

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 928

April 2008

Interpreting Long-Horizon Estimates in Predictive Regressions

Erik Hjalmarsson

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at [www.federalreserve.gov/pubs/ifdp/](http://www.federalreserve.gov/pubs/ifdp/).

# Interpreting long-horizon estimates in predictive regressions

Erik Hjalmarsson\*

Division of International Finance

Federal Reserve Board, Mail Stop 20, Washington, DC 20551, USA

March 2008

## Abstract

This paper analyzes the asymptotic properties of long-horizon estimators under both the null hypothesis and an alternative of predictability. Asymptotically, under the null of no predictability, the long-run estimator is an increasing deterministic function of the short-run estimate and the forecasting horizon. Under the alternative of predictability, the conditional distribution of the long-run estimator, given the short-run estimate, is no longer degenerate and the expected pattern of coefficient estimates across horizons differs from that under the null. Importantly, however, under the alternative, highly endogenous regressors, such as the dividend-price ratio, tend to deviate much less than exogenous regressors, such as the short interest rate, from the pattern expected under the null, making it more difficult to distinguish between the null and the alternative.

*JEL classification:* C22, G1.

*Keywords:* Predictive regressions; Long-horizon regressions; Stock return predictability.

---

\*This paper has benefitted from comments by Lennart Hjalmarsson, Randi Hjalmarsson and Jonathan Wright, as well as an anonymous referee. Tel.: +1-202-452-2426; fax: +1-202-263-4850; email: erik.hjalmarsson@frb.gov. The views in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

# 1 Introduction

Long-run regressions were made popular by influential articles such as Fama and French (1988) and Campbell and Shiller (1988). When attempting to predict stock returns over longer horizons, often covering several years, rather than on a month-to-month basis, the evidence in favour of predictability generally appears much stronger. For instance, the estimated regression coefficients tend to increase almost linearly with the forecasting horizon. However, in a recent paper, Boudoukh, Richardson, and Whitelaw (2006) (BRW hereafter) show that this pattern is exactly the one to be expected in the *absence* of predictability. Although it is well known that the estimated slope coefficient will increase with the horizon when there *is* predictability, BRW appear to be the first ones to observe that the same is also true under the null hypothesis. They interpret their findings as a strong critique of the widespread belief that long-horizon regressions provide solid evidence of predictability (e.g. Cochrane, 2001).<sup>1</sup>

The aim of the current paper is to understand in more detail the properties of long-horizon estimates under an alternative of predictability. The practical purpose of this is to establish the pattern of estimated coefficients that may be expected both under the null and the alternative, across different forecasting horizons. Although the generally increasing pattern of coefficients in long-horizon regressions is well established, the exact asymptotic sampling properties of long-run estimators under an alternative of predictability are not previously well understood.

I derive the asymptotic distribution of the long-run OLS estimator, with overlapping observations, under the assumptions that the true data generating process is given by the standard linear predictive regression model and that the regressors are highly persistent variables. Under the alternative of predictability, the sampling properties of the long-run estimator are fundamentally different than under the null hypothesis, and the limiting distribution is highly non-standard. From a practical perspective, this result is of individual interest. It shows that confidence intervals for long-run estimates, based on inverting a test statistic that is valid under the null hypothesis, will not be correctly sized under the alternative, given the non-standard distribution.

The theoretical results allow for an exact characterization of the conditional distribution of the long-run estimator, given the short-run estimate. Under the null hypothesis, the long-run estimator is, asymptotically, completely determined once the short-run estimate is given. Importantly, however, this is not true under the alternative of predictability. In fact, the degree to which the long-run estimate can vary independently of the short-run estimate is determined by the degree of endogeneity of the regressors.<sup>2</sup> Long-run estimates for

---

<sup>1</sup>Earlier studies discussing other inferential issues in long-horizon regressions include Hansen and Hodrick (1980), Richardson and Stock (1989), Richardson and Smith (1991, 1994), Hodrick (1992), Goetzman and Jorion (1993), Nelson and Kim (1993), Richardson (1993), Kirby (1997), and Valkanov (2003).

<sup>2</sup>A predictive regressor is generally referred to as endogenous if the innovations to the returns are contemporaneously corre-

highly endogenous regressors, such as the dividend-price ratio, are also almost completely pinned down by the short-run estimate under the alternative hypothesis. On the other hand, for near exogenous regressors such as the short interest rate, the long-run estimator has a relatively large independent component given the short-run estimate.<sup>3</sup> Given the short-run estimate, under the alternative of predictability, the implications are that: (i) the highly endogenous regressor will have a more predictable pattern for long-horizons than the near exogenous one, and (ii) this pattern will closely resemble that under the null hypothesis.

These results provide an alternative interpretation of the empirical findings in BRW. BRW interpret the findings that the dividend-price ratio has a pattern very similar to that predicted under the null, whereas the short interest rate does not, as evidence against the predictive ability of the dividend-price ratio and in favor of the predictive ability of the short rate. Given the results in this paper, however, their findings could merely reflect the fact that there is more independent variation in the long-run estimates for fairly exogenous regressors.

## 2 Model and assumptions

Let  $r_{t+1}$  denote the one period stock return from  $t$  to  $t+1$  and let  $r_{t+q}(q) = \sum_{j=1}^q r_{t+j}$  be the corresponding  $q$ -period return from  $t$  to  $t+q$ . The standard long-run forecasting regression is specified as follows,

$$r_{t+q}(q) = \alpha_q + \beta_q x_t + u_{t+q}(q), \quad (1)$$

where long-run future returns are regressed onto a one period predictor,  $x_t$ . Let the OLS estimator of  $\beta_q$  in equation (1), using overlapping observations, be denoted by  $\hat{\beta}_q$ . The primary focus of interest will be the properties of  $\hat{\beta}_q$  for different values of  $q$ , and in particular the relationship between  $\hat{\beta}_1$  and  $\hat{\beta}_q$  for  $q > 1$ .

In order to formally analyze the sampling properties of  $\hat{\beta}_q$ , the data generating process for  $r_t$  and  $x_t$  need to be explicitly specified. Following Nelson and Kim (1993) and Campbell (2001), I assume that  $r_t$  and  $x_t$  satisfy:

$$r_{t+1} = \alpha + \beta x_t + u_{t+1}, \quad (2)$$

$$x_{t+1} = \gamma + \rho x_t + v_{t+1}. \quad (3)$$

---

lated with the innovations to the regressor. When the regressor is strictly stationary, such endogeneity has no impact on the properties of the estimator, but when the regressor is persistent in some manner, the properties of the estimator will be affected (e.g. Stambaugh, 1999).

