A unifying view of some predictability tests

by

Stanislav Anatolyev*

New Economic School, Moscow

Abstract

We show that many existing tests for time-series predictability are special cases of a general nonparametric test based on the OLS estimator of the slope coefficient in a bivariate linear regression of certain type. By manipulating the features of this regression one can construct numerous new predictability tests. It turns out that some of the tests existing in the literature are asymptotically equivalent, and differ only by what kind of pivotization is applied to the core statistic. In addition, we show that the same tests may be constructed via reverse regressions. Among other things, we pay special attention to the issue of correct pivotization, discuss interpretation of regression-based tests, and argue against some widespread misconceptions.

Key Words and Phrases: Testing, time series, mean predictability, sign predictability, market timing.

JEL codes: C12, C22, C32, C53.

*Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Pr., 47, Moscow, 117418 Russia. E-mail: sanatoly@nes.ru.

1 Introduction

Applied researchers in macroeconomics and finance routinely utilize various tests for time-series predictability. Often these tests are used to evaluate the quality of a particular predictor, in other cases are they used to verify whether a particular series is mean or sign predictable, sometimes they are used to test for serial dependence. There are many of such tests that are dispersed throughout the literature: among them are market timing tests of Henriksson and Merton (1981), Cumby and Modest (1987), Breen, Glosten and Jagannathan (1989), and Bossaerts and Hillion (1999), the directional accuracy test of Pesaran and Timmermann (1992), the excess profitability test developed recently in Anatolyev and Gerko (2005), and a generalization of the Henriksson and Merton test proposed by Pesaran and Timmermann (1994). Some of these tests are based on testing for the nullity of slope coefficients in linear regressions, some are Hausman type tests based on contrasts, some are based on contingency tables. Some of the tests are designed for the null of conditional mean independence, others – for the null of conditional sign independence, although most are used to detect predictability in general rather than that of one or another specific type.

In this paper, we show that most of aforementioned tests are special cases of a nonparametric test based on the OLS estimator of the slope coefficient in a certain bivariate linear regression. The left side variable in this regression is determined by what feature of the given series is tested for conditional independence, while the right side variable can be chosen as an arbitrarily function of the predictor. This implies that many new predictability (or "market timing") tests can be designed (i) by manipulating this function whose choice has an impact only on the power of the test, and (ii) by manipulating the feature tested for conditional independence. In addition, it turns out that some of the tests existing in the literature are asymptotically equivalent and differ only by what kind of pivotization is used for the core statistic. In particular, the directional accuracy test of Pesaran and Timmermann (1992) is asymptotically equivalent to market timing tests of Henriksson and Merton (1981) and

Breen, Glosten and Jagannathan (1989), and the (suitably robustified) excess profitability test of Anatolyev and Gerko (2005) is asymptotically equivalent to the market timing test of Cumby and Modest (1987). Moreover, we show that the same tests may be constructed via reverse regressions, i.e. those where the left side and right side variables switch places, but the pivotization must be done very carefully. Theoretically such equivalences of various tests are very important, while in practice they may facilitate testing via use of standard regression packages.

We also provide an extension to multiple null hypotheses and, respectively, tests based on multiple regressions. In this context, we thoroughly analyze tests for independence in (larger than 2×2) contingency tables, both the classical χ^2-tests and the one considered in Pesaran and Timmermann (1994). It turns out that these tests can also be interpreted as tests for the nullity of coefficients in linear regressions, which may be straight or reverse, but now these are multiple and multivariate. As another example, we discuss joint testing for conditional mean and conditional variance independence of returns, also considered in Marquering and Verbeek (2004).

Throughout, we pay special attention to the important issue of constructing standard errors for the OLS estimate, whether it should or may not be a heteroskedasticity and autocorrelation consistent (HAC) one. As a by-product, we prove that the directional accuracy test (originally developed under the assumption of total independence between predictands and predictors) is robust to serial correlation and conditional heteroskedasticity, and derive a HAC version of the excess profitability test (also originally developed under the same independence assumption). Finally, among other things, we devote a separate section to a discussion of issues pertinent to the regression-based tests, such as their interpretation and nonparametric nature, and argue against some widespread misconceptions.

The paper is organized as follows. We start in Section 2 by considering existing tests based on contrasts, and provide a generalization to other null hypotheses. In Section 3 we develop regression-based tests, show that they are asymptotically equivalent to contrast-based tests of Section 2, and provide a further generalization. In Section 4 we discuss tests based on reverse regressions. Section 5 contains an extension to multiple null hypotheses and tests based on multiple regressions. In Section 6 we present a discussion of some conceptual testing issues. Finally, Section 7 concludes. All proofs and lengthy derivations are contained in appendices.

2 Contrast-based tests

Let y_t represent some stationary economic variable, such as GNP growth, change in some commodity's log price, log return from some financial market, etc. Let x_t be a continuously distributed forecast of y_t that depends only on the data from $\mathcal{I}_{t-1} = \{y_{t-1}, y_{t-2}, \cdots\}$, or, more generally, from the extended information set $\mathcal{I}_{t-1} \supset \{y_{t-1}, y_{t-2}, \cdots\}$ which may include other historical variables. Let us introduce the following notation for future use assuming that these moments exist:

$$
\begin{aligned}
m_x &= E\left[\operatorname{sign}(x_t)\right], \\
m_y &= E\left[\operatorname{sign}(y_t)\right], \\
M_y &= E\left[y_t\right], \\
V_y &= \operatorname{var}\left[y_t\right].
\end{aligned}
$$

We start by reviewing two market timing tests which are known in the literature and have a similar structure. We will discuss what null hypotheses these tests assume, and how their asymptotic distributions are derived. Then we will extend the structure of these two tests to a more general framework that nests both tests. In subsequent sections, we will generalize this class of tests further, and look at it at various other angles.

The directional accuracy (DA) test of Pesaran and Timmermann (1992) is routinely used as a predictive-failure test in constructing forecasting models; see, among others, Pesaran and Timmermann (1995), Gençay (1998), Qi (1999), Franses and van Dijk (2000), Granger and Pesaran (2000), Qi and Wu (2003). Formally, the DA test is a test for sign predictability, i.e. for the null

$$H_0^{DA} : E\left[\text{sign}(y_t)|\mathcal{I}_{t-1}\right] = \text{const.}$$

It is shown in Anatolyev and Gerko (2005) that the original DA test statistic (based on the contrast between the empirical frequency of correct directional predictions and an efficient estimate of the frequency under independence between y_t and x_t) is asymptotically equivalent to the following one where recentered and renormalized components are used[1]:

$$DA = \sqrt{T}\frac{A_{DA} - B_{DA}}{\sqrt{\hat{V}_{DA}}} \xrightarrow{d} \text{N}(0,1), \tag{2.1}$$

where

$$A_{DA} - B_{DA} = \frac{1}{T}\sum_t \text{sign}\left(x_t\right)\text{sign}\left(y_t\right) - \left(\frac{1}{T}\sum_t \text{sign}\left(x_t\right)\right)\left(\frac{1}{T}\sum_t \text{sign}\left(y_t\right)\right), \tag{2.2}$$

$$\hat{V}_{DA} = \left(1 - \hat{m}_x^2\right)\left(1 - \hat{m}_y^2\right),$$

and

$$\hat{m}_y = \frac{1}{T}\sum_t \text{sign}(y_t) \xrightarrow{p} m_y, \quad \hat{m}_x = \frac{1}{T}\sum_t \text{sign}(x_t) \xrightarrow{p} m_x.$$

Pesaran and Timmermann (1992) also indicate that the DA test is asymptotically equivalent to the well-known Henriksson and Merton (1981) test of market timing, which is an exact test for independence in a 2×2 contingency table (for more about such relations, see Sections 4 and 5). The latter popular test has been used in a deal of finance and macroeconomic papers on forecasts evaluation (e.g., Havenner and Modjtahedi, 1988, Lai, 1990), sometimes together with the DA test (e.g., Ash, Smyth and Heravi, 1998, Greer, 2003). Interestingly, Granger and Pesaran (2000) show that the DA test statistic is also a standardized version of "Kuipers score" used in the meteorological literature.

An important fact is that even though the DA test is a test for sign predictability, the asymptotic distribution of the DA test statistic is derived under a much stronger presumption

[1]This transformation entails a linear operation of switching from indicators to signs, and omitting a term of higher order of negligence in the variance formula.

that under H_0^{DA} there is independence of y_t and x_t at all lags and leads (see the manipulations below equation (4) in Pesaran and Timmermann, 1992, p. 462; these manipulations aim at computation of finite-sample variances of "predecessors" of A_{DA} and B_{DA}). This independence evidently does not hold whenever a predictor is related to the series being predicted. For example, a simple moving average forecast contains lags of y_t and hence is correlated with lags of the series itself. It will not be an exaggeration to say that this presumption is *never* satisfied in practice.

The similar in spirit excess profitability (EP) test for mean predictability, i.e. for the null

$$H_0^{EP} : E\left[y_t|\mathcal{I}_{t-1}\right] = \text{const},$$

is developed in Anatolyev and Gerko (2005), and is based on the contrast

$$A_{EP} - B_{EP} = \frac{1}{T}\sum_t \text{sign}(x_t)y_t - \left(\frac{1}{T}\sum_t \text{sign}(x_t)\right)\left(\frac{1}{T}\sum_t y_t\right). \tag{2.3}$$

The EP test statistic and its asymptotic distribution are

$$EP = \sqrt{T}\frac{A_{EP} - B_{EP}}{\sqrt{\hat{V}_{EP}}} \xrightarrow{d} \mathrm{N}(0,1), \tag{2.4}$$

where

$$\hat{V}_{EP} = \left(1 - \hat{m}_x^2\right)\hat{V}_y,$$

and

$$\hat{V}_y = \frac{1}{T}\sum_t (y_t - \bar{y})^2 \xrightarrow{p} V_y.$$

When y_t represents a logarithmic return on some financial asset or index, the EP statistic can be interpreted as a normalized return of the position implied by a simple trading strategy that issues a buy signal if a forecast of next period return is positive and a sell signal otherwise, over a certain benchmark (see Anatolyev and Gerko, 2005 for details).

