not followed from any of our previous assumptions. Note that $(u_i^{*\delta})$ may no longer be close to a sequence of independent standard normals, since the time change is obtained from the quadratic variation of Y and now $\beta \neq 0$. If the process X generating (x_i) is of bounded variation, the quadratic variation of Y is the same as that of process generating the regression error (u_i). Therefore, if we assume that the generating process of (u_i) is a continuous martingale, then $(u_i^{*\delta})$ reduces to an approximate iid sequence of normals. However, the generating process X of covariate (x_i) may not be of bounded variation. In this case, the quadratic variation of Y becomes in general different from that of the generating process of (u_i).

It seems obvious that condition (b) holds for all applications related to return predictability. The condition is easy to check, since $(x_i^{*\delta})$ is observed. As we mentioned earlier, for each of the cases where $(x_i^{*\delta})$ is stationary, nearly-nonstationary and nonstationary with fat-tailed innovations, its convergence rate is given by $N, N^{3/2}$ and $N^{(3+\alpha)/2\alpha}\ell(N)$. Therefore, it is easy to see that the condition is met in all these cases.

5 Monte Carlo Simulations

In this section, we begin to examine the effects that nonstationary volatility can have on existing tests as well as on our proposed test. We start with a general model encompassing previous econometric models of predictive regressions. We then examine six different nonstationary volatility structures. First, we introduce a simple structural break; that is, a single change in the volatility of errors and innovations. We also consider a regime switching model with high and low volatility regimes. Finally, we consider the more complex stochastic volatility models of exponential stochastic volatility and the Heston model.

Since our methodology and its theory are developed in a continuous time framework, our Monte Carlo simulations are conducted based on continuous time models. Therefore, we specify the return and predictor processes, Y and X. Further, we assume that these processes are observed at δ-intervals. Throughout our simulations, we set $\delta = 1/250$ so that $n = T/\delta$, implying that the daily observations are available for both Y and X. For the process Y, we consider

$$dY_t = \sigma_t dW_t, \tag{13}$$

where W is a standard Brownian motion. In the typical model of predictability, σ_t is a constant. For more complex and realistic models the exact structure of volatility will vary from model to model. Later in this section, we discuss the volatility processes σ_t that we analyze.

For X, we use

$$dX_t = \kappa(\mu - X_t)dt + \sigma_t dV_t \tag{14}$$

with V a Brownian motion, i.e., an Ornstein-Uhlenbeck process. The Brownian motions W and V have correlation ρ_{uv}. For the purposes of simulation, we let $\rho = -0.98$, or the approximate empirical correlation of returns with differenced predictive ratios. Here, we always assume that $\mu = 0$.

For a direct comparison between the more general format we use here, and the existing literature, we introduce a second distinct interval. While δ is a frequency at which the observations are available (here we use daily), let D be the frequency of observations at which the regression is run (generally monthly, quarterly, or yearly) and $N = T/D$.If X is observed at discrete D-intervals, then the corresponding AR regression of this process becomes

$$X_{iD} = (1 + e^{\kappa D})X_{(i-1)D} + \upsilon_{iD} \tag{15}$$

for $i = 1, \ldots, N$ where $\upsilon_t \sim \mathbb{N}(0, \sigma^2(1 - e^{2\kappa D} - 1)/2\kappa)$. Thus, this model nests the discrete time formulations of previous studies like Campbell and Yogo (2006) and Chen and Deo (2008), which instead use a constant value of σ_t. We set κ such that $e^{-\kappa D} \approx 1 - \kappa D = 1 - c/N$. Indeed, if κ is selected in this way and $\sigma_t = 1$ for all t, then (15) gives the exact model used by Campbell and Yogo (2006) and Chen and Deo (2008).

Turning now to the various volatility structures, we first consider the most simple case of nonstationary volatility: a single break in the standard deviation of errors and innovations. The return and predictor processes are defined as above, but the volatility process follows

$$\sigma_t = \begin{cases} \sigma_1 & \text{for } t \in [0, T/2] \\ \sigma_2 & \text{for } t \in (T/2, T]. \end{cases}$$

The existence of structural breaks in return volatility has been widely hypothesized. Furthermore, as has been shown previously in Kim, Leybourne, and Newbold (2004) as well as Cavaliere and Taylor (2007), unit root tests can be highly affected by persistent changes in volatility (i.e., nonstationary volatility), which directly implies that standard tests of predictability will also be affected. For this experiment, we consider a three sizes of breaks ($\sigma_1 = 1.5$, 3, or 4.5 and $\sigma_2 = 1$) in order to isolate the effects specific to the change in volatility. These standard deviation ratios are typical of the Kim, Leybourne, and Newbold (2004).

Next, we turn from deterministic changes in volatility to stochastic changes. We address three separate stochastic volatility models. The first stochastic volatility model examined is a simple regime shift in volatility with two possible regimes. Again, this process follows equations (13) and (14) with σ_t defined in the following way. Let S_t be the indicator function of the current state of the world. In the low volatility state, $S_t = 0$, while in the high volatility state, $S_t = 1$. As in Veronesi (1999), the process will or will not transition to another volatility state in any given time period with probabilities given by

$$\mathbb{P}\left[S_{t+h} = 0 | S_t = 0\right] = p_L h,$$
$$\mathbb{P}\left[S_{t+h} = 1 | S_t = 1\right] = p_H h,$$

and so

$$\sigma_t = (1 - S_t)\sigma_L + S_t \sigma_H.$$

Schaller and Van Norden (1997) find that for monthly stock return data the probability of remaining in a low volatility regime to be 0.9908 and a 0.9411 probability of staying in a high volatility regime, with low and high regime standard deviations $\sigma_L = 0.0392$, and $\sigma_H = 0.1180$, respectively. Correspondingly, we assign these values to standard deviation and the transition probabilities to $p_L = 0.9996$ and $p_H = 0.9976$ for our daily data.