<sup>3</sup>The contemporaneous correlation between the innovations to the returns and the innovations to the regressor determines the endogeneity of a predictive regressor (see previous footnote). Campbell and Yogo (2006) show that for valuation ratios, such as the dividend-price ratio, this correlation is large and often greater than 0.9 in absolute magnitude. For interest rate variables, however, the correlation is close to zero and, from the perspective of a predictive regression, these variables are thus nearly exogenous.

Thus, the one-period regression in equation (1) coincides with the true data generating process and for any horizon,  $x_t$  will be the optimal forecaster of returns given current information at time  $t$ .<sup>4</sup> To capture the near persistence found in most forecasting variables, such as interest rates or valuation ratios, it is further assumed that the auto-regressive root,  $\rho$ , is close to one in a local sense. In particular, it is assumed that  $\rho = 1 + c/T$ , where  $c$  is some finite parameter and  $T$  is the sample size with  $t = 1, \dots, T$ . This captures the near unit-root, or highly persistent, behavior of many predictor variables, but is less restrictive than a pure unit-root assumption. The near unit-root construction, where the autoregressive root drifts closer to unity as the sample size increases, is used as a tool to enable an asymptotic analysis where the persistence in the data remains large relative to the sample size, also as the sample size increases to infinity. That is, if  $\rho$  is treated as fixed and strictly less than unity, then as the sample size grows, the process  $x_t$  will behave as a strictly stationary process asymptotically and the standard first order asymptotic results will not provide a good guide to the actual small sample properties of the model. For  $\rho = 1$ , the usual unit-root asymptotics apply to the model, but this is clearly a restrictive assumption for most potential predictor variables. Instead, by letting  $\rho = 1 + c/T$ , the effects from the high persistence in the regressor will appear also in the asymptotic results, but without imposing the strict assumption of a unit root. Cavanagh et al. (1995), Lanne (2002), Valkanov (2003), Torous et al. (2004), and Campbell and Yogo (2006) all use similar models, with a near unit-root construct, to analyze the predictability of stock returns.

The error processes are assumed to satisfy a martingale difference sequence with finite fourth order moments. That is, let  $w_t = (u_t, v_t)'$  and  $\mathcal{F}_t = \{w_s | s \leq t\}$  be the filtration generated by  $w_t$ . Then  $E[w_t | \mathcal{F}_{t-1}] = 0$ ,  $E[w_t w_t'] = \Sigma = [(\sigma_{11}, \sigma_{12}), (\sigma_{21}, \sigma_{22})]$ ,  $\sup_t E[u_t^4] < \infty$ , and  $\sup_t E[v_t^4] < \infty$ .

By standard arguments,  $T^{-1/2} \sum_{t=1}^{[Tr]} w_t \Rightarrow B(r) = BM(\Sigma)(r)$ , where  $B(\cdot) = (B_1(\cdot), B_2(\cdot))'$  denotes a two dimensional Brownian motion and  $\Rightarrow$  denotes weak convergence of the associated probability measures. Further, as  $T \rightarrow \infty$ ,  $T^{-1/2} x_{[Tr]} \Rightarrow J_c(r) = \int_0^r e^{(r-s)c} dB_2(s)$  and an analogous result holds for the demeaned variables  $\underline{x}_t = x_t - T^{-1} \sum_{t=1}^n x_t$ , with the limiting process  $J_c$  replaced by  $\underline{J}_c = J_c - \int_0^1 J_c$  (Phillips, 1987, 1988). Let  $W_1(\cdot)$  and  $W_2(\cdot)$  be the standardized Brownian motions, with unit variance and correlation  $\delta = \sigma_{12}(\sigma_{11}\sigma_{22})^{-1/2}$ , that correspond to  $B_1(\cdot)$  and  $B_2(\cdot)$ , respectively. By the properties of conditional normal distributions,  $W_2 = \sqrt{1 - \delta^2} W_{2.1} + \delta W_1$ , where  $W_{2.1}$  is a Brownian motion with unit variance and orthogonal to  $W_1$ . Further, let  $J_c^W$  be the standardized version of  $J_c$ .

---

<sup>4</sup>There are two primary reasons why the analysis of the long-run regression in equation (1) is of interest under the assumption that the true model is given by the short-run equation (2). First, there is the long standing popular belief that predictability is more evident in the long run, and that there may therefore be power gains to analyzing the long-horizon regression, even if the short-run specification given by equation (2) is correct; for instance, Campbell (2001) analyzes the power of long-run tests under the same specification that is used in this paper. Alternatively, since there appears to be no other data generating processes that are widely used for modeling return predictability, in the short- or long-run, the results derived under the data generating process given by equations (2) and (3) can be viewed as a benchmark against which to compare results from other specifications.

To ease the notation, define

$$\xi_1 \equiv \left( \int_0^1 dW_1 \underline{J}_c^W \right) \left( \int_0^1 (\underline{J}_c^W)^2 \right)^{-1} \quad \text{and} \quad \xi_2 \equiv \left( \int_0^1 dW_2 \underline{J}_c^W \right) \left( \int_0^1 (\underline{J}_c^W)^2 \right)^{-1}, \quad (4)$$

and write  $\xi_2 = \sqrt{1 - \delta^2} \xi_{2.1} + \delta \xi_1$ , where  $\xi_{2.1} \equiv \left( \int_0^1 dW_{2.1} \underline{J}_c^W \right) \left( \int_0^1 (\underline{J}_c^W)^2 \right)^{-1}$ .

### 3 Asymptotic distributions under the null and the alternative

#### 3.1 The limiting distribution of $\hat{\beta}_q$

The foundations for the subsequent analysis is given in the following theorem, which outlines the asymptotic properties of  $\hat{\beta}_q$  under both the null hypothesis of no predictability and the alternative of predictability.

**Theorem 1** *Suppose the data are generated by equations (2) and (3).*

1. *Under the null hypothesis that  $\beta = 0$ , as  $T \rightarrow \infty$ ,*

$$T \left( \hat{\beta}_q - 0 \right) \Rightarrow q \left( \int_0^1 dB_1 \underline{J}_c \right) \left( \int_0^1 \underline{J}_c^2 \right)^{-1} = q \sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \xi_1. \quad (5)$$

2. *Under the alternative hypothesis that  $\beta \neq 0$ , as  $T \rightarrow \infty$ ,*

$$T \left( \hat{\beta}_q - \beta_q \right) \Rightarrow \left( q \int_0^1 dB_1 J_c + \beta \eta(\rho, q) \int_0^1 dB_2 J_c \right) \left( \int_0^1 J_c^2 \right)^{-1} = q \sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \xi_1 + \beta \eta(\rho, q) \xi_2, \quad (6)$$

where  $\beta_q \equiv \beta \nu(\rho, q)$  with  $\nu(\rho, q) = (1 + \rho + \dots + \rho^{q-1}) = q + O(T^{-1})$ , and  $\eta(\rho, q) \equiv \sum_{h=1}^{q-1} \sum_{p=h}^{q-1} \rho^{p-h} = \frac{q(q-1)}{2} + O(T^{-1})$ . For  $\rho = 1$ , it holds exactly that  $\nu(\rho, q) = q$ , and  $\eta(\rho, q) = \frac{q(q-1)}{2}$ .

