Understanding Spurious Regression in Financial Economics*

Ai Deng[†]

Boston University

First version: March, 2005. This Version: Dec, 2005.

Abstract

This paper provides an asymptotic theory for the spurious regression analyzed by Ferson, Sarkissian and Simin (2003). The asymptotic framework developed by Nabeya and Perron (1994) is used to provide approximations for the various estimates and statistics. Also, using a fixed-bandwidth asymptotic framework, a convergent t test is constructed, following Sun (2005). These are shown to be accurate and to explain the simulation findings in Ferson et al. (2003). Monte Carlo studies show that our asymptotic distribution provides a very good finite sample approximation for sample sizes often encountered in finance. Our analysis also reveals an important potential problem in the theoretical hypothesis testing literature on predictability. A possible reconciling interpretation is provided.

Keywords: spurious regression, observational equivalence, Nabeya-Perron asymptotics, fixed-b asymptotics, data mining, nearly integrated, nearly white noise (NINW)

*This is a subtantially revised version of the previously circulated paper with the same title. I thank Professor Timothy T. Simin for providing the data set and Professor Eiji Kurozumi as well as two anonymous referees for comments on earlier draft. Thanks also go to the econometrics seminar participants at Boston University. My advisor Professor Pierre Perron encouraged me to finish the paper and gave me constructive comments throughout the project. My advisor Professor Guido Kuersteiner also helped me to improve the exposition. However, all remaining errors are mine.

[†]Department of Economics, Boston University, Boston, 02215, USA. Email: dengai@bu.edu.

1 Introduction

A recent paper by Ferson, Sarkissian and Simin (2003a), hereafter FSS, points out the possibility of spurious rejection in financial predictive regression inference, or a type II spurious predictive regression as defined in Chiarella and Gao (2002) where type I and type II spurious regressions are defined as "spurious acceptance" and "spurious rejection", respectively. The authors use extensive simulations to show that if the expected (excess) return follows a persistent process and the so-called true R^2, to be defined in next section is not extremely low, then spurious regression bias will emerge in finite samples so that one spuriously rejects the hypothesis of no predictability too often. The key mechanism in their study is that stock returns can be decomposed into a persistent component (expected return) and a large noise component. In an empirical investigation, they show the t statistic and regression R^2 from some previous studies that would be significant by standard criteria are no longer significant. This research has drawn the financial profession's attention, see for example, among others, Torous, Valkanov and Yan (2005), and Amihud and Hurvich (2004).

FSS comment that their finding of spurious regression in financial economics is "well outside the classic setting of Yule (1926) and Granger and Newbold (1974)"[1], in part because the dependent variable in predictive regression, i.e. the stock returns "are much less persistent than the levels of most economic time series." They remarked "even though stock returns are not highly autocorrelated ... thus one may think that spurious regression problems are unlikely. However ... there is a spurious regression bias...". Indeed, as will be further discussed below, the classic spurious regression theory developed in Phillips (1986, 1998) is not able to provide a satisfactory explanation for FSS's finite sample results.

The purpose of this paper is to provide an asymptotic framework in which FSS type spurious regression can be studied analytically. As in their paper, this paper studies only the short horizon predictive regression. Such an analytical framework is important in understanding the underlying generating mechanism of spurious bias. It allows us to find out "what exactly goes wrong" in the FSS Monte Carlo set-up. It is also important if we would like to generalize their finite sample results and provide guidance on when one should be alerted. In fact, our analytical analysis not only shows that many of FSS's findings are quite predictable with our proposed asymptotics, but also reveals other important implications for the predictive regression inference literature.

The motivation of our proposed framework comes from the $ARMA(1,1)$ time series

[1]In Granger and Newbold's (1974) study it is shown that regression between two independent I (1) processes would result in spurious statistical significance of t statistic.

representation of stock returns under FSS's assumptions. It will be shown analytically, under the assumptions of persistent expected returns and large noise component in (observed) stock returns, this $ARMA(1,1)$ representation has both the AR root and the MA root close to unity. This observation immediately implies that the classic spurious regression theory in which the dependent variable is characterized by a (nearly) integrated process asymptotically is not expected to deliver useful finite sample implications. However, Nabeya and Perron (1994)'s (hereafter NP) nearly white noise, nearly integrated asymptotic framework seems particularly relevant in this case. Specifically, their asymptotics postulate the following $ARMA(1,1)$ process,

$$\begin{aligned} y_t &= (1 + c/T)\, y_{t-1} + u_t, \\ u_t &= e_t + \gamma_T e_{t-1}, \\ \gamma_T &= -1 + \delta/\sqrt{T} \end{aligned}$$

where the rate of which the AR and MA coefficients approach the boundary values are specified so that non-degenerate asymptotic theory can be obtained. As is obvious, in this framework, both AR and MA roots are local to 1, hence this representation mimics the behavior of return process even asymptotically. To see more of the intuition behind this framework, we note y_t can be equivalently written as

$$y_t = a_T e_t + b_T X_t$$

where $a_T = \left(1 - \delta T^{-1/2}\right) \exp\left(-c/T\right) \to 1, b_T = 1 - \exp\left(-c/T\right)\left(1 - \delta T^{-1/2}\right)$, so $T^{1/2} b_T \to \delta$ and $X_t = \sum_{j=1}^{t} \exp\left(\left(t-j\right) c/T\right) e_{t-j}$ is a near-integrated process. This representation is interesting because it decomposes stock returns into two components, one of which is the noise component $a_T e_t$, and the other is persistent component $b_T X_t$, completely analogous to the model in FSS. Moreover, the fact that $b_T = O\left(\frac{1}{\sqrt{T}}\right)$ implies that as time accumulates, the persistent component X_t will not dominate in stock returns. I think of this implication as a specially appealing property of this asymptotic framework since it is both theoretically and empirically unreasonable that stock returns behave more and more like nearly integrated series as is true when $b_T X$ is unbounded. The usefulness of this framework is confirmed in our simulation studies. Notice this representation is interestingly related to Torous and Valkanov (2000) where they also proposed an asymptotic framework that captures the large noise component in stock returns. A comparison between our results and the results under their asymptotics is the author's ongoing project.

With the NP asymptotics, the asymptotic distributions of regression coefficients and various statistics of interest are derived. Simulation shows the asymptotics provide very good finite sample approximations. Among other results, I show the R^2 has a non-degenerate limit distribution, implying it is an inadequate measure of goodness of fit, and also the simple t statistic diverges at rate $\sqrt{T}$, which is the same rate of divergence one would get from classic spurious regression. The divergent t statistic has important implications for the predictive regression literature because many popular inference theories rely on a convergent (simple) t statistics, the limiting distribution of which involves Dickey Fuller distribution from the test of unit root in the predicting variable. This result hence has the following implication: suppose the stock returns are indeed predictable by some observable highly persistent variable, then, caution must be taken if another irrelevant highly persistent variable is used in hypothesis testing using the currently available testing procedures because a rejection of null could be spurious, a point alluded by FSS (2003). [2]

The rest of the paper is organized as follows. Section 2 provides a brief summary of FSS (2003) findings. Section 3.1 presents easy-to-derive analytical results for their model. Section 3.2 derives asymptotic theory for spurious regression in this setting. Section 4 reports the simulation results concerning the finite sample approximation of various asymptotic distributions as well as the behavior of a convergent t test obtained by employing the fixed-b asymptotics, namely, by making the bandwidth M in HAC variance estimator as a fixed proportion of the sample size T. Discussion about the interpretation of FSS's inference as well as other existing procedures in predictive regression is contained in Section 6. Section 7 concludes. Proofs are relegated to a technical appendix.

2 Simulation results from FSS (2003a, b)

FSS's (2003a, b) simulation investigation is based on the following standard model of stock return. Let r_{t+1} be the future stock return, Z_t^* be the unobserved latent variable in DGP which is interpreted as the (unobserved) expected stock return. The DGP for future stock return from FSS (2003a) is[3]

$$r_{t+1} = \mu + Z_t^* + u_{t+1} \tag{A}$$

[2]We note that Valkanov (2003) obtained similar asymptotic results for t statistic and R^2 in the long horizon regression framework. His insight is that with long horizon regressions, assuming the dependent or independent variables are sums over nontrivial proportion of the whole sample, these variables behave as integrated processes. His theoretical results, however, do not take into account the large noise component in the stock returns and do not extend to the short horizon regression case here.

[3]Cochrane's textbook (2001) presents a very similar model and obtain a result similar to our lemma 1 below. And this model can be well thought of as a unobserved component model.

[4]where u_{t+1} is mean zero white noise with variance σ_u^2. The predictive regression, in stead, is obtained by regressing r_{t+1} on a constant and a lagged (observed) predictor variable Z_t, that is

$$r_{t+1} = \hat{\alpha} + \hat{\beta} Z_t + \hat{v}_{t+1}$$

They also assume the DGP for Z_t and Z_t^* to be[5]

$$\begin{pmatrix} Z_t^* \\ Z_t \end{pmatrix} = \begin{pmatrix} \rho^* & 0 \\ 0 & \rho \end{pmatrix} \begin{pmatrix} Z_{t-1}^* \\ Z_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^* \\ \varepsilon_t \end{pmatrix}, Var\left(\varepsilon_t^*\right) = \sigma_*^2, \varepsilon_t^* \perp \varepsilon_t$$

In words, they are considering a situation where researchers, in hope for predicting stock returns, come up with a completely irrelevant predictor Z_t, which is highly autocorrelated when ρ is close to 1. Naturally, what a researcher wants from performing a significance test on regression parameter $\hat{\beta}$ is an insignificant statistic (irrespective of the degree of autocorrelation in Z_t). However, just like the classic spurious regression problem demonstrated by Granger and Newbold (1974), FSS show using simulation the spurious significance of t statistic in the above described financial economics context, when both ρ^* and ρ are *close* to 1. What they further consider is the interaction between data mining and spurious regression bias, which I describe below.

Different from the original work of Granger and Newbold (1974), FSS assume their hypothetical analyst uses the popular HAC t statistic based on Newey-West standard error estimator when examining the statistical significance of the slope estimate, i.e.

$$t^{HAC} = \frac{\hat{\beta} - \beta}{\hat{S}}$$

where the ordinary least squares estimate of β is given by

$$\hat{\beta} = \frac{\sum_{t=1}^{T} \left(Z_t - \bar{Z}\right)\left(r_{t+1} - \bar{r}\right)}{\sum_{t=1}^{T} \left(Z_t - \bar{Z}\right)^2}$$

where $\bar{Z} = \sum_{t=1}^{T} Z_t / T$ and $\bar{r} = \sum_{t=1}^{T} r_t / T$ and the Newey West estimator $\hat{S}$ is given by

$$\hat{S}^2 = T\hat{\Omega} \left(\sum_{t=1}^{T} \left(Z_t - \bar{Z}\right)^2 \right)^{-2}$$

[4]The intercept term in FSS calibration is 0.0029, very close to 0. So in our theoretical development, we will assume a zero mean in the DGP for the dependent variable.

[5]As mentioned above, the same model has been proposed in Conrad and Kaul (1988).

where

$$\hat{\Omega} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{M}\right)\hat{\Gamma}(j),$$

$$\hat{\Gamma}(j) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-j}\left(Z_{t+j}-\bar{Z}\right)\hat{v}_{t+j}\hat{v}_t\left(Z_t-\bar{Z}\right) & \text{for } j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T}\left(Z_{t+j}-\bar{Z}\right)\hat{v}_{t+j}\hat{v}_t\left(Z_t-\bar{Z}\right) & \text{for } j < 0 \end{cases}$$

and $k(.)$ is the kernel function and M is the bandwidth. Notice, to get a consistent estimator of the HAC standard error, one necessary condition is $M \to \infty$, $M/T \to 0$ as $T \to \infty$. In case of Newey-West estimator, the kernel function is as follows,

$$k(x) = \begin{cases} 1-|x| & \text{for } |x| \leq 1, \\ 0 & \text{for } |x| > 1, \end{cases}$$

In most of FSS's simulation, the truncation lag M in the above variance estimator is selected based on the number of statistically significant residual autocorrelations. FSS's (2003) main findings are summarized as follows.

1. the OLS estimate of β is "well-behaved". They found $\hat{\beta}$ are all very close to 0.