Similarly to the DA test, the asymptotic distribution of the EP test in the original paper is derived under the same strong presumption that x_t is independent from y_t for all lags and leads, by computing the finite-sample variances of A_{EP} and B_{EP}. The consequences of this presumption can be clearly seen in the expressions for $\hat{V}_{DA}$ and $\hat{V}_{EP}$ the components of which do not take the form of autocorrelation and even heteroskedasticity consistent variance estimators.

To emphasize that the DA and EP tests test different nulls which is important for further discussion, let us look at two examples of data generating processes implying different types of nonpredictability. These examples are inspired by the analysis in Christoffersen and Diebold (2003).

Example 1 *In a pure ARCH model $y_t = \mu + \varepsilon_t$, where $\varepsilon_t|I_{t-1} \sim N(0, \sigma_t^2)$, conditional mean independence follows because $E[y_t|I_{t-1}] = \mu = \text{const}$. There is sign predictability because*

$$\begin{aligned} E[\text{sign}(y_t)|I_{t-1}] &= \Pr[y_t > 0|I_{t-1}] - \Pr[y_t < 0|I_{t-1}] \\ &= \Pr\left[\frac{\varepsilon_t}{\sigma_t} > -\frac{\mu}{\sigma_t}|I_{t-1}\right] - \Pr\left[\frac{\varepsilon_t}{\sigma_t} < -\frac{\mu}{\sigma_t}|I_{t-1}\right] \\ &= 2\Phi\left(\frac{\mu}{\sigma_t}\right) - 1 \neq \text{const}. \end{aligned}$$

Example 2 *In a simple variant of an ARCH-M model $y_t = \gamma\sigma_t + \varepsilon_t$, where $\gamma > 0$ and $\varepsilon_t|I_{t-1} \sim N(0, \sigma_t^2)$, there is mean predictability because $E[y_t|I_{t-1}] = \gamma\sigma_t \neq \text{const}$. Conditional sign independence follows because*

$$\begin{aligned} E[\text{sign}(y_t)|I_{t-1}] &= \Pr[y_t > 0|I_{t-1}] - \Pr[y_t < 0|I_{t-1}] \\ &= \Pr\left[\frac{\varepsilon_t}{\sigma_t} > -\gamma|I_{t-1}\right] - \Pr\left[\frac{\varepsilon_t}{\sigma_t} < -\gamma|I_{t-1}\right] \\ &= 2\Phi(\gamma) - 1 = \text{const}. \end{aligned}$$

In what follows, we will be concerned with conditional independence and predictability of a more general type. In the rest of this Section, we

a) consider a class of predictability tests that contains DA and EP as special cases, and

b) derive an autocorrelation and heteroskedasticity consistent version of this general test.

Let us consider testing the null hypothesis

$$H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const},$$

where $g(u)$ is a given stationary function. In contrast to the two preceding papers, we require that only the conditional independence of $g(y_t)$, together with certain regularity conditions to be specified below, can be used in deriving the distribution of statistics of interest under H_0^g. By analogy to the DA and EP tests, consider the contrast

$$A_g - B_g \equiv \frac{1}{T}\sum_t \text{sign}(x_t)g(y_t) - \left(\frac{1}{T}\sum_t \text{sign}(x_t)\right)\left(\frac{1}{T}\sum_t g(y_t)\right). \tag{2.5}$$

Setting $g(u) = \text{sign}(u)$ leads to testing for conditional sign predictability and the DA test statistic, while setting $g(u) = u$ leads to testing to conditional mean predictability and the EP test statistic. Let us additionally introduce the notation

$$\begin{aligned} M_g &= E\left[g(y_t)\right], \\ V_g &= \text{var}\left[g(y_t)\right], \end{aligned}$$

and impose

Assumption 1

(i) The series y_t and its forecast x_t are continuously distributed, strictly stationary, and strongly mixing with mixing coefficients $\alpha(j)$ satisfying $\sum_{j=1}^{\infty} \alpha(j)^{1-1/\nu} < \infty$ for some $\nu > 1$.
(ii) The forecast x_t is $\mathcal{I}_{t-1}$-measurable.
(iii) The function $g(u)$ is measurable, and $E\left[g(y_t)^{2\nu}\right]$ exists and is finite for ν from (i).

The asymptotic distribution of the contrast (2.5) is given in the following theorem.

Theorem 1 *Suppose the regularity conditions specified in Assumption 1 hold. Then, under $H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const}$, we have*

$$\sqrt{T}\left(A_g - B_g\right) \xrightarrow{d} \text{N}(0, V^g),$$

where

$$V^g = \left(1 - m_x^2\right) V_g - 2m_x \text{cov}\left[\text{sign}(x_t), g(y_t)^2\right].$$

Beside the DA and EP tests, Theorem 1 may be used to construct other tests for predictability of various moments. For example, a null of conditional second moment independence and a corresponding test for second moment predictability will result if one takes $g(u) = u^2$.

A remarkable feature of the result in Theorem 1 is that the expression for the asymptotic variances does *not* contain long-run variances and/or long-run covariances despite their presence in the asymptotic variances of A_g and B_g separately. Why this is so will be transparent in Section 3. The asymptotic variance V^g does, however, include a term associated with a contemporaneous correlation of predictor's sign and (the square of) the predicted feature.

Specialization of Theorem 1 to the two special cases of interest yields

Corollary 1

(i) Under the null of conditional sign independence, i.e. $H_0^{DA}: E\left[\text{sign}(y_t)|\mathcal{I}_{t-1}\right] = \text{const}$,

$$\sqrt{T}\left(A_{DA} - B_{DA}\right) \xrightarrow{d} \text{N}(0, V_{DA}),$$

where

$$V_{DA} = \left(1 - m_x^2\right)\left(1 - m_y^2\right).$$

(ii) Under the null of conditional mean independence, i.e. $H_0^{EP}: E\left[y_t|\mathcal{I}_{t-1}\right] = \text{const}$,

$$\sqrt{T}\left(A_{EP} - B_{EP}\right) \xrightarrow{d} \text{N}(0, V_{EP}),$$

where

$$V_{EP} = \left(1 - m_x^2\right)V_y - 2m_x \,\text{cov}\left[\text{sign}(x_t), y_t^2\right].$$

It immediately follows that the directional accuracy statistic (2.1) is robust to serial dependence between the predicted series and the series of predictands, and has correct asymptotic size under such departures from the strong assumptions under which this test was initially derived. In contrast, the EP test (2.4) will have wrong asymptotic size unless

$\text{cov}\left[\text{sign}(x_t), y_t^2\right] = 0$, which is not likely to hold under conditional heteroskedasticity, and unless $m_x = 0$. The form of the asymptotic variance V_{EP} suggests that a heteroskedasticity-robust version of the EP test statistic can be easily constructed by replacing $\hat{V}_{EP}$ in (2.4) by

$$\hat{V}_{EP} = \left(1 - \hat{m}_x^2\right)\hat{V}_y - 2\hat{m}_x\hat{C},$$

where, for example,

$$\hat{C} = \frac{1}{T}\sum_t \left(\text{sign}(x_t) - \hat{m}_x\right) y_t^2.$$

Remark 1 *The power calculations in Anatolyev and Gerko (2005, section 3) stay correct for the robust version of the EP statistic because the additional term in $\hat{V}_{EP}$ tends to zero under the local alternatives asymptotics due to vanishing $\hat{m}_x$.*

3 Regression-based tests

Note that $A_g - B_g$ in (2.5) is a sample covariance between $g(y_t)$ and $\text{sign}(x_t)$. A sample covariance arises in a formula for the OLS estimator of a slope coefficient in a bivariate regression with a constant. Hence, consider the regression

$$g(y_t) = \alpha_g + \beta_g \,\text{sign}(x_t) + \eta_t. \tag{3.1}$$

The OLS estimator of β_g,

$$\hat{\beta}_g = \frac{T^{-1}\sum_t \text{sign}(x_t) g(y_t) - \left(T^{-1}\sum_t \text{sign}(x_t)\right)\left(T^{-1}\sum_t g(y_t)\right)}{1 - \left(T^{-1}\sum_t \text{sign}(x_t)\right)^2},$$

is proportional to $A_g - B_g$ in (2.5). The t-ratio for $\hat{\beta}_g$ is in turn proportional to $\hat{\beta}_g$. Hence, a valid t-test for $\beta_g = 0$ is asymptotically equivalent to an appropriate test of section 2, although the test statistics may be different in value in finite samples. Conversely, the test statistic developed in Section 2, of which the DA and EP statistics are special cases, can be interpreted as (possibly up to a multiplicative term that is asymptotically equal to unity) a t-ratio in bivariate predictive regression (3.1). In particular, such interpretation immediately provides a rationale to the observation made in Section 2 that the asymptotic variance of $A_g - B_g$ does not contain long-run variances and/or long-run covariances. Indeed, when $E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const}$, a linear regression of $g(y_t)$ on any $\mathcal{I}_{t-1}$-measurable variable will provide a zero slope coefficient and does not require autocorrelation-consistent estimation of asymptotic variance because the problem is a single-period one.

