Centro de Investigación
y Docencia Económicas, A. C.

CIDE

Stationary and Trend Stationary Processes with Level and Trend Breaks: Central Limit Theorems and Asymptotic Distribution

Tesina

que para obtener el grado de

MAESTRO EN ECONOMÍA

presenta

ALFONSO BASULTO MUÑOZ

Director de la Tesina:
DR. DANIEL VENTOSA-SANTAULÀRIA

Ciudad de México Junio, 2019

"Opinion is the medium between knowledge and ignorance"- Plato, The Republic

Abstract

This thesis provides an analytical study of the limiting distribution of the typical statistic employed in the central limit theorem, albeit with a minor twist. An asymptotic theory is developed for the aforementioned statistic involving data generated from stationary and trend stationary processes with level and trend breaks. Four case-specific theorems are proved and Monte Carlo simulation is utilized in order to both confirm empirically that these results hold and to provide evidence of how variations in the parameterization of the underlying data generating processes affect the normality of the central limit theorem statistic. Finally, two possible applications of these theoretical results, pertaining to a context of ordinary least squares linear regression, are presented and discussed in detail. These applications exploit intermediate results obtained in the process of proving the main theorems and the contrapositives of the theorem statements so as to bring about derivations of the limiting distribution of regression coefficients and a potential independent variable exogeneity test statistic.

Keywords: time series, stationary, trend stationary, asymptotic theory, central limit theorem, limiting distribution

Acknowledgements

To Dr. Daniel Ventosa-Santaulària, without whose support and guidance the idea for this thesis would have been neither conceived nor consummated.

To my first classmates, roommates and now friends, Jenny and Javier, with whom I shared nights of endless jokes and laughter during the last two years.

To my entire family and, in particular, to my mom, Graciela, my sister, Gloria, and my brother, Raphael, for the constant emotional support and encouragement.

To José Luís and Guilda who have been there for me ever since my return to Mexico. I wouldn't be here without you. And to their son, Adrián, and daughter, Guida, for being my second family.

To Melissa Floca, for being my mentor during the last three years and adding structure to my academic and professional goals.

And finally, to my dog, Halle, for being herself.

Thank you all for existing.

Contents

List of Tables

Introduction

Correlation does not imply causation, or so goes the now ubiquitous saying. Indeed, much can be said about statistical relationships between two seemingly unrelated variables which not only turn out to be specious but, in many instances, also whimsical. Such is the case with the apparent dependence between variables such as *the number of letters in the winning word of the Scripps National Spelling Bee* and *the number of people killed by venomous spiders*. Examples like these abound, some less deceptive and beguiling than others, and can be found in works such as *Spurious Correlations* (Vigen, 2015). This phenomenon is undeniably pervasive and, as such, has been studied in myriad ways.

The recent history of research into spurious regressions in econometrics can be traced back to work regarding spurious models where sufficient care is not taken over the autocorrelation structure of the errors and its formulation in the regression equation (Box and Newbold, 1971). Earlier simulation findings that relate integrated random processes shed light on the ease with which spurious relations can arise in regressions involving two non stationary processes (Granger and Newbold, 1974). More concretely, Granger and Newbold studied how simple linear regressions between two independent AR(1) processes resulted in statistically significant coefficients, large R^2's and small Durbin-Watson statistics more often than should be expected by chance. Twelve years later, Phillips used Granger and Newbold's framework as the basis for the development of asymptotic theory for regression coefficients and for conventional significance tests in the context of ordinary least squares. He showed that the usual t and F test statistics do not possess limiting distributions and diverge as the sample size tends to infinity (Phillips, 1986). Evidently, this translates into the rejection of the null hypothesis of no relation between the explanatory and dependent variable more often than what would be expected in a linear regression relating two independent stochastic processes. Phillips concludes that the nonstationarity of these time series is the underlying cause of spurious regression.

Finally, Tsay and Chung extend the theoretical analysis of spurious regressions from the I(1) processes to the long memory fractionally integrated processes. They find that regressions between two long memory fractionally integrated processes, whether they're stationary or not, result in divergent t ratios and spurious effects (Tsay and Chung, 2000). These two authors conclude that the cause of spurious effects is not the nonstationarity of the processes but the long memory. It's from the spirit of the work done by these aforementioned authors that the following work gets its motivation.

Despite the fact that the fundamental purpose of the results that will be proved here is purely theoretical, they are based on a similar result proved by my thesis advisor which was later applied in a paper on financial econometrics (Osterrieder et al., 2018). Notwithstanding, I'd like to emphasize that the results presented here were derived with no particular application in mind. In *The VIX, the Variance Premium, and Expected Returns*, it was proved that if you multiply a fractionally integrated stochastic process by an independent $I(0)$ noise, then the mean of that product multiplied by the root of the sample size will have an asymptotically gaussian distribution. That is quite a strong result and was applied in the context of instrumental variable regression for the estimation of the risk-return relation. While this result holds under the assumption of a fractionally integrated data generating process, it need not hold when this assumption is loosened. In fact, every particular assumption about the underlying data generating process requires case-specific asymptotic theory. Hence, the objective of this work is to do precisely this for four other data generating stochastic processes and add our grain of sand to the ever-growing repertoire of time series asymptotic theory.