**Remark 1.1** Note that the asymptotic distribution of  $\hat{\beta}_1$  is identical under the null and the alternative.

That is, for any value of  $\beta$ ,

$$T \left( \hat{\beta}_1 - \beta \right) \Rightarrow \sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \xi_1. \quad (7)$$

This follows both from standard asymptotic theory applied to the one-period OLS estimator, but also from plugging in  $q = 1$  in equations (5) and (6).

**Remark 1.2** Under the null hypothesis, the asymptotic distribution of  $\hat{\beta}_q$  is identical to that of  $\hat{\beta}_1$ , apart from a scaling factor. This result follows from the persistent nature of the regressors; the intuition behind it is discussed in more detail in Hjalmarsson (2007). For the purposes of this paper, the implications are that the short-run and long-run estimators are perfectly correlated asymptotically. In fact, from equation (5) it

follows that,

$$\hat{\beta}_q \sim \frac{q}{T} \sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \xi_1 \text{ and } \hat{\beta}_1 \sim \frac{1}{T} \sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \xi_1, \quad (8)$$

where  $\sim$  is used to denote an approximate distributional equivalence. Asymptotically, therefore, the conditional distribution of  $\hat{\beta}_q$  given  $\hat{\beta}_1$  satisfies  $\hat{\beta}_q | \hat{\beta}_1 = q\hat{\beta}_1$ . Given  $\hat{\beta}_1$ , the estimator  $\hat{\beta}_q$  is thus asymptotically a deterministic linear function of the forecasting horizon. This is similar to the result in BRW, which is derived under the assumption of a fixed autoregressive root  $\rho$  that is strictly less than unity.

**Remark 1.3** Under the alternative hypothesis, the long-run estimator converges to the parameter  $\beta_q = \beta\nu(\rho, q)$ , which for  $\rho = 1 + c/T$  implies that  $\beta_q = q\beta + O(T^{-1})$ . That is, the ‘true’ parameter value,  $\beta_q$ , as well as the estimator  $\hat{\beta}_q$ , grows approximately one-to-one with the forecasting horizon.

**Remark 1.4** Under the alternative hypothesis of predictability, the distribution of  $\hat{\beta}_q$  is quite different than under the null hypothesis. To understand the intuition behind this result, note first that the true model is given by equations (2) and (3). The long-run regression equation is thus a fitted regression, rather than the data generating process. As shown in Appendix A, the long-run returns  $r_{t+q}(q)$  actually satisfy the following relationship when ignoring the constant, derived from equations (2) and (3):

$$r_{t+q}(q) = \beta_q x_t + u_{t+q}(q) + \beta \sum_{h=1}^{q-1} \left( \sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h}. \quad (9)$$

There are now two error terms, the usual  $u_{t+q}(q)$  plus the additional term  $\beta \sum_{h=1}^{q-1} \left( \sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h}$ , which stems from the fact that at time  $t$  there is uncertainty regarding the path of  $x_{t+j}$  for  $j = 1, \dots, q-1$ . That is, since the true model is given by equations (2) and (3), there is uncertainty regarding both the future realizations of the returns as well as of the predictor variable when forming  $q$ -period ahead forecasts. The first error term,  $u_{t+q}(q)$ , corresponds to the asymptotic  $\xi_1$  term in the limiting distribution and the second error term,  $\beta \sum_{h=1}^{q-1} \left( \sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h}$ , corresponds to the  $\xi_2$  term. For large  $q$ , the second error term in (9) will clearly dominate the asymptotic properties since it is of an order of magnitude larger than the first one, a result reflected in the weights on  $\xi_1$  and  $\xi_2$  in equation (6). However, the weight on  $\xi_2$  in (6) also depends on  $\beta$ . Thus for  $\beta$  close to zero,  $\xi_1$  will still be important for relatively large  $q$ .

**Remark 1.5** Following the analysis in BRW, the results are derived under the assumption that  $q$  is fixed as  $T \rightarrow \infty$ . However, it is easy to show that the results remain similar if  $q$  increases with the sample size  $T$ , but at a slower pace, such that  $q/T \rightarrow 0$ , as  $q, T \rightarrow \infty$ . Under this assumption, it follows easily under the null hypothesis that  $\frac{T}{q} (\hat{\beta}_q - 0) \Rightarrow \sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \xi_1$ . Under the alternative hypothesis,  $\frac{T}{\eta(\rho, q)} (\hat{\beta}_q - \beta_q) =$

$\frac{T}{q} \left( \frac{\hat{\beta}_q}{q} - \frac{\beta_q}{q} \right) + o_p(1) \Rightarrow \beta \xi_2$ , where the first term in (6) now disappears asymptotically. As discussed in the previous remark, however, the first term in (6) will still be important for relatively large  $q$ , provided  $\beta$  is small; this can be achieved also for asymptotically large  $q$  by treating  $\beta$  as small in a local sense. All the results in this paper therefore hold also under the general assumption that  $q$  grows with the sample size but at a slower pace. As shown in Hjalmarrsson (2007), asymptotic results derived under this assumption seem to provide good approximations of the finite sample properties of  $\hat{\beta}_q$  for forecasting horizons spanning upwards of 15 to 20 percent of the sample size. For completeness, however, Appendix B presents the results for the case where  $q$  is asymptotically large relative to  $T$ , in a manner such that  $q/T = \lambda \in (0, 1)$ , as  $T \rightarrow \infty$ ; i.e. when  $q$  grows at the same pace as the sample size.

### 3.2 Finite sample adjustments

In the analysis of BRW, it follows that under the null hypothesis,  $\hat{\beta}_q \sim \nu(\rho, q) \hat{\beta}_1$  (see equation 6 in BRW), rather than  $\hat{\beta}_q \sim q \hat{\beta}_1$ , as found here, where  $\nu(\rho, q)$  is defined in Theorem 1.<sup>5</sup> However, under the current assumption of  $\rho = 1 + c/T$ , it follows that  $\nu(\rho, q) = q + O(T^{-1})$ . Thus, for the local-to-unity specification of  $\rho$  that is used here,  $\nu(\rho, q)$  and  $q$  are asymptotically indistinguishable, and replacing  $q$  by  $\nu(\rho, q)$  does not affect the asymptotic arguments but merely provides a finite sample adjustment. As the analysis of BRW implies, along with simulation results that are not reported here, the rate of growth of  $\hat{\beta}_q$  under the null hypothesis seems to correspond best to  $\nu(\rho, q)$ , rather than  $q$ , in finite samples.