Lastly, we consider two additional, more sophisticated, models of stochastic volatility. Specifically, we consider exponential stochastic volatility and the Heston model of stochastic volatility, so that

$$d\sigma_t^2 = d\exp(\theta Z_t), \tag{16}$$

$$d\sigma_t^2 = \lambda(\omega - \sigma_t^2)dt + \upsilon\sigma_t dZ_t, \tag{17}$$

where W and Z are Brownian motions. Models with stochastic volatility driven by the exponential of a nonstationary or nearly-nonstationary process, as in (16), have long been a feature of the finance literature (e.g., Hull and White (1987)). On the other hand, Heston (1993) proposed modeling volatility as a Ornstein-Uhlenbeck process, as in (17) and is currently widely used in the literature. For this final volatility process, the parameter values are set as follows. We let

$$\operatorname{cov}\begin{pmatrix} V \\ W \\ Z \end{pmatrix} = \begin{pmatrix} 1 & -0.98 & 0.539 \\ -0.98 & 1 & -0.55 \\ 0.539 & -0.55 & 1 \end{pmatrix}. \tag{18}$$

For σ_t^2, we set $\lambda = 1, \omega = 0.04$ and $\upsilon = 0.4$ in volatility model (a), and $\theta = 1/\sqrt{T}$ for volatility model (b). These parameter values are the approximate average parameter estimates given by Bakshi, Cao, and Chen (1997) which compares various methods of estimation.

Before preceding, since we must evaluate performance after a time change, the standard sizes of the Bonferroni Q-test and the RLRT warrant discussion. Viewing the data in volatility time gauruntees that returns will have a standard normal distribution, giving the ideal situation in which to apply these two tests. However, even under ideal circumstances; cases for which these tests were specifically developed, they will sometimes be mis-sized. The Bonferroni Q-test will have approximately double the nominal size for samples of 50 and will be too small when applied to samples of more than 500. This test will only be truly properly sized if a sample size happens to fall within this window. The RLRT will generally be oversized when the regressor is non-stationary, but quickly converges to the proper size as c gets further from zero. All of these statements simply summerize results reported in Campbell and Yogo (2006) and Chen and Deo (2008). So, while these tests are not always properly sized even after the time change, they perform as well as they possibly can given their limitiations. To avoid focussing on the less-than-perfect aspects of these tests, it is also worth restating that while the Cauchy t-test will generally have the correct size, it is clearly dominated, in terms of power, by both of these tests, both before and after the time change.

Table 3 and 4 show the effects of a single structural break in volatility. The tables give the size of the standard OLS t-test, Cauchy t-test, the Bonferroni Q-test, and the RLRT at a 5% nominal size when performed on standard calendar time data. Since, as demonstrated previously, the distribution of the existing test statistics depends critically on the structure of volatility, we would expect such a structural break to be distortionary. OLS, which is already inappropriate due to the persistence in the regressor and the correlation in innovations, becomes further distorted. In the most extreme case in which the predictor is nonstationary and there is a large break, most tests will reject a true null. For a relatively small break, the Bonferroni Q-test is over sized for small sample sizes and undersized for larger sample sizes. This result is similar to those in the simple, stationary variance case. However, the distortion increases with the size of the break. If $c = 0$ with 5 years of observations and there is a large change in volatility, the 5% test is 20%. The RLRT seems to suffers from distortions as well. It is more heavily impacted by a larger break in the standard deviations of volatility. It is interesting that the size distortion decreases and then increases as the process goes from nonstationary to more and more stationary. On the other hand, the Cauchy t-test performs quite well in all cases despite the break.

Though, existing tests performed poorly when there is a break in volatility, we find quite a different result when using our new technique. Table 5 gives the sizes of the same test performed on the time-changed data at the 5% nominal level. The OLS t-statistic remains distorted. This is due to the fact that, despite the time change, there remains a correlation in innovations and persistence in the regressor. The efficient Bonferroni Q-test and the RLRT perform much better. Verifying the theoretical result, the sizes are nearly identical to the sizes generated from data with normal errors with a constant variance. The RLRT, as in the standard case, remains slightly too large when the the regressor has a unit root. Likewise, the Bonferroni Q-test, again as in the standard case, begins too large and the gradually becomes slightly too small as the sample size increases. The size of the Cauchy t-statistics on the time changed data is very nearly ideal.

Next, we apply the same techniques to stochastic volatility models in calendar time. The results of these experiments is given by Table 6. Again, the Cauchy t-ratio is very close to optimally sized. Surprisingly, the size of the Bonferroni Q-test often *improves* versus the standard constant volatility models. This should not be understood as an improvement in the quality of the test. Indeed, this demonstrates the distortion to the distribution of the test statistic. For the regime switching and and exponential stochastic volatility models, the RLRT also suffers severe distortion. However, for the Heston model, the performance of both the Bonferroni Q-test and the RLRT is very similar to that in the standard case.