2. HAC t test is spuriously biased towards rejection and the magnitude of spurious regression bias in HAC t depends on several parameters: ρ, ρ^* and true R^2, defined as

$$\text{true } R^2 = \frac{Var(Z^*)}{Var(Z^*) + \sigma_u^2}$$

   Recalling the definition of Z^* and σ_u^2, this quantity is interpreted as the measure of fit if one actually observes the true underlying expected return. *They find as ρ and ρ^**[6] *get closer to 1 and true R^2 gets larger, the HAC t test is more and more biased. And fixing ρ and ρ^* at values close to 1, the bias is increasing in magnitude of true R^2, and likewise, fixing a true R^2, the bias is increasing in ρ and ρ^*.*

3. The (first order) regression residual autocorrelation is not highly inflated.

4. In large sample size, a large number of lags in construction of Newey-West variance estimator "can" control the spurious regression problem. Of particular interest is a simulation reported in their footnote 7. with sample size $T = 5000$, their choice of

[6] In most of their simulations, they set ρ and ρ^* equal.

truncation lag $M = 240$ in the construction of HAC estimator yields a well-behaved t statistic 2.23, very close to 1.96. But they do not recommend this because of the arbitrariness and lack of theoretical support involved in this construction.

5. The data mining and spurious regression reenforce each other. They show again through simulation in a set of to-be-mined predicting instruments, more persistent variable is more likely to be chosen based on largest R^2 criterion. Hence, when analyst has searched among many potential regressors for one that produces a largest R^2 in the predicting regression, he or she is more likely to run into the problem of spurious regression bias.

The rest of the paper will address all these findings. Specifically, an asymptotic theory which could explain all these results will be developed For result 4 above, our simulation reveals an interesting explanation.

3 Theory

3.1 Analytical results and observational equivalence

In this section, two simple lemmas are derived directly from FSS's stock return model. And then based on these two lemmas, it is shown how the quantitative finite sample prediction from the classic asymptotic theory is not adequate, hence there is a need for an alternative theory.

The first lemma shows that this setup implies (obviously) an ARMA(1,1) representation for the stock returns, which is a well-known result. Following footnote 8, I will set $\mu = 0$ from now on.

Lemma 1 *The model specified above implies the following ARMA(1,1)*[7] *representation for* r_{t+1}

$$r_{t+1} = \rho^* r_t + \eta_{t+1} + \theta\eta_t$$

[7] An ARMA(1,1) representation for stock return has a long history in financial economics. Fama and French (1988) presents a stock price model which implies the return has an ARMA(1,1) representation. See also the discussion in Perron and Vodounou (1998). Cochrane (2001) also derives a similar univarite representation for stock return. However, this model has been challenged by many others recently. Shively (2000) and Khil and Lee (2002) propose an ARMA(2,2) representation for stock returns which captures the empirical finding of positive short-horizon autocorrelation and negative long-horizon autocorrelation.

where η_t is serially uncorrelated and the variance of which σ_η^2 and θ satisfy the following system, imposing an invertability condition,

$$\begin{cases} \sigma_*^2 + (1+\rho_*^2)\,\sigma_u^2 = \left(1+\theta^2\right)\sigma_\eta^2 \\ \theta = -\frac{\rho^*\sigma_u^2}{\sigma_\eta^2}, |\theta| < 1 \end{cases} \tag{1}$$

P roof. The proof is simple by observational equivalence argument, hence omitted. ■

The lemma first says stock return will inherit the first order autoregressive persistence from the expected stock return Z^*. If Z^* is highly persistent as FSS conjectures, so is the 1st order AR persistence of stock returns. However, given the strict positivity of persistence parameter ρ^* and σ_u^2, the MA coefficient in the ARMA representation is strictly negative. θ hence generates cancelling effect on the observational persistence of stock returns, which is closely related to the well-known (nearly) observational equivalence in unit root testing. See Campbell and Perron (1991) for an excellent non-technical introduction. To match the weak autocorrelation feature of stock returns, ρ^* and $|\theta|$ must be close to each other. And the ARMA representation is then said to be observationally equivalent (with respect to the covariance structure) to the observed stock returns behavior. FSS note that the difference between their model and the "classic" spurious regression as studied in Phillips (1986) is that they allow nonzero and possibly large variance of u_t in order to "accommodate the large noise component of stock returns". Equivalently, As shown above, nonzero variance σ_u^2 is needed to accommodate the fact that the observed stock return is (observationally) not first order persistent.

Lemma 2 $\frac{\sigma_u^2}{\sigma_\eta^2} \uparrow 1$ *as* $\sigma_u^2 \to \infty$

P roof. The proof follows from straightforward manipulation of system (1) and hence omitted. ■

Notice this lemma further implies that $|\theta| \uparrow \rho^*$ as $\sigma_u^2 \to \infty$. Hence, the larger the σ_u^2, the stronger the cancellation between θ and ρ^* will be. And in the limit, the stock return r_{t+1} will be serially uncorrelated. The intuition behind this lemma is simple. As the variability of noise term becomes dominant in returns ($\sigma_u^2 \to \infty$), the return itself will behave more like "white noise"[8]. Notice the connection with the true R^2 defined in FSS (2003a, 2003b). In their setup, $Var\left(Z^*\right)$ is a fixed constant[9], therefore, controlling true R^2 as FSS (2003a, 2003b)

[8]See also Campbell (2001).

[9]They set it to equal the sample variance of the S&P 500 return, in excess of a one-month Treasury bill return, multiplied by 0.10.

did is equivalent to controlling σ_u^2 and hence to controlling the observational persistence of r_{t+1}. To get a sense of the relative magnitude of ρ^* and the implied θ. Table I reports some of the θ values corresponding to various ρ^* and the true R^2 used in FSS (2003a, b). As we can see from Table I, the behavior of θ is exactly the same as described above.

[*Insert Table I Here*]

With the two lemmas, it should be noted that the asymptotic theory of Phillips (1986, 1988) is not appropriate in our situation, primarily because the cancelling effect between the AR and MA roots in the postulated stock return process. A prominent example where the classic theory fails to produce useful small sample prediction is that the highly autocorrelated regression residuals, prescribed by Phillips' theory is not at all observed in FSS (2003).

3.2 Alternative asymptotic theory: nearly integrated, nearly white noise

One of the reasons that Phillips' (1986) asymptotics is not expect to be adequate in explaining FSS's findings is because his asymptotic framework does not capture the fact that dependent variable (the stock returns) behaves like nearly white noise, but has strong AR persistence, with cancelling MA persistence. However, this feature fits well in the asymptotic framework of nearly white noise developed in Nabeya and Perron (1994), see also Perron and Ng (1996). Recall, the asymptotics are based on the following local-to-unity specification. Assuming $\{e_t\}$ to be *i.i.d.* $(0, \sigma_e^2)$,

$$\begin{aligned} y_t &= (1 + c/T)\, y_{t-1} + w_t, \\ w_t &= e_t + \gamma_T e_{t-1}, \\ \gamma_T &= -1 + \delta/\sqrt{T} \end{aligned}$$

y_t is then nearly integrated in finite sample, but a white noise in the limit. The focus here is the issue of spurious regression in this framework. Specifically, I develop an asymptotic theory when a nearly white noise process y_t is regressed on an independent (nearly) integrated process x_t and an intercept term.

Remark 1 *With the dependent variable, the stock returns, decomposed into persistent and noise components, it seems possible to consider the asymptotic framework whereby the regressand is simply a nearly integrated process plus white noise. But this is not expected to deliver satisfactory finite sample approximation because it is not capable of capturing the relative variance between the white noise error term and the nearly integrated expected returns. In* *fact, in such an asymptotic framework, nearly integrated component will always dominate in the limit, implying stock return behaves more and more like an integrated process. Our nearly white noise asymptotics avoid such undesirable feature, and keep the (nearly) stationary behavior of stock return in the limit. Torous and Valkanov (2000) propose a simple modification to this framework, that has close relationship with ours. Details can be found in Deng (2005).*

To specify the appropriate model for the regressor, recall the alternative representation of stock returns,

$$y_t = a_T e_t + b_T X_t$$

Following the simulation setup of FSS, it is natural to set the regressor x to be $b_T \widetilde{X}_t$ where $\widetilde{X}_t$ is an independent (nearly) integrated process generated by i.i.d. variable e_t^* with variance $\sigma_{e^*}^2$, say, and b_T as defined for y_t. The behavior of several statistics and estimates are studied. Results are collected in the following theorem. Notice, importantly, the qualitatively same results can be easily obtained when the regressor x_t is exactly integrated. In what follows, let $e_\infty(r) = \lim_{T\to\infty} e_{[Tr]}/\sigma_e$. And only the result on residual first order autocorrelation will be presented, but higher order autocorrelations can be easily obtained following the proof in the Appendix. Standard notations are used throughout. $\Rightarrow$ signifies weak convergence of probability measure. $W(t)$ and $V(t)$ are independent Wiener processes and $W_{c_x}(t)$ and $V_{c_y}(t)$ are independent diffusion processes associated with regressor x_t and regressand y_t respectively and $W_{c_x}(t) = \int_0^t \exp(c(r-s))\, dW(s), V_{c_y}(t) = \int_0^t \exp(c(r-s))\, dV(s)$. Whenever no confusion should be caused, $\int W_c$ denotes $\int_0^1 W_c(r)\, dr$ and $\int W_c^2$ denotes $\int W_c^2$.

Theorem 1 *1. (slope coefficient)*

$$\hat{\beta} \Rightarrow \frac{\sigma_{e^*}\sigma_e \left\{ \int_0^1 V_c W_{c_x} - \int_0^1 V_c \int_0^1 W_{c_x} \right\}}{\sigma_{e^*}^2 \left\{ \int_0^1 (W_{c_x})^2 - \left(\int_0^1 W_{c_x} \right)^2 \right\}} \triangleq \frac{\sigma_e}{\sigma_{e^*}} \kappa$$

2. (intercept)

$$\hat{\alpha} \Rightarrow \delta\sigma_e \int_0^1 V_c - \sigma_e \kappa \delta \int_0^1 W_{c_x}$$

3. (naive t test)

$$T^{-1/2} t_\beta \Rightarrow \mu / v^{1/2}$$

where,

$$
\begin{aligned}
\mu &= \delta\left\{\int_0^1 V_c W_{c_x} - \int_0^1 V_c \int_0^1 W_{c_x}\right\} \\
v &= \left(1+\delta^2 \int_0^1 V_c^2 dr - \delta^2 \left(\int_0^1 V_c\right)^2\right) \\
&\quad \times \left(\int_0^1 W_{c_x}^2 - \left(\int_0^1 W_{c_x}\right)^2\right) - \delta^2 \left\{\int_0^1 V_c W_{c_x} - \int_0^1 V_c \int_0^1 W_{c_x}\right\}^2
\end{aligned}
$$

4. *(regression R^2)*

$$
R^2 \Rightarrow \frac{\kappa^2\delta^2\left\{\int_0^1 W_{c_x}^2 dt - \left(\int_0^1 W_{c_x}\right)^2\right\}}{1+\delta^2\int_0^1 V_c^2 - \delta^2\left(\int_0^1 V_c\right)^2}
$$

5. *(Durbin Watson)*

$$
DW \Rightarrow \frac{1+\kappa^2}{\left(1+\delta^2\int_0^1 V_c^2 - \delta^2\left(\int_0^1 V_c\right)^2\right) - \kappa^2\delta^2\left\{\int_0^1 W_{c_x}^2 - \left(\int_0^1 W_{c_x}\right)^2\right\}}
$$

6. *(residual serial correlation)*

$$
\begin{aligned}
r_1 &= \frac{\sum_2^T \hat{u}_t \hat{u}_{t-1}}{\sum_1^T \hat{u}_t^2} \Rightarrow \frac{B}{D}, \ \textit{where} \\
B &= \delta^2\left[\int_0^1 (V_c)^2 - \left(\int_0^1 V_c\right)^2\right] - 2\kappa\delta^2\left\{\int_0^1 V_c W_{c_x} - \int_0^1 V_c \int_0^1 W_{c_x}\right\} \\
&\quad +\kappa^2\delta^2\left[\int_0^1 (W_{c_x})^2 - \left(\int_0^1 W_{c_x}\right)^2\right] \\
D &= \left(1+\delta^2\int_0^1 (V_c)^2 dr - \delta^2\left(\int_0^1 V_c\right)^2\right) - \kappa^2\delta^2\left\{\int_0^1 (W_{c_x})^2 - \left(\int_0^1 W_{c_x}\right)^2\right\}
\end{aligned}
$$

7. *(HAC t test)*

$$
\sqrt{\frac{M}{T}} t_\beta^{HAC} = \frac{\hat{\beta}}{\sqrt{\frac{T}{M}}\sigma_{HAC}} \Rightarrow \frac{\kappa}{\sqrt{F^{-2}\int_0^1\int_0^1 H(r)\,G(r)\,k(r-s)\,G(s)\,H(s)\,drds}}
$$

where

$$
\begin{aligned}
F &= \int_0^1 (W_{c_x})^2 - \left(\int_0^1 W_{c_x}\right)^2 \\
G(r) &= \left(e_\infty(r) + \delta V_c(r) - \delta \int_0^1 V_c\right) - \kappa\delta\left(W_{c_x}(r) - \int_0^1 W_{c_x}\right) \\
H(r) &= \delta\left(W_{c_x}(r) - \int_0^1 W_{c_x}\right)
\end{aligned}
$$

The above results are interesting. They show some both quantitatively and qualitatively different properties from most of the existing spurious regression theory. It may be useful to summarize the classic asymptotic theory of spurious regression of Phillips (1986, 1998) in the following lemma for easy comparison. Therefore, the following results pertain to the regression where an $I(1)$ (or near I(1)) variable is regressed on an independent $I(1)$ (or near I(1)) variable (and a constant).