Chapter 1

Central Limit Theorems

The central limit theorem is a key concept in probability theory that provides conditions under which the adequately scaled sum of random variables limiting standard gaussian distribution. This result, together with the law of large numbers, is fundamental to most of mathematical and inferential statistics. The statement of the classical central limit theorem goes as follows:

Theorem (Classical Central Limit Theorem)**.** *Let $X_1, ..., X_T$ be independent and identically distributed with mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Let $\bar{X}_T = T^{-1} \sum_{t=1}^{T} X_t$. Then*

$$Z_T = \frac{\sqrt{T}\left(\bar{X}_T - \mu\right)}{\sigma} \rightsquigarrow Z$$

where $Z \sim N(0, 1)$. In other words,

$$\lim_{T\to\infty} \mathbb{P}\left(Z_T \leq z\right) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

The previous theorem can be interpreted as establishing which probability statements which can be made about about $\bar{X}_T$ using the normal distribution (Wasserman, 2004). There exist various other versions of the central limit theorem for cases in which the assumptions of independence, identical distribution, and finite moments of the underlying random variables are not met. For example, there are central limit theorems specific to each of the four following conditions: independent identically distributed observations, independent heterogeneously distributed observations, dependent identically distributed observations, and dependent heterogeneously distributed observations.

Analogous too all these versions of the central limit theorem, but with much stronger assumptions regarding the data generating process, is Lemma 1 of *The VIX, the Variance Premium, and Expected Returns* which goes as follows:

Lemma. *(Osterrieder et al., 2018)*
Let a_t and b_t be two independent processes given by $a_t = \phi(L)\epsilon_t$ and $b_t = (1-L)^{-d}\eta_t$ where $\phi(L) = \sum_{t=0}^{\infty} \phi_i L^i$ with $\sum_{i=0}^{\infty} i|\phi_i| < \infty$, $\phi(1) \neq 0$ and $(1-L)^d = \sum_{i=0}^{\infty} \gamma_i L^i$ with $\gamma_i = \Gamma(i+d) / (\Gamma(d)\Gamma(i+1))$, $0 \leq d < \frac{1}{2}$ and $\epsilon_t \sim i.i.d.(0, \sigma_\epsilon^2)$, $\eta_t \sim i.i.d.(0, \sigma_\eta^2)$. Define $z_t = a_t b_t$. Then

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} z_t / \bar{\sigma}_T \rightsquigarrow N(0,1)$$

where

$$\bar{\sigma}_T^2 = Var\left(T^{-\frac{1}{2}}\sum_{t=1}^{T} z_t\right) \to \bar{\sigma}^2 \; as \; T \to \infty$$

As can be seen, whenever the assumptions about the underlying data generating process change, so does the central limit theorem that's applied within that frame work. That is, different conditions will apply to different kinds of economic data. In this work, we derive four additional versions of the central limit theorem. The statement of all the central limit theorems that will be considered have the following form:

Theorem (Generic CLT)**.** *Given restrictions on the moments, dependence, and heterogeneity of a scalar sequence* $\{z_t\}$*,*

$$\left(\bar{z}_T - \mu\right)/(\bar{\sigma}_n/\sqrt{T}) = \sqrt{T}\left(\bar{z}_n - \mu\right)/\bar{\sigma}_n \rightsquigarrow N(0,1),$$

where $\mu = \mathbb{E}[z_t]$ *and* $\bar{\sigma}_T^2 = Var\left(\bar{z}_T\right)$

This can be summed up as saying that the sample average, under general conditions, has a limiting standard gaussian distribution. As has been stated in other works, there are trade-offs to the restrictions and conditions imposed and assumed in these theorems. "Typically, greater dependence or heterogeneity is allowed at the expense of imposing more stringent moment requirements" (White, 2001). In this paper, we define four time series stochastic processes and develop central limit theorems for each of them, albeit with a small twist as the reader will notice. These four processes can all be used to represent a vast variety of economic and financial data and potential applications for these results will be presented in the final section of this work.

Chapter 2

Main Results

Four Central Limit Theorems

We now present the results which comprise the core of this paper. What follows is a set of four central limit theorems each one for a different time series stochastic processes. The processes that will be considered are: an uncentered white noise, an uncentered white noise with level breaks, a trend stationary process, and a trend stationary process with level and trend breaks. As the reader will notice, the central limit theorems obtained are not concerning the sample mean of the data generating process, but instead pertain to the sample mean of the product between the original stochastic process and a white noise independent from this data generating process.

The subsequent section to this prelude will be organized as follows: a definition of each stochastic process x_t and its transformation $z_t = x_t \epsilon_t$ followed by a statement of the theorem derived for this new transformed variable.