Linear predictive regressions similar to (3.1) have been developed and extensively used in the finance literature. Breen, Glosten and Jagannathan (1989, section II, bottom of subsection A) proposed running a bivariate regression of the indicator that a realized return is positive on the indicator that a return prediction is positive. Provided that forecasts do not have a probability atom at zero, up to a linear transformation of variables this is equivalent to the regression (3.1) with $g(u) = \text{sign}(u)$, and, as we already know, the t-test for a zero slope coefficient (BGJ test henceforth) is asymptotically equivalent to the DA test. Cumby and Modest (1987) in turn proposed running a bivariate regression of a realized return itself on the indicator that a return prediction is positive. Analogously, up to a linear transformation of the right-hand side variable this is equivalent to the regression (3.1) with $g(u) = u$, and, as we already know, the t-test for a zero slope coefficient (CM test henceforth) is asymptotically equivalent to the EP test. To our knowledge, the asymptotic equivalence of the BGJ and CM tests to the DA and EP tests, respectively, has not yet been recognized in the literature. The BGJ and CM tests have been applied to many financial time series like stock index returns, as well as to macroeconomic series such as GNP growth (Stekler and Petrei, 2003) and petroleum price changes (Sadorsky, 2002). Along with $g(u) = \text{sign}(u)$, the literature displays an interest to the choice $g(u) = \text{sign}(u - \kappa)$ for some fixed known constant κ. For example, Schnader and Stekler (1990) use $g(u) = \text{sign}(u - \kappa)$ with several choices of κ to see if forecasts can distinguish between periods of low and high levels of GNP growth.

The need to correct for heteroskedasticity was recognized in the finance literature pretty early; see Breen, Jagannathan and Ofer (1986) who emphasized the importance of correcting for heteroskedasticity in predictive regressions, and found that using conventional standard errors may lead to severe size distortions. At the same time, some papers, e.g., Breen, Glosten and Jagannathan (1989), used autocorrelation-consistent variance estimators even though there is no need for that correction, while others, e.g., Cumby and Modest (1987), correct only for heteroskedasticity.

Remark 2 *Consider again the special case* $g(u) = \text{sign}(u)$ *leading to the DA test. Because under the null* $E\left[\text{sign}(y_t)|\mathcal{I}_{t-1}\right] = \text{const}$ *we have*

$$\text{var}\left[\text{sign}(y_t)|\mathcal{I}_{t-1}\right] = E\left[\text{sign}(y_t)^2|\mathcal{I}_{t-1}\right] - E\left[\text{sign}(y_t)|\mathcal{I}_{t-1}\right]^2 = 1 - m_y^2 = \text{const},$$

the regression is conditionally homoskedastic. This implies that the White and conventional forms of asymptotic variance are equivalent. This conclusion parallels the one made in Section 2 about the robustness of the DA test to conditional heteroskedasticity. Hence, one may use conventional standard errors to construct a regression-based test for sign predictability. Thus, in C-panels of their tables IV and V, Breen, Glosten and Jagannathan (1989) did not have to use correction for heteroskedasticity in constructing the standard errors. This also explains why some authors obtained nearly identical conventional and White t-statistics (e.g., Marquering and Verbeek, 2004).

Now we take a further step in extending the constructed class of tests. One may think that it may not be optimal to use $\text{sign}(x_t)$ as a regressor in the predictive regression (3.1) as this variable may not be a strong predictor when there is predictability. Indeed, a regression on a more general regressor $h(x_t)$ can be used to construct an analogous test:

$$g(y_t) = \alpha_{g,h} + \beta_{g,h} h(x_t) + \eta_t. \tag{3.2}$$

In applied literature, it is a common practice to use predictive regressions (3.2) with $g(u) = h(u) = u$. For instance, see a finance application in Hartzmark (1991) where the author calls this technology testing for the "big hit" forecast ability. Such predictive regressions are also familiar from the literature on testing whether one macroeconomic variable has information about another, as in, for example, Hansen and Hodrick (1980) and Mishkin (1990), and from the literature on time-series model selection, as in, for example, Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999). When the task is testing for predictability, the null is $H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const}$, as before. We replace Assumption 1(iii) with

Assumption 2 *The functions $g(u)$ and $h(u)$ are measurable, and $E\left[|g(y_t)|^{2\nu q}\right]$ and $E\left[|h(x_t)|^{2\nu p}\right]$ exist and are finite for ν from Assumption 1, and for some q and p such that $q^{-1} + p^{-1} = 1$.*

The moment condition is sufficient, but not necessary. With a choice of bounded $h(u)$, as is the case for the DA and EP tests, it is possible to set $p = \infty$ and $q = 1$, so that the moment condition on $g(y_t)$ is quite mild.

For the regression (3.2) an appropriate test statistic is the White-corrected t-ratio for the OLS estimator $\hat{\beta}_{g,h}$ of $\beta_{g,h}$:

$$\hat{\beta}_{g,h} = \frac{T^{-1}\sum_t h(x_t)g(y_t) - \left(T^{-1}\sum_t h(x_t)\right)\left(T^{-1}\sum_t g(y_t)\right)}{T^{-1}\sum_t h(x_t)^2 - \left(T^{-1}\sum_t h(x_t)\right)^2}. \tag{3.3}$$

Theorem 2 *Suppose the regularity conditions specified in Assumptions 1 and 2 hold. Consider the regression (3.2), the OLS estimator (3.3) of $\beta_{g,h}$, and the corresponding White-corrected t-ratio $t_{g,h}$. Then under $H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const}$, we have*

$$\sqrt{T}\hat{\beta}_{g,h} \xrightarrow{d} \mathrm{N}\left(0, \frac{V^{g,h}}{V_h^2}\right),$$

where

$$V^{g,h} = V_h V_g + \text{cov}\left[h(x_t)^2, g(y_t)^2\right] - 2M_h \text{cov}\left[h(x_t), g(y_t)^2\right], \tag{3.4}$$

and

$$t_{g,h} \xrightarrow{d} \mathrm{N}(0,1).$$

Note that the formula for $V^{g,h}$ has one more term compared to that of V^g which is non-zero under conditional heteroskedasticity.

As we already know, an alternative (asymptotically equivalent) form of the same test can be directly based on the contrast

$$A_{g,h} - B_{g,h} = \frac{1}{T}\sum_t h(x_t)g(y_t) - \left(\frac{1}{T}\sum_t h(x_t)\right)\left(\frac{1}{T}\sum_t g(y_t)\right).$$

To construct the test statistic, this contrast may be pivotized using

$$\hat{V}^{g,h} = \hat{V}_h\hat{V}_g + \hat{C}_1 - 2\hat{M}_h\hat{C}_2,$$

where, for example,

$$\begin{aligned}
\hat{M}_h &= \frac{1}{T}\sum_t h(x_t), \quad \hat{V}_h = \frac{1}{T}\sum_t \left(h(x_t) - \hat{M}_h\right)^2, \\
\hat{M}_g &= \frac{1}{T}\sum_t g(y_t), \quad \hat{V}_g = \frac{1}{T}\sum_t \left(g(y_t) - \hat{M}_g\right)^2, \\
\hat{C}_1 &= \frac{1}{T}\sum_t h(x_t)^2 g(y_t)^2 - \frac{1}{T}\sum_t h(x_t)^2 \frac{1}{T}\sum_t g(y_t)^2, \\
\hat{C}_2 &= \frac{1}{T}\sum_t \left(h(x_t) - \hat{M}_h\right) g(y_t)^2.
\end{aligned}$$

We want to emphasize that the comparable contrast-based and regression-based tests are asymptotically equivalent to the degree that the difference between the two test statistics is $o_p(1)$.

It is interesting to know which function $h(u)$ one should use in (3.2) to get as much as possible from this test, that is, for the emerging test to be as powerful as possible. To this end, we analyze the power of the test under sequences of local alternatives $H_\delta^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \delta\left(x_t\right)/\sqrt{T}$. It is possible to show (see Appendix B) that the maximal local power is reached when $h(x_t) - M_h$ is proportional to

$$\frac{\delta\left(x_t\right) - M_\delta}{\text{var}\left[g(y_t)|\mathcal{I}_{t-1}\right]},$$

where $M_\delta = E\left[\delta\left(x_t\right)\right].$ In particular, if $g(y_t)$ is conditionally homoskedastic (as we know is the case when $g(u) = \text{sign}(u)$, see Remark 2), to construct the most powerful test, one should take the regressor in (3.2) as close as possible to the direction of a suspected deviation from non-predictability. Application of these results in practice, however, is problematic because $\delta(x_t)$ is unknown; a model selection approach is possible but lies beyond the scope of this paper.

Before concluding this section, we discuss the following question which also concerns the previous and all subsequent sections. From the start, we have been assuming, and will be assuming further, that the frequency of the data coincides with the forecast horizon, so that $y_t \in \mathcal{I}_t$, and the regression errors in (3.2) are non-overlapping. This is not the case in numerous papers on forecasts evaluation, such as Havenner and Modjtahedi (1988), Ash, Smyth and Heravi (1998), Greer (2003), Stekler and Petrei (2003), among others, where long-term forecasts are evaluated so that the spacing of the data falls short of the forecast horizon, and the regression errors in (3.2) are overlapping. The consequence is that some of the tests in the cited and many other papers become asymptotically wrongly sized. These mistakes probably would not have happened if the authors utilized the regression-based approach where the community recognized the necessity of using HAC correction long ago, instead of blindly copying the tests (the exact Henriksson and Merton, 1981, and the DA test of Pesaran and Timmermann, 1992) created for a different environment, and where the need of correcting for autocorrelation is concealed.

4 Reverse regression-based tests

To test for forecasting ability, Breen, Glosten and Jagannathan (1989, section II, subsection B) advise to run a linear regression of $\text{sign}(x_t)$ on $\text{sign}(y_t)$, i.e. a reverse to (3.2) regression with $h(u) = g(u) = \text{sign}(u)$. The authors note that testing for a zero slope coefficient in such reverse regression is asymptotically equivalent to the asymptotic version of the Henriksson and Merton (1981) test of market timing (referred to as HM henceforth). In this Section, we clarify whether this and similar suggestions are valid, and how tests so constructed relate to the class of tests under consideration.

