Understanding spurious regression in financial economics

Ai Deng*

Boston University

Preliminary version: 25 Feb, 2005.

Abstract

In view of the fact that classic asymptotic theory can not provide satisfactory explanation for Ferson, Sarkissian and Simin's (2003a, 2003b) simulation findings on spurious regression in the context of financial economics, we develop an alternative distributional theory. Closely related is the well-known (nearly) observational equivalence issue in unit root testing literature. This study employs Nabeya-Perron type asymptotics and shows their simulation results can be well predicted. We hence re-enforce the fact that autocorrelation of dependent variable can not be used as an indication of spurious regression bias. Further, a convergent t test based on fix-b asymptotics following Sun (2005) is studied. Our simulation studies reveal further interesting result which explains and generalizes an illustrative simulation finding in FSS (2003a) and shows the asymptotic distribution of the convergent t statistic can be very close to standard normal if one chooses b properly. The interaction between spurious regression effect and data mining is also discussed.

Keywords: spurious regression, observational equivalence, Nabeya-Perron asymptotics, fix-b asymptotics, data mining.

*Address: Department of Economics, Boston University, 270 Bay State Road, Boston, MA, 02215. Email: dengai@bu.edu. I thank Professor Simin for providing me a data set and Professor Kurozumi for helpful discussion. Special thanks go to my advisor Professor Pierre Perron.

1 Introduction

A recent paper by Ferson, Sarkissian and Simin (2003a)[1], hereafter FSS, points out the possibility of type II spurious predictive regression[2] in financial economics. The authors use various simulations to show that if the expected (excess) return follows a persistent process, for which they provide some theoretical evidence and the so-called true R^2, to be defined in next section is not extremely low, then spurious regression bias will emerge so that one spuriously rejects the hypothesis of no predictability too often. This calls some of the predictability results in the literature into question. In a companion paper by the same authors (Ferson, Sarkissian and Simin (2003b)), they also discuss some possible solutions to this problem. This research has drawn the financial professionals' attention, e.g. Torous, Valkanov and Yan (2005), in their study of predictive regression using nearly integrated variables, explicitly excludes the possibility of the spurious regression of FSS (2003a, b) type; Amihud and Hurvich (2004) also explicitly make clear that their model is different and hence immuned from the spurious regression of their type. It is also cited in recent survey papers and empirical studies, see, for example, Rey (2004) and Wetherilt and Wells (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)"[3], 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 further comment "even though stock prices are not highly autocorrelated,..., thus one may think that spurious regression problems are unlikely. However, ..., there is a spurious regression bias..."

The first part of this paper shows that their spurious regression has close relation with classic spurious regression in an analytical manner and more importantly, how alternative asymptotic theory is needed to provide more insights into the problem at hand. The arguments are closely related to the well-known observational equivalence issue in econometrics, in particular, from the unit root literature. In doing so, we will first point out the difficulties of Phillips' (1986, 1988) results in explaining their simulations. Then, we propose to use Nabeya-Perron's (1994) asymptotic framework. It turns out the asymptotic theory explains all the simulation results reported in FSS (2003a, 2003b). Our results re-enforce their point, i.e. autocorrelation in dependent variable in a regression can not be used as an indication of the danger of spurious result. We also studied a convergent t test following Sun (2005). Interestingly, our Monte Carlo reveals explanation for one interesting simulation result reported in FSS (2003a). Specifically, FSS (2003a) shows that with other parameters, when sample size $T = 5000$, their choice of truncation lag $M = 240$ in the construction of HAC estimator yields a well-behaved t statistic 2.23, very close to 1.96. We show in this paper what is important is the proportion b defined as $b = M/T$. It is found in our simulation that

[1] As their paper, we only concern ourselves with short horizon predictive regression throughout the paper. Extension to long horizon regression is beyond the scope of the current project and left for future research.

[2] For a definition of the type I and type II spurious regression, see Chiarella and Gao (2002).

[3] where it has been shown that regression between two independent I (1) processes would result in spurious statistical significance of t statistic.

$b = 0.05$ yields closest approximation to standard normal, which can be seen to explain FSS (2003a) by noticing $240/5000 = 0.048 \approx 0.05$. We hence generalize their result. Though we show the convergent t test, by itself, does not completely solve the problem at hand, it does alleviate the problem significantly in large sample and shed more light on it. In the second section on Monte Carlo, we discuss further the issue of data mining and spurious regression bias. Our result confirms that the "largest R^2" data mining criterion is probabilistically associated with the more persistent stochastic variables. This, again, explains FSS's (2003a, b) simulation result. *Importantly, in this paper, we do not inquiry the empirical validity of their model, but rather we study their properties, taken their model as given, though some comments will be given concerning this.*

The rest of the paper is organized as follows. Section 2 provide a brief summary of FSS (2003) findings. Some of the results mentioned above will be restated and emphasized. Section 3 then will provide some classic spurious regression theory with nearly integrated series fully derived in Phillips (1988) as our background theory. Section 4.1 presents easily-derived analytical results for their model. Section 4.2 derives asymptotic theory suitable for spurious regression in this setting and the theory is of interest in its own right. Section 5 reports the simulation results concerning the behavior of the convergent t test. Section 6 concludes. 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 $W_{c_x}(t) = \int_0^t \exp(c(r-s))\,dW(s), V_{c_y}(t) = \int_0^t \exp(c(r-s))\,dV(s)$. 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[4]

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

[5]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

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

[4]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.

[5]The intercept term in FSS calibration is very close to 0. So in our theoretical development, we will assume a zero mean in the DGP for the dependent variable. Nevertheless, our qualitative results remain the same regardless of inclusion of a constant term.

They also assume the DGP for Z_t and Z_t^* to be[6]

$$\begin{pmatrix} Z_t^* \\ Z_t \end{pmatrix} = \begin{pmatrix} \rho^* & 0 \\ 0 & \rho \end{pmatrix} \begin{pmatrix} Z_{t-1}^* \\ Z_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^* \\ \varepsilon_t \end{pmatrix}$$

Where ε_t^* (variance of which is σ_*^2) is independent of ε_t. In words, they are considering a situation where researchers, in hope for predicting stock returns, come up with a completely irrelavent 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 shows using simulation the spurious significance of t statistic in the above decribed 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 we further 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

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

where

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

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

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, they used a testing procedure to determine the truncation lag M in the above variance estimator. Since we will frequently refer to FSS's simulation results, we summarize their main findings as follows.

1. the OLS estimate of β is well-behaved. In their case, $\hat{\beta}$ 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^**[7] *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 *huge* 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 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 re-enforce 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, we will provide asymptotic theory which could explain all these results. For result 4 above, our simulation reveals interesting explanation as mentioned in the introduction section of the paper.