Lemma 3 *(Phillips (1988, 1998)) Under certain regularity conditions,*

$$
\begin{aligned}
\hat{\beta} &= O_p(1); \hat{\alpha} = O_p\left(\sqrt{T}\right); T^{-1/2}t_\beta = O_p(1); R^2 = O_p(1); \\
TDW(\textit{Durbin Watson}) &= O_P(1); T(\gamma_s - 1) = O_p(1); \sqrt{M/T}t_\beta^{HAC} = O_p(1).
\end{aligned}
$$

Result 1 and 2 in Theorem 1 contrast with Phillips' results by showing OLS estimate $\hat{\beta}$ is still inconsistent but now the intercept estimate no longer diverges. Recall FSS claims $\hat{\beta}$ is "well behaved". Our asymptotic result suggests this claim should be qualified. The simulation reported in next section reveals that the standard deviation of the slope distribution is much smaller than a standard normal. A small standard deviation can lead to a conclusion that the slope estimate is consistent.

Result 3 says that the conventional t statistic will still diverge at the rate $\sqrt{T}$ resulting in spurious rejection. In view of result 1, the reason for diverging t statistic is not just the poor estimate of the variance term as believed in FSS in large sample. An important issue arising from this result is that our model does not produce a convergent simple t statistic on which many existing testing theory in predictive regression relies crucially. The implications of our result on the current literature will be further discussed below.

Result 4 shows R^2 is in fact a random variable in large sample, allowing us to provide some probabilistic explanation for the FSS's finding that data mining and spurious bias reenforce each other. The result is reported in the Monte Carlo section. Result 5 shows that DW statistic has a nondegenerate limiting distribution, which is a complicated function of nuisance parameter and functional of diffusion and Wiener processes. This is also different from previous result.

Result 6 shows that the residual first order autocorrelation converges to a limiting random variable, unlike the Phillips' result that says autocorrelation converges to 1 at very fast rate T. Recall FSS find that the autocorrelation not inflated and also points out this result is not compatible with Phillips' (1986) theory. Here, It is shown here under our asymptotic framework their Monte Carlo result can be explained as r no longer converges.

Result 7 shows some similarity with the classic spurious regression. It shows t test based on HAC variance estimator is still divergent given the well-known HAC estimation consistency requirement $M \to \infty$, $M/T \to 0$ as $T \to \infty$. The same rate of divergence has been obtained in Phillips (1998) for classic spurious regression model.

Based on the above results, the spurious regression bias of this kind may be more serious than others, especially compared with the classic spurious regression bias in Phillips (1986, 1988) and Durlauf and Phillips (1988). This is because several commonly used statistics (like DW, r_1) have nondegenerate distributions. Thus, there is no simple "rule of thumb" for us to even get alerted by using conventional statistics.

3.3 A convergent t test

Recently, a type of long run variance estimator is becoming popular. This class of estimator uses the kernel-based method, but without truncation, i.e. the bandwidth is equal to sample size, i.e. $M = T$ or more generally with truncation lag being a fixed proportion of sample size, i.e. $M = bT$ where b is a fixed constant, see Kiefer and Vogelsang (2002). It is shown in many situations that such an estimator can improve the performance of a test statistic. Sun (2002, 2005) considers the spurious regression issue using standard asymptotics, where the author shows once such an variance estimator is used, t statistic will have a well-defined distribution under null. He also shows by simulation using a properly selected b could alleviate the spurious regression problem in the context of his interest. I will follow these ideas and develop the corresponding theory in our case. This study is relevant because as FSS find out when they used a very long lag in case of a large sample (which could potentially corresponds to a fixed b.), the spurious regression problem can be reduced[10], but they do not recommend it due to the difficulties (or arbitrariness) in deciding how "long" the lag should be. Further comment on this piece of their finding is given in next section. Let $t_{\beta,b}$

[10]See FSS (2003a) footnote 7.

denote this version of t statistic to signify its dependence on b. It is stated as a corollary to theorem 1 with the following condition on the kernels.

- Assumption 1 (Sun (2005)): Kernel condition

  The following conditions on the kernels used in construction of HAC variance estimator is imposed to ensure positive definiteness, i.e. the kernels belong to the following class,

$$K = \left\{ k\left(.\right) : \left[-1,1\right] \rightarrow \left[0,1\right] | k\left(x\right) = k\left(-x\right), k\left(0\right) = 1, \text{ and } \int_{-1}^{1} k\left(x\right) e^{-i\lambda x} dx \geq 0 \ \forall \lambda \in R \right\}$$

Also let M to be the truncation lag (or bandwidth). Obviously, the Bartlet Kernel used in Newey-West estimator satisfies this condition.

Corollary 2 *Let $M = bT$, for some $b \in (0,1]$ then*

$$t_{\beta,b} \Rightarrow \frac{\kappa}{\sqrt{F^{-2} \int_0^1 \int_0^1 H\left(r\right) G\left(r\right) k\left(\frac{r-s}{b}\right) G\left(s\right) H\left(s\right) drds}}$$

where F, H, and G are defined as in Theorem 1.

It shows the t statistic then has a well-defined distribution, which is what to expect given result 6 in Theorem 1. Before moving on to the next section, note that the above derived limiting distributions could be potentially used for performing hypothesis testing. The difficulty lies in the localizing parameter c and δ since they can not consistently estimated without strong assumptions[11], see for example, Elliott and Stock (1994). However, conservative bound test can also be constructed as in Cavanagh et al (1995), Torous et al. (2005), Campbell et al. (2003) etc.. But the problem at hand may be a little more involved since there are two localizing parameters instead of one, which could result in very conservative procedures. I do not pursue this direction in the present paper. In the next section on Monte Carlo, the simulation results concerning the asymptotic and finite sample distributions of slope coefficient, regression R^2 and simple t statistics will be reported. The convergent t test and the interaction between data mining and spurious regression bias are also evaluated in simulation.

[11]Phillips and Moon (2000) propose a way to consistently estimate c in panel data, but with restrictive assumption that the localizing parameters in each cross section series be the same. Phillips et al (2001) propose a new block local to unity model in which c can be consistently estimated.

4 Monte Carlo

4.1 Finite sample approximation

This section presents the results on the finite sample approximation provided by the asymptotic distributions of three statistics, the slope coefficient, the regression R^2 and simple t statistics. The purpose of this exercise is to see how informative our asymptotic theory is when sample size is finite. The experimental design is the following.

1. $T : T \in (300, 800)$
2. $\rho : \rho \in (0.98, 0.99)$
3. True R^2 : true $R^2 \in (0.10, 0.15)$ [12]

To generate the corresponding asymptotic distribution, it is necessary to compute the localization parameters c and δ. They are calculated as in Nabeya and Perron (1994) as $c = T(\rho - 1)$, and $\delta = \sqrt{T}(1 + \theta)$, where θ is the implied MA coefficient. The variance of e_t is calculated as σ_u^2, the variance of u_t, divided by a_T^2 for obvious reason. The asymptotic and finite sample distributions of slope estimate, regression R^2 and the (normalized) simple t statistic are plotted in Figure 1, 2 and 3 respectively. The densities are calculated using kernel estimate. All simulation is done with 5000 replications and 5000 steps are used to approximate the stochastic integrals.

$$[Insert\ Figures\ 1, 2, 3\ Here]$$

The simulations show that our asymptotic distributions are very good finite sample approximations. For all three statistics, when the sample size is 300, there is some small discrepancy between the finite sample and asymptotic densities, mainly at the mode of the distributions. When the sample size is increased to 800, the asymptotic and finite sample densities are very close. Referring to our previous comment on the claim FSS made about the well-behaved slope coefficient $\widehat{\beta}$, the simulation here shows the impression of a consistent slope estimate could well result from the small standard deviation of slope coefficient's distribution. In summary, these Monte Carlo exercises illustrate the usefulness of our proposed asymptotic framework as well as the implied asymptotic distributions.

[12] We have done simulations for $T \in (100, 300, 500, 800)$; $\rho \in (0.90, 0.95, 0.98, 0.99)$; true $R^2 \in (0.01, 0.05, 0.10, 0.15)$. The results not reported here are available upon request.

4.2 Spurious regression and data mining

FSS (2003a, b) also discusses the interaction between pure spurious regression and data mining, the latter of which has also been studied in Foster et al (1997). Their simulation finding is that spurious regression effect interacts with data mining such that in a set of to-be-mined instruments, the more persistent instrument variables will be chosen based on the "largest R^2" criterion. Hence it worsens the spurious effect. This section establishes a theoretical justification for this Monte Carlo result. Using our asymptotic theory, the mean and median values of the R^2 as a function of localizing parameters c and δ are simulated[13]. These are plotted in Figure 4 and Figure 5.

$$[Insert\ Figure\ 4,5\ Here]$$

The distribution of R^2 is unimodal. Hence, the mean and median are two relevant measures of the location of large probability mass. The finding provides surpportive explanation for their result. That is, the mean and median of R^2 are monotonically decreasing in $|c|$ uniformly in δ considered. Therefore, our results imply more persistent "predicting variables" are with higher probability to produce higher R^2. This provides a clear picture of the interaction between data mining and spurious regression bias.

4.3 Convergent t statistic

In this subsection, the following question is investigated: can we reduce the size distortion resulted from spurious regression by using the convergent t statistic with normal critical values[14]? A yes to this question means we can be agnostic about the spurious bias and conventional distribution can be used with some caution. The answer to this question certainly depends on the kernel and the fixed proportion b one chooses. A simulation study is then conducted using the Bartlet kernel as in Sun (2005)[15]. Again 5000 replications are used. [16] The other parameter values used in simulation are:

1. $T : T \in (66, 824)$

[13] We set the localization parameters for regressor and regresand the same in these simulations.

[14] This question follows from Sun (2005), where the author shows using convergent t test with normal critical value performs better than the naive non-HAC t test.

[15] Other popular kernels, in particular, the recently proposed sharp origin kernel in Phillips et al. (2003) are also tried. It turns out no improvement can be found using this and other kernels.

[16] We have replicated some of FSS (2003a)'s simulation results and find, though with a smaller number of replications, our simulated t values differ theirs only at 4th decimal. Hence, we can safely compare our simulation results with theirs.

2. $\rho : \rho \in (0.9, 0.95, 0.98, 0.99)$

3. True R^2 : true $R^2 \in (0.01, 0.05, 0.10, 0.15)$

4. $b : b \in (0.01, 0.025, 0.05, 0.10, 0.20, ..., 0.90)$

This selection is based on theoretical consideration as well as the purpose of an easy comparison with FSS (2003a, b) results and Sun (2002, 2005). The goal here is to find the "best" b in the sense that the approximation of the standard normal is the best. Two experiments are conducted:

1. $M = bT$ with Bartlet window for various values of b, T, true R^2, and ρ;

2. For $T = 5000$, true $R^2 = 10\%$, and $\rho = 0.98$ with Bartlet window with various b.

Experiment 2 is a response to the the comment made by FSS (2003a), see their footnote 7, also mentioned above. In fact, this experiment can be regarded as simulating the (approximate) asymptotic distribution given the large sample size $T = 5000$.