Finally, we apply a time change to the data tested in Table 6. These results are

Table 3: Sizes in calendar time. This table demonstrates the effects of a structural break in volatility on the OLS and Cauchy t-tests, the Bonferroni Q-test, and the RLRT. In all cases $\sigma_2 = 1$. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.

		$c = 0$			$c = -2$			$c = -20$		
		5	20	50	5	20	50	5	20	50
$\sigma_1 = 1.5$	OLS	30.27	29.47	29.23	19.49	18.23	19.17	7.54	7.55	7.48
	C-t	4.88	5.13	5.27	5.10	4.73	4.95	5.12	5.26	5.64
	B-Q	9.50	5.78	4.67	9.40	5.63	4.85	11.25	4.26	3.75
	REML	9.24	9.68	9.40	5.90	6.06	6.28	5.90	6.54	6.43
$\sigma_1 = 3$	OLS	30.75	29.53	29.63	21.63	20.98	20.57	8.32	7.96	8.28
	C-t	4.81	5.41	5.02	5.05	4.95	4.79	4.87	5.00	4.68
	B-Q	12.00	7.26	6.14	11.70	8.10	7.17	11.57	5.68	5.64
	REML	13.70	13.64	14.21	9.59	9.76	9.83	9.99	11.65	10.53
$\sigma_1 = 4.5$	OLS	29.88	29.14	29.05	23.45	20.00	21.32	8.82	8.48	8.58
	C-t	5.07	4.97	4.92	5.24	5.01	4.77	4.93	4.98	5.03
	B-Q	14.03	8.54	7.33	13.83	10.19	9.01	12.27	6.96	6.98
	REML	16.02	15.52	15.82	10.64	12.02	11.59	11.17	12.77	13.13

Table 4: Sizes in calendar time. This table demonstrates the effects of a structural break in volatility on the OLS and Cauchy t-tests, the Bonferroni Q-test, and the RLRT. In all cases $\sigma_1 = 1$. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.

		$c = 0$			$c = -2$			$c = -20$		
		5	20	50	5	20	50	5	20	50
$\sigma_2 = 1.5$	OLS	37.19	36.77	36.29	9.79	7.58	7.00	6.71	6.68	6.67
	C-t	4.96	4.86	4.75	5.03	4.93	5.47	5.42	5.33	4.89
	B-Q	9.40	5.94	4.65	4.19	5.89	5.11	5.43	5.19	4.25
	REML	9.01	9.37	8.75	4.69	4.79	4.80	6.11	6.31	6.34
$\sigma_2 = 3$	OLS	49.75	49.72	49.58	13.59	12.98	12.21	12.32	12.88	12.29
	C-t	4.77	4.76	5.14	5.04	5.00	4.74	5.05	4.97	5.14
	B-Q	12.73	9.12	7.75	12.42	9.95	8.27	19.13	9.51	8.01
	REML	14.07	13.00	13.18	5.77	6.39	6.32	10.43	10.37	10.04
$\sigma_2 = 4.5$	OLS	53.37	53.72	53.68	14.81	14.27	14.11	13.53	14.47	15.02
	C-t	4.42	4.97	5.23	4.92	5.09	5.12	4.95	4.85	5.09
	B-Q	15.03	12.38	10.89	14.42	12.75	11.19	20.96	12.21	10.33
	REML	15.88	15.94	15.37	6.82	6.93	6.45	12.12	11.92	11.97

Table 5: Sizes in volatility time. This table demonstrates the detrimental effects of a structural break in volatility on the OLS and Cauchy t-tests, the Bonferroni Q-test, and the RLRT when applied to the volatility time data. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.

		$c = 0$			$c = -2$			$c = -20$		
		5	20	50	5	20	50	5	20	50
$\sigma_1 = 1.5$	OLS	30.27	29.47	29.23	19.49	18.23	19.17	7.54	7.55	7.48
	C-t	4.88	5.13	5.27	5.10	4.73	4.95	5.12	5.26	5.64
	B-Q	9.45	5.52	4.86	8.83	5.39	4.69	11.61	4.37	3.42
	REML	8.38	8.43	8.67	5.14	4.99	5.02	5.20	4.61	4.78
$\sigma_1 = 3$	OLS	30.75	29.53	29.63	21.63	20.98	20.57	8.32	7.96	8.28
	C-t	4.81	5.41	5.02	5.05	4.95	4.79	4.87	5.00	4.68
	B-Q	9.03	5.65	4.85	9.07	5.73	4.90	12.64	4.03	3.76
	REML	8.70	8.28	8.89	6.18	6.59	6.53	4.91	4.80	4.34
$\sigma_1 = 4.5$	OLS	29.88	29.14	29.05	23.45	20.00	21.32	8.82	8.48	8.58
	C-t	5.07	4.97	4.92	5.24	5.01	4.77	4.93	4.98	5.03
	B-Q	8.86	5.80	4.52	9.04	5.73	4.43	12.36	4.39	3.56
	REML	8.93	7.99	8.24	5.66	7.42	8.06	5.02	4.31	4.84

Table 6: Sizes in calendar time. This table demonstrates the effects of nonstationary volatility on the OLS and Cauchy t-tests, the Bonferroni Q-test, and the RLRT. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.