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

3 Background theory: nearly integrated case

For reasons that will be clear soon, we first state some asymptotic results on spurious regression involving independent nearly integrated processes, which are special cases of Phillips (1988). We also state the asymptotic behavior of t statistic with HAC variance estimator. Since this is what is used in FSS simulation, it will be more important in the context than the naive t statistic. Specifically, we consider the regression as described in last section, using different but more general notations,

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{u}_t,\ t = 1, ..., T$$

and the DGP's generating y_t and x_t are

$$\begin{aligned} y_t &= \exp\left(\frac{c_y}{T}\right) y_{t-1} + v_t,\ t = 1, 2, ... \\ x_t &= \exp\left(\frac{c_x}{T}\right) x_{t-1} + w_t,\ t = 1, 2, ... \end{aligned}$$

$$\exp\left(\frac{c}{T}\right) \sim 1 + c/T$$

where v_t and w_t satisfies the assumption stated below. Let $\{\zeta_t\}_1^\infty$ be a sequence of random n-vectors defined on measure space (Ω, B, P) and $S_t = \sum_{j=1}^t \zeta_j$ be the partial sum process and set $S_0 = 0$. In our case, $\zeta_t = (v_t, w_t)'$. We use γ_s the sth order autocorrelation.

- Assumption 1 (Phillips(1986)): Error condition

  (1) $E(\zeta_t) = 0$ for all t;

  (2) $\sup_{i,t} E\left|\zeta_{i,t}\right|^{\beta+\epsilon} < \infty$ for some $\beta > 2$ and $\epsilon > 0$;

  (3) $\Sigma = \lim_{T\to\infty} T^{-1} E(S_T S_T')$ exists and is positive definite;

  (4) $\{\zeta_t\}_1^\infty$ is strong mixing with mixing numbers α_m satisfying $\sum_1^\infty \alpha_m^{1-2/\beta} < \infty$

Notice, importantly, the covariance matrix between v_t and w_t in spurious regression setup is diagonal so that the dependent and independent variables are independent with each other.

We will state the stochastic order of the statistics considered, for limiting distribution, see Phillips (1988, 1998).

- Assumption 2 (Sun (2005)): Kernel condition

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

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

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

Lemma 1 [8]*(Phillips (1988, 1998))*

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

Notice, the last result is reported in Phillips (1998) while the rest can be found in Phillips (1988). These results imply that for nearly integrated processes, spurious regression still exists. In particular, a conventional (HAC) t test on the regression coefficient will result in spurious significance[9]. Also notice that in this case DW statistics converges to 0 at rate T and residual autocrrelation converges to 1 at rate T as well. Therefore, Caution must be taken when the processes are not exactly integrated (of order 1), but rather close to integrated. This implication itself should not be a surprise at all, since nearly integrated process is asymptotically integrated (see Elliott and Stock, 1994).

Remark 1 *These results, to some extent, also explain in a local asymptotic framework some of the simulation results in Granger, Hyung, and Jeon (2001), where they consider the spurious regression issue with stationary series.*

4 Theory

4.1 Analytical results and observational equivalence

In this section, we try to achieve two goals. First, we will establish two simple lemmas derived directly from FSS's stock return model. And then based on these two lemmas, we show how the classic asymptotic theory introduced in the previous section can shed light on FSS's simulation results. However, we will also point out most of the interesting findings in FSS (2003a,b) can not be explained by the classic theory. Overall, we show there is a definite need for an alternative theory.

The following two lemmas provide easily derived analytical results and explain many of FSS simulation findings together with classic 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, We will set $\mu = 0$ from now on.

Lemma 2 *The model specified above implies the following ARMA(1,1)*[10] *representation for* r_{t+1}

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

[8] We ignore the results on OLS estimate of α since it is not the primary concern of our study.

[9] Recall the necessary condition on the relative rate of M and T for a consistent HAC estimator.

[10] 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, where we also impose an invertibility condition,

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

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

Remark 2 *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, the stylized fact that stock return itself has far less first order autocorrelation than the commonly used predicting variables, like dividend price ratio[11], implies there is more to say. Indeed, 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[12]. To match the feature of stock returns, ρ^* and $|\theta|$ must be close to each other. And the ARMA representation can be said to be observationally equivalent (with respect to the covariance structure) to the observed stock returns behavior. FSS notice that the difference between their model and the "classic" one 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". But as we have shown above, nonzero variance σ_u^2 is needed to accommodate the fact that the observed stock return is (observationally) not first order persistent, otherwise, this model can not be justified in the first place. However, what is true is that the process r_{t+1} still has a near unit root when the ρ^* is conjectured near unity. If ρ is also close to unity, then the regression of r_{t+1} on Z_t will incur the spurious bias. This is nothing but a direct implication of classic spurious regression results and lemma 1 developed above. And very intuitively, the bias depends on both ρ^* and $|\theta|$. The closer the two are, the less the bias (provides ρ^* is sufficiently close to 1, i.e. close to "problem region").*

Lemma 3 $\frac{\sigma_u^2}{\sigma_\eta^2} \uparrow 1$ *as* $\sigma_u^2 \to \infty$

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

Remark 3 *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*

[11] Campell, Lo and Mackinlay (1997)'s table 2.4 reports some estimated autocorrelations of the CRSP indexes on daily, weekly, and monthly frequencies. The estimated maximum first order autocorrelation 0.43 for CRSP equal-weighted index on 62:07:03-78:10:27 sample period, while at weekly and monthly frequencies (FSS studies monthly returns), the autocorrelation is a lot less with a maximum of 0.21.

[12] This should remind us of the well-known (nearly) observational equivalence in unit root testing. See Campbell and Perron (1991) for an excellent non-technical introduction.

variability of noise term becomes dominant in returns ($\sigma_u^2 \to \infty$), the return itself will behave more like "white noise"[13]*. Notice the connection with the true R^2 defined in FSS (2003a, 2003b). In their setup, $Var(Z^*)$ is a fixed constant*[14]*, therefore, controlling true R^2 as FSS (2003a, 2003b) did is equivalent to controlling σ_u^2 and hence to controlling the observational persistence of r_{t+1}, hence together with ρ^*, the likelihood of spurious regression. Let us sum up what we have learned—the smaller the true R^2 implies the larger the σ_u^2, the closer θ and $-\rho^*$, the stronger the cancellation, the less autocorrelation should be observed in stock returns, and the less the spurious bias (given ρ^* is close to 1). This is partially justified by what they find out through simulation. "...spurious regression bias does not arise to any serious degree, provided ρ^* is 0.90 or less, and the true R^2 is one percent or less". The first qualification is easy to understand in our analytical framework as well. Obviously, when ρ^* is not close to 1, r_{t+1} is not a nearly integrated process (so is the regressor Z_t since in most cases, they set $\rho^* = \rho$), hence spurious regression bias is not expected to occur*[15]*.*

To get a sense of the relative magnitude of ρ^* and the implied θ, in Table I, we report 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 we describe above.