The data series are simulated the same way as in FSS (2003a) did. Table II records some representative entries for the 97.5% critical $t_{\beta}^{w/o}$(i.e. the statistic with $M = T$) and Table III those for $t_{\beta,b}$ for $b = 0.01, 0.025, 0.05, 0.1, 0.2, 0.3$. The kernel smoothed density functions for a variety of parameter values are plotted in Figure 4 [17]. Results for other values of b are available upon request.

$[Insert\ Tables\ I,\ II\ Here]$

$[Insert\ Figure\ 6\ Here]$

The general conclusion from these exercises is that all the distributions have heavier tails than the standard normal, hence lead to spurious rejection. In particular, the distribution with $T = M$ represents a substantial departure from the standard normal, a lot worse than the HAC with truncation in FSS (2003a, 2003b). Turning to the comparison of the relative closeness of these distributions to the standard normal, there appear to be somewhat mixed results. For sample size $T = 824$, $b = 0.10$ and $b = 0.05$ seem to be very close to each other and both are the closest to the standard normal *on average*, especially so when true R^2 is large and ρ is close to 1, i.e. the "problematic region", see Tables II and III together with

[17] These densities are shown to be symmetric, therefore it makes sense to report our 97.5% critical value alone. Also, the normal density is not plotted to ensure a better picture of the densities of interest.

Figure 6[18]. Comparing the t values with $b = 0.05$ with FSS Table II, it is found that ours are uniformly closer to 1.96. The most striking difference is when $T = 824$, $\rho = 0.99$ and true $R^2 = 0.15$ where FSS HAC $t = 4.9151$, while ours is 3.9845, see our Table III. That $b = 0.10$ (also $b = 0.05$) is preferable in large sample surprisingly coincides with the finding in Sun (2005), where it is found that the same b produces the most desirable result in a fractional integration context. However, with $T = 66$, smaller b seems to be slightly better. This is not surprising, though, because for a small sample of this size, more lags basically add more noise to the estimation, see also FSS (2003a, b). Furthermore, since financial data is usually much larger than 66, this seems to be a less troublesome result. It is therefore concluded from this limited set of simulations, with a reasonably large sample size, $b = 0.05$ (or $b = 0.1$ or any value in between) is a sensible choice to potentially alleviate the spurious regression bias.

To get some idea on the magnitude of corresponding size distortions incurred by using the convergent t test, three representative sample sizes 1000, 1250, 1500 are studied with the following specification—true $R^2 = 0.10$, $\rho = 0.98$. It is found that the t statistic decreases somewhat slowly, for $T = 1000, 1250, 1500$, the actual sizes are 11.28%, 10.84% and 10.20% respectively at 5% nominal level for $b = 0.05$. It is also verified that for these sample sizes that $b = 0.05$ and 0.1 are again the preferable choices.

What is really interesting is that FSS reported a well-behaved t value 2.23[19] when $T = 5000$, true $R^2 = 0.10$, $\rho = 0.98$ and if they choose $M = 240$, which corresponds exactly to $b = 240/5000 = 0.048 \approx 0.05$. Our experiment 2 shows $b = 0.10$ yields a value of 2.38 and $b = 0.025$ yields a value of 2.29, which are slightly higher. It is also verified that other b values did not give a closer value, for example, $b = 0.2$ yields a value of 2.6428, $b = 0.4$ a value of 3.2882. Figure 5 plots some of the densities with different $b\prime s$.

[*Insert Figure* 7 *Here*]

Figure 7 reveals that when $b = 0.05$ and 0.10, the distributions are very close to standard normal. Thus this "large sample" experiment also confirms the preferable choice of $b = 0.05$ (or $b = 0.1$ as the difference is very weak.) as in $T = 824$. Combining our results for several different samples, it is concluded that these values of b are systematic.

Although by itself, the simple rule of thumb does not resolve the problem, it does greatly alleviate in large sample and shed more light on the problem at hand. Notice, when $b = 0.05$

[18] Especially for true $R^2 = 0.15$ and $\rho = 0.98$, i.e. figure 3, they are just on top of each other. This implies for b values between 0.05 and 0.1, the difference is minor.

[19] Our simulation gave the same value.

and $b = 0.10$, the densities are very close to standard normal in large sample, see figure 5.

The simulation findings reported above are summarized follows.

a The Bartlet window with b in the range of $[0.05, 0.1]$ seems to produce the best approximation to standard normal in large sample.

b The small sample spurious regression problem remains by using the new convergent t test unless one has a large sample.

5 Relation with the literature on predictive regression inference

FSS (2003) have already pointed out that when performing a conventional test, the null hypothesis is that the slope coefficient is zero, hence ruling out the persistent expected return. Thus, spurious regression presents no problem from this perspective. FSS' simulation and the theory provided in this paper further imply that the conventional inference theories may be better viewed as tests conditional on the belief that the stock returns are white noise (or simply serially uncorrelated). Otherwise, it is hard to interpret the situation where the test rejects the zero slope hypothesis of more than one persistent predictors. This is because a first rejection may suggest the NINW representation of the stock returns, which will, in turn, produce a divergent, hence "significant" t statistic for any (other) irrelevant persistent predictor (for example, the second predictor). Should one believe both rejections are really evidence of statistical significance? A possible explanation to reconcile the problem presented above is that all conventional tests are a conditional test of $\beta = 0$ given r_t is white noise. Campbell et al. (1997) and Campbell (2001) present a model where returns can be white noise even when they are predicted by a persistent variable. In the current framework, this amounts to have the error term u_t be a composite error (shock to expectations of future dividends and shock to expectations of future returns), see, for example, Campbell (2001). Aside from the concern whether such situation is generic or not, it shows it is possible to maintain the white noise assumption under both null and alternative. The above discussion and the finding of a divergent t statistic imply for conventional model a rejection can not be interpreted as sole evidence against $\beta = 0$. Interestingly, this interpretation implies the testing idea in Lanne (2000) is fundamentally different from all others because his test exploits the fact that r_t's univariate property is dependent on the testing outcome, a feature that could result in "inconsistent" inference. The observation Lanne (2000) made is that under the alternative, stock return has a unit root and so he proposed to test the stationarity of the stock returns as a general test to see if stock returns can be predicted by *any* highly persistent variable.[20]

The discussion so far has been directing our attention to the univariate property of stock returns. In other words, our results provide a testable condition on the possible spurious regression bias discussed in Ferson et al. (2000). But distinguishing white noise and NINW may not be easy. One reason is that, as far as we know, all available tests against ARMA(1,1) alternative have power against other forms of serial correlation, see, for example, Andrews and Ploberger (1996) and King and McAleer (1987). Therefore, a rejection by those tests can not be simply interpreted as an indication of ARMA behavior. Nevertheless, a careful investigation to present some evidence whether stock returns can be better characterized as NINW or weakly stationary process is an interesting avenue of future research. [21] [22]

Now, it is not difficult to see the difference in FSS approach. Similar to conventional inference, FSS also present a conditional test problem. But their conditional hypothesis is now $\beta = 0$ given r_t is NINW. Hence, a rejection by conventional tests does not necessarily imply the significance of the predictor, but could also result from the fact that stock returns are NINW under which the t statistic is divergent.

6 Conclusions

The predicting variable in return predictive regression in finance is usually highly autocorrelated. By postulating a persistent expected return as well as a large noise component, FSS (2003a, 2003b) used simulation to show these two facts together are an indication of possible spurious regression in financial economics. They discovered some new finite sample results, which lead them to comment that their finding is well outside of the classic setting. Motivated by the implications of their model on the univariate time series property of stock

[20]Due to simple observational equivalence argument, our analysis suggests this approach will have low power because the important large noise component which produces a NINW stock return process under the alternative is not fully realized. Similar comments are made also in Campbell and Yogo (2005)

[21]Khil et al. (2002) is a recent study on the univariate behavior of stock returns (on a daily and monthly basis). Their advocate an ARMA(2,2), which does not support either of the assumptions, though.

[22]Some preliminary results of our study show, for example, the supLM test by Andrews and Ploberger (1996) has non-trivial local power against NINW process. Applying the supLM test on the SP500 monthly stock returns data in Ferson et al. (2000), we get a rejection for full sample and some selected subsamples. But, also notice, the test has power against other forms of serial correlation as well. We also applied the LM test for testing AR(1) against ARMA(1,1) (see King and McAleer (1987)) to the same data set. We obtain highly significant statistics using full sample. And the full sample estimation of ARMA(1,1) model in Matlab gives AR root 0.999 and MA root 0.983. Some subsample estimates can be smaller. We will report all these studies in a subsequent paper.

returns, an alternative new asymptotic theory for their Monte Carlo results is developed. The use of a new asymptotic framework, namely, the nearly integrated/nearly white noise framework allows us to obtain analytical asymptotic distributions of various statistics and slope coefficient. It turns out that the finite sample behavior of several statistics can be well predicted by our asymptotic theory and simulations show our asymptotic distributions provide very good finite sample approximation. One conclusion that transpires from our study is that the autocorrelation of the dependent variable should not be taken to be indicative of spurious regression bias. This observation is important in applied work and is essentially a restatement of the well known observational equivalence issue in unit root testing (see, for example, Campbell and Perron (1991)). A convergent t statistic is also constructed and properties studied in the Monte Carlo simulation. The findings reported here allow us to explain an interesting simulation result in FSS (2003a, b). A probabilistic interpretation is provided for the interaction between data mining and spurious regression bias. The implications from this study are general in and outside of the financial econometrics framework. In the last section, further implications of the FSS's spurious regression problem is provided. The importance of understanding univariate behavior of stock return is emphasized and it is acknowledged that the difficulty in that area remains.

References

1. Amihud, Yakov., and Clifford Hurvich, 2004, Predictive regressions: A reduced-bias estimation method, Journal of Financial and Quantative Analysis 39 (4), 813-841

2. Andrews, Donald W. K. and Werner Ploberger, 1996, Testing for serial correlation against an ARMA (1,1) process, Journal of the American Statistical Association 91, 1331-1342

3. Campbell, John, 2001, Why long horizons? A study of power against persistent alternatives, Journal of Empirical Finance 8, 459-491.

4. Campbell, John Y., Andrew W. Lo and A Craig Mackinlay, 1997, The Econometrics of Financial Markets (Princeton University Press)

5. Campbell, John, and Pierre Perron, 1991, Pitfalls and opportunities: What macroeconomists should know about unit root, NBER working paper.

6. Campell, John and Motohiro Yogo, 2005, Efficient tests of stock return predictability, Journal of financial economics, 2006, forthcoming

7. Cavanagh, Christopher, Graham Elliott, and James Stock, 1995, Inference in models with nearly integrated regressors. Econometric theory 11, 1131-1147.

8. Chiarella, Carl and Gao Shenhuai, 2002, Type I spurious regression in econometrics. School of Finance and Economics, University of Technology, Sydney, Working Paper Number: 114.

9. Cochrane, J. 2001. Asset Pricing, Princeton University Press.

10. Granger, Clive W.J. and Paul Newbold, 1974, Spurious regression in economics, Journal of Econometrics 4, 111-120

11. Deng, Ai, 2005, Theoretical and Monte Carlo comparison of two asymptotics in predictive regression framework, work in progress, Department of Economics, Boston University

12. Durlauf, Steven N and Phillips, P.C.B., 1988, Trends versus random walks in time series analysis, Econometrica, Vol.56, No.6, 1333-1354

13. Elliott, Graham and Stock, James H. 1994, Inference in time series regression when the order of integration of a regressor is unknown. Econometric theory 10, 672-700.

14. Ferson, Wayne, Sergei Sarkissian and Tim, Simin, 2003a, Spurious regression in financial economics? Journal of Finance, Vol LVIII, No. 4

15. Ferson, Wayne, Sergei Sarkissian and Tim Simin, 2003b, Is stock return predictability spurious?, Journal of Investment Management.

16. Khil, Jaeuk, and Bong-Soo Lee, 2002, A time series model of stock returns with a positive short-term correlation and a negative long-term correlation. Review of Quantitative Finance and Accounting, 18, 381-404.