Note that the sample covariance between $g(y_t)$ and $h(x_t)$ in (3.3) can also arise in a formula for the OLS estimator of a slope coefficient in a bivariate regression of $h(x_t)$ on $g(y_t)$, instead of a bivariate regression of $g(y_t)$ on $h(x_t)$, with a constant. Consider this reverse regression:

$$h(x_t) = \alpha_{h,g} + \beta_{h,g} g(y_t) + \theta_t. \tag{4.1}$$

The OLS estimator $\hat{\beta}_{h,g}$ of $\beta_{h,g}$ is

$$\hat{\beta}_{h,g} = \frac{T^{-1}\sum_t h(x_t)g(y_t) - \left(T^{-1}\sum_t h(x_t)\right)\left(T^{-1}\sum_t g(y_t)\right)}{T^{-1}\sum_t g(y_t)^2 - \left(T^{-1}\sum_t g(y_t)\right)^2}. \tag{4.2}$$

For the reverse regression, we obtain the following interesting results.

Theorem 3 *Suppose $h(u)$ and $g(u)$ satisfy the regularity conditions specified in Assumptions 1 and 2. Consider the reverse regression (4.1), the OLS estimator (4.2) of $\beta_{h,g}$, and the corresponding t-ratio $t_{h,g}$. Then under $H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const}$, we have:*
(i)

$$\sqrt{T}\hat{\beta}_{h,g} \xrightarrow{d} \mathrm{N}\left(0, \frac{V^{g,h}}{V_g^2}\right),$$

where $V^{g,h}$ is given in (3.4);
(ii) if the t-ratio $t_{h,g}$ for $\beta_{h,g}$ is constructed using White-corrected standard errors, then

$$t_{h,g} \xrightarrow{d} \mathrm{N}(0, 1);$$

(iii) if the t-ratio $t_{h,g}$ for $\beta_{h,g}$ is constructed using conventional standard errors, then $t_{h,g}$ is **not** *asymptotically standard normal unless $V^{g,h} = V_h V_g$;*
(iv) if the t-ratio $t_{h,g}$ for $\beta_{h,g}$ is constructed using a heteroskedasticity and autocorrelation consistent formula, then $t_{h,g}$ is **not** *asymptotically standard normal unless $h(x_t)$ has an martingale difference structure.*

Perhaps surprisingly, running the predictability test based on the reverse regression (4.1), that of a (function of) *predictor* on a (function of) *predictand*, is as valid as running the predictability test based on the straight regression (3.2). Another surprise comes from the fact that, unless in special circumstances, in constructing the standard errors it is erroneous to use a HAC asymptotic variance estimate, and only correction for heteroskedasticity is needed to provide the correct standard errors. This seemingly contradicts the common experience that autocorrelation-consistent forms are necessary when the estimated regression is not a single-period one; the reverse regression (4.1) is not a single-period one if $h(x_t)$ exhibits serial correlation, which is likely to be the case with non-trivial predictors. The explanation is that the reverse regression (4.1) is not a regression in a strict sense even under the null hypothesis.

Returning to the suggestion in Breen, Glosten and Jagannathan (1989, section II, subsection B), we can now say that it is valid provided that the standard errors are appropriately constructed. In their reverse regressions, however, Breen, Glosten and Jagannathan (1989) used the autocorrelation consistent asymptotic variance estimate when constructing standard errors (see D-panels of their tables IV and V), which, as follows from Theorem 3(iv), is incorrect. The standard errors there should have been either heteroskedasticity-corrected according to Theorem 3(ii), or conventional, as in the special case $h(u) = g(u) = \text{sign}(u)$ there is conditional homoskedasticity and $V^{g,h}$ equals $V_h V_g$, according to Theorem 3(iii).

We also want to emphasize that the regression-based and corresponding reverse regression-based tests are asymptotically equivalent to the degree that the difference between the two test statistics is $o_p(1)$. In spite of this asymptotic equivalence, sometimes both test statistics are reported even though they are indeed very close in value (e.g., Sadorsky, 2002).

5 Extension to multiple hypotheses

So far we considered the tests where a researcher tests one feature for conditional independence using one function of a predictor. In this section we extend this framework to testing for conditional independence of more than one feature of y_t, as well as to using more than one function of a predictor. In addition to the importance of this extension per se, we will see that this extended framework encompasses some of tests that exist in the literature and have not yet been considered here so far in this paper.

Consider the following multiple multivariate linear regression

$$\mathbf{g}(y_t) = \boldsymbol{\alpha}_{\mathbf{g},\mathbf{h}} + \mathrm{B}_{\mathbf{g},\mathbf{h}}\mathbf{h}(x_t) + \boldsymbol{\eta}_t, \tag{5.1}$$

where boldface characters represent vectors: $\mathbf{g}(y_t)$, $\boldsymbol{\alpha}_{\mathbf{g},\mathbf{h}}$ and $\boldsymbol{\eta}_t$ are $\ell \times 1$, and $\mathbf{h}(x_t)$ is $k \times 1$. The $\ell \times k$ matrix of coefficients $\mathrm{B}_{\mathbf{g},\mathbf{h}}$ equals

$$\mathrm{B}_{\mathbf{g},\mathbf{h}} = \mathrm{cov}\left[\mathbf{g}(y_t), \mathbf{h}(x_t)\right] \left(\mathrm{var}\left[\mathbf{h}(x_t)\right]\right)^{-1}.$$

The null hypothesis of interest

$$H_0^{\mathbf{g}} : E\left[\mathbf{g}(y_t)|\mathcal{I}_{t-1}\right] = \mathbf{const} \tag{5.2}$$

can be tested by testing if some or all coefficients in $\mathrm{B}_{\mathbf{g},\mathbf{h}}$ are zero using the Wald test and equation-by-equation OLS estimates. The OLS estimator equals

$$\begin{aligned}\hat{\mathrm{B}}_{\mathbf{g},\mathbf{h}} &= \left(\frac{1}{T}\sum_t \mathbf{g}(y_t)\mathbf{h}(x_t)' - \frac{1}{T}\sum_t \mathbf{g}(y_t)\frac{1}{T}\sum_t \mathbf{h}(x_t)'\right) \\ &\qquad \times \left(\frac{1}{T}\sum_t \mathbf{h}(x_t)\mathbf{h}(x_t)' - \frac{1}{T}\sum_t \mathbf{h}(x_t)\frac{1}{T}\sum_t \mathbf{h}(x_t)'\right)^{-1}.\end{aligned}$$

The first factor in $\hat{\mathrm{B}}_{\mathbf{g},\mathbf{h}}$ is a familiar contrast, which is now an $\ell \times k$ matrix.

Previously, we have discussed the case $\ell = k = 1$. Using $\ell > 1$ and keeping $k = 1$, one can test for conditional independence of more than one feature of y_t. For example, it can be a joint test for conditional mean and conditional variance independence, to which we will return shortly. Using $k > 1$ and keeping $\ell = 1$, one can increase the power of the tests discussed previously, as several regressors have a better chance to be collinear with the direction implied by the alternative hypothesis. Below we pay special attention to a particular case where $\ell = k$ and $\mathbf{g}(u) = \mathbf{h}(u)$, but, of course, the relationship between ℓ and k may be any, and $\mathbf{g}(u)$ and $\mathbf{h}(u)$ need not coincide.

A leading example of tests with $\ell = k > 1$ that can be encountered in the literature corresponds to

$$\mathbf{g}(u) = \mathbf{h}(u) = \begin{pmatrix} \text{sign}(u - \kappa_1) \\ \text{sign}(u - \kappa_2) \\ \vdots \\ \text{sign}(u - \kappa_\ell) \end{pmatrix} \tag{5.3}$$

for a fixed set of constants $\kappa_1 < \kappa_2 < \cdots < \kappa_\ell$. Testing in this context can be mapped to χ^2-tests for independence (or, more precisely, for no association) using the following $(\ell+1) \times (\ell+1)$ contingency table with identical categorizations (for a review of the theory of contingency tables see, for instance, Kendall and Stuart, 1973, chapter 33):

		y_t				
		$\kappa_0 < y_t \leq \kappa_1$	$\kappa_1 < y_t \leq \kappa_2$	$\cdots$	$\kappa_{\ell-1} < y_t \leq \kappa_\ell$	$\kappa_\ell < y_t \leq \kappa_{\ell+1}$
	$\kappa_0 < x_t \leq \kappa_1$	p_{11}	p_{12}	$\cdots$	$p_{1\ell}$	$p_{1,\ell+1}$
	$\kappa_1 < x_t \leq \kappa_2$	p_{21}	p_{22}	$\cdots$	$p_{2\ell}$	$p_{2,\ell+1}$
x_t	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	$\kappa_{\ell-1} < x_t \leq \kappa_\ell$	$p_{\ell 1}$	$p_{\ell 2}$	$\cdots$	$p_{\ell\ell}$	$p_{\ell,\ell+1}$
	$\kappa_\ell < x_t \leq \kappa_{\ell+1}$	$p_{\ell+1,1}$	$p_{\ell+1,2}$	$\cdots$	$p_{\ell+1,\ell}$	$p_{\ell+1,\ell+1}$

where $\kappa_0 \equiv -\infty$, $\kappa_{\ell+1} \equiv +\infty$, and

$$p_{ij} = \frac{1}{T} \sum_t \mathbb{I} \left(\kappa_{i-1} < x_t \leq \kappa_i, \kappa_{j-1} < y_t \leq \kappa_j \right).$$

Let us also denote

$$\begin{aligned} p_{i\cdot} &= \frac{1}{T} \sum_t \mathbb{I} \left(\kappa_{i-1} < x_t \leq \kappa_i \right), \\ p_{\cdot j} &= \frac{1}{T} \sum_t \mathbb{I} \left(\kappa_{j-1} < y_t \leq \kappa_j \right), \end{aligned}$$

and

$$m_{x,i} = E \left[\text{sign} \left(x_t - \kappa_i \right) \right], \quad m_{y,j} = E \left[\text{sign} \left(y_t - \kappa_j \right) \right].$$

Note that if $i_1 < i_2$, then

$$\begin{aligned} \operatorname{cov}\left[\operatorname{sign}\left(x_t - \kappa_{i_1}\right), \operatorname{sign}\left(x_t - \kappa_{i_2}\right)\right] &= E\left[\operatorname{sign}\left(x_t - \kappa_{i_1}\right)\operatorname{sign}\left(x_t - \kappa_{i_2}\right)\right] \\ &\quad - E\left[\operatorname{sign}\left(x_t - \kappa_{i_1}\right)\right] E\left[\operatorname{sign}\left(x_t - \kappa_{i_2}\right)\right] \\ &= \left(1 - m_{x,i_1}\right)\left(1 + m_{x,i_2}\right), \end{aligned}$$

therefore the matrix $\operatorname{var}\left[\mathbf{h}(x_t)\right]$ consists of such products for various i_1 and i_2.