Likewise, in Part 2 of Theorem 1, the factor  $q$  in front of  $\xi_1$  can be replaced by  $\nu(\rho, q)$ , since this multiplier arises in an identical manner to the one in Part 1. That is, under the alternative hypothesis, one can write,

$$T \left( \hat{\beta}_q - \beta_q \right) \Rightarrow \nu(\rho, q) \sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \xi_1 + \beta \eta(\rho, q) \xi_2. \quad (10)$$

In the analysis in the next section, I use these finite sample adjusted results. This does not qualitatively change any of the results, and for  $\rho = 1$  it holds exactly that  $q = \nu(\rho, q)$ .

## 4 The relationship between the long-run and the short-run

The results in the previous section provide the necessary building blocks for understanding the properties of, and relationship between, the long- and short-run estimators both under the null hypothesis and under the alternative of predictability. In this section, I consider the implications of these results through an informal

---

<sup>5</sup>I use the ‘ $\sim$ ’ sign here because in the framework of BRW, the asymptotic distribution of  $\hat{\beta}_q$  is not entirely pinned down by  $\hat{\beta}_1$ .



analysis. For ease of notation, it is assumed that  $\sigma_{11} = \sigma_{22} = 1$ .

Under both the null and the alternative, the short-run estimator satisfies,  $T(\hat{\beta}_1 - \beta_1) \Rightarrow \xi_1$ , and one can write informally,

$$\hat{\beta}_1 \sim \beta + \frac{1}{T}\xi_1. \quad (11)$$

Similarly, under the null with  $\beta = 0$ , the long-run estimator satisfies

$$\hat{\beta}_q \sim \nu(\rho, q) \frac{1}{T}\xi_1 = \nu(\rho, q) \hat{\beta}_1. \quad (12)$$

Thus, as noted above, under the null-hypothesis,  $\hat{\beta}_1$  and  $\hat{\beta}_q$  are perfectly asymptotically correlated.

Under the alternative of predictability,

$$\hat{\beta}_q \sim \beta_q + \frac{\nu(\rho, q)}{T}\xi_1 + \beta \frac{\eta(\rho, q)}{T}\xi_2 = \nu(\rho, q) \hat{\beta}_1 + \beta \frac{\eta(\rho, q)}{T} \left( \sqrt{1 - \delta^2} \xi_{2,1} + \delta \xi_1 \right). \quad (13)$$

The distribution of  $\hat{\beta}_q$  is now a function of  $\hat{\beta}_1$ , as well as an additional term. Note, however, that given  $\hat{\beta}_1$ , the random variable  $\xi_1$  is fixed, and the only independent information in  $\hat{\beta}_q$ , given  $\hat{\beta}_1$ , derives from the  $\xi_{2,1}$  variable.

To better understand the properties of  $\hat{\beta}_q$  under the alternative hypothesis, it is useful to consider the two special cases of  $\delta = 0$  and  $\delta$  close to  $-1$ . The case of  $\delta$  close to  $1$  will be symmetrical to that of  $\delta$  close to  $-1$ , but the latter is much more common in stock return applications. To more easily understand the variation in  $\hat{\beta}_q$ , Figure 1 shows the density plots for  $\xi_1$ , for different values of  $\delta$ , and  $\xi_2$  for the case of  $c = 0$  ( $\rho = 1$ ); the density of  $\xi_{2,1}$  is identical to that of  $\xi_2$ .

As is seen in Figure 1,  $\xi_2$ , and hence  $\xi_{2,1}$ , is almost always negative, a fact which will be used in the discussion below. To see this analytically, consider the case when  $c = 0$  and note that one can then write  $\xi_2 = \left( \frac{1}{2} \left( W_2(1)^2 - 1 \right) - W_2(1) \int_0^1 W_2(r) dr \right) \left( \int_0^1 W_2^2 \right)^{-1}$ . Since  $W_2(1)^2$  is distributed as a  $\chi_1^2$  variable, there is an approximately two-thirds probability that the first term in the numerator will be negative. The second term will also tend to be negative, since the correlation between  $W_2(1)$  and  $W_2(r)$  is positive. Further, when  $W_2(1)^2$  is large, the denominator will also be large, skewing the distribution further to the left. The ratio will therefore be negative most of the time and have a negative mean. A similar argument can be made for  $c \neq 0$ .

#### 4.1 The case of $\delta = 0$

When  $\delta = 0$ ,

$$\hat{\beta}_q \sim \nu(\rho, q) \hat{\beta}_1 + \beta \frac{\eta(\rho, q)}{T} \xi_{2,1}, \quad (14)$$

where the second term is stochastically independent of  $\hat{\beta}_1$ . Since  $E[\xi_{2,1}] < 0$ , as is apparent from Figure 1 and the discussion above,  $\hat{\beta}_q$  will tend to be below the curve  $\hat{\beta}_1\nu(\rho, q)$ ; the 5th, 50th and 95th percentile of  $\xi_{2,1}$ , for  $c = 0$ , are given by  $-14.05$ ,  $-4.37$ , and  $-0.13$ , respectively. To further understand the relationship between  $\hat{\beta}_1$  and  $\hat{\beta}_q$  in this case, consider a simple example. Suppose  $T = 500$ ,  $q = 50$ ,  $c = 0$ , and  $\hat{\beta}_1 = 0.015$ , which is a typical estimate of the short-run slope parameter in a regression with monthly standardized data such that  $\sigma_{11} = \sigma_{22} = 1$  (e.g. Campbell and Yogo, 2006). If the true value of  $\beta$  is equal to zero, then asymptotically,  $\hat{\beta}_q = \nu(\rho, q)\hat{\beta}_1 = 0.75$ . On the other hand, if  $\beta = 0.015$ , so that the short-run estimate is equal to the true value, then conditional on  $\hat{\beta}_1$ , the 5th, 50th and 95th percentiles of  $\hat{\beta}_q$  are equal to 0.234, 0.589, and 0.745, respectively, based on equation (14) and the percentiles of  $\xi_{2,1}$ .

## 4.2 The case of $\delta \approx -1$

As  $\delta \downarrow -1$ ,

$$\hat{\beta}_q \sim \nu(\rho, q)\hat{\beta}_1 - \beta \frac{\eta(\rho, q)}{T} \xi_1. \quad (15)$$

For  $\delta$  close to minus one,  $\xi_1$  is almost always positive, and  $\hat{\beta}_q$  will tend to be smaller than  $\nu(\rho, q)\hat{\beta}_1$ . Note also, that once  $\hat{\beta}_1$  is determined, there is no additional variance left in the estimator  $\hat{\beta}_q$ . That is, since  $\hat{\beta}_1 \sim \beta + \xi_1/T$ , for a given  $\hat{\beta}_1$  and  $\beta$ ,  $\xi_1$  is pinned down, and hence  $\hat{\beta}_q$  as well.