The message of the above discussion is that FSS spurious regression may be well expected from our background theory since the variables involved are all close to the "problem" region.

With the knowledge we gain from above discussion, it is intuitively not surprising to learn some of the rest of their simulation results. For example, the spurious regression bias becomes severe when a more plausible $\rho \geq 0.95$ (since the predictor variables, like dividend yield, are usually highly autocorrelated.) and true $R^2 = 10\%$ is used in simulation. Also, it is not difficult to understand why FSS finds out "only mildly inflated residual autocorrelations...samples as large as $T = 2000$" whereas Phillips (1986) shows that the sample autocorrelation converges to 1 at rate T. FSS also provides an intuitive explanation of this observation. In fact, it can be better understood in current framework. As their simulation shows the slope estimate is relatively well-behaved, the residual inherits the ARMA structure of stock returns and due to the cancellation effect of θ and ρ, one should *not* expect to see much *observational* first order autocorrelation[16].

Although the FSS framework is seen to be to closely related to the classic setting (as the nearly integrated case), the asymptotic theory developed by Phillips (1986) is not expected to be an appropriate approximation in finite sample of their case. In particular, classic asymptotic theory is unable to provide an explanation for the following two major findings in FSS (2003a, b): (1) the OLS estimate of slope parameter $\hat{\beta}$ is well-behaved, (2) the residual autocorrelation is not highly inflated (given Phillips (1986) shows it converges to

[13]See also Campbell (2001).

[14]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.

[15]FSS also find that when $\rho^* = 0$ there is no spurious regression bias. This result also follows straightforwardly from the above discussion since then r_{t+1} becomes stationary.

[16]Thus, this model has pretty strong implication concerning the (excess) stock return process. Hence, their model could be checked by estimating such a model from the data series.

1). In next section, we will provide an alternative asymptotic theory for this and other simulation results.

4.2 More asymptotic theory: nearly integrated, nearly white noise

One of the reasons that Phillips' (1986) asymptotics may not be adequate in explaining FSS's findings is because his asymptotics do 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 Ng and Perron (1996). The asymptotics are based on the following local-to-unity specification. Assuming $\{e_t\}$ to be $i.i.d.\,(0, \sigma_e^2)$,

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

y_t is then nearly integrated in finite sample, but a white noise in the limit. What we will study here is the issue of spurious regression in this framework. Specifically, we develop asymptotic theory when a nearly white noise process y_t is regressed on an independent (nearly) integrated process x_t[17] and an intercept term, the estimate of which is denoted $\hat{\alpha}$. We study the behavior of several statistics. Results are collected in the following theorem. Once again, since FSS used the HAC-based t test in their simulation, we will report this particularly important asymptotic result as well. And these results are of independent interest, too. 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 we only present the result on residual first order autocorrelation, but as we show in the appendix, higher order autocorrelations can be easily obtained.

Theorem 1 *1.*

$$\sqrt{T}\hat{\beta} \Rightarrow \frac{\sigma_w \sigma_e \left\{ \int_0^1 \left(e_\infty(r) + \delta V_c(r)\right) W_{c_x}(r)\, dr - \delta \int_0^1 V_c(r)\, dr \int_0^1 W_{c_x}(t)\, dt \right\}}{\sigma_w^2 \left\{ \int_0^1 W_{c_x}(t)^2\, dt - \left(\int_0^1 W_{c_x}(t)\, dt \right)^2 \right\}} \triangleq \frac{\sigma_e}{\sigma_w} \kappa$$

2.

$$\hat{\alpha} \Rightarrow \sigma_e \left(\delta \int_0^1 V_c(r)\, dr - \kappa \int_0^1 W_{c_x}(t)\, dt \right)$$

[17]Durlauf and Phillips (1988) study the asymptotic regression theory of y_t being stationary, and the regressors include x_t being integrated, an intercept and a first order time trend.

3.

$$
\begin{aligned}
T^{-1/2}t_{\beta} &\Rightarrow \mu/v^{1/2} \text{ where,}\\
\mu &= \int_0^1 \left(e_{\infty}(r)+\delta V_c(r)\right)W_{c_x}(r)\,dr-\delta\int_0^1 V_c(r)\,dr\int_0^1 W_{c_x}(t)\,dt\\
v &= \left(1+\delta^2\int_0^1 V_c(r)^2\,dr-\delta^2\left(\int_0^1 V_c(r)\,dr\right)^2\right)\left(\int_0^1 W_{c_x}(t)^2\,dt-\left(\int_0^1 W_{c_x}(t)\,dt\right)^2\right)\\
&\quad -\left\{\int_0^1 \left(e_{\infty}(r)+\delta V_c(r)\right)W_{c_x}(r)\,dr-\delta\int_0^1 V_c(r)\,dr\int_0^1 W_{c_x}(t)\,dt\right\}^2
\end{aligned}
$$

4.

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

5.

$$
\begin{aligned}
DW &\Rightarrow \varsigma/\tau, \text{ where,}\\
\varsigma &= 2\\
\tau &= \left(1+\delta^2\int_0^1 V_c(r)^2\,dr\right)-\kappa^2\left\{\int_0^1 W_{c_x}(t)^2\,dt-\left(\int_0^1 W_{c_x}(t)\,dt\right)^2\right\}
\end{aligned}
$$

6.

$$
\begin{aligned}
\gamma_1 &= \frac{\sum_2^T \hat{u}_t\hat{u}_{t-1}}{\sum_1^T \hat{u}_t^2} \Rightarrow \frac{B}{D}, \text{ where}\\
B &= \delta^2\int_0^1 V_c(r)^2\,dr-\delta^2\left(\int_0^1 V_c(r)\,dr\right)^2\\
&\quad -2\kappa\left\{\int_0^1 \left(e_{\infty}(r)+\delta V_c(r)\right)W_{c_x}(r)\,dr-\delta\int_0^1 V_c(r)\,dr\int_0^1 W_{c_x}(t)\,dt\right\}\\
&\quad +\kappa^2\left[\int_0^1 W_{c_x}(r)^2\,dr+\int_0^1 W_{c_x}(r)\,dW(r)-\left(\int_0^1 W_{c_x}(r)\,dr\right)^2\right]\\
D &= \left(1+\delta^2\int_0^1 V_c(r)^2\,dr-\delta^2\left(\int_0^1 V_c(r)\,dr\right)^2\right)-\kappa^2\left\{\int_0^1 W_{c_x}(t)^2\,dt-\left(\int_0^1 W_{c_x}(t)\,dt\right)^2\right\}
\end{aligned}
$$

7.