		$c = 0$			$c = -2$			$c = -20$		
		5	20	50	5	20	50	5	20	50
Regime	OLS	28.84	28.96	29.03	30.43	28.78	28.45	25.95	23.89	24.46
	C-t	5.03	4.90	5.04	4.94	4.54	5.09	5.29	5.04	5.49
	B-Q	7.39	6.50	6.10	7.89	7.51	6.67	9.54	9.90	8.97
	REML	7.11	7.26	6.36	7.15	7.78	7.18	8.48	11.81	11.54
Exp. SV	OLS	12.43	12.32	12.32	12.67	11.88	11.92	11.35	10.85	10.62
	C-t	4.88	4.87	5.01	5.02	4.90	4.63	5.10	5.05	5.23
	B-Q	6.48	6.59	7.05	6.25	6.26	6.56	7.09	7.08	10.53
	REML	9.67	9.70	9.49	9.43	9.02	9.00	10.25	10.52	10.53
Heston	OLS	6.31	5.41	5.64	6.33	5.21	5.49	5.89	5.23	4.90
	C-t	5.21	4.97	5.27	5.07	4.81	5.16	5.06	5.30	4.85
	B-Q	7.38	6.49	5.91	7.29	5.86	5.66	8.11	5.73	5.90
	REML	5.75	4.89	5.40	5.56	4.90	5.26	5.27	5.11	4.81

Table 7: Sizes in volatility time. This table demonstrates the effects of nonstationary volatility on the OLS and Cauchy t-tests, the Bonferroni Q-test, and the RLRT when applied to the time changed data. The nominal level is 5% and all values are given in terms of percentages. The experiment is replicated 10,000 times.

		$c=0$			$c=-2$			$c=-20$		
		5	20	50	5	20	50	5	20	50
Regime	OLS	29.72	29.02	29.55	19.16	18.89	19.27	7.77	8.10	7.80
	C-t	5.06	5.04	5.25	5.38	5.27	4.66	5.37	5.04	4.02
	B-Q	9.87	5.39	4.79	9.20	5.72	4.47	11.81	4.33	3.84
	REML	8.47	8.04	8.75	5.47	5.42	5.82	5.38	5.49	5.56
Exp. SV	OLS	17.30	17.26	16.79	14.77	14.27	13.87	7.80	7.54	7.11
	C-t	4.83	4.66	4.70	4.67	4.70	4.76	4.98	5.07	4.78
	B-Q	9.49	5.64	4.78	9.40	5.40	4.69	12.11	4.34	3.81
	REML	7.43	7.50	7.49	7.09	7.28	6.76	6.22	6.19	6.10
Heston	OLS	10.47	9.62	9.59	8.92	8.22	7.60	6.12	5.83	5.60
	C-t	4.96	5.06	5.23	5.07	4.68	5.07	5.33	4.88	5.10
	B-Q	9.16	5.71	5.03	9.07	5.70	4.85	11.62	4.13	3.42
	REML	8.73	9.12	9.42	7.30	7.75	7.42	5.42	5.43	5.44

given in Table 7. The improvements to the calendar time tests is clear. While the Bonferroni Q-test and the RLRT are not always close to the nominal size of 5%, they are indistinguishable from their sizes in the standard case they were designed to address. That is, they perform as designed since the data now meets their basic assumptions. The Cauchy t-test keeps its outstanding performance in terms of size. Of course, as stated before, this comes at the cost of some power.

As can be clearly seen, when applied to volatility time data, existing tests perform as designed. Conversely, if the data is not well behaved, their performance quickly deteriorates. The existing more efficient tests do have a power advantage over the Cauchy t-test. Since volatility time returns have a normal distribution, we would expect that their power differentials should remain unchanged. Indeed, this is precisely what we find in that the powers are virtually identical to Figure 3.

6 Empirical Results

In our model, we let dY_t be the instantaneous excess return at time t. That is, the (risky) stock return less the (riskless) short-term interest rate. Moreover, we let X_t be the value of the earnings-price ratio, dividend-price ratio, or some other suspected predictive variable available at time t.

The data set used in our application consists of

- Stock returns from CRSP: NYSE/AMEX value-weighted index over the period 1926/12/01 - 2002/12/31, daily, monthly, quarterly and annual frequencies.
- Risk free rates from CRSP: One month and three month treasury bond rates.

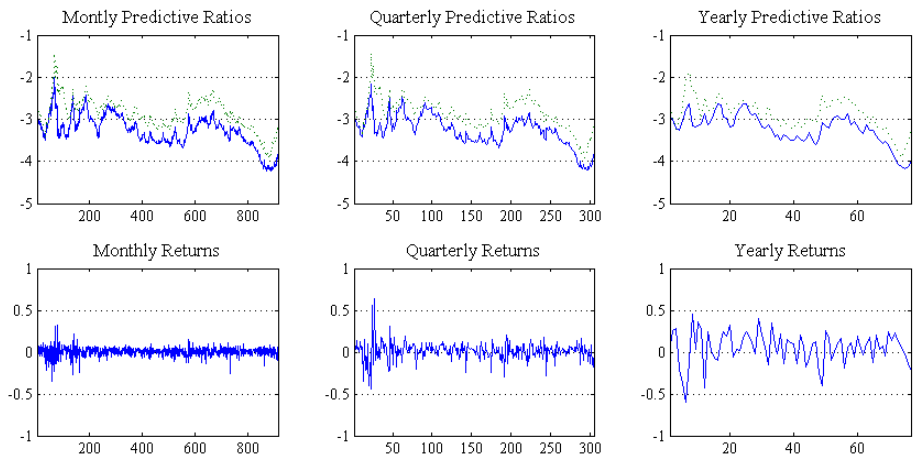

Figure 4: Data plot. This figure plots the data, with monthly, quarterly, and yearly data going from left to right. The regressors are displayed on the top row and the stock returns are displayed on the bottom. For the regressors, the solid blue line represents the dividend-price ratio, while the dotted green line represents the earnings price ratio.