Consider a similar thought experiment to that above. Again, suppose  $T = 500$ ,  $q = 50$ ,  $c = 0$ , and  $\hat{\beta}_1 = 0.015$ . If  $\beta = 0$ , then  $\hat{\beta}_1 = \xi_1/T$ , which implies that  $\xi_1 = 7.5$  and  $\hat{\beta}_q = q\hat{\beta}_1 = 0.75$ . Now, if  $\beta = 0.01$ , then  $\hat{\beta}_1 = \beta + \xi_1/T$  implies that  $\xi_1 = 2.5$ , and  $\hat{\beta}_q = 0.689$ . If  $\beta = 0.015$ , then  $\xi_1 = 0$ , and  $\hat{\beta}_q = 0.75$ . To the extent that  $\beta$  is greater than or equal to zero, a large negative correlation  $\delta$  severely limits the range of probable values that  $\hat{\beta}_q$  can attain once  $\hat{\beta}_1$  is fixed. (The 5th, 50th and 95th percentile of  $\xi_1$  for  $\delta = -0.99$  and  $c = 0$ , are given by 0.10, 4.32, and 14.03, respectively.)

## 4.3 Implied long-horizon estimates

Thus, when the predictor is exogenous, so that  $\delta = 0$ , somewhat substantial deviations in  $\hat{\beta}_q$  from that predicted under the null are possible and to some extent expected. When the regressors are highly endogenous, and  $\delta$  is close to minus one, the range of possibilities also under the alternative is more restricted and large deviations from that predicted under the null are not likely.<sup>6</sup>

Figure 2 further illustrates this last point. Using equations (14) and (15), it plots potential outcomes of

<sup>6</sup> Apart from increasing slope coefficients, increasing  $R^2$ s have also been used as an argument in favor of long-run predictability. In Appendix C, I first replicate BRW's finding that  $R^2$  increases with the forecast horizon under the null hypothesis. In addition, I show that the asymptotic properties of  $R^2$  under the alternative hypothesis are not a function of the degree of endogeneity of the regressor and that  $R^2$  still increases almost linearly with the forecasting horizon. Thus, unlike for  $\hat{\beta}_q$ , there is no systematic difference in the asymptotic properties of  $R^2$  for exogenous and endogenous regressors under the alternative hypothesis.

$\hat{\beta}_q$ , given  $\hat{\beta}_1 = 0.015$ , for four different values of the true  $\beta = 0, 0.005, 0.010, 0.015$ . The same parameters as in the examples above are used, with  $T = 500$  and  $c = 0$ . For  $\delta \approx -1$ , once  $\hat{\beta}_1$  and  $\beta$  are fixed, the outcome of  $\hat{\beta}_q$  is fully determined and there is thus no range of possibilities. For  $\delta = 0$ , there is independent variation left in  $\hat{\beta}_q$ , given  $\hat{\beta}_1$  and  $\beta$ , in the form of  $\xi_{2,1}$ . The graphs for  $\delta = 0$  show the lower bound of  $\hat{\beta}_q$ , based on the 5th percentile of  $\xi_{2,1}$ . The upper bound, based on the 95th percentile of  $\xi_{2,1}$  is virtually identical to  $\nu(\rho, q)\hat{\beta}_1$ , since the 95th percentile of  $\xi_{2,1}$  is almost equal to zero.

The graphs clearly demonstrate the limited range of plausible outcomes for  $\hat{\beta}_q$  given a typical one-period estimate of  $\hat{\beta}_1$ , when the regressor is highly endogenous. Indeed, when the estimate  $\hat{\beta}_1$  is in fact identical to the true  $\beta$ , the outcome is observationally equivalent to that under the null hypothesis. When the predictor is exogenous, the range of outcomes is obviously much larger, and there is a fair chance of detecting patterns that deviate substantially from those expected under the null.

In their empirical analysis, BRW show that the coefficients for the dividend-price ratio, which is highly endogenous, are nearly linear in the forecasting horizon whereas those for the short interest rate, which is nearly exogenous, grow at a much slower pace. In light of Figure 2, these findings are suggestive of predictive ability in the short interest rate, but can say little or nothing regarding the predictive ability of the dividend-price ratio.<sup>7</sup>

## 5 Summary and conclusion

To sum up, under the null hypothesis, the long-run estimator is asymptotically completely determined by the one-period estimate and the persistence in the regressor. Under the alternative hypothesis, the degree to which the long-run estimates can vary independently of the one period ones is determined by the degree of endogeneity in the regressors. Nearly exogenous predictors, such as the short interest rate, allow for more independent variation than highly endogenous predictors such as the earnings-price ratio. Unfortunately, long-run estimates therefore provide additional information in cases where short-run inference is relatively straightforward but adds little in the case of endogenous regressors where short-run inference is fraught with difficulties (i.e. Stambaugh, 1999, and Campbell and Yogo, 2006).

Finally, it is worth pointing out, that the asymptotic framework used in this paper delivers an asymptotically degenerate distribution of  $\hat{\beta}_q$  given  $\hat{\beta}_1$ , under the null hypothesis. This prevents the construction

---

<sup>7</sup>The results in this paper are all based on the assumption that the regressor follows a near unit-root process. For a fixed autoregressive root  $\rho$  strictly less than unity, the effects arising from endogeneity would not appear in the asymptotic analysis, although one could perhaps obtain some similar results using the finite sample bias derived in Stambaugh (1999). In practice, of course, the near unit-root construction is designed to asymptotically capture the finite sample bias that arises from highly persistent and endogenous regressors, which are not necessarily unit-root processes. For regressors with very low persistence, i.e.  $\rho \ll 1$ , there will be no effects from endogeneity and there should thus be no systematic difference between endogenous and exogenous regressors in the behavior of either the short-run or the long-run estimators. However, most relevant predictors of stock returns tend to be highly persistent.

of formal, and asymptotically meaningful, tests on the joint distribution of  $\hat{\beta}_1$  and  $\hat{\beta}_q$  under the null. BRW devise a joint test of  $\beta_1 = \beta_2 = \dots = \beta_q = 0$  under the assumption of a fixed  $\rho$  strictly less than one. However, under these assumptions, the standard asymptotic distribution of the test is not likely to be well satisfied due to the standard complications in inference with endogenous and persistent variables. Monte Carlo simulations not reported in this paper confirm that for a large negative  $\delta$ , the BRW test will tend to severely over reject. It is also interesting to note that the Wald statistic of BRW is scaled by  $(1 - \rho)^{-1}$ , which diverges to infinity as  $\rho \rightarrow 1$ . The degenerate case encountered in the current asymptotics thus follows as a limiting case in their analysis. The construction of an asymptotically valid and correctly sized joint test of  $\beta_1$  and  $\beta_q$  for  $\rho$  close to unity is thus left unresolved.