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

where

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

The above results are interesting. They show some both qualitatively and quantitatively different properties from most of the existing spurious regression theory.

Remark 4 *Result 1 and 2 contrast with others by showing that OLS estimate $\hat{\beta}$ is still consistent at usual rate $\sqrt{T}$ and now the intercept estimate has a nondegenerate limit distribution. This explains why FSS simulation finds $\hat{\beta}$ to be actually well-behaved. See their Table II and related discussion.*

Remark 5 *Result 3 says that the conventional t statistics will still diverge at the rate $\sqrt{T}$ resulting in spurious rejection. In view of result 1, we know the exclusive reason for diverging t statistic is the inconsistent estimate of the variance term. This also justifies FSS's suggestion that in finite sample, variance estimate, but not OLS slope estimate, is the major concern. Therefore, both simulation and asymptotic theory indicates the way to solve the problem of spurious regression is to get a well-behaved variance estimate.*

Remark 6 *Result 4 shows R^2 has a nondegenerate limiting distribution and hence FSS's theoretical R^2 derived from the F distribution is not appropriate. 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 different from previous result.*

Remark 7 *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 finds that the autocorrelation not inflated and also points out this result is not compatible with Phillips (1986)'s theory. Here, we show under our asymptotic framework their Monte Carlo result can be explained as r no longer converges.*

Remark 8 *Result 7 shows further 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. This result is relevant here because it explains FSS spurious rejection result.*

From the above results, we believe 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.

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. It is shown in many situations that such an estimator can improve the performance of a test statistic, which is how such estimator was motivated. 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. We 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[18], *but they don't recommend it due to extreme large sample they believe it requires and the difficulties (or arbitrariness) in deciding how "long" the lag should be.* We will further comment on this piece of their finding in next section. We use $t_{\beta,b}$ to denote this version of t statistic and signify the dependence of it on b. It is stated as a corollary to theorem 1.

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

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

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, we 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't consistently estimated without strong assumptions[19], 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 we have two localizing parameters in stead of one, which could result in very conservative procedures. We do not pursue this direction in the present paper, but hope to investigate these in future

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

[19]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.

research. Another possible extension is to study the finite sample approximation of these asymptotic distributions. It can be done by either simulation or numerical integration. In the next section on Monte Carlo, we will report the simulation concerning convergent t test and the interaction between data mining and spurious regression bias. Further interesting finding will emerge.

5 Monte Carlo

5.1 Convergent t statistic

In this subsection, we try to answer the following question: can we reduce the size of spurious problem by using this statistic with normal critical values[20]? A yes to this question means we can be agnostic about the spurious bias and conventional distribution is (approximately) valid. The answer to this question certainly depends on the kernel and the fixed proportion b one chooses. We hence conduct a simulation study. We will be using the Bartlet kernel as in Sun (2005) and leave the other choices of kernel as future research. All simulation is done with 5000[21] replications in programming language Matlab. The other parameter values we use in simulation are:

1. $T : T \in (66, 824)$
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. $b : b \in (0.01, 0.025, 0.05, 0.10, 0.20, ..., 0.9)$

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 to standard normal is the best. Specifically, we accomplish the following two experiments.

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 Barlet 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 $T = 5000$.

[20] 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.

[21] 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 1000th decimal. Hence, we can safely compare our simulation results with theirs.

We simulate the data series the same way as FSS (2003a) did. We record some representative entries for the 97.5% critical $t_\beta^{w/o}$(i.e. the statistic with $M = T$) in Table II and those for $t_{\beta,b}$ in Table III for $b = 0.01, 0.025, 0.05, 0.1, 0.2, 0.3$. We also plotted the kernel smoothed density functions for a variety of parameter values in Figure 6.1 through 6.4[22]. Results for other values of b are available upon request.

The general conclusion is that all the distributions have heavier tails than the standard normal, hence leads 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), in that this distribution has a lot thicker tail. On a micro scale, we want to compare the relative closeness of these distributions to the standard normal. We find 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 Figures 1 through 4[23]. If we compare the t values with $b = 0.05$ with FSS Table II, we find 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, though in terms of testing, these two values are equally "bad". That $b = 0.10$ (also $b = 0.05$) is preferable in large sample is surprisingly the same as Sun (2005), who found the same b produces the most desirable result in his fractional integration context. However, with $T = 66$, smaller b seems to be slightly better. This is not a surprise, though, because for a 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 and relevant result. We therefore conclude from this limited set of simulations, with a large enough 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 provide further guidance, we try to provide a sense of largeness and the corresponding size distortion incured by using the convergent t test. We thus studied three representative sample sizes 1000, 1250, 1500 with the following specification—true $R^2 = 0.10$, $\rho = 0.98$. We find 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 $b = 0.05$. We also verify that for these sample sizes that $b = 0.05$ and 0.1 are indeed the preferable choices.

What is really interesting is that FSS reported a well-behaved t value 2.23[24] 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. We also verified 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

[22]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.

[23]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.

[24]Our simulation gave the same value.

of 3.2882. Figure 6.5 plots some of the densities with different b/s. And we can see that when $b = 0.05$ and 0.10, the distributions are very close to standard normal. Further, with $T = 5000$, the "finite" sample distribution would not be very different from the asymptotic distribution. 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, we believe these values of b are indeed systematic.

Thanks to our finding, we have explained the simulation FSS conducted that their t statistic is close to 1.96 because they happened to choose $b = 0.05$ which is the proportion that makes the distribution in Corollary 1 close to standard normal in large sample. Although by itself, the simple rule of thumb doesn't resolve the problem completely, it does greatly alleviate in large sample and shed more light on the problem at hand. Notice, when $b = 0.05$ and $b = 0.10$, the densities are very close to standard normal in large sample, see figure 5.

Finally, we note that in usual practice, the maximum lag is commonly set to be $Int(12 \times (T/100)^{1/4})$. This rule of thumb maximum lag is very small compared with the deterministic rule of $b = 0.05$ when sample size is large, say over 400, which is common in financial data. In the case of $T = 824$, fixed b-0.05 rule will dictate a lag value of 41, while conventional method implies a *maximum* lag value of 21.