The χ^2-test statistic for no association equals

$$X^2 = \sum_{i=1}^{\ell+1}\sum_{j=1}^{\ell+1} \frac{\left(p_{ij} - p_{i\cdot}p_{\cdot j}\right)^2}{p_{i\cdot}p_{\cdot j}},$$

and is asymptotically distributed as $\chi^2\left(\ell^2\right)$. Using the identity

$$\mathbb{I}\left(u_1 < u \leq u_2\right) = \frac{1}{2}\left(\operatorname{sign}\left(u - u_1\right) - \operatorname{sign}\left(u - u_2\right)\right),$$

we can obtain that

$$X^2 \stackrel{A}{=} \sum_{i=1}^{\ell+1}\sum_{j=1}^{\ell+1} \frac{\left(T^{-1}\sum_t \Delta\operatorname{sign}_i\left(x_t\right)\Delta\operatorname{sign}_j\left(y_t\right) - T^{-1}\sum_t \Delta\operatorname{sign}_i\left(x_t\right) T^{-1}\sum_t \Delta\operatorname{sign}_j\left(y_t\right)\right)^2}{4\left(m_{x,i-1} - m_{x,i}\right)\left(m_{y,j-1} - m_{y,j}\right)},$$

where we use the notational shortcut

$$\Delta\operatorname{sign}_i(u) \equiv \operatorname{sign}\left(u - \kappa_{i-1}\right) - \operatorname{sign}\left(u - \kappa_i\right).$$

One can see that X^2 (asymptotically) has the form of a Wald test for the nullity of all slope coefficients in the multiple multivariate regression having $\ell + 1$ equations, the j^{th} of which has $\Delta\operatorname{sign}_j\left(y_t\right)$ as a left side variable, and $\Delta\operatorname{sign}_1\left(x_t\right), \cdots, \Delta\operatorname{sign}_{\ell+1}\left(x_t\right)$, along with a constant, as right side variables. There are $(\ell+1)^2$ such slope coefficients, but by construction one of the equations is a linear combination of the others, and one regressor is a linear combination of the others, which results in only ℓ^2 degrees of freedom. Moreover, by a linear transformation, the nullity of all slope coefficients in the system just described is equivalent to the nullity of all slope coefficients in the system of ℓ equations the j^{th} of which has $\operatorname{sign}\left(y_t - \kappa_j\right)$ as a left side variable, and $\operatorname{sign}\left(x_t - \kappa_1\right), \cdots, \operatorname{sign}\left(x_t - \kappa_\ell\right)$, along with a constant, as right side variables, i.e. in the original system (5.1) with (5.3). Thus we can conclude that the classical χ^2-test for no association is equivalent to the test for predictability of particular signs of shifted arguments via testing for the nullity of all slope coefficients in the multiple multivariate linear regression of the type (5.1).

Pesaran and Timmermann (1992, Section 2, and 1994) propose a generalization of the DA test that was discussed previously. After its components are recentered and renormalized as before, their test statistic is based on the sum of $\ell + 1$ contrasts:

$$\sum_{i=1}^{\ell+1} \left(p_{ii} - p_{i\cdot}p_{\cdot i}\right). \tag{5.4}$$

Again, switching from indicators to signs yields that

$$\sum_{i=1}^{\ell+1} \left(p_{ii} - p_{i\cdot}p_{\cdot i}\right) = \sum_{i=1}^{\ell+1} \left(\frac{1}{T}\sum_t \Delta\,\mathrm{sign}_i\left(x_t\right)\Delta\,\mathrm{sign}_i\left(y_t\right) - \frac{1}{T}\sum_t \Delta\,\mathrm{sign}_i\left(x_t\right)\frac{1}{T}\sum_t \Delta\,\mathrm{sign}_i\left(y_t\right)\right). \tag{5.5}$$

It follows that this test tests for the nullity of the following linear combination of coefficients in $\mathrm{B}_{\mathbf{g},\mathbf{h}}$:

$$b = \mathrm{tr}\left(\left\|\mathrm{cov}\left[\Delta\,\mathrm{sign}_{i_1}\left(y_t\right), \Delta\,\mathrm{sign}_{i_2}\left(x_t\right)\right]\right\|_{i_1,i_2=1}^{\ell+1}\right)$$

The covariance matrix inside the tr operator is

$$\left\|\mathrm{cov}\left[\Delta\,\mathrm{sign}_{i_1}\left(y_t\right), \Delta\,\mathrm{sign}_{i_2}\left(x_t\right)\right]\right\|_{i_1,i_2=1}^{\ell+1} = \Xi\left\|\mathrm{cov}\left[\mathrm{sign}\left(y_t - \kappa_{i_1}\right), \mathrm{sign}\left(x_t - \kappa_{i_2}\right)\right]\right\|_{i_1,i_2=1}^{\ell}\Xi',$$

where

$$\Xi = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}.$$

Hence,

$$b = \mathrm{tr}\left(\Xi\mathrm{B}_{\mathbf{g},\mathbf{h}}\mathrm{var}\left[\mathbf{h}(x_t)\right]\Xi'\right) = \mathrm{tr}\left(\mathrm{B}_{\mathbf{g},\mathbf{h}}\mathrm{var}\left[\mathbf{h}(x_t)\right]\Psi\right),$$

where

$$\Psi = \Xi'\Xi = \begin{bmatrix} 2 & -1 & \cdots & 0 & 0 \\ -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & \cdots & -1 & 2 \end{bmatrix},$$

and the composition of the matrix $\mathrm{var}\left[\mathbf{h}(x_t)\right]$ was discussed above. Why this particular linear combination is used for testing the null

$$H_0^{\mathrm{sign}} : E\left[\mathrm{sign}(y_t - \kappa_j)|\mathcal{I}_{t-1}\right] = \mathrm{const} \text{ for } j = 1, \cdots, \ell,$$

is unclear. For example, when $\ell = 2$, this combination equals

$$\begin{aligned} b &= (1 - m_{x,1})(1 + 2m_{x,1} - m_{x,2})\beta_{11} + (1 + m_{x,2})(1 - 2m_{x,1} + m_{x,2})\beta_{12} \\ &\quad + (1 - m_{x,1})(1 + 2m_{x,2} - m_{x,1})\beta_{21} + (1 + m_{x,2})(1 - 2m_{x,2} + m_{x,1})\beta_{22}. \end{aligned}$$

There does not seem to be anything special about this combination. A more natural test would be the χ^2-test for nullity of all coefficients in $\mathrm{B}_{\mathbf{g},\mathbf{h}}$ if the target is indeed testing for H_0^{sign}, as discussed above.

Now we will discuss, as another illustration, the construction of a joint test for conditional mean and conditional variance independence of returns. The null $H_0^{E,var} : E\left[y_t|\mathcal{I}_{t-1}\right] = \mathrm{const}$, $\mathrm{var}\left[y_t|\mathcal{I}_{t-1}\right] = \mathrm{const}$ can be reformulated as $H_0^{E,var} : E\left[y_t|\mathcal{I}_{t-1}\right] = \mathrm{const}$, $E\left[y_t^2|\mathcal{I}_{t-1}\right] = \mathrm{const}$. Then the above framework applies with $\mathbf{g}(y_t) = \left(y_t, y_t^2\right)'$. If, for example, one naturally chooses $\mathbf{h}(x_t) = \left(x_t, x_t^2\right)'$, the null hypothesis $H_0^{E,var}$ can be tested by running OLS on the system

$$y_t = \alpha_y + \beta_{y,x}x_t + \beta_{y,x^2}x_t^2 + \eta_{y,t},$$
$$y_t^2 = \alpha_{y^2} + \beta_{y^2,x}x_t + \beta_{y^2,x^2}x_t^2 + \eta_{y^2,t},$$

and testing $\beta_{y,x} = \beta_{y,x^2} = \beta_{y^2,x} = \beta_{y^2,x^2} = 0$ using the Wald test. In a similar manner one can construct a joint test for directional non-predictability of returns and their volatility; then $\mathbf{g}(y_t) = \left(\mathrm{sign}\left(y_t\right), \mathrm{sign}\left(y_t^2 - \kappa\right)\right)'$, where κ is a chosen positive benchmark, and a natural choice for $\mathbf{h}(x_t)$ then is $\mathbf{h}(x_t) = (\text{sign}\,(x_t)\,, \text{sign}\,(x_t^2 - \kappa))'$. The second entry in both vectors is not exactly the sign of shifted volatility, but it is quite close to it given that for a near-zero mean series r_t, $\text{var}\,[r_t] = E\left[r_t^2\right] - E\left[r_t\right]^2 \approx E\left[r_t^2\right]$. Recently, Marquering and Verbeek (2004) have developed an analog of the BGJ test aimed at testing jointly for directional non-predictability of returns and their volatility, which may also be implemented by using the chi-square test discussed above for a 4×4 contingency table. The authors suggest a 4-equation multiple regression of indicators of events related to both returns and observed volatility, on indicators of events related to both return and volatility predictors. In order to implement such test, one needs an observable measure of volatility; the authors use a monthly realized volatility measure calculated from daily returns. Note that in our approach one need not observe the volatility to test essentially the same directional hypothesis; and in the case of mean independence of returns and their squared the procedure is totally clean of any approximations.