- dividend-price ratios from CRSP: Dividends over the past year divided by the current price.
- earnings-price ratios from Global Financial Data : A moving average of earnings over the past ten years divided by the current price.

We investigate the return predictability at monthly, quarterly and annual frequencies. Daily returns adjusted by one month treasury bond rates are used to compute the realized variance of errors. Plots of the data are provided in Figure 4.

For a time change to be appropriate, we must have high frequency data. Daily price data is available, but CRSP dividend-price ratio data is only available at a monthly frequency. To construct a daily dividend-price ratio for day t, we divide the most recent monthly dividend data by the daily price at day t. To provide results corresponding to those reported by Campbell and Yogo (2006), we select Δ so that the number of post-time change observations n are equal to the number of observations in the appropriate Campbell and Yogo (2006) data set. That is, we compare standard time monthly data (for instance) with random time data with volatility equal to the average monthly volatility as measured by realized variance.

We make one practical modification to a time change based strictly on the DDS definition. As described earlier, the limiting variance of the random time data is exactly Δ. However, if we assign the stopping times based on $\langle Y \rangle_t^\delta > t$, then the actual variance will always be above Δ. This reality is addressed by instead selecting Δ such that we are as close to the limiting variance as possible. That is, we select the stopping times so as to minimize the distance between the realized variance and Δ. Of course, we add the restriction that the time may not repeat; it must go up by at least one period. This technique has been used and detailed previously in Jacewitz, Kim, and Park (2008).

Figure 5 presents the estimated error distributions for the samples collected at a fixed frequency and under the time change. Recall that we expect the time deformed return process to be normally distributed. The solid lines show the kernel density estimate

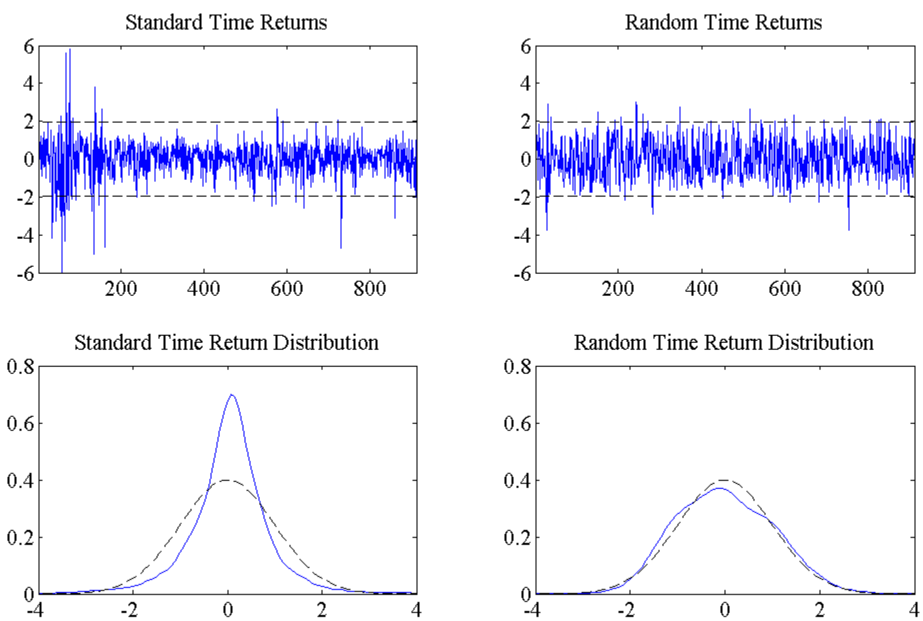

Figure 5: Data plot and kernel estimates of return distributions before and after the time change. The dashed line represents a standard normal distribution and the blue line gives the estimated distributions for innovations normalized by their respective standard deviations.

Table 8: Results of selecting parameter values which achieve the highest degree of normality. "CvM" refers to the Cramér-von Mises distance from a standard normal distribution.

		Standard Time			Random Time		
Frequencies	Covariate	μ	ρ	CvM	μ	ρ	CvM
Monthly	D/P	0.3265	1.1000	0.9155	-0.3191	1.0710	0.0316
	E/P	-0.6586	0.7711	0.4813	-0.3187	1.0707	0.0325
Quarterly	D/P	-0.1000	0.6949	0.1588	-0.2115	1.0807	0.0330
	E/P	0.2832	1.1000	0.9250	-0.2268	1.0862	0.0327
Annual	D/P	-0.0529	0.9889	0.0378	0.0814	0.9555	0.0311
	E/P	-0.5455	0.8166	0.0304	0.0822	0.9556	0.0306

of the distribution of returns, while the dotted lines in the graphs show the standard normal density. The kernel density estimates are retrieved from the `ksdensity` function in MATLAB. As can be clearly seen in, the error distributions change drastically after the time change. The error distribution from the time-changed regression becomes close to normal, while the distribution of errors from the conventional regression are far from being normally distributed. For the time-changed data, the Kolmogorov-Smirnov statistic is 0.0257 implying a probability of 0.5797, so normality cannot be rejected.