We summarize our simulation findings as follows.

a The distributions of t statistic with $M = T$ represents a substantial departure from the standard normal.

b 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.

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

5.2 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. In this section, we establish a theoretical justification for this Monte Carlo result. Using our asymptotic theory, we simulate the mean and median values of the R^2 as a function of localizing parameters c and δ. These are plotted in Figure 6.6 and Figure 6.7. In unreported simulation, we find the distribution of R^2 is unimodal, looking like a χ^2 distribution. Hence, the mean and median are two relevant measures of the location of large probability mass. What we find provides surportive 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.

6 Conclusion

The predicting variable in return predictive regression in finance is usually highly autocorrelated. By postulating the expected return to be persistent as well, 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 results different from previous spurious regression theory, which leads them to comment that their finding is well outside of the classic setting. In this paper, we derived simple implications of their model using observational equivalence argument and provide new asymptotic theory for their Monte Carlo results. In particular, the use of a new asymptotic framework, namely, nearly integrated-nearly white noise framework, shows that finite sample behavior of several statistics can be predicted by asymptotic theory. Altogether, we provide a unifying way of looking at their results and the "classic" spurious regression results. The main conclusion is the autocorrelation of 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 doomed low power of unit root test, see discussion on observational equivalence in Campbell and Perron (1991). A convergent t statistic is also constructed, whose properties are studied in Monte Carlo simulation. Our finding allows us to be able to explain an interesting simulation result in FSS (2003a, b). A probablistic interpretation is provided for the interaction between data mining and spurious regression bias. The implications from this study are general in and outside of financial econometrics framework, in particular, it may have implications for monetary macroeconomics. Further econometric extensions include the study of alternative kernels, and see if any other could make the distribution closer to standard normal hence eventually eliminate the spurious regression problem. Of particular interest is the recently proposed sharp orgin kernel. This is the subject of ongoing research.

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

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

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

5. Campell, John and Motohiro Yogo, 2004, Efficient tests of stock return predictability, manuscript, Harvard University.

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

7. 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.

8. Cochrane, J. 2001. Asset Pricing (Princeton University Press).

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

10. Granger, Clive W.J., Namwon, Hyung, and Yongil, Jeon, 2001, Spurious regression with stationary series, Applied economics 33, 899-904

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

12. 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.

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

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

15. 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.

16. 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

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

18. 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

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

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

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

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

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

24. 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.

25. Rey, David, 2004, Stock market predictability: Is it there? A Critical Review. WWZ/Department of Finance, Working Paper No. 12/03

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

27. ..., ... 2005, A convergent t-statistic in spurious regressions, Forthcoming in Econometric Theory

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

29. Wetherilt, Anne Vila and Simon Wells, 2004. Long horizon equity return predictability: some new evidence for the United Kingdom, Bank of England, working paper no. 244.

30. 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

$X_t = \sum_{j=1}^{t} \exp\left(\left(t-j\right) c/T\right) e_t$ is a near-integrated process. 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 $e_\infty(r) = \lim_{T\to\infty} e_{[Tr]}/\sigma_e$. Then as $T \to \infty$, (a) $T^{-1}\sum_{t=1}^{T} y_{t-1}^2 \Rightarrow \sigma_e^2\left(1 + \delta^2 \int_0^1 V_c(r)^2\, dr\right)$; (b) $T^{-1}\sum_{t=1}^{T} y_{t-1}u_t \Rightarrow -\sigma_e^2$, (c) $y_{[Tr]} \Rightarrow \sigma_e^2\left(e_{\infty r} + \delta V_c(r)\right)$, and (d) $T^{-1}\sum_{t=1}^{T} u_t^2 \Rightarrow 2\sigma_e^2$.*

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

Proof of Lemma (A.2).

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

where the last relation follows from an LLN for $\{e_t\}$ and convergence results for a_T and b_T ■

Proof of theorem 1.

$$\begin{aligned}
T^{1/2}\hat{\beta} &= \frac{T^{-3/2}\sum y_t x_t - T^{-1/2}\bar{y}\bar{x}}{T^{-2}\sum\left(x_t - \bar{x}\right)^2} \\
&\Rightarrow \frac{\sigma_w \sigma_e \left\{\int_0^1 \left(e_\infty(r) + \delta V_c(r)\right) W_{c_x}(r)\, dr - \delta \int_0^1 V_c(r)\, dr \int_0^1 W_{c_x}(t)\, dt\right\}}{\sigma_w^2 \left\{\int_0^1 W_{c_x}(t)^2\, dt - \left(\int_0^1 W_{c_x}(t)\, dt\right)^2\right\}} \triangleq \frac{\sigma_e}{\sigma_w}\kappa
\end{aligned}$$

because of standard asymptotic results for x_t.

For the estimated intercept term, we have,

$$\begin{aligned}
\hat{\alpha} &= T^{-1}\sum y_t - \hat{\beta} T^{-1}\sum x_t \\
&= \bar{y} - T^{1/2}\hat{\beta} T^{-3/2}\sum x_t \\
&\Rightarrow \delta\sigma_e \int_0^1 V_c(r)\, dr - \sigma_e\kappa \int_0^1 W_{c_x}(t)\, dt
\end{aligned}$$

For (3), first consider

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

Now,

$$
\begin{aligned}
T^{-1/2}t_\beta &= \frac{T^{1/2}\hat{\beta}\left(T^{-2}\sum (x_t-\bar{x})^2\right)^{1/2}}{s} \\
&\Rightarrow \frac{\sigma_e\left\{\int_0^1 (e_\infty(r)+\delta V_c(r))\,W_{c_x}(r)\,dr-\delta\int_0^1 V_c(r)\,dr\int_0^1 W_{c_x}(t)\,dt\right\}}{\sigma_w\left\{\int_0^1 W_{c_x}(t)^2\,dt-\left(\int_0^1 W_{c_x}(t)\,dt\right)^2\right\}} \\
&\quad \times\sigma_w\left\{\int_0^1 W_{c_x}(t)^2\,dt-\left(\int_0^1 W_{c_x}(t)\,dt\right)^2\right\}^{1/2} \\
&\quad \div\sigma_e\left\{\left\{1+\delta^2\int_0^1 V_c(r)^2\,dr-\delta^2\left(\int_0^1 V_c(r)\,dr\right)^2\right\}\right. \\
&\quad \left.-\kappa^2\left\{\int_0^1 W_{c_x}(t)^2\,dt-\left(\int_0^1 W_{c_x}(t)\,dt\right)^2\right\}\right\}^{1/2} \\
&= \left\{\int_0^1 (e_\infty(r)+\delta V_c(r))\,W_{c_x}(r)\,dr-\delta\int_0^1 V_c(r)\,dr\int_0^1 W_{c_x}(t)\,dt\right\} \\
&\quad \div\left\{\left(1+\delta^2\int_0^1 V_c(r)^2\,dr-\delta^2\left(\int_0^1 V_c(r)\,dr\right)^2\right)\left(\int_0^1 W_{c_x}(t)^2\,dt-\left(\int_0^1 W_{c_x}(t)\,dt\right)^2\right)\right. \\
&\quad \left.-\left\{\int_0^1 (e_\infty(r)+\delta V_c(r))\,W_{c_x}(r)\,dr-\delta\int_0^1 V_c(r)\,dr\int_0^1 W_{c_x}(t)\,dt\right\}^2\right\}^{1/2} \\
&= \mu/v^{1/2}
\end{aligned}
$$