Returning to the general regression formulation (5.1), we can also see by the logic of Section 4 that the null (5.2) can also be tested via the Wald test in the reverse regression

$$\mathbf{h}(x_t) = \boldsymbol{\alpha}_{\mathbf{h},\mathbf{g}} + \mathrm{B}_{\mathbf{h},\mathbf{g}}\mathbf{g}(y_t) + \boldsymbol{\theta}_t \tag{5.6}$$

and the OLS estimate of the slope coefficients

$$\begin{aligned}\hat{\mathrm{B}}_{\mathbf{h},\mathbf{g}} &= \left(\frac{1}{T}\sum_t \mathbf{g}(y_t)\mathbf{h}(x_t)' - \frac{1}{T}\sum_t \mathbf{g}(y_t)\frac{1}{T}\sum_t \mathbf{h}(x_t)'\right)' \\ &\qquad \times \left(\frac{1}{T}\sum_t \mathbf{g}(y_t)\mathbf{g}(y_t)' - \frac{1}{T}\sum_t \mathbf{g}(y_t)\frac{1}{T}\sum_t \mathbf{g}(y_t)'\right)^{-1},\end{aligned}$$

because both $\hat{\mathrm{B}}_{\mathbf{g},\mathbf{h}}$ and $\hat{\mathrm{B}}_{\mathbf{h},\mathbf{g}}$ are based on the same contrast.

6 Discussion

In this Section we discuss some issues pertinent to the regression-based tests such as their interpretation and nonparametric nature, and argue against some widespread misconceptions.

We have established that the three ways of constructing a test for conditional independence of $g(y_t)$ are asymptotically equivalent, and their numerical values depend, holding the choice of $h(x_t)$ constant, only on how the contrast $A_{g,h} - B_{g,h}$ is pivotized. This implies, in particular, that any valid interpretation given to such test, that is reliant on the form of pivotization, can also be extended to other forms. The same is true regarding the choice of $h(x_t)$, as this choice impacts only the power of the test. On the other hand, any interpretation should not rely on the relationship not implied by the null hypothesis of conditional independence. Let us, for example, recall how the simple Henriksson and Merton (1981) test of market timing corresponding to $h(u) = g(u) = \text{sign}(u)$, is interpreted in the literature:

1. Rejection means that returns forecasts have value to an investor (Merton, 1981).

2. Rejection means that an investor's prior probability density over returns is changed when the investor obtains a returns forecast (Henriksson and Merton, 1981).

3. Rejection means that return forecasts are independent of observed returns (Pesaran and Timmermann, 1992).

4. Rejection means that a set of return forecasts differs significantly from a naive model that consistently predicts up or consistently predicts down (Schnader and Stekler, 1990).

Some of these interpretations, however, assume much more than just the relationship implied by the null hypothesis $H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const}$.

Another warning concerns the out-of-sample interpretation of tests. The EP test of Anatolyev and Gerko (2005) was constructed as explicitly tied to a virtual investor's simple trading strategy. The DA test of Pesaran and Timmermann (1992) can possibly also be tied an analogous, albeit a more complicated, trading strategy involving options in the spirit of Breen, Glosten and Jagannathan (1989, section 2B). In spite of these facts, the EP or DA tests should not be given an interpretation that they test out-of-sample predictive ability of an underlying forecasting model. As follows from their asymptotic equivalence to corresponding regression-based tests, no comparison of out-of-sample measures is involved in their construction.

In the rest of this Section, we comment on a series of misconceptions that often accompany the use of the tests under consideration in the applied literature.

The first misconception concerns a sometimes awkward comparison of tests that belong to different null hypotheses. Often one can encounter claims that successful sign predictions may be of smaller value than successful mean predictions for an investor concerned with maximizing profits (Skouras, 2000, Stekler and Petrei, 2003, Anatolyev and Gerko, 2005). These claims are perfectly valid. A related claim, however, that this implies that a test for one null is more powerful than a test for another null, is questionable. Power of tests can be compared if they test the same feature, as in our analysis of local power at the end of Section 3 where we hold the null $H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const}$ fixed and vary choices of $h(x_t)$. Which null to test depends on the researcher's objective, be it profit maximization or something else, but given the null, it is legitimate to compare only tests that test this particular null. Hence, the critique of the HM test in Cumby and Modest (1987, pp.177–178) should actually be viewed as a critique of the null that is tested, not of the HM test itself.

Regarding the power, claims that one test (say, the EP or CM test) is powerful and another test (say, the DA or HM test) is weak as done, for example, in Cumby and Modest (1987, pp.175–177) and repeated in many applied papers, are also questionable. Indeed, in example 1 of Section 2 where the series exhibits conditional mean independence but sign predictability, the EP (CM) test will not have power at all, while the DA (HM) test will. Analogously, in example 2 of Section 2 where the series exhibits conditional sign independence but mean predictability, the DA (HM) test will not have power at all, while the EP (CM) test will.

One may, on the other hand, view predictability in principle (i.e. absence of serial dependence) as a "universal" null hypothesis, and compare ability of different tests to detect this predictability. One should make sure, however, that the predictability extends in directions away from all the nulls assumed by the tests under consideration. Such exercise is performed in Anatolyev and Gerko (2005, section 3) who compare the DA and EP tests in this way. Implementation of a test for the absence of serial dependence by running simultaneous tests with different choices of $g(u)$, however, does not seem effective[2]; probably other tools should be used instead, e.g. the Ljung–Box test for linear predictability, the BDS test (Brock, Dechert, Scheinkman and LeBaron, 1996) for nonlinear predictability.

Another serious misconception concerns attempts in the literature to give predictive regressions (3.2) and (4.1) a status of a parametric model [cf., for example, the following passages: "The Henriksson–Merton test ... treats realized returns as a dichotomous variable..." and "the more restrictive assumption that the forecast is independent of the magnitude of subsequent realized return" in Cumby and Modest (1987, p.178), or "... CM assumes that the magnitude of the price change ... depends linearly on the forecast" in Hartzmark (1991, p. 54).] As follows from our analysis, the choice of the function $g(u)$ is dictated only by what feature a researcher wants to test, and the choice of the function $h(u)$ is pretty arbitrary, although it has an impact on the power and interpretability. Although (3.2) or (4.1) deceivingly look as regressions, in fact they are not (except that (3.2) *is* a regression but only under the null H_0^g), and they do not admit any "causality" or "dependence" interpretation. The fact that we estimate linear equation (3.2) or (4.1) does not mean that we assume some structural linear relationship between $g(y_t)$ and $h(x_t)$. In reality, all tests we consider in this paper are nonparametric tests, even though some may appear as tests on coefficients in a linear "parametric" regression. This appearance is only a part of a convenient device that allows one to apply the regression theory for one very specific purpose – to test for zeroness of slope coefficients. In particular, this device facilitates testing in practice via use of standard

[2]Schnader and Stekler (1990) repeat the HM procedure changing $g(u) = \text{sign}(u)$ to $g(u) = \text{sign}(u - \kappa)$ with several choices of κ to determine whether the results of the HM test are robust. A more appropriate procedure is one described in Section 5.

regression packages such as, for example, *Econometric Views*.

7 Conclusion

We have discovered that many existing tests for time-series predictability are special cases of a general nonparametric test based on the OLS estimator of the slope coefficient in a bivariate linear regression of certain type. The class of such tests is big and extends in at least two directions: one is indexed by the feature whose conditional expectation is tested for independence, and the other is indexed by the function of the given predictor. Of course, this class, albeit big, is restricted in several respects; for example, all variables are assumed stationary. The class considered does not include some other existing tests. For example, the BDS test statistics (Brock, Dechert, Scheinkman and LeBaron, 1996) used as a portmanteau test for neglected nonlinearity, also has a form of a pivotized contrast, but the components of this contrast are Wilcoxon-type averages rather than simple averages. Sometimes literature suggests using even more complicated statistics like the Spearman rank correlation coefficient as in Chance and Hemler (2001), or a certain U-statistic as in Jiang (2003). It is possible that such tests may also be outcomes of some estimators in linear regressions, but those estimators do not take a form of a function of simple empirical averages but rather take some fancier forms.

Acknowledgements

I would like to thank participants of the "Predictability testing" session at the Econometric Society World Congress 2005 in London, and seminar audiences at the Helsinki Center of Economic Research, Stockholm University, and especialy Research Division of Sveriges Riksbank.

References

Anatolyev, S. and A. Gerko (2005) A trading approach to testing for predictability. *Journal of Business and Economic Statistics* 23, 455–461.

Ash, J.C.K., D.J. Smyth and S.M. Heravi (1998) Are OECD forecasts rational and useful? A directional analysis. *International Journal of Forecasting* 14, 381–391.

Bossaerts, P. and P. Hillion (1999) Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies* 12, 405–428.

Breen, W., L.R. Glosten and R. Jagannathan (1989) Economic significance of predictable variations in stock index returns. *Journal of Finance* 44, 1177–1189.