17. Kiefer, Nicholas and Timothy Vogelsang, 2002, Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size, Econometric Theory, 18, 1350-1366

18. King, Maxwell and Micheal McAleer, 1987, Further results on testing AR(1) against MA(1) disturbances in the linear regression models, Review of Economic Studies, 54, 649-663

19. Lanne, Markku, 2002, Testing the predictability of stock returns, Review of Economics and Statistics, 84, 407-415.

20. Lo, Andrew W., and A. Craig MacKinlay, 1988, Stock market prices do not follow random walks: evidence from a simple specification test, Review of Financial Studies 1, 41-66

21. Nabeya, S. and P. Perron 1994, Local asymptotic distributions related to the AR(1) model with dependent errors, Journal of Econometrics, 62, 229-264.

22. Perron, P. and Serena Ng, 1996, Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. The review of economic studies, Vol63(3), 435-463

23. Phillips, P.C.B, 1986, Understanding spurious regressions in econometrics. Journal of econometrics 33, 311-340.

24. Phillips, P.C.B, 1988, Regression theory for nearly integrated series, Econometrica 56, 1021-1043

25. Phillips, P.C.B, 1998, New tools for understanding spurious regressions. Econometrica 66, No.6 1299-1325

26. Phillips, P.C.B, and Hyungsik R Moon, 2000, Estimation of autoregressive roots near unity using panel data. Econometric Theory, 16, 927-997.

27. Phillips, Peter C B, Hyungsik Roger Moon and Zhijie Xiao , 2001. How to estimate autoregressive roots near unity, Econometric Theory, 17(1), 29-69.

28. Phillips, Peter C B, Yixiao Sun and Sainan Jin, 2003. Consistent HAC Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation," Cowles Foundation Discussion Papers 1407, Cowles Foundation, Yale University.

29. Philip A. Shively, 2000, Stationary components in stock prices: an exact pointwise most powerful invariant test. Journal of Business and Economic statistic 18, 4 489-496.

30. Sun, Yixiao, 2002, A convergent t-statistic in spurious regressions, working paper, Department of Economics, UCSB

31. Sun, Yixiao, 2005, A convergent t-statistic in spurious regressions, Forthcoming in Econometric Theory

32. Torous, Walter, and Rossen Valkanov, 2000, Boundaries of predictability: Noisy predictive regressions, working paper, UCLA.

33. Torous, Walter, Rossen Valkanov, and Shu Yan, 2005, On predicting stock returns with nearly integrated explanatory variables, Journal of Business 78, forthcoming.

34. Yule, George U., 1926, Why do we sometimes get nonsense correlations between time series? A study in sampling and the nature of time series, Journal of the Royal Statistical Society 89, 1-64

Technical Appendix

As is shown in Nabeya and Perron (1994), under the assumptions in the text, one can write

$$y_t = a_T e_t + b_T X_t$$

where $a_T = \left(1 - \delta T^{-1/2}\right) \exp\left(-c/T\right) \to 1, b_T = 1 - \exp\left(-c/T\right)\left(1 - \delta T^{-1/2}\right)$, so $T^{1/2} b_T \to \delta$ and

$X_t = \sum_{j=1}^{t} \exp\left(\left(t-j\right) c/T\right) e_{t-j}$ is a near-integrated process. And the regressor is $b_T X_t^*$, where $X_t = \sum_{j=1}^{t} \exp\left(\left(t-j\right) c/T\right) e_{t-j}^*$. And e_{t-j}^*, with variance $\sigma_{e^*}^2$ is independent of e_{t-j} for all t and j. We first state the following useful lemma in Perron and Ng (1998) collected from Nabeya and Perron (1994) whose proof is based on the above representation. Lemma A.2 is a direct result as well.

Lemma 4 (A.1) *(Perron and Ng (1998)). Let $\{y_t\}$ be generated as nearly white noise and let $e_\infty\left(r\right) = \lim_{T\to\infty} e_{[Tr]}/\sigma_e$. Then as $T \to \infty$, (a) $T^{-1}\sum_{t=1}^{T} y_{t-1}^2 \Rightarrow \sigma_e^2\left(1 + \delta^2 \int_0^1 V_c{}^2\right)$; (b) $T^{-1}\sum_{t=1}^{T} y_{t-1} w_t \Rightarrow -\sigma_e^2$, (c) $y_{[Tr]} \Rightarrow \sigma_e^2\left(e_{\infty r} + \delta V_c\left(r\right)\right)$, and (d) $T^{-1}\sum_{t=1}^{T} w_t^2 \Rightarrow 2\sigma_e^2$.*

Lemma 5 (A.2) $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t \Rightarrow \delta\sigma_e \int_0^1 V_c$

Proof of Lemma (A.2).

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} y_t &= a_T T^{-1}\sum_{t=1}^{T} e_t + b_T T^{-1}\sum_{t=1}^{T} X_t \\ &= a_T T^{-1}\sum_{t=1}^{T} e_t + T^{1/2} b_T T^{-1/2} T^{-1}\sum_{t=1}^{T} X_t \\ &\Rightarrow \delta\sigma_e \int_0^1 V_c \end{aligned}$$

where the last relation follows from an LLN for $\{e_t\}$ and convergence results for a_T and b_T. Also notice under this specification, for the regressor x_t, we have

$$\begin{aligned} \bar{x} &\to \sigma_{e^*}\delta \int W_{c_x} \\ T^{-1}\sum\left(x_t - \bar{x}\right)^2 &\to \delta^2\sigma_{e^*}^2\left(\int \left(W_{c_x}\right)^2 - \left(\int W_{c_x}\right)^2\right) \end{aligned}$$

■

Proof of theorem 1.

$$
\begin{aligned}
\hat{\beta} &= \frac{T^{-1}\sum y_t\left(x_t-\bar{x}\right)}{T^{-1}\sum\left(x_t-\bar{x}\right)^2} = \frac{T^{-1}\sum y_t\left(x_t-\bar{x}\right)}{T^{-1}\sum\left(x_t-\bar{x}\right)^2} \\
&\Rightarrow \frac{\sigma_{e^*}\sigma_e\left\{\int_0^1 V_c W_{c_x} - \int_0^1 V_c \int_0^1 W_{c_x}\right\}}{\sigma_{e^*}^2\left\{\int_0^1 \left(W_{c_x}\right)^2 - \left(\int_0^1 W_{c_x}\right)^2\right\}} \triangleq \frac{\sigma_e}{\sigma_{e^*}}\kappa
\end{aligned}
$$

because of standard asymptotic results for x_t.

For the estimated intercept term, we have,

$$
\begin{aligned}
\hat{\alpha} &= T^{-1}\sum y_t - \hat{\beta}T^{-1}\sum x_t \\
&\Rightarrow \sigma_e\left(\delta\int_0^1 V_c - \kappa\delta\int_0^1 W_{c_x}\right)
\end{aligned}
$$

For (3), first consider

$$
\begin{aligned}
s^2 &= T^{-1}\sum\left(y_t-\bar{y}\right)^2 - \hat{\beta}^2 T^{-1}\sum\left(x_t-\bar{x}\right)^2 \\
&\Rightarrow \sigma_e^2\left\{\left[1+\delta^2\int_0^1 \left(V_c\right)^2 - \delta^2\left(\int_0^1 V_c\right)^2\right] - \kappa^2\delta^2\left[\int_0^1 {W_{c_x}}^2 - \left(\int_0^1 W_{c_x}\right)^2\right]\right\}
\end{aligned}
$$

Now,

$$
\begin{aligned}
T^{-1/2}t_\beta &= \frac{\hat{\beta}\left(T^{-1}\sum\left(x_t-\bar{x}\right)^2\right)^{1/2}}{s} \\
&\Rightarrow \frac{\sigma_e\left\{\int_0^1 V_c W_{c_x} - \int_0^1 V_c\int_0^1 W_{c_x}\right\}}{\sigma_{e^*}\left\{\int_0^1\left(W_{c_x}\right)^2 - \left(\int_0^1 W_{c_x}\right)^2\right\}} \\
&\quad \times\sigma_{e^*}\delta\left\{\int_0^1 {W_{c_x}}^2 - \left(\int_0^1 W_{c_x}\right)^2\right\}^{1/2} \\
&\quad \div\left\{\sigma_e^2\left\{1+\delta^2\int_0^1 {V_c}^2 - \delta^2\left(\int_0^1 V_c\right)^2\right\}\right. \\
&\quad \left. -\kappa^2\delta^2\sigma_{e^*}^2\left\{\int_0^1 W_{c_x}^2\,dt - \left(\int_0^1 W_{c_x}\right)^2\right\}\right\}^{1/2}
\end{aligned}
$$

$$
\begin{aligned}
&= \delta\left\{\int_0^1 V_c W_{c_x} - \int_0^1 V_c \int_0^1 W_{c_x}\right\} \\
&\quad \div \left\{\left(1+\delta^2\int_0^1 V_c^2 dr - \delta^2\left(\int_0^1 V_c\right)^2\right)\left(\int_0^1 W_{c_x}^2 - \left(\int_0^1 W_{c_x}\right)^2\right)\right. \\
&\quad \left. -\delta^2\left\{\int_0^1 V_c W_{c_x} - \int_0^1 V_c \int_0^1 W_{c_x}\right\}^2\right\}^{1/2} \\
&= \mu/v^{1/2}
\end{aligned}
$$

, proving (b). Next,

$$
\begin{aligned}
R^2 &= \frac{\sum(\hat{y}_t-\bar{y})^2}{\sum(y_t-\bar{y})^2} = \frac{\hat{\beta}^2 T^{-1}\sum(x_t-\bar{x})^2}{T^{-1}\sum(y_t-\bar{y})^2} \\
&\Rightarrow \frac{\kappa^2\delta^2\left\{\int_0^1 W_{c_x}^2 dt - \left(\int_0^1 W_{c_x}\right)^2\right\}}{1+\delta^2\int_0^1 V_c^2 - \delta^2\left(\int_0^1 V_c\right)^2}
\end{aligned}
$$

Next consider Durbin Watson statistic,

$$
DW = \frac{\sum_2^T(\hat{u}_t-\hat{u}_{t-1})^2}{\sum_1^T \hat{u}_t^2} = \frac{T^{-1}\sum_1^T\left(y_t-y_{t-1}-\hat{\beta}(x_t-x_{t-1})\right)^2}{T^{-1}\sum_1^T\left(y_t-\bar{y}-\hat{\beta}(x_t-\bar{x})\right)^2}
$$

$$
\begin{aligned}
T^{-1}\sum_1^T\left(y_t-y_{t-1}-\hat{\beta}(x_t-x_{t-1})\right)^2 &= T^{-1}\sum_1^T\left(\frac{c_y}{T}y_{t-1}+w_t-\hat{\beta}\left(\frac{c_x}{T}x_{t-1}+e_t^*\right)\right)^2 \\
&= T^{-1}\sum_1^T w_t^2 + T^{-1}\hat{\beta}^2\sum_1^T e_t^{*2} - 2\hat{\beta}T^{-1}\sum_1^T w_t e_t^* \\
&\to {}_p\sigma_e^2 + \kappa^2\sigma_e^2
\end{aligned}
$$

since the last two terms go to 0. Therefore

$$
DW \Rightarrow \frac{1+\kappa^2}{\left(1+\delta^2\int_0^1 V_c^2 - \delta^2\left(\int_0^1 V_c\right)^2\right) - \kappa^2\delta^2\left\{\int_0^1 W_{c_x}^2 - \left(\int_0^1 W_{c_x}\right)^2\right\}}
$$

Now, consider r_1. We have shown that

$$
s^2 \Rightarrow \sigma_e^2\left(1+\delta^2\int_0^1 V_c^2 - \delta^2\left(\int_0^1 V_c\right)^2\right) - \kappa^2\sigma_e^2\delta^2\left\{\int_0^1 W_{c_x}^2 - \left(\int_0^1 W_{c_x}\right)^2\right\}
$$