, proving (b). Next,

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

We next consider Durbin Watson statistic,

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

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

since the last two terms go to 0. Therefore

$$DW = \frac{2}{\left(1 + \delta^2 \int_0^1 V_c(r)^2 dr - \delta^2 \left(\int_0^1 V_c(r) dr\right)^2\right) - \kappa^2 \left\{\int_0^1 W_{c_x}(t)^2 dt - \left(\int_0^1 W_{c_x}(t) dt\right)^2\right\}}$$

Now, consider r_1. We have shown that

$$\begin{aligned} s^2 &= T^{-1}\sum_1^T \hat{u}_t^2 = T^{-1}\sum_1^T \left(y_t - \bar{y} - \hat{\beta}(x_t - \bar{x})\right)^2 \\ &\Rightarrow \sigma_e^2 \left(1 + \delta^2 \int_0^1 V_c(r)^2 dr - \delta^2 \left(\int_0^1 V_c(r) dr\right)^2\right) - \kappa^2 \sigma_e^2 \left\{\int_0^1 W_{c_x}(t)^2 dt - \left(\int_0^1 W_{c_x}(t) dt\right)^2\right\} \end{aligned}$$

Consider

$$\begin{aligned} T^{-1}\sum_2^T \hat{u}_t \hat{u}_{t-1} &= T^{-1}\sum_2^T \left(y_t - \bar{y} - \hat{\beta}(x_t - \bar{x})\right)\left(y_{t-1} - \bar{y} - \hat{\beta}(x_{t-1} - \bar{x})\right) \\ &= T^{-1}\sum_2^T (y_t - \bar{y})(y_{t-1} - \bar{y}) - T^{-1}\sum_2^T \hat{\beta}(x_{t-1} - \bar{x})(y_t - \bar{y}) \\ &\quad - T^{-1}\sum_2^T \hat{\beta}(x_t - \bar{x})(y_{t-1} - \bar{y}) + T^{-1}\sum_2^T \hat{\beta}^2 (x_t - \bar{x})(x_{t-1} - \bar{x}) - - - (A) \end{aligned}$$

The first term in (A):

$$\begin{aligned} T^{-1}\sum_2^T (y_t - \bar{y})(y_{t-1} - \bar{y}) &= T^{-1}\sum_2^T y_t y_{t-1} - T^{-1}\sum_2^T y_t \bar{y} - T^{-1}\sum_2^T y_{t-1}\bar{y} + T^{-1}\sum_2^T \bar{y}^2 \\ &= T^{-1}\sum_2^T y_t y_{t-1} - \bar{y}^2 \end{aligned}$$

So consider

$$\begin{aligned}
T^{-1}\sum_{2}^{T} y_t y_{t-1} &= T^{-1}\sum_{2}^{T}\left(y_{t-1}+u_t\right)y_{t-1} = T^{-1}\sum_{2}^{T} y_{t-1}^2 + T^{-1}\sum_{2}^{T} u_t y_{t-1} \\
&\Rightarrow \sigma_e^2\left(1+\delta^2\int_0^1 V_c\left(r\right)^2 dr\right) - \sigma_e^2 = \sigma_e^2\delta^2\int_0^1 V_c\left(r\right)^2 dr
\end{aligned}$$

The Second term in (A):

$$\begin{aligned}
T^{-1}\sum_{2}^{T}\hat{\beta}\left(x_{t-1}-\bar{x}\right)\left(y_t-\bar{y}\right) &\rightarrow \left(T^{1/2}\hat{\beta}\right)T^{-3/2}\sum_{2}^{T}\left(x_{t-1}-\bar{x}\right)y_t - \left(T^{1/2}\hat{\beta}\right)T^{-3/2}\bar{y}\sum_{2}^{T}\left(x_{t-1}-\bar{x}\right) \\
&= \left(T^{1/2}\hat{\beta}\right)\left[T^{-3/2}\sum_{2}^{T} x_{t-1}y_t - T^{-1/2}\bar{x}T^{-1}\sum_{2}^{T} y_t\right. \\
&\qquad \left. -\bar{y}T^{-3/2}\sum_{2}^{T} x_{t-1} + \bar{y}T^{1/2}\bar{x}\right] \\
&= \sigma_e^2\kappa\left\{\int_0^1\left(e_\infty\left(r\right)+\delta V_c\left(r\right)\right)W_{c_x}\left(r\right)dr - \delta\int_0^1 V_c\left(r\right)dr\int_0^1 W_{c_x}\left(t\right)dt\right\}
\end{aligned}$$

The Third term in (A):

$$T^{-1}\sum_{2}^{T}\hat{\beta}\left(x_t-\bar{x}\right)\left(y_{t-1}-\bar{y}\right) \Rightarrow \sigma_e^2\kappa\left\{\int_0^1\left(e_\infty\left(r\right)+\delta V_c\left(r\right)\right)W_{c_x}\left(r\right)dr - \delta\int_0^1 V_c\left(r\right)dr\int_0^1 W_{c_x}\left(t\right)dt\right\}$$

The last term in (A):

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

So,

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

Finally,

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

higher order autocorrelation can be easily obtained analogously.

Next, consider the HAC t test. First,

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

Then, following Sun (2005), we write $\frac{T^2}{M} \sigma_{HAC}^2$ as

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

where

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

Therefore,

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

■

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

Bartlet Kernel

$T = 66$

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 $t_{\not\beta,b}$ with truncation

Bartlet Kernel									
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.