As an additional advantage to the time change, innovations in the regressor become much more normal as well as returns. Normality in innovations is another common and simplifying assumption made in econometric modeling. The normality of innovations when the time change is based on returns is due to a common volatility factor in returns and dividend- or earnings-price ratio. This is obvious since price appears in both the returns and the price ratios. Therefore, since dividends and earnings change relatively little, volatility in returns implies volatility in the predictive ratio. To illustrate this fact, we select parameters μ and ρ which makes $x_{t+1} - \rho x_t$ as close as possible, in the Cramér-von Mises sense, to $\mathbb{N}(\mu, \sigma)$, for any μ and σ. We require parameters to be "reasonable", that is if we impose that $\mu \in [-1, 1]$, and $\rho \in [0.5, 1.1]$. Figure 6 provides the graphical results for the dividend-price ratio. Numerical results are given in Table 8. These show that regardless of the parameter values, the most normal standard time innovations can possibly be is still less normal than the random time innovations. In most cases, the random time distributions are closer to normal than the best possible standard time distributions. The innovations of the earnings-price ratio is marginally closer to normal than for random time for the annual frequency.

Our main results are given in Table 9. For comparison, the results by Campbell and Yogo (2006) are replicated and reported here. Overall, our estimates of the slope parameter are roughly the same in terms of magnitude as those in previous papers, although ours are more often negative. The important difference is found in the t-statistics. We find that absolutely none of the estimated parameters are significant. Once the time change has been applied, the Cauchy estimates clearly do not support predictability. Moreover, Table 10 provides similar results for the subsample periods matching those in Campbell and Yogo (2006). Again, there is no evidence whatsoever for predictability in any subsample.

Although the Cauchy t-test is extremely robust and the time change allows for a wide variety of diffusion models, the asymptotic theory does require a continuous sample path

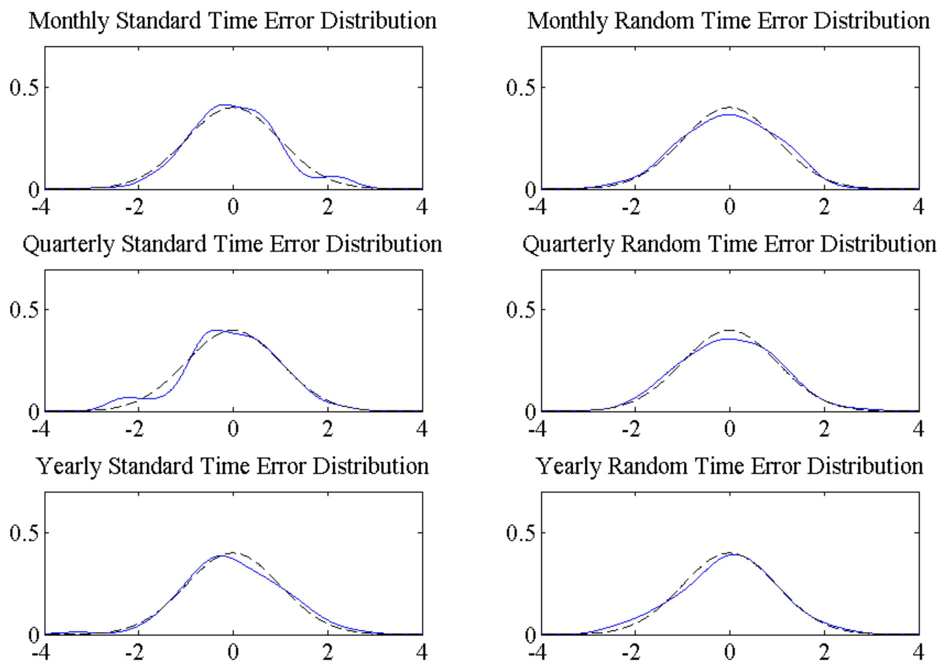

Figure 6: Kernel estimates of innovation distributions for the parameter values which achieve the most normal innovations in the regressor under standard time sampling and under random time sampling. These are distributions for the dividend-price ratio. For the earnings-price ratio, the results are entirely similar. The dashed line represents a standard normal distribution and the blue line gives the estimated distributions for innovations normalized by their respective standard deviations.

Table 9: Main empirical results for return predictability. D/P and E/P represents the dividend- and earnings-price ratios, respectively. TC indicates a time change was employed.

		Campbell and Yogo		TC Cauchy Regression	
Frequencies	Covariate	$\hat{\beta}_N$	$\tau(\hat{\beta}_N)$	$\tilde{\beta}_N^{*\delta}$	$\tau(\tilde{\beta}_N^{*\delta})$
Monthly	D/P	0.009	1.71	0.0209	0.0463
	E/P	0.014	2.66	0.5533	1.1820
Quarterly	D/P	0.034	2.06	-0.7326	-0.5195
	E/P	0.049	2.91	2.1820	1.4949
Annual	D/P	0.125	2.53	-3.3871	-0.5779
	E/P	0.169	2.77	2.3213	0.3871

Table 10: Subsample empirical results for return predictability.