## A Proof of Theorem 1

**Proof.** For ease of notation the case with no intercept is treated. The results generalize immediately to regressions with fitted intercepts by replacing all variables by their demeaned versions. Part 1 is proved in Hjalmarrsson (2007), but is repeated here for completeness.

1. Under the null hypothesis,

$$\frac{T}{q} \left( \hat{\beta}_q - 0 \right) = \left( \frac{1}{qT} \sum_{t=1}^{T-q} u_{t+q}(q) x_t \right) \left( \frac{1}{T^2} \sum_{t=1}^{T-q} x_t^2 \right)^{-1} = \left( \frac{1}{qT} \sum_{t=1}^{T-q} \sum_{j=1}^q u_{t+j} x_t \right) \left( \frac{1}{T^2} \sum_{t=1}^{T-q} x_t^2 \right)^{-1}.$$

By standard arguments,  $\frac{1}{qT} \sum_{t=1}^{T-q} \sum_{j=1}^q u_{t+j} x_t = \frac{1}{qT} \sum_{t=1}^{T-q} (u_{t+1} x_t + \dots + u_{t+q} x_t) \Rightarrow \int_0^1 dB_1 J_c$ , as  $T \rightarrow \infty$ , since for any  $h > 0$ ,  $\frac{1}{T} \sum_{t=1}^T u_{t+h} x_t \Rightarrow \int_0^1 dB_1 J_c$ . Therefore,  $T \left( \hat{\beta}_q - 0 \right) \Rightarrow q \left( \int_0^1 dB_1 J_c \right) \left( \int_0^1 J_c^2 \right)^{-1}$ .

2. By summing up on both sides in equation (2),

$$\begin{aligned} r_{t+q}(q) &= \beta (x_t + x_{t+1} + \dots + x_{t+q-1}) + u_{t+q}(q) \\ &= \beta \left( (x_t + \rho x_t + \dots + \rho^{q-1} x_t) + v_{t+1} + (\rho v_{t+1} + v_{t+2}) + \dots + \sum_{p=2}^q \rho^{q-p} v_{t+p-1} \right) + u_{t+q}(q) \\ &= \beta_q x_t + \beta \sum_{h=1}^{q-1} \left( \sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h} + u_{t+q}(q), \end{aligned}$$

where  $\beta_q = \beta (1 + \rho + \dots + \rho^{q-1}) = \beta \nu(\rho, q)$ . Thus,

$$r_{t+q}(q) = \beta_q x_t + \beta \sum_{h=1}^{q-1} \left( \sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h} + u_{t+q}(q) = \beta \nu(\rho, q) x_t + \beta \sum_{h=1}^{q-1} \left( \sum_{p=h}^{q-1} \rho^{p-h} \right) v_{t+h} + u_{t+q}(q),$$

and

$$T \left( \hat{\beta}_q - \beta_q \right) = \left( \beta \sum_{h=1}^{q-1} \left( \sum_{p=h}^{q-1} \rho^{p-h} \right) \frac{1}{T} \sum_{t=1}^T v_{t+h} x_t + \frac{1}{T} \sum_{t=1}^T u_{t+q}(q) x_t \right) \left( \frac{1}{T^2} \sum_{t=1}^T x_t^2 \right)^{-1}.$$

Observe that  $\frac{1}{T} \sum_{t=1}^T v_{t+h} x_t \Rightarrow \int_0^1 dB_2 J_c$  for all  $h$ . Let  $\eta(\rho, q) = \sum_{h=1}^{q-1} \sum_{p=h}^{q-1} \rho^{p-h}$  and it follows that, as  $T \rightarrow \infty$ ,  $\sum_{h=1}^{q-1} \left( \sum_{p=h}^{q-1} \rho^{p-h} \right) \frac{1}{T} \sum_{t=1}^T v_{t+h} x_t \Rightarrow \eta(\rho, q) \int_0^1 dB_2 J_c$ . By the results in Part 1.,  $\frac{1}{T} \sum_{t=1}^T u_{t+q}(q) x_t \Rightarrow q \int_0^1 dB_1 J_c$ , as  $T \rightarrow \infty$ , and the desired result follows. Note that,

$$\eta(\rho, q) = \sum_{h=1}^{q-1} \sum_{p=h}^{q-1} \left( 1 + \frac{c}{T} \right)^{p-h} = \sum_{h=1}^{q-1} \sum_{p=h}^{q-1} \left( 1 + \frac{c(p-h)}{T} \right) + O(T^{-2}) = \frac{1}{2} q(q-1) + O(T^{-1}).$$

■

## B Results for the case when $q/T = \lambda \in (0, 1)$ as $T \rightarrow \infty$

Some of the literature on long-horizon regressions has analyzed the case where  $q$  is asymptotically large relative to  $T$ , such that  $q/T = \lambda \in (0, 1)$ , as  $T \rightarrow \infty$ . In the context of this study, such asymptotics are less useful because the long-run OLS estimator will *not* converge to a properly defined long-run coefficient. Since the current focus is on the distribution of the long-run estimator conditional on the short-run estimator, it makes more sense to consider the case when the long-run estimator does converge. Nevertheless, it is still interesting to see if any of the results derived in the main text continue to hold under this assumption.

Again treating the case without an intercept, Valkanov (2003) shows that under the null hypothesis, with  $q/T = \lambda$  as  $T \rightarrow \infty$ ,  $\hat{\beta}_q \Rightarrow \left( \int_0^{1-\lambda} B_1(r; \lambda) J_c(r) \right) \left( \int_0^{1-\lambda} J_c^2 \right)^{-1}$ , where  $B_1(r; \lambda) \equiv B_1(r + \lambda) - B_1(r)$ . Under the alternative,  $\frac{\hat{\beta}_q}{T} \Rightarrow \beta \left( \int_0^{1-\lambda} J_c(r; \lambda) J_c(r) \right) \left( \int_0^{1-\lambda} J_c^2 \right)^{-1}$ , where  $J_c(r; \lambda) \equiv \int_r^{r+\lambda} J_c(r)$ . Thus,  $\hat{\beta}_q$  no longer converges to a constant and it is therefore not surprising that the strong connection between the asymptotic distributions for the short-run and long-run estimators is no longer apparent. Indeed, given the highly non-standard limiting distributions, it is difficult to get a grasp of the properties of  $\hat{\beta}_q$ . A very rough approximation, however, can provide some guidelines.