Consider

$$
\begin{aligned}
T^{-1}\sum_{2}^{T}\hat{u}_t\hat{u}_{t-1} &= T^{-1}\sum_{2}^{T}\left(y_t-\bar{y}-\hat{\beta}\left(x_t-\bar{x}\right)\right)\left(y_{t-1}-\bar{y}-\hat{\beta}\left(x_{t-1}-\bar{x}\right)\right)\\
&= T^{-1}\sum_{2}^{T}\left(y_t-\bar{y}\right)\left(y_{t-1}-\bar{y}\right)-T^{-1}\sum_{2}^{T}\hat{\beta}\left(x_{t-1}-\bar{x}\right)\left(y_t-\bar{y}\right)\\
&\quad -T^{-1}\sum_{2}^{T}\hat{\beta}\left(x_t-\bar{x}\right)\left(y_{t-1}-\bar{y}\right)+T^{-1}\sum_{2}^{T}\hat{\beta}^2\left(x_t-\bar{x}\right)\left(x_{t-1}-\bar{x}\right)\\
&= I+II+III+IV
\end{aligned}
$$

$$
\begin{aligned}
I &= T^{-1}\sum_{2}^{T}\left(y_t-\bar{y}\right)\left(y_{t-1}-\bar{y}\right)=T^{-1}\sum_{2}^{T}y_ty_{t-1}-T^{-1}\sum_{2}^{T}y_t\bar{y}-T^{-1}\sum_{2}^{T}y_{t-1}\bar{y}+T^{-1}\sum_{2}^{T}\bar{y}^2\\
&= T^{-1}\sum_{2}^{T}y_ty_{t-1}-\bar{y}^2
\end{aligned}
$$

So consider

$$
\begin{aligned}
T^{-1}\sum_{2}^{T}y_ty_{t-1} &\sim T^{-1}\sum_{2}^{T}\left(y_{t-1}+w_t\right)y_{t-1}=T^{-1}\sum_{2}^{T}y_{t-1}^2+T^{-1}\sum_{2}^{T}w_ty_{t-1}\\
&\Rightarrow \sigma_e^2\left(1+\delta^2\int_0^1 V_c^2\right)-\sigma_e^2=\sigma_e^2\delta^2\int_0^1 V_c^2
\end{aligned}
$$

The second term

$$
\begin{aligned}
II &= \hat{\beta}T^{-1}\sum_{2}^{T}\left(x_{t-1}-\bar{x}\right)y_t\to\left(T^{1/2}\hat{\beta}\right)T^{-3/2}\sum_{2}^{T}\left(x_{t-1}-\bar{x}\right)y_t-\left(T^{1/2}\hat{\beta}\right)T^{-3/2}\bar{y}\sum_{2}^{T}\left(x_{t-1}-\bar{x}\right)\\
&= \sigma_e^2\kappa\delta^2\left\{\int_0^1 V_cW_{c_x}-\int_0^1 V_c\int_0^1 W_{c_x}\right\}
\end{aligned}
$$

The Third term

$$
III=T^{-1}\sum_{2}^{T}\hat{\beta}\left(x_t-\bar{x}\right)\left(y_{t-1}-\bar{y}\right)\Rightarrow\sigma_e^2\kappa\delta^2\left\{\int_0^1 V_cW_{c_x}-\int_0^1 V_c\int_0^1 W_{c_x}\right\}
$$

The last term

$$\begin{aligned}
IV &= T^{-1}\sum_{2}^{T}\hat{\beta}^{2}\left(x_t-\bar{x}\right)\left(x_{t-1}-\bar{x}\right)=\left(\sqrt{T}\hat{\beta}\right)^{2}T^{-2}\sum_{2}^{T}\left(x_t-\bar{x}\right)\left(x_{t-1}-\bar{x}\right)\\
&= \left(\sqrt{T}\hat{\beta}\right)^{2}\left[T^{-2}\sum_{2}^{T}x_t x_{t-1}-T^{-2}\bar{x}\sum_{2}^{T}x_t\right.\\
&\quad \left. -T^{-2}\bar{x}\sum_{2}^{T}x_{t-1}+T^{-2}\sum_{2}^{T}\bar{x}^{2}\right]\\
&\Rightarrow \left(\sigma_e\kappa\delta\right)^{2}\left[\int_0^1\left(W_{c_x}\right)^2-\left(\int_0^1 W_{c_x}\right)^2\right]
\end{aligned}$$

So,

$$\begin{aligned}
T^{-1}\sum_{2}^{T}\hat{u}_t\hat{u}_{t-1} &= \sigma_e^2\delta^2\left(\int_0^1\left(V_c\right)^2-\left(\int_0^1 V_c\right)^2\right)\\
&\quad -2\sigma_e^2\kappa\delta^2\left\{\int_0^1 V_cW_{c_x}-\int_0^1 V_c\int_0^1 W_{c_x}\right\}\\
&\quad +\left(\sigma_e^2\kappa\delta\right)^2\left[\int_0^1\left(W_{c_x}\right)^2-\left(\int_0^1 W_{c_x}\right)^2\right]
\end{aligned}$$

Finally,

$$\begin{aligned}
r_1 &= \frac{\sum_2^T\hat{u}_t\hat{u}_{t-1}}{\sum_1^T\hat{u}_t^2}\Rightarrow\frac{B}{D},\ \text{where}\\
B &= \sigma_e^2\delta^2\left[\int_0^1\left(V_c\right)^2-\left(\int_0^1 V_c\right)^2\right]-2\kappa\delta^2\sigma_e^2\left\{\int_0^1 V_cW_{c_x}-\int_0^1 V_c\int_0^1 W_{c_x}\right\}\\
&\quad +\sigma_e^2\kappa^2\delta^2\left[\int_0^1\left(W_{c_x}\right)^2-\left(\int_0^1 W_{c_x}\right)^2\right]\\
D &= \sigma_e^2\left(1+\delta^2\int_0^1\left(V_c\right)^2dr-\delta^2\left(\int_0^1 V_c\right)^2\right)-\kappa^2\delta^2\sigma_e^2\left\{\int_0^1\left(W_{c_x}\right)^2-\left(\int_0^1 W_{c_x}\right)^2\right\}
\end{aligned}$$

higher order autocorrelation can be easily obtained analogously.

Next, consider the HAC t test. First,

$$\begin{aligned}
\hat{u}_{[Tr]} &= \left(y_{[Tr]}-\bar{y}\right)-\hat{\beta}\left(x_{[Tr]}-\bar{x}\right)\\
&\Rightarrow \sigma_e\left(e_\infty\left(r\right)+\delta V_c\left(r\right)-\delta\int_0^1 V_c\right)-\sigma_e\kappa\left(W_{c_x}\left(r\right)-\int_0^1 W_{c_x}\right)
\end{aligned}$$

Then, following Sun (2005), we write $\frac{T}{M}\sigma_{HAC}^2$ as, using the notations that $x_t^{\#} = \sqrt{T}x_t$, $\bar{x}^{\#} = \sqrt{T}\bar{x}_t$,

$$\begin{aligned}
&\left(\frac{1}{T}\sum_{t=1}^{T}(x_t-\bar{x})^2\right)^{-2}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{\left(x_t^{\#}-\bar{x}^{\#}\right)}{\sqrt{T}}\hat{u}_t k\left(\frac{t-s}{T}\right)\hat{u}_s\frac{\left(x_s^{\#}-\bar{x}^{\#}\right)}{\sqrt{T}} \\
\Rightarrow\;& \sigma_e^2\sigma_{e^*}^{-2}F^{-2}\int_0^1\int_0^1 H(r)\,G(r)\,k(r-s)\,G(s)\,H(s)\,drds
\end{aligned}$$

where

$$\begin{aligned}
F &= \int_0^1 (W_{c_x})^2 - \left(\int_0^1 W_{c_x}\right)^2 \\
G(r) &= \left(e_\infty(r) + \delta V_c(r) - \delta\int_0^1 V_c\right) - \kappa\delta\left(W_{c_x}(r) - \int_0^1 W_{c_x}\right) \\
H(r) &= \delta\left(W_{c_x}(r) - \int_0^1 W_{c_x}\right)
\end{aligned}$$

Therefore,

$$\sqrt{\frac{M}{T}}t_\beta^{HAC} = \frac{\hat{\beta}}{\sqrt{\frac{T}{M}}\sigma_{HAC}} \Rightarrow \frac{\kappa}{\sqrt{F^{-2}\int_0^1\int_0^1 H(r)\,G(r)\,k(r-s)\,G(s)\,H(s)\,drds}}$$

■

Proof of Corollary 1. The proof follows directly those of result 6 in theorem 1.

■

Table I: the implied θ

ρ^*/R^2	0.01	0.05	0.10	0.15
0.9	-0.8914	-0.8616	-0.8306	-0.8029
0.95	-0.9412	-0.9143	-0.8883	-0.8660
0.98	-0.9718	-0.9509	-0.9324	-0.9169
0.99	-0.9827	-0.9665	-0.9528	-0.9414

Note: In computing these values, we also used the calibrated $Var(Z^*)$ as described in FSS(2003a)

Table II: 97.5% Critical $t_\beta^{w/o}$

without truncation				
$T = 66$				
R^2/ρ^*	0.9	0.95	0.98	0.99
0.01	5.8056	6.0040	5.8316	5.8462
0.05	6.4760	6.0823	6.1775	6.2418
0.10	6.5238	6.7656	6.8125	6.6303
0.15	6.6186	6.7727	6.7057	6.4297
$T = 824$				
R^2/ρ^*	0.9	0.95	0.98	0.99
0.01	5.1163	5.2940	5.5422	5.8563
0.05	5.2390	5.1917	5.8121	6.5784
0.10	5.2941	5.2216	7.1039	7.2997
0.15	5.1271	5.4731	6.3576	7.4835

Table III: 97.5% Critical $t\!\!/_{\beta,b}$ with truncation

b=0.01								
$T = 66$					$T = 824$			
R^2/ρ^*	0.9	0.95	0.98	0.99	R^2/ρ^*	0.9	0.95	0.98	0.99
0.01	2.2745	2.2505	2.2118	2.2513	0.01	2.0296	2.0916	2.3673	2.3953
0.05	2.5212	2.4558	2.4436	2.3651	0.05	2.3191	2.4171	3.1651	3.7966
0.10	2.6392	2.6252	2.5951	2.4346	0.10	2.2691	2.6859	3.6640	4.3644
0.15	2.8249	2.9414	2.6986	2.5792	0.15	2.4115	2.9307	4.0783	5.0238
b=0.025									
$T = 66$					$T = 824$				
R^2/ρ^*	0.9	0.95	0.98	0.99	R^2/ρ^*	0.9	0.95	0.98	0.99
0.01	2.3740	2.3188	2.3121	2.3167	0.01	2.0716	2.1399	2.4084	2.4952
0.05	2.5790	2.4959	2.4760	2.4358	0.05	2.3280	2.3190	2.9500	3.5695
0.10	2.6332	2.6923	2.6557	2.5329	0.10	2.1949	2.4717	3.3066	3.8937
0.15	2.8342	2.9352	2.8628	2.6762	0.15	2.2878	2.6293	3.4925	4.2945
b=0.05									
$T = 66$					$T = 824$				
R^2/ρ^*	0.9	0.95	0.98	0.99	R^2/ρ^*	0.9	0.95	0.98	0.99
0.01	2.4323	2.4050	2.4113	2.3960	0.01	2.1446	2.2121	2.5079	2.5639
0.05	2.6295	2.5958	2.5598	2.5146	0.05	2.3681	2.3405	2.8822	3.3979
0.10	2.6598	2.7334	2.7109	2.6128	0.10	2.2340	2.3980	3.0940	3.6061
0.15	2.9221	2.9837	2.9556	2.7274	0.15	2.2996	2.5694	3.2672	3.9845

Table III Continued: b=0.10

	$T = 66$					$T = 824$			
R^2/ρ^*	0.9	0.95	0.98	0.99	R^2/ρ^*	0.9	0.95	0.98	0.99
0.01	2.5851	2.6751	2.6531	2.6148	0.01	2.3736	2.4176	2.6094	2.8401
0.05	2.7599	2.7800	2.8257	2.8617	0.05	2.3952	2.4680	2.9625	3.4444
0.10	2.9623	3.0564	3.0828	2.9116	0.10	2.4149	2.6457	3.1360	3.8080
0.15	3.1468	3.2835	3.1851	2.9767	0.15	2.4382	2.6300	3.2601	4.0143

b=0.20

	$T = 66$					$T = 824$			
R^2/ρ^*	0.9	0.95	0.98	0.99	R^2/ρ^*	0.9	0.95	0.98	0.99
0.01	2.9634	3.0769	3.1090	3.1044	0.01	2.6123	2.8022	2.8317	3.2855
0.05	3.0695	3.3850	3.3537	3.2298	0.05	2.7121	2.9024	3.2624	3.8889
0.10	3.4254	3.5046	3.5650	3.3415	0.10	2.7421	2.9670	3.5548	3.7708
0.15	3.3997	3.6490	3.7471	3.5316	0.15	2.6950	2.9734	3.5922	4.0648

b=0.30

	$T = 66$					$T = 824$			
R^2/ρ^*	0.9	0.95	0.98	0.99	R^2/ρ^*	0.9	0.95	0.98	0.99
0.01	3.3862	3.5111	3.5533	3.5536	0.01	3.0563	3.0883	3.3946	3.4640
0.05	3.6554	3.6235	3.7013	3.6690	0.05	2.9838	3.1759	3.8025	4.3192
0.10	3.5426	3.8976	3.8257	4.1001	0.10	2.9875	3.4395	3.7616	4.6522
0.15	3.8838	4.1978	4.2413	3.9934	0.15	3.1186	3.2724	3.8732	4.7828

Note: Results for $b > 0.30$ are available upon request.