Subsample Range		1926 - 1994		1952 - 2002	
Frequencies	Covariate	$\tilde{\beta}_N^{*\delta}$	$\tau(\tilde{\beta}_N^{*\delta})$	$\tilde{\beta}_N^{*\delta}$	$\tau(\tilde{\beta}_N^{*\delta})$
Monthly	D/P	-0.016721	-0.02899	-0.025263	-0.03695
	E/P	0.72858	1.2156	1.1314	1.5878
Quarterly	D/P	-0.81486	-0.44024	0.44003	0.19657
	E/P	2.908	1.5105	3.8907	1.6675
Annual	D/P	-3.0416	-0.37796	10.4761	0.97535
	E/P	4.532	0.53225	17.702	1.6269

in the stock price process. This explicitly precludes jumps and other discontinuities. We test for the presence of jumps using bipower variation as described in Barndorff-Nielsen and Shephard (2006) rejecting as discontinuous any interval which displays a test statistic that exceeds the 10% significance level. The offending interval is then eliminated from our analysis. The regression is then run on the remaining data. The results of these procedures are given in Table 11. The linear bipower variation test detected 37 and 6 jumps and the ratio test detected 6 and 1 jumps for each interval, respectively. At the 5% level, the linear bipower variation test detected 24 and 5 jumps and the ratio test detected 3 and 0 jumps for each interval, respectively. Finally, at the 1% level, the linear bipower variation test detected 17 and 2 jumps and the ratio test detected no jumps for either interval, respectively. There were no jumps detected at yearly intervals. Still, we find no significant changes in our empirical results. They remain robust to the presence of jumps in the price process.

7 Conclusion

Stock return predictability is one of the most prolific and important topics in financial economics and finance, and has been so for decades. Routinely, tests of many financial ratios, and most commonly the dividend-price and earnings-price ratios, are found to be highly useful predictors of future stock returns. However, it is possible that this is merely an econometric artifact. There are some widely recognized features of return and predictive ratio data that can seriously distort standard hypothesis testing. Two characteristics, persistence in the regressor and a correlation between regressand and innovations in the regressor, have been extensively addressed, though no solution has garnered broad adoption. A third widely accepted characteristic, time varying stochastic volatility, has thus far been ignored in the predictive regression literature. These three characteristics can explain the ubiquitous finding of stock return predictability. Our main contribution is to provide a specific technique that is uniquely suited to each of these issues. In the preceding sections, we have offered a new way to test for predictability using a time change to volatility time and the Cauchy estimator. It is robust, has the exactly correct size, and has power comparable to the most advanced existing techniques.

This new technique has the highly desirable characteristic that, regardless of any of the econometric complications commonly found in the related data, the t-ratio will always have a standard normal limiting distribution. The simple technique consists of

Table 11: Jump robustness check. This table provides estimates of the predictability parameter when intervals which have test statistics below the 10% critical value are removed.

Frequencies	Covariate	Linear $\tilde{\beta}_N^{*\delta}$	Linear $\tau(\tilde{\beta}_N^{*\delta})$	Ratio $\tilde{\beta}_N^{*\delta}$	Ratio $\tau(\tilde{\beta}_N^{*\delta})$
		10%			
Monthly	D/P	-0.0078939	-0.57068	-0.0083818	-0.64655
	E/P	-0.0061447	-0.48864	-0.0058486	-0.50149
Quarterly	D/P	-0.021698	-0.28862	-0.021947	-0.29636
	E/P	-0.019398	-0.2813	-0.01729	-0.25649
		5%			
Monthly	D/P	-0.0082209	-0.60318	-0.0082877	-0.6411
	E/P	-0.0061745	-0.50529	-0.0062952	-0.54155
Quarterly	D/P	-0.020763	-0.27715	-	-
	E/P	-0.018859	-0.27445	-	-
		1%			
Monthly	D/P	-0.0080578	-0.60092	-	-
	E/P	-0.0060236	-0.49953	-	-
Quarterly	D/P	-0.02218 1	-0.29786	-	-
	E/P	-0.017502	-0.25805	-	-

constructing a volatility time using realized volatility and then estimating the model using the Cauchy estimator. The Cauchy estimator itself has many useful characteristics provided that the error is independent and normally distributed. The random time sampling ensures that this will always be true.

We demonstrated that these widely hypothesized data characteristics can cause severe over rejection of a true null. With Monte Carlo simulations, we precisely quantified the distortionary effects of stochastic volatility on tests of predictaility. To empirically apply our technique, we used data covering the same time periods and the same regressors as Campbell and Yogo (2006), who do find evidence of predictability. However, we found no evidence, whatsoever, of predictability in stock returns using the most commonly used predictive ratios. Many may find this results unsurprising, from a theoretical standpoint. Empirically, it is striking. The vast majority of previous papers find strong evidence in favor of predictability, to the extent that it has become a stylized fact of stock returns. In our results, there is no support for predictability at any frequency for any of the predictors examined during any of the periods or subperiods. The results are strong and unambiguous. It seems clear that the return predictability disappears, if the characteristics of the data are properly addressed.

References

Ahn, S., S. Fotopoulos, and L. He (2001): "Unit Root Tests With Infinite Variance Errors," *Econometric Reviews*, 20(4), 461–483.

Aït-Sahalia, Y. (2004): "Disentangling Diffusion from Jumps," *Journal of Financial Economics*, 74(3), 487 – 528.

Bakshi, G., C. Cao, and Z. Chen (1997): "Empirical Performance of Alternative Option Pricing Models," *The Journal of Finance*, 52(5), 2003–2049.

Balvers, R. J., T. F. Cosimano, and B. McDonald (1990): "Predicting Stock Returns in an Efficient Market," *The Journal of Finance*, 45(4), 1109–1128.

Barndorff-Nielsen, O. E., and N. Shephard (2006): "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4(1), 1–30.