Breen, W., R. Jagannathan and A.R. Ofer (1986) Correcting for heteroskedasticity in tests for market timing ability. *Journal of Business* 59, 585–598.

Brock, W., D. Dechert, J. Scheinkman and B. LeBaron (1996) A test for independence based on the correlation dimension. *Econometric Reviews* 15, 197–235.

Chance, D.M. and M.L. Hemler (2001) The performance of professional market timers: daily evidence from executed strategies. *Journal of Financial Economics* 62, 377–411.

Christoffersen, P.F. and F.X. Diebold (2003) Financial asset returns, direction-of-change forecasting, and volatility dynamics. Manuscript, University of Pennsylvania.

Cumby, R.E., and D.M. Modest (1987) Testing for market timing ability: a framework for forecast evaluation. *Journal of Financial Economics* 19, 169–89.

Franses, P. and D. van Dijk (2000) Nonlinear Time Series Models in Empirical Finance. Cambridge University Press.

Gençay, R. (1998) Optimization of technical trading strategies and the profitability in security markets. *Economics Letters* 59, 249–254.

Granger, C.W.J. and H.M. Pesaran (2000) Economic and Statistical Measures of Forecast Accuracy. *Journal of Forecasting* 19, 537–560.

Greer, M. (2003) Directional accuracy tests of long-term interest rate forecasts. *Interna-*

tional Journal of Forecasting 19, 291–298.

Hansen, L.P. and R.J. Hodrick (1980) Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. *The Journal of Political Economy* 88, 829–853.

Hartzmark, M.L. (1991) Luck versus forecast ability: determinants of trader performance in futures markets. *Journal of Business* 64, 49–74.

Havenner, A. and B. Modjtahedi (1988) Foreign exchange rates: A multiple currency and maturity analysis. *Journal of Econometrics* 37, 251–264.

Henriksson, R.D. and R.C. Merton (1981) On market timing and investment performance II: statistical procedures for evaluating forecasting skills. *Journal of Business* 54, 513–533.

Ibragimov, I.A. (1962) Some limit theorems for stationary processes. *Theory of Probability and its Applications* 7, 349–382.

Jiang, W. (2003) A nonparametric test of market timing. *Journal of Empirical Finance* 10, 399–425.

Kendall, M.G. and A. Stuart (1973) *The Advanced Theory of Statistics.* Griffin: London.

Lai, K.S. (1990) An evaluation of survey exchange rate forecasts. *Economics Letters* 32, 61–65.

Marquering, W. and M. Verbeek (2004) A multivariate nonparametric test for return and volatility timing. *Finance Reearch Letters* 1, 250–260.

Merton, R.C. (1981) On market timing and investment performance I: an equilibrium theory of value for market forecasts. *Journal of Business* 54, 363–406.

Mishkin, F. (1990) What does the term structure tell us about future inflation? *Journal of Monetary Economics* 25, 77–95.

Pesaran, M.H. and A. Timmermann (1992) A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* 10, 561–565.

Pesaran, M.H. and A. Timmermann (1994) A generalization of the nonparametric Henriksson–Merton test of market timing. *Economics Letters* 4, 1–7.

Pesaran, M.H. and A. Timmermann (1995) Predictability of stock returns: robustness and economic significance. *Journal of Finance* 50, 1201–1228.

Qi, M. (1999) Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business and Economic Statistics* 17, 419–429.

Qi, M. and Y. Wu (2003) Nonlinear prediction of exchange rates with monetary fundamentals. *Journal of Empirical Finance* 10, 623–640.

Sadorsky, P. (2002) Time-varying risk premiums in petroleum futures prices. *Energy Economics* 24, 539–556.

Schnader, M.H. and H.O. Stekler (1990) Evaluating predictions of change. *Journal of Business* 63, 99–107.

Skouras, S. (2000) Risk neutral forecasting. Unpublished manuscript, Santa Fe Institute.

Stekler, H.O. and G. Petrei (2003) Diagnostics for evaluating the value and rationality of economic forecasts. *International Journal of Forecasting* 19, 735–742.

A Appendix: proofs

Lemma 1 *Under assumptions 1 and 2, the infinite summation*

$$\sum_{j=-\infty}^{+\infty} E\left[h(x_t)g(y_t), h(x_{t+j})g(y_{t+j})\right]$$

absolutely converges.

Proof. Using Ibragimov's (1962) and Hölder's inequalities, for all $j > 0$, the quantity $\left|E\left[h(x_t)g(y_t)h(x_{t+j})g(y_{t+j})\right]\right|$ equals

$$\begin{aligned}
& \left|E\left[h(x_t)g(y_t)\left(E\left[h(x_{t+j})g(y_{t+j})|I_t\right] - E\left[h(x_{t+j})g(y_{t+j})\right]\right)\right]\right| \\
\leq\ & \left(E\left[|h(x_t)g(y_t)|^{2\nu}\right]\right)^{1/2\nu}\left(E\left[\left|E\left[h(x_{t+j})g(y_{t+j})|I_t\right] - E\left[h(x_{t+j})g(y_{t+j})\right]\right|^{2\nu/(2\nu-1)}\right]\right)^{1-1/2\nu} \\
\leq\ & \left(E\left[|h(x_t)g(y_t)|^{2\nu}\right]\right)^{1/2\nu} \cdot 8\alpha\left(j\right)^{1-1/\nu}\left(E\left[|h(x_t)g(y_t)|^{2\nu}\right]\right)^{1/2\nu} \\
\leq\ & 8\alpha\left(j\right)^{1-1/\nu}\left(E\left[|h(x_t)|^{2\nu p}\right]\right)^{1/\nu p}\left(E\left[|g(y_t)|^{2\nu q}\right]\right)^{1/\nu q}.
\end{aligned}$$

Hence,

$$\sum_{j=-\infty}^{+\infty} |E[h(x_t)g(y_t)h(x_{t+j})g(y_{t+j})]| \leq 8 \left(E\left[|h(x_t)|^{2\nu p}\right]\right)^{1/\nu p} \left(E\left[|g(y_t)|^{2\nu q}\right]\right)^{1/\nu q} \sum_{j=-\infty}^{+\infty} \alpha(|j|)^{1-1/\nu} < \infty.$$

■

Lemma 2 *Suppose $h(u)$ and $g(u)$ satisfy the regularity conditions specified in Assumptions 1 and 2. Consider the contrast*

$$A_{g,h} - B_{g,h} \equiv \frac{1}{T}\sum_t h(x_t)g(y_t) - \left(\frac{1}{T}\sum_t h(x_t)\right)\left(\frac{1}{T}\sum_t g(y_t)\right). \tag{A.1}$$

Under $H_0^g : E[g(y_t)|\mathcal{I}_{t-1}] = \text{const}$,

$$\sqrt{T}(A_{g,h} - B_{g,h}) \xrightarrow{d} \mathrm{N}(0, V^{g,h}),$$

where

$$V^{g,h} = V_h V_g + \operatorname{cov}\left[h(x_t)^2, g(y_t)^2\right] - 2M_h \operatorname{cov}\left[h(x_t), g(y_t)^2\right].$$

Proof. Note that

$$\begin{aligned}\sqrt{T}(A_{g,h} - B_{g,h}) &= \sqrt{T}\left(T^{-1}\sum_t h(x_t)g(y_t) - M_h M_g\right) \\ &\quad - T^{-1}\sum_t h(x_t)\sqrt{T}\left(T^{-1}\sum_t g(y_t) - M_g\right) \\ &\quad - M_g\sqrt{T}\left(T^{-1}\sum_t h(x_t) - M_h\right).\end{aligned}$$

Under the conditions of the theorem,

$$\sqrt{T}\begin{pmatrix} T^{-1}\sum_t h(x_t)g(y_t) - M_h M_g \\ T^{-1}\sum_t g(y_t) - M_g \\ T^{-1}\sum_t h(x_t) - M_h \end{pmatrix} \xrightarrow{d} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{12} & V_{22} & V_{23} \\ V_{13} & V_{23} & V_{33} \end{pmatrix}\right),$$

where, using repeatedly the CLT for stationary and mixing sequences, the result of Lemma 1, the condition $E\left[g(y_t)|\mathcal{I}_{t-1}\right] = M_g$, and the law of iterated expectations,

$$\begin{aligned}
V_{11} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), h(x_{t+j})g(y_{t+j})\right] \\
&= \operatorname{var}\left[h(x_t)g(y_t)\right] + 2M_g \sum_{j=1}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), h(x_{t+j})\right], \\
V_{22} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[g(y_t), g(y_{t+j})\right] = V_g, \\
V_{33} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[h(x_t), h(x_{t+j})\right] = V_h + 2\sum_{j=1}^{+\infty} \operatorname{cov}\left[h(x_t), h(x_{t+j})\right], \\
V_{12} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), g(y_{t+j})\right] \\
&= \operatorname{cov}\left[h(x_t), g(y_t)^2\right] + M_h V_g + M_g \sum_{j=1}^{+\infty} \operatorname{cov}\left[g(y_t), h(x_{t+j})\right], \\
V_{13} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), h(x_{t+j})\right] \\
&= M_g V_h + M_g \sum_{j=1}^{+\infty} \operatorname{cov}\left[h(x_t), h(x_{t+j})\right] + \sum_{j=1}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), h(x_{t+j})\right], \\
V_{23} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[g(y_t), h(x_{t+j})\right] = \sum_{j=1}^{+\infty} \operatorname{cov}\left[g(y_t), h(x_{t+j})\right].
\end{aligned}$$