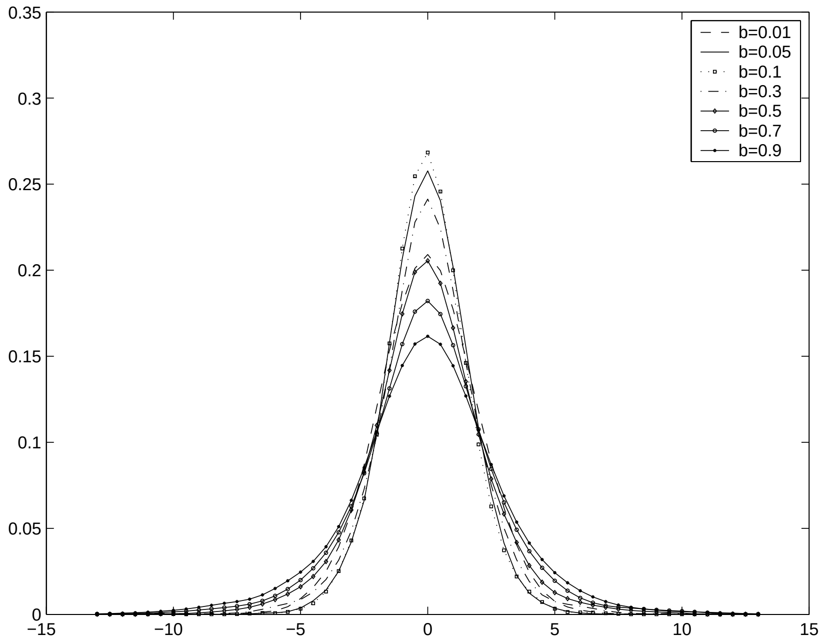

Figure 6.1: $T = 824$: kernel estimates of densities of $t_{\beta,b}$ when true $R^2 = 0.10, \rho = 0.98$

Note: *in figure 4.1 to 4.4, the bench mark normal density is not plotted to ensure a clear picture of the densities of the statistic. It is useful to note that the standard normal density evaluated at the mean value 0 is 0.398, a lot higher than the densities shown in these four figures. But note in Figure 4.5, normal density is plotted.*

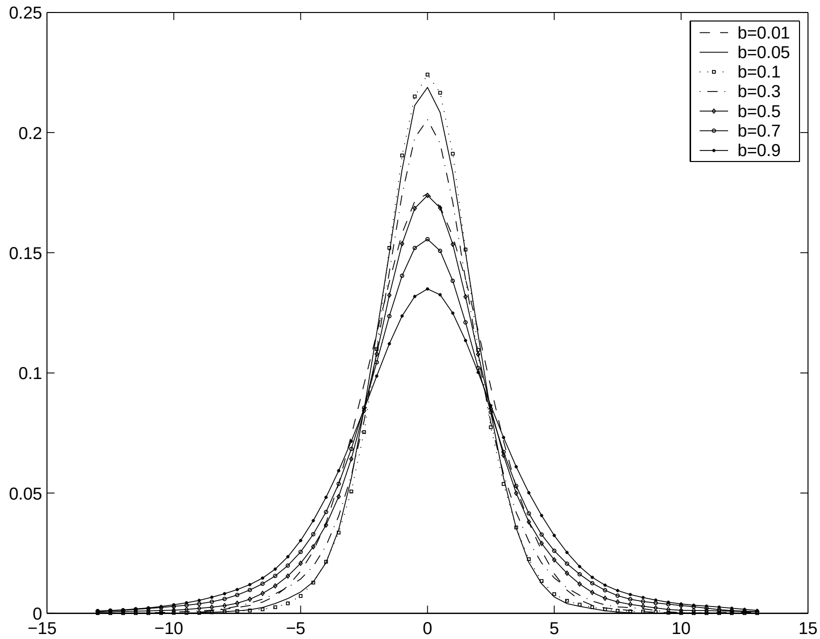

Figure 6.2: $T = 824$: kernel estimates of densities of $t_{\beta,b}$ when true $R^2 = 0.10, \rho = 0.99$

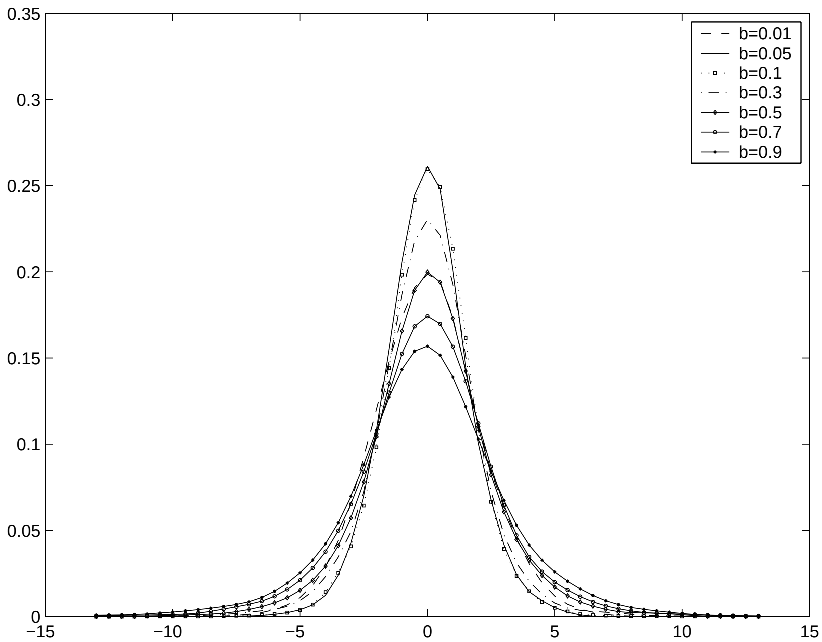

Figure 6.3: $T = 824$: kernel estimates of densities of $t_{\beta,b}$ when true $R^2 = 0.15, \rho = 0.98$

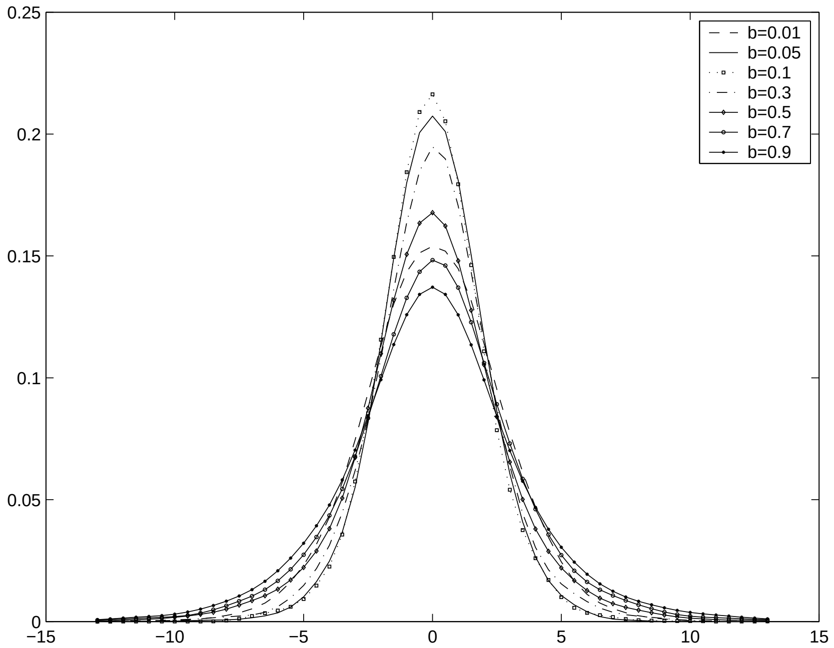

Figure 6.4: $T = 824$: kernel estimates of densities of $t_{\beta,b}$ when true $R^2 = 0.15, \rho = 0.99$