Borodin, A. N., and I. A. Ibragimov (1995): "Limit Theorems for Functionals of Random Walks," *Proceedings of the Steklov Institute of Mathematics*, 195(2).

Bossaerts, P., and P. Hillion (1999): "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *The Review of Financial Studies*, 12(2), 405–428.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *The Journal of Political Economy*, 107(2), 205–251.

Campbell, J. Y., and M. Yogo (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81(1), 27 – 60.

Cavaliere, G. (2004): "Unit Root Tests Under Time-Varying Variances," *Econometric Reviews*, 23(3), 259–292.

Cavaliere, G., and A. R. Taylor (2007): "Testing for Unit Roots in Time Series Models with Non-Stationary Volatility," *Journal of Econometrics*, 140(2), 919 – 947.

Chan, Y. L., and L. Kogan (2002): "Catching up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices," *The Journal of Political Economy*, 110(6), 1255–1285.

Chang, Y. (2002): "Nonlinear IV unit root tests in panels with cross-sectional dependency," *Journal of Econometrics*, 110(2), 261–292.

Chen, W. W., and R. S. Deo (2008): "Bias Reduction and Likelihood Based Almost-Exactly Sized Hypothesis Testing in Predictive Regressions Using the Restricted Likelihood," Working Paper.

Chung, H., and J. Y. Park (2007): "Nonstationary Nonlinear Heteroskedasticity in Regression," *Journal of Econometrics*, 137(1), 230–259.

Cochrane, J. H. (1999): "New Facts in Finance," *Economic Perspectives*, 23, 36–58.

Cochrane, J. H. (2005): *Asset Pricing, (rev. ed.)*. Princeton University Press.

Elliott, G., and J. H. Stock (1994): "Inference in Time Series Regression When the Order of Integration of a Regressor Is Unknown," *Econometric Theory*, 10(3/4), 672–700.

Ferson, W. E., S. Sarkissian, and T. T. Simin (2003): "Is Stock Return Predictability Spurious?," *Journal of Investment Management*, 1(3), 1–10.

Goyal, A., and P. S. Clara (2003): "Idiosyncratic Risk Matters!," *The Journal of Finance*, 58(3), 975–1007.

Goyal, A., and I. Welch (2003): "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49(5), 639–654.

Hall, P., and C. C. Heyde (1980): *Martingale Limit Theory and Its Application.* New York: Academic Press.

Heston, S. L. (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *The Review of Financial Studies*, 6(2), 327–343.

Hull, J., and A. White (1987): "The Pricing of Options on Assets with Stochastic Volatilities," *The Journal of Finance*, 42(2), 281–300.

Jacewitz, S., H. Kim, and J. Y. Park (2008): "Testing for No Arbitrage in Continuous Time: A Resolution to the Forward Premium Anomaly," Working Paper.

Jacquier, E., N. G. Polson, and P. E. Rossi (2004): "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122(1), 185–212.

Kim, T.-H., S. Leybourne, and P. Newbold (2004): "Behaviour of Dickey-Fuller Unit-Root Tests Under Trend Misspecification," *Journal of Time Series Analysis*, 25(5), 755–764.

Lettau, M., and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *The Journal of Finance*, 56(3), 815–849.

Lettau, M., and S. Van Nieuwerburgh (2008): "Reconciling the Return Predictability Evidence," *Rev. Financ. Stud.*, 21(4), 1607–1652.

Loretan, M., and P. C. B. Phillips (1994): "Testing the Covariance Stationarity of Heavy-tailed Time Series: An Overview of the Theory with Applications to Several Financial Datasets," *Journal of Empirical Finance*, 1(2), 211–248.

Menzly, L., T. Santos, and P. Veronesi (2004): "Understanding Predictability," *Journal of Political Economy*, 112(1), 1–47.

Park, J. Y. (2002): "Nonlinear Nonstationary Heteroskedasticity," *Journal of Econometrics*, 110, 383–415.

——— (2008): "Martingale Regression and Time Change," Working Paper.

Park, J. Y., and P. C. B. Phillips (1999): "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory*, 15(03), 269–298.

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74(3), 535–547.

REVUS, D., AND M. YOR (2005): *Continuous Martingales and Brownian Motion (3rd ed.)*. New York: Springer-Verlag.

SCHALLER, H., AND S. VAN NORDEN (1997): "Regime switching in stock market returns," *Applied Financial Economics*, 7(2), 177–191.

SO, B. S., AND D. W. SHIN (1999): "Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals," *Econometric Theory*, 15(02), 165–176.

STAMBAUGH, R. F. (1999): "Predictive Regressions," *Journal of Financial Economics*, 54(3), 375–421.

STÄRICĂ, C., AND C. GRANGER (2005): "Nonstationarities in Stock Returns," *Review of Economics and Statistics*, 87(3), 503–522.

TOROUS, W., R. VALKANOV, AND S. YAN (2004): "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," *The Journal of Business*, 77(4), 937–966.

VERONESI, P. (1999): "Sotck Market Overreactions to Bad News in Good Times: A Rational Expectations Equilibrium Model," *Review of Financial Studies*, 12(5), 975–1007.

WELCH, I., AND A. GOYAL (2008): "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Rev. Financ. Stud.*, 21(4), 1455–1508.

WILCOX, S. E. (2007): "The Adjusted Earnings Yield," *Financial Analysts Journal*, 63(5), 54–68.