Now, using the Delta method and Slutsky's lemma,

$$\begin{aligned}
\sqrt{T}\left(A_g - B_g\right) &\overset{A}{\sim} \begin{pmatrix} 1 \\ -M_h \\ -M_g \end{pmatrix}' \sqrt{T} \begin{pmatrix} T^{-1}\sum_t h(x_t)g(y_t) - M_h M_g \\ T^{-1}\sum_t g(y_t) - M_g \\ T^{-1}\sum_t h(x_t) - M_h \end{pmatrix} \\
&\xrightarrow{d} \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -M_h \\ -M_g \end{pmatrix}' \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{12} & V_{22} & V_{23} \\ V_{13} & V_{23} & V_{33} \end{pmatrix} \begin{pmatrix} 1 \\ -M_h \\ -M_g \end{pmatrix} \right) \\
&\sim \mathrm{N}(0, V^{g,h}),
\end{aligned}$$

where

$$\begin{aligned} V^{g,h} &= \text{cov}\left[h(x_t)^2, g(y_t)^2\right] + V_h V_g + M_h^2 V_g + M_g^2 V_h \\ &\quad -2M_h \text{cov}\left[h(x_t), g(y_t)^2\right] - M_h^2 V_g - M_g^2 V_h \\ &= V_h V_g + \text{cov}\left[h(x_t)^2, g(y_t)^2\right] - 2M_h \text{cov}\left[h(x_t), g(y_t)^2\right], \end{aligned}$$

because

$$\begin{aligned} \text{var}\left[h(x_t)g(y_t)\right] &= \text{E}\left[h(x_t)^2 g(y_t)^2\right] - M_h^2 M_g^2 \\ &= \text{cov}\left[h(x_t)^2, g(y_t)^2\right] + V_h V_g + M_h^2 V_g + M_g^2 V_h. \end{aligned}$$

■

Proof. [of Theorem 1] Specialization of Lemma 2 to the case $h(u) = \text{sign}(u)$ gives $V_h = 1 - m_x^2$, $M_h = m_x$, $\text{cov}\left[h(x_t)^2, g(y_t)^2\right] = 0$, $\text{cov}\left[h(x_t)^2, g(y_t)\right] = 0$, so

$$V^{g,h} = \left(1 - m_x^2\right) V_g - 2m_x \text{cov}\left[\text{sign}(x_t), g(y_t)^2\right] = V^g.$$

■

Proof. [of Corollary 1] (i) Substitute $g(u) = \text{sign}(u)$, then

$$V_g = \text{var}\left[\text{sign}(y_t)\right] = E\left[\text{sign}(y_t)^2\right] - E\left[\text{sign}(y_t)\right]^2 = 1 - m_y^2 = (1 - m_y)(1 + m_y).$$

Next,

$$\text{cov}\left[\text{sign}(x_t), \text{sign}(y_t)^2\right] = \text{cov}\left[\text{sign}(x_t), 1\right] = 0.$$

The desired result follows.

(ii) Substitute $g(u) = u$, then the desired result follows. ■

Proof. [of Theorem 2] By Lemma 2, the asymptotics of $\hat{\beta}_{g,h}$ under the null H_0^g is

$$\sqrt{T}\hat{\beta}_{g,h} \xrightarrow{d} \frac{\text{N}(0, V^{g,h})}{V_h}.$$

Under H_0^g, (3.2) is a single-period regression on stationary and mixing variables. From the regression theory, the t-ratio using White-corrected standard errors is asymptotically standard normal. ■

Proof. [of Theorem 3] By Lemma 2, under the null H_0^g, $\hat{\beta}_{h,g}$ converges in probability to zero, and

$$\sqrt{T}\hat{\beta}_{h,g} \xrightarrow{d} \frac{\mathrm{N}(0, V^{g,h})}{V_g}.$$

Similarly to the previous lemma, the White-corrected asymptotic variance estimate for $\hat{\beta}_{h,g}$ is

$$\hat{V}_{h,g} = \frac{T^{-1}\sum_t g(y_t)^2\hat{\theta}_t^2 + \left(T^{-1}\sum_t g(y_t)\right)^2 T^{-1}\sum_t \hat{\theta}_t^2 - 2\left(T^{-1}\sum_t g(y_t)\right)\left(T^{-1}\sum_t g(y_t)\hat{\theta}_t^2\right)}{\left(T^{-1}\sum_t g(y_t)^2 - \left(T^{-1}\sum_t g(y_t)\right)^2\right)^2}.$$

The denominator converges in probability to V_g^2. The numerator – to

$$E\left[g(y_t)^2\left(h(x_t) - M_h\right)^2\right] + M_g^2 V_h - 2M_g E\left[g(y_t)\left(h(x_t) - M_h\right)^2\right].$$

Straightforward computations yield that under H_0^g this is exactly $V^{g,h}$. Similarly, it is straightforward to show that when conventional standard errors are used, $\hat{V}_{h,g} \xrightarrow{p} V_h/V_g$.When a HAC asymptotic variance is used, the numerator in $\hat{V}_{h,g}$ contains in addition sample correlations that are zeros only when $h(x_t)$ has is a martingale difference. ■

B Appendix: power computations

Let $\zeta_t = g(y_t) - \delta(x_t)/\sqrt{T}$. We have $E[\zeta_t|\mathcal{I}_{t-1}] = 0$ under H_δ^g. Let us first look at the asymptotics of the normalized contrast $\sqrt{T}(A_{g,h} - B_{g,h})$. We have

$$\begin{aligned}\sqrt{T}A_{g,h} &= \frac{1}{\sqrt{T}}\sum_t h(x_t)g(y_t) = \frac{1}{\sqrt{T}}\sum_t h(x_t)\left(\frac{\delta(x_t)}{\sqrt{T}} + \zeta_t\right) \\ &\xrightarrow{d} \mathrm{N}(E[\delta(x_t)h(x_t)], E\left[h(x_t)^2\zeta_t^2\right]),\end{aligned}$$

$$\begin{aligned}\sqrt{T}B_{g,h} &= \sqrt{T}\left(\frac{1}{T}\sum_t h(x_t)\right)\left(\frac{1}{T}\sum_t g(y_t)\right) = \left(\frac{1}{T}\sum_t h(x_t)\right)\left(\frac{1}{T}\sum_t \delta(x_t) + \frac{1}{\sqrt{T}}\sum_t \zeta_t\right) \\ &\xrightarrow{d} E[h(x_t)]\,\mathrm{N}(E[\delta(x_t)], E\left[\zeta_t^2\right]),\end{aligned}$$

and

$$\sqrt{T}\begin{pmatrix} T^{-1}\sum_t h(x_t)g(y_t) \\ T^{-1}\sum_t g(y_t) \end{pmatrix} \xrightarrow{d} \mathrm{N}\left(\begin{pmatrix} E\left[\delta\left(x_t\right)h(x_t)\right] \\ E\left[\delta(x_t)\right] \end{pmatrix}, \begin{pmatrix} E\left[h(x_t)^2\zeta_t^2\right] & E\left[h(x_t)\zeta_t^2\right] \\ E\left[h(x_t)\zeta_t^2\right] & E\left[\zeta_t^2\right] \end{pmatrix}\right).$$

Together, these give

$$\sqrt{T}\left(A_{g,h}-B_{g,h}\right) \xrightarrow{d} \mathrm{N}(E\left[\delta\left(x_t\right)h(x_t)\right]-E\left[h(x_t)\right]E\left[\delta(x_t)\right], E\left[\left(h(x_t)-M_h\right)^2\zeta_t^2\right]).$$

Next, the pivotization factor:

$$\begin{aligned} V^{g,h} &= E\left[\left(h(x_t)-M_h\right)^2\right]E\left[\zeta_t^2\right]+E\left[h(x_t)^2\zeta_t^2\right]-E\left[h(x_t)^2\right]E\left[\zeta_t^2\right] \\ &\quad -2M_hE\left[h(x_t)\zeta_t^2\right]+2M_h^2E\left[\zeta_t^2\right] \\ &= E\left[\left(h(x_t)-M_h\right)^2\zeta_t^2\right]. \end{aligned}$$

Hence, in total

$$T \xrightarrow{d} \mathrm{N}\left(\frac{\mathrm{cov}\left[\delta\left(x_t\right), h(x_t)\right]}{\sqrt{E\left[\left(h(x_t)-M_h\right)^2\zeta_t^2\right]}}, 1\right).$$

The local power is maximized when the square of the noncentrality parameter is maximized. Denote $\sigma_{t-1}^2 = \mathrm{var}\left[g(y_t)|I_{t-1}\right]$. Then

$$\begin{aligned} \frac{\mathrm{cov}\left[\delta\left(x_t\right), h(x_t)\right]^2}{E\left[\left(h(x_t)-M_h\right)^2\zeta_t^2\right]} &= \frac{\mathrm{cov}\left[\frac{\delta\left(x_t\right)-M_\delta}{\sigma_{t-1}}, \left(h(x_t)-M_h\right)\sigma_{t-1}\right]^2}{E\left[\left(h(x_t)-M_h\right)^2\sigma_{t-1}^2\right]} \\ &\leq \frac{\mathrm{var}\left[\frac{\delta\left(x_t\right)-M_\delta}{\sigma_{t-1}}\right]\mathrm{var}\left[\left(h(x_t)-M_h\right)\sigma_{t-1}\right]}{E\left[\left(h(x_t)-M_h\right)^2\sigma_{t-1}^2\right]} \\ &= \mathrm{var}\left[\frac{\delta\left(x_t\right)-M_\delta}{\sigma_{t-1}}\right]. \end{aligned}$$

This bound does not depend on $h(x_t)$ and is attained when the Cauchy–Schwatz inequality binds, that is, when $\left(h(x_t)-M_h\right)\sigma_{t-1}$ and $\frac{\delta\left(x_t\right)-M_\delta}{\sigma_{t-1}}$ are linearly dependent. Since the means of both random variables are zero, this implies that the maximal local power is reached when $h(x_t)-M_h$ is proportional to $\frac{\delta\left(x_t\right)-M_\delta}{\sigma_{t-1}^2}$.