Note that  $B_1(r + \lambda) - B_1(r) = \int_r^{r+\lambda} dB_1(s) \approx \lambda dB_1(r)$  and  $\int_r^{r+\lambda} J_c(r) \approx \lambda J_c(r)$ . Under the null hypothesis, it follows that  $\hat{\beta}_q \sim \lambda \left( \int_0^{1-\lambda} dB_1 J_c \right) \left( \int_0^{1-\lambda} J_c^2 \right)^{-1} \approx \lambda T \hat{\beta}_1 = q \hat{\beta}_1$ , and under the alternative hypothesis,  $\hat{\beta}_q \sim T \lambda \beta = q \beta$ . Thus, also for  $q/T = \lambda$ , it would appear that  $\hat{\beta}_q$  will grow with the forecasting horizon. Under the null hypothesis, there is still some indication of the relationship between  $\hat{\beta}_q$  and  $\hat{\beta}_1$ , but under the alternative hypothesis the more subtle connections between the short-run and the long-run estimator are no longer evident, as might be expected given the lack of a consistent long-run estimator.

## C Asymptotic properties of $R^2$

Let  $R_q^2$  be the coefficient of determination from the  $q$ -period regression in equation (1). Using the same arguments as in the proof of Theorem 1, it follows that under the null hypothesis,  $TR_1^2 \Rightarrow \frac{1}{\sigma_{11}} \left( \int_0^1 dB_1 J_c \right)^2 \left( \int_0^1 J_c^2 \right)^{-1}$ . Similarly,  $TR_q^2 \Rightarrow \frac{q}{\sigma_{11}} \left( \int_0^1 dB_1 J_c \right)^2 \left( \int_0^1 J_c^2 \right)^{-1}$ , using the result that  $\frac{1}{qT} \sum_{t=1}^T u_{t+q}^2 \rightarrow_p \sigma_{11}$ , which is derived in Hjalmarsson (2007). Thus, asymptotically, under the null hypothesis,  $R_q^2 | R_1^2 = qR_1^2$ .

Because  $x_t$  is a near-integrated regressor, it follows easily that  $R_q^2 \rightarrow_p 1$  as  $T \rightarrow \infty$ . This is not very useful from a practical perspective and it is more interesting to analyze the properties under a local alternative,  $\beta = b/\sqrt{T}$ . Under this alternative, using similar arguments as before,  $R_1^2 \Rightarrow 1 - \frac{\sigma_{11}}{\sigma_{11} + b^2 \left( \int_0^1 J_c^2 \right)}$ , and  $R_q^2 \Rightarrow 1 - \frac{\sigma_{11}}{\sigma_{11} + qb^2 \left( \int_0^1 J_c^2 \right)}$ . Standardizing so that  $\sigma_{11} = 1$ , and using the approximation that  $(1+x)^{-1} \approx 1-x$ , it follows that  $R_1^2 \approx b^2 \left( \int_0^1 J_c^2 \right)$  and  $R_q^2 \approx qb^2 \left( \int_0^1 J_c^2 \right) \approx qR_1^2$ .

## References

- [1] Boudoukh J., M. Richardson, and R.F. Whitelaw, 2006. The myth of long-horizon predictability, *Review of Financial Studies*, forthcoming.
- [2] Campbell, J.Y., 2001. Why long horizons? A study of power against persistent alternatives, *Journal of Empirical Finance* 8, 459-491.
- [3] Campbell, J.Y., and R. Shiller, 1988. Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661-676.
- [4] Campbell, J.Y., and M. Yogo, 2006. Efficient Tests of Stock Return Predictability, *Journal of Financial Economics* 81, 27-60.
- [5] Cavanagh, C., G. Elliot, and J. Stock, 1995. Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131-1147.
- [6] Cochrane, J., 2001. *Asset Pricing*, Princeton, Princeton University Press.
- [7] Fama, E.F., and K.R. French, 1988. Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3-25.
- [8] Goetzman W.N., and P. Jorion, 1993. Testing the Predictive Power of Dividend Yields, *Journal of Finance* 48, 663-679.
- [9] Hansen, L.P., and R.J. Hodrick, 1980. Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis, *Journal of Political Economy* 88, 829-853.
- [10] Hjalmarsson, E., 2007. Inference in Long-Horizon Regressions, Working Paper, Federal Reserve Board.
- [11] Hodrick, R.J., 1992. Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Financial Studies* 5, 357-386.
- [12] Lanne, M., 2002. Testing the Predictability of Stock Returns, *Review of Economics and Statistics* 84, 407-415.
- [13] Kirby, C., 1997. Measuring the Predictable Variation in Stock and Bond Returns, *Review of Financial Studies* 10, 579-630.
- [14] Nelson, C.R., and M.J. Kim, 1993. Predictable Stock Returns: The Role of Small Sample Bias, *Journal of Finance* 48, 641-661.

- [15] Phillips, P.C.B, 1987. Towards a Unified Asymptotic Theory of Autoregression, *Biometrika* 74, 535-547.
- [16] Phillips, P.C.B, 1988. Regression Theory for Near-Integrated Time Series, *Econometrica* 56, 1021-1043.
- [17] Richardson, M., 1993. Temporary Components of Stock Prices: A Skeptic's View, *Journal of Business and Economics Statistics* 11, 199-207.
- [18] Richardson, M., and T. Smith, 1991. Tests of Financial Models in the Presence of Overlapping Observations, *Review of Financial Studies* 4, 227-254.
- [19] Richardson, M., and T. Smith, 1994. A Unified Approach to Testing for Serial Correlation in Stock Returns, *Journal of Business* 67, 371-399.
- [20] Richardson, M., and J.H. Stock, 1989. Drawing Inferences from Statistics Based on Multiyear Asset Returns, *Journal of Financial Economics* 25, 323-348.
- [21] Stambaugh, R., 1999. Predictive regressions, *Journal of Financial Economics* 54, 375-421.
- [22] Torous, W., R. Valkanov, and S. Yan, 2004. On Predicting Stock Returns with Nearly Integrated Explanatory Variables, *Journal of Business* 77, 937-966.
- [23] Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications, *Journal of Financial Economics* 68, 201-232.