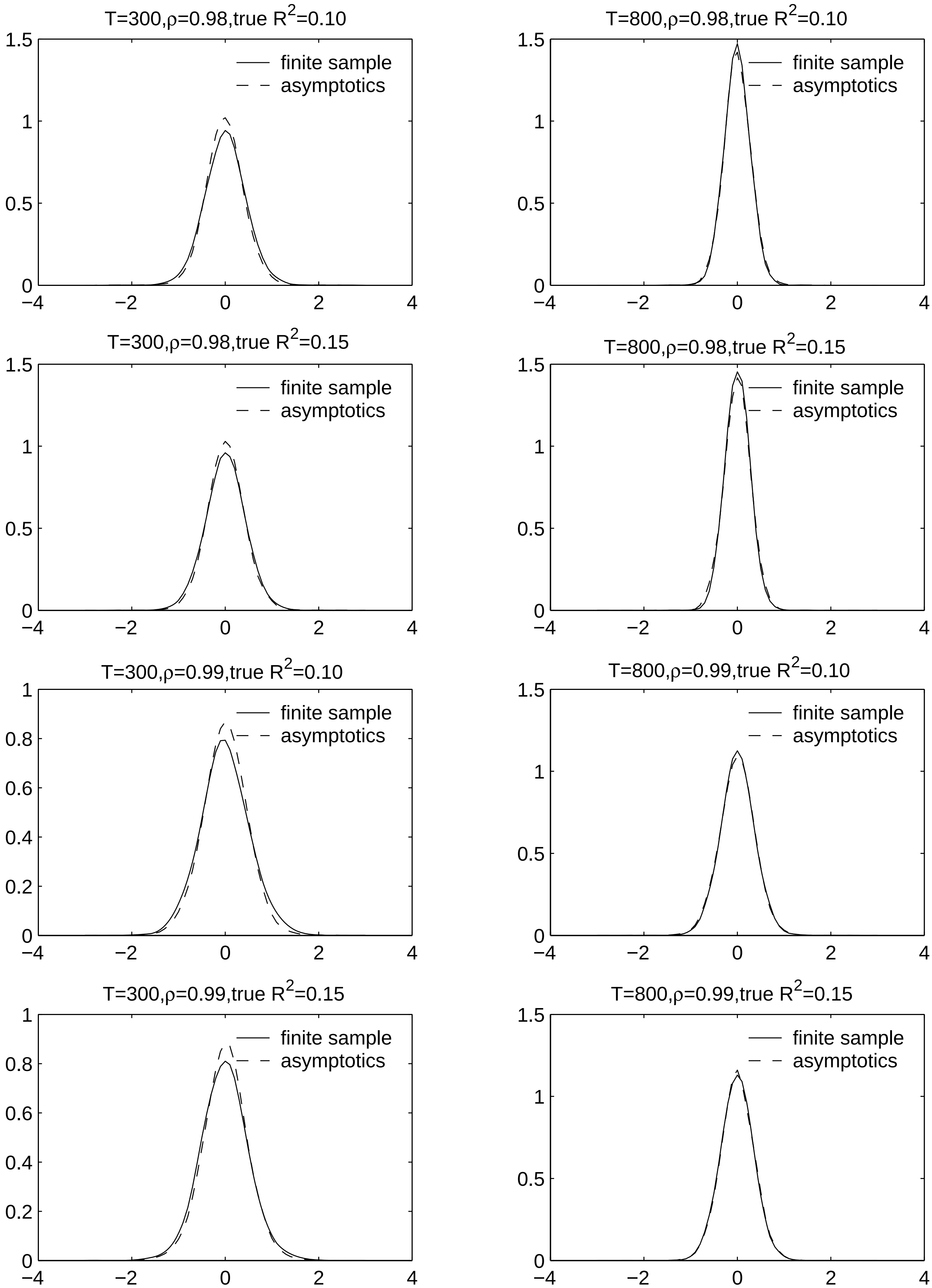

Figure 1: Finite sample distributions versus asymptotic distributions: slope estimate. The parameters for asymptotic distributions are calculated as $c = T(\rho - 1), \delta = \sqrt{T}\,(1 + \theta)$ and $var(e) = \sigma_u^2 / a_T^2$

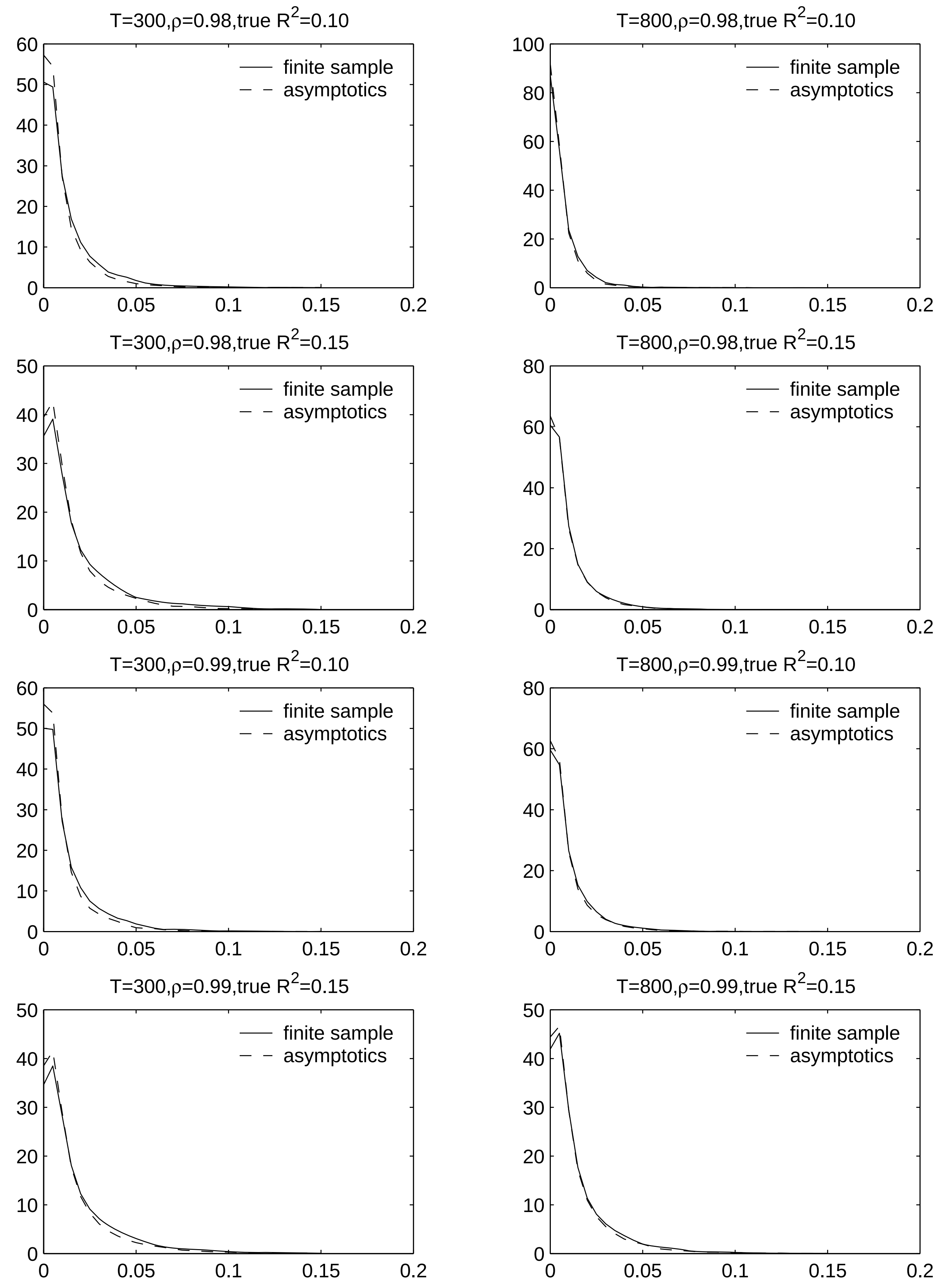

Figure 2: Finite sample distributions versus asymptotic distributions: regression R^2. The parameters for asymptotic distributions are calculated as $c = T(\rho - 1), \delta = \sqrt{T}\,(1 + \theta)$ and $var(e) = \sigma_u^2/a_T^2$

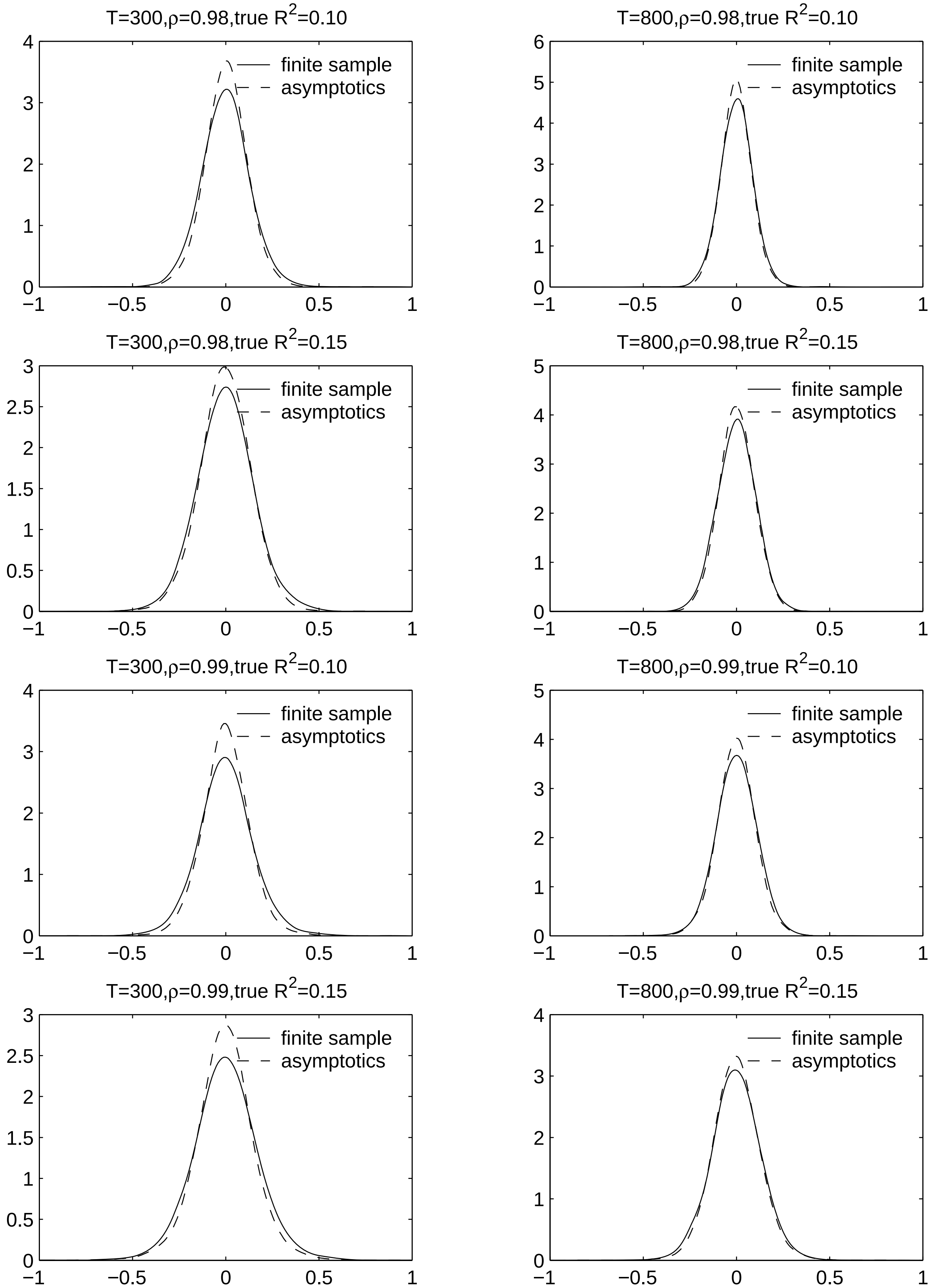

Figure 3: Finite sample distribution versus asymptotic distribution: normalized simple t statistics ($T^{-1/2}t_\beta$). The parameters for asymptotic distributions are calculated as $c = T(\rho - 1), \delta = \sqrt{T}\,(1+\theta)$ and $var(e) = \sigma_u^2/a_T^2$