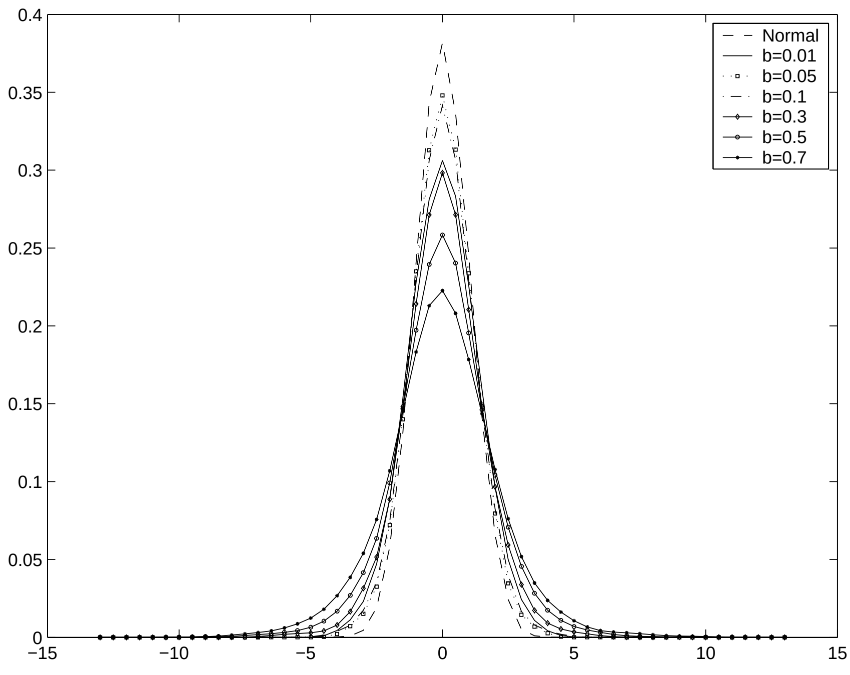

Figure 6.5: $T = 5000$: true $R^2 = 0.10, \rho = 0.98$

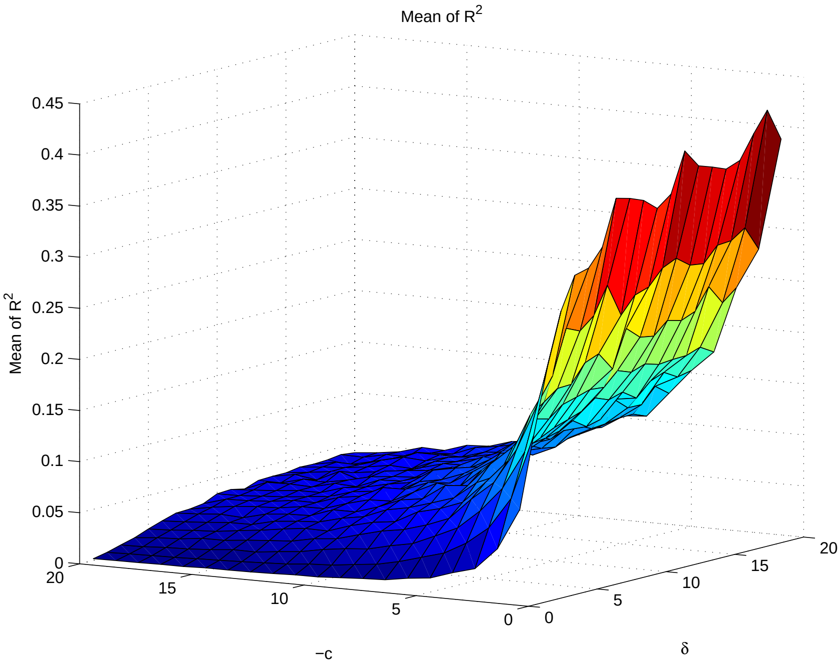

Figure 6.6: Mean of R^2 as a function of $-c$ and δ

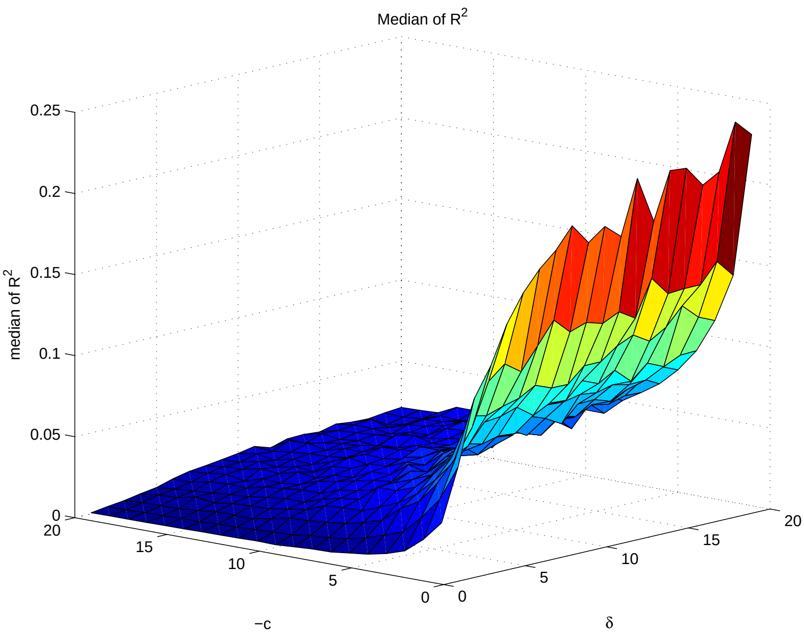

Figure 6.7: Median of R^2 as a function of $-c$ and δ