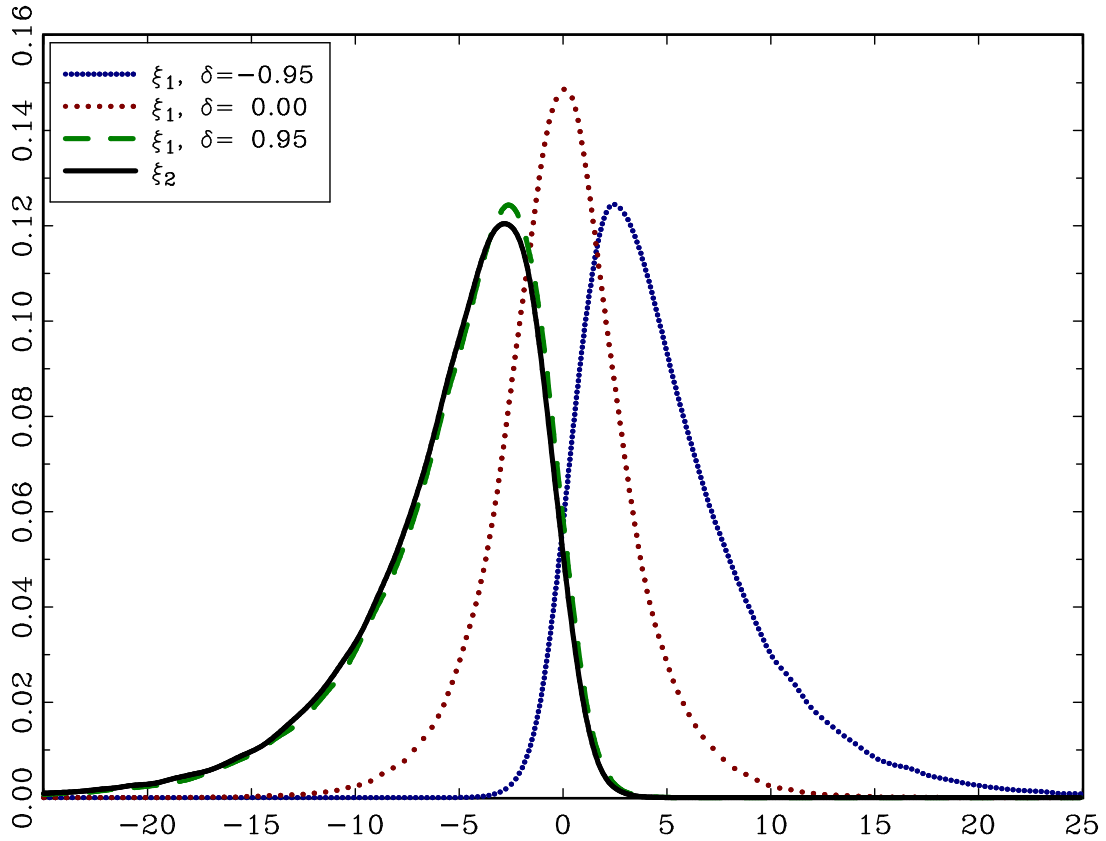


Figure 1: Density plots of  $\xi_1$  and  $\xi_2$  for  $c = 0$ . The graphs show the densities for the Brownian functionals  $\xi_1$ , for different values of  $\delta$ , and  $\xi_2$ , obtained by kernel estimation of simulated data using 100,000 repetitions and a sample size of 500 in each repetition. The shape of the density of  $\xi_{2,1}$  is identical to that of  $\xi_2$ .

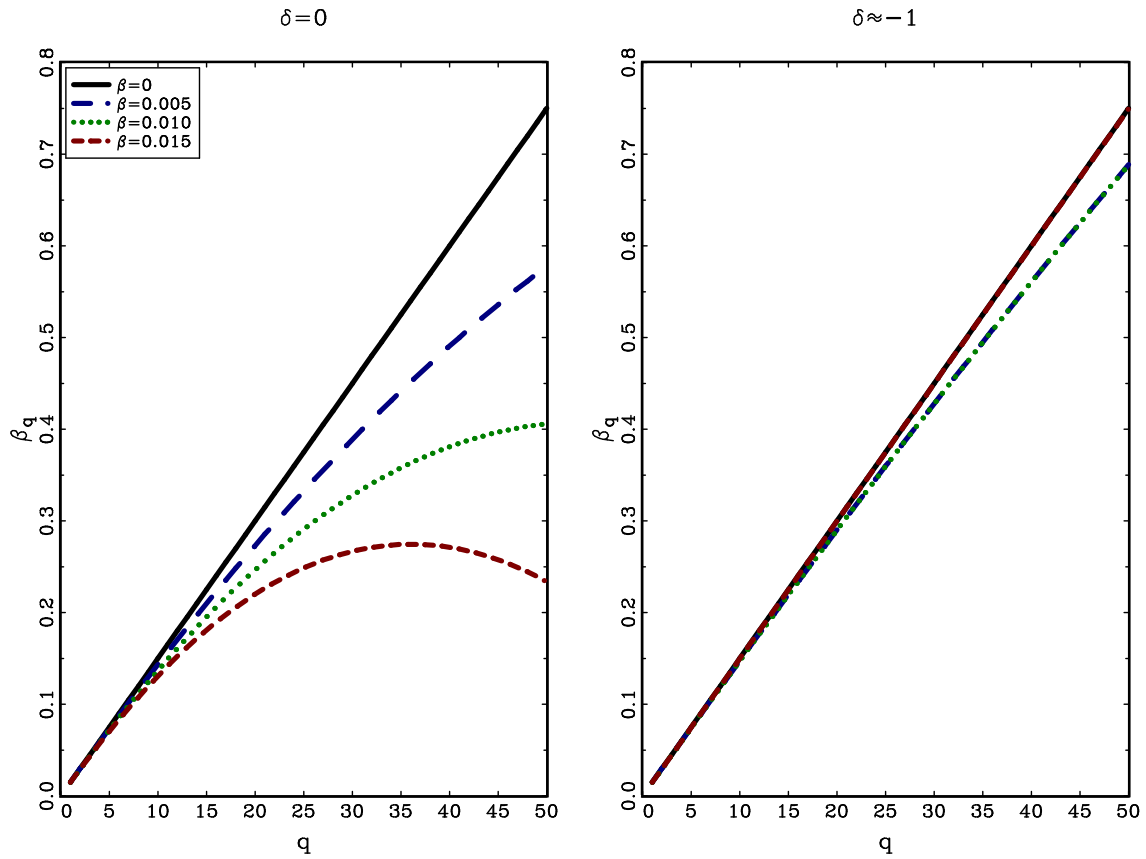


Figure 2: Possible outcomes of  $\hat{\beta}_q$  given  $\hat{\beta}_1$  and  $\beta$ . The graphs show potential outcomes of  $\hat{\beta}_q$ , given that  $\hat{\beta}_1 = 0.015$ , for  $\beta = 0, 0.005, 0.01, 0.015$ . The left hand side panel shows the case for exogenous regressors with  $\delta = 0$  and the right hand side shows the case with highly endogenous regressors. The plots are formed using equations (14) and (15), letting  $T = 500$  and  $c = 0$ . For  $\delta \approx -1$ , once  $\hat{\beta}_1$  and  $\beta$  are fixed, the outcome of  $\hat{\beta}_q$  is fully determined and there is thus no range of possibilities. The graphs for  $\delta = 0$  show the lower bound of  $\hat{\beta}_q$ , based on the 5th percentile of  $\xi_{2,1}$ . The upper bound, based on the 95th percentile of  $\xi_{2,1}$ , is virtually identical to  $\nu(\rho, q)\hat{\beta}_1$ ; i.e. the line for  $\beta = 0$ .