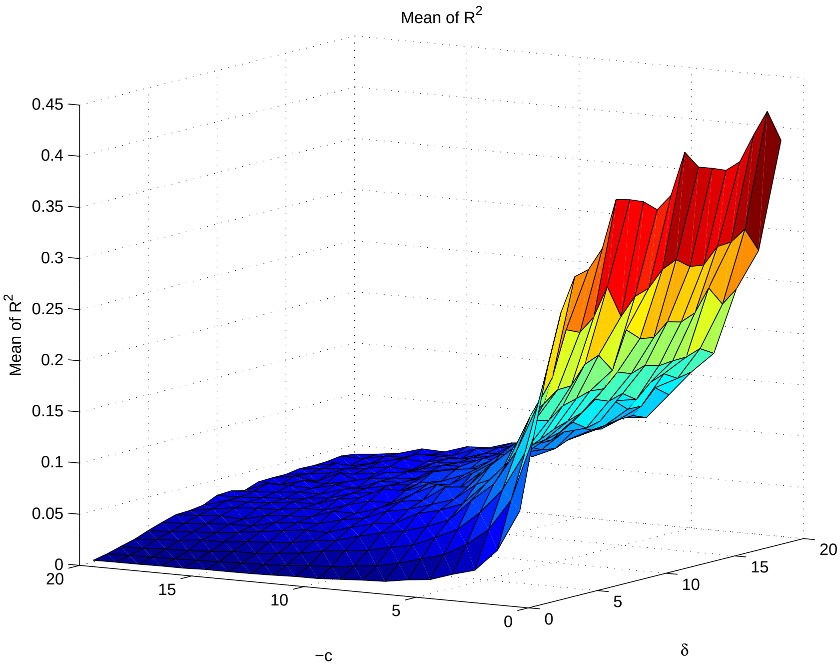

Figure 4: Mean of R^2 as a function of $-c$ and δ

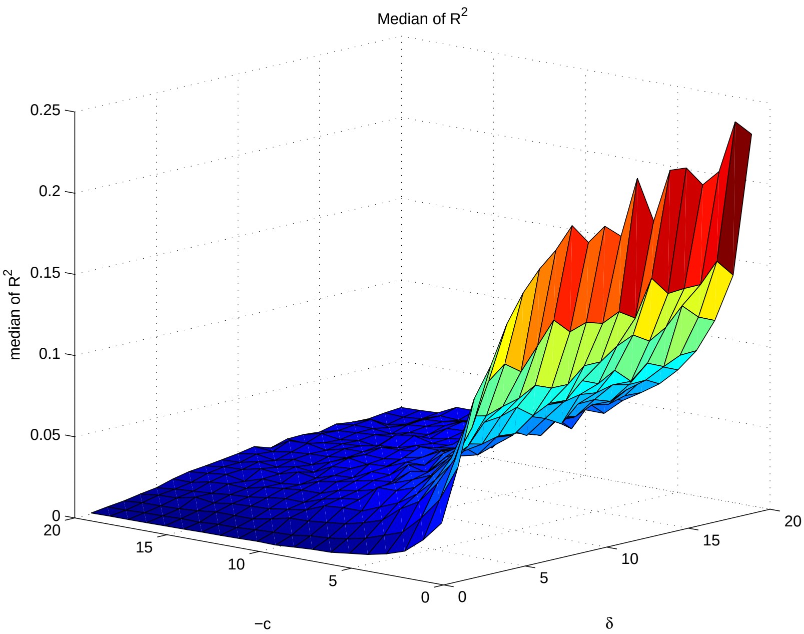

Figure 5: Median of R^2 as a function of $-c$ and δ

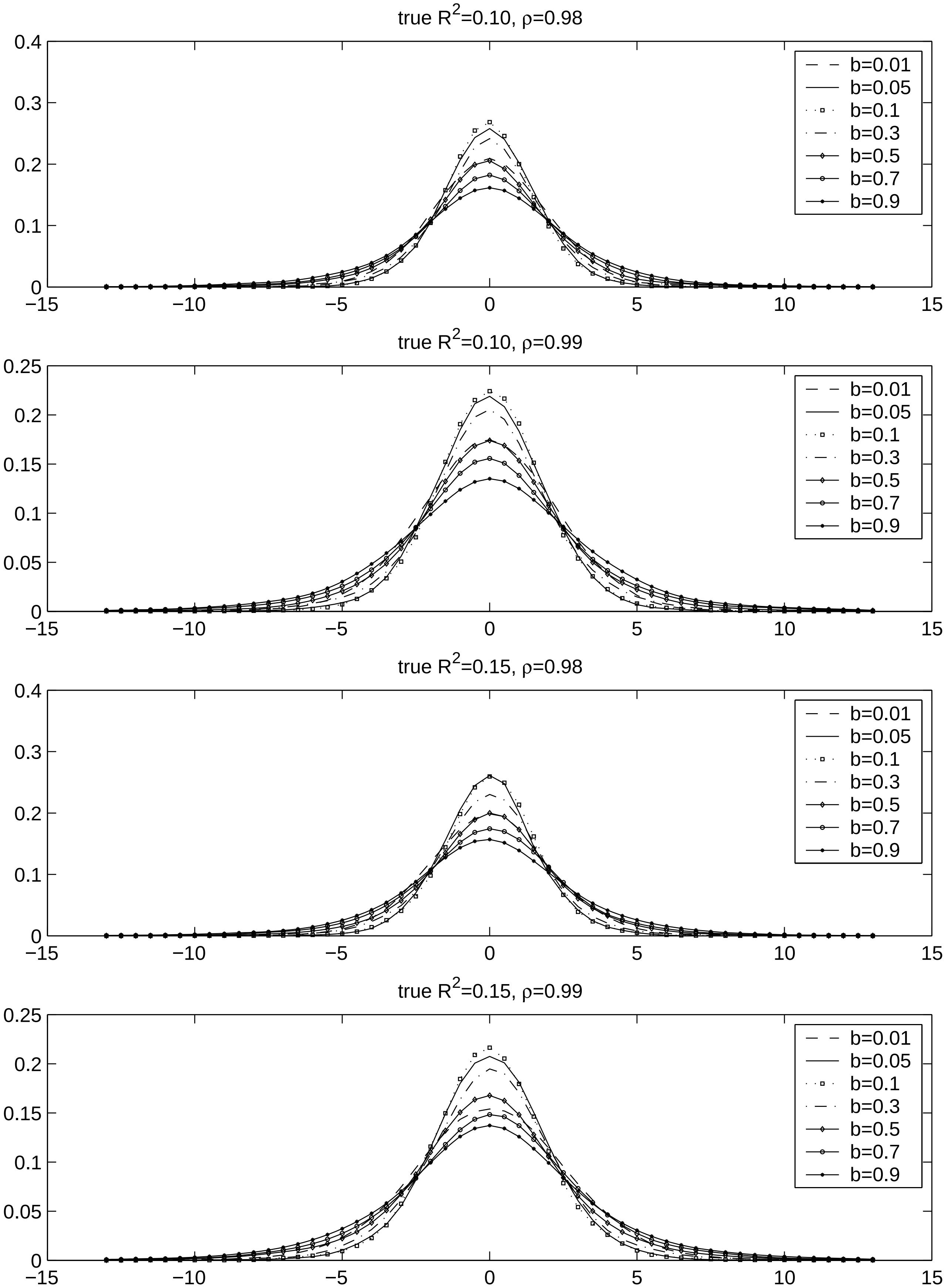

Figure 6: Convergent t distributions with varying b, true R^2 and ρ: $T = 824$

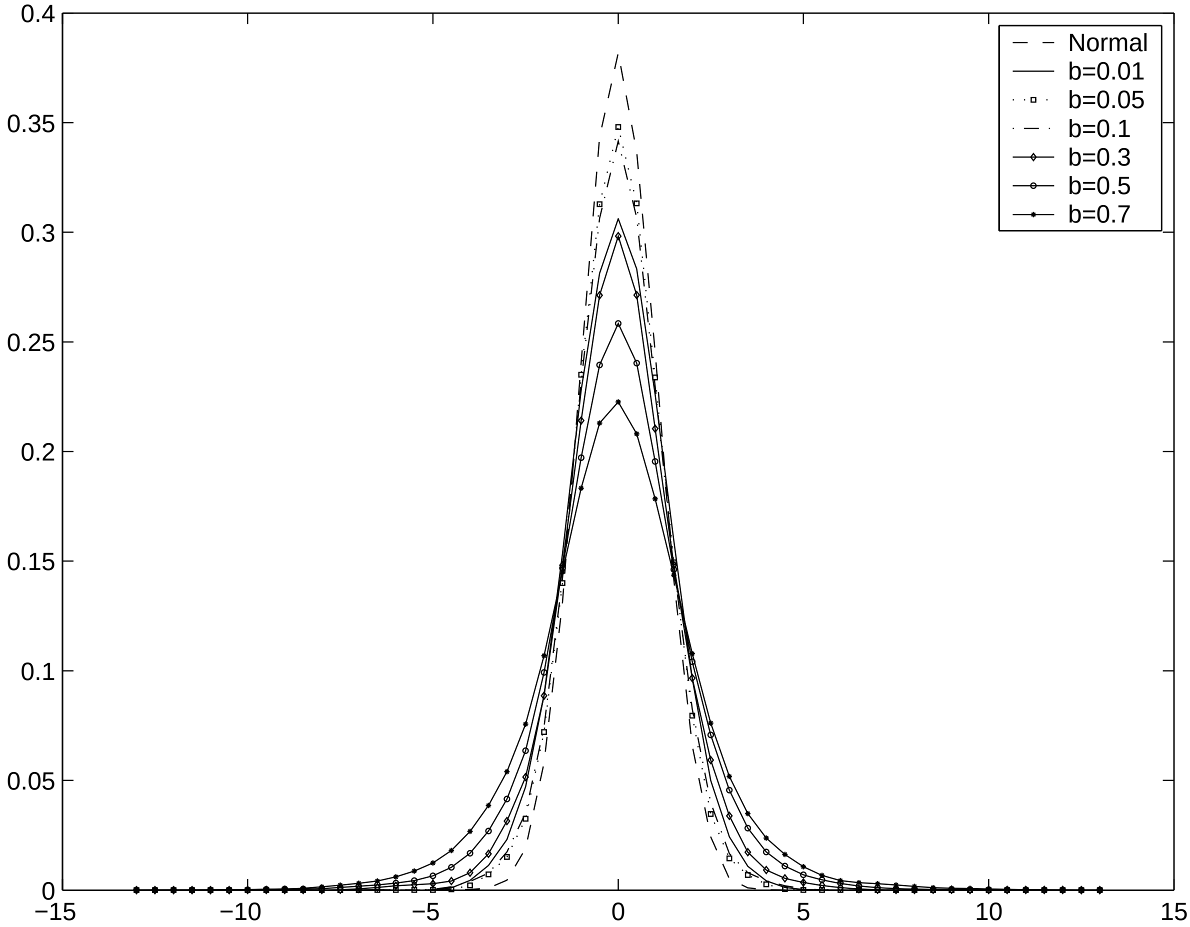

Figure 7: $T = 5000$: true $R^2 = 0.10, \rho = 0.98$.