All the proofs have been placed in the appendix, so the inquisitive reader can refer to the end of the document for a detailed, step by step proof of each of the four results.

2.1. Uncentered White Noise

Theorem 2.1. *Let* $\{x_t\}_{t=1}^T$ *be a stochastic process defined as* $x_t = \mu + u_t$ *where* μ *is a constant and* $u_t \sim iid(0, \sigma_u^2)$ *is a white noise. Let* $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ *be another white noise independent from* u_t. *Define* $z_t = x_t\epsilon_t = \mu\epsilon_t + u_t\epsilon_t$. *Let* $\hat{\sigma}_z^2 = \frac{\sum_{t=1}^T (z_t - \bar{z}_T)^2}{T-1}$ *be the unbiased and consistent estimator of* $Var(z_t)$, *then*

$$\frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z} \rightsquigarrow N(0,1)$$

Proof: See Appendix A.

2.2. Uncentered White Noise with Level Breaks

Theorem 2.2. *Let* $\{x_t\}_{t=1}^T$ *be a stochastic process defined as* $x_t = \mu + \theta DU_t + u_t$ *where* μ *is a constant,* $u_t \sim iid(0, \sigma_u^2)$ *is a white noise and* DU_t *is a level dummy variable defined as:*

$$DU_t = \begin{cases} 0 & if\ t \leq T_u \\ 1 & if\ t > T_u \end{cases}$$

where $T_u = \lfloor \lambda_u T \rfloor$ *is the break time* ($\lfloor \cdot \rfloor$ *is the floor function) and* $\lambda_u \in (0,1)$. *Let* $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ *be another white noise independent from* u_t *and assume that both* u_t *and* ϵ_t *have finite third moments. Define* $z_t = x_t\epsilon_t = \mu\epsilon_t + \theta DU_t\epsilon_t + u_t\epsilon_t$.
Then

$$\frac{\sum_{t=1}^T z_t}{\sqrt{Var(\sum_{t=1}^T z_t)}} = \frac{T^{-1}\sum_{t=1}^T z_t}{\sqrt{Var(T^{-1}\sum_{t=1}^T z_t)}} = \frac{\sqrt{T}\bar{z}}{\hat{\sigma}_z} \rightsquigarrow N(0,1)$$

where

$$Var(T^{-1}\sum_{t=1}^{T} z_t) = T^{-1}\left[\lambda_u\sigma_{z,B}^2 + (1-\lambda_u)\sigma_{z,A}^2\right], \textit{ for large } T$$

Proof: See Appendix B.

2.3. Trend Stationary Process

Theorem 2.3. *Let* $\{x_t\}_{t=1}^T$ *be a stochastic process defined as* $x_t = \mu + \beta t + u_t$ *where* μ *and* β *are constants,* $u_t \sim iid(0, \sigma_u^2)$ *is a white noise and* $t = 1, ..., T$. *Let* $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ *be another white noise independent from* u_t. *Define* $z_t = x_t\epsilon_t = (\mu + \beta t + u_t)\epsilon_t$. *Then*

$$\frac{T^{-1}\sum_{t=1}^T z_t}{\sqrt{Var\left(T^{-1}\sum_{t=1}^T z_t\right)}} \rightsquigarrow \sqrt{3}\left[\omega_\epsilon(1) - \int_0^1 \omega_\epsilon(r)dr\right] \sim N(0,1)$$

where $\omega_\epsilon(r)$ *is standard brownian motion (Wiener process).*

Proof: See Appendix C.

2.4. Trend Stationary Process with Trend and Level Breaks

Theorem 2.4. *Let* $\{x_t\}_{t=1}^{T}$ *be a stochastic process defined as* $x_t = \mu + \theta DU_t + \beta t + \gamma DT_t + u_t$ *where* μ, θ, β *and* γ *are constants and* $u_t \sim iid(0, \sigma_u^2)$ *is a white noise. The level dummy and trend dummy variables,* DU_t *and* DT_t*, respectively, are defined as:*

$$DU_t = \begin{cases} 0 & \textit{if } t \leq T_u \\ 1 & \textit{if } t > T_u \end{cases}$$

where $T_u = \lfloor \lambda_u T \rfloor$ *is the level break time and* $\lambda_u \in (0, 1)$*. The trend dummy variable is*

$$DT_t = \begin{cases} 0 & \textit{if } t \leq T_\tau \\ t - T_\tau & \textit{if } t > T_\tau \end{cases}$$

where $T_\tau = \lfloor \lambda_\tau T \rfloor$ *is the trend break time and* $\lambda_\tau \in (0, 1)$*. Let* $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ *be another white noise independent from* u_t*. Define*

$$z_t = x_t \epsilon_t = (\mu + \theta DU_t + \beta t + \gamma DT_t + u_t)\epsilon_t$$

. *Then*

$$\frac{T^{-1}\sum_{t=1}^{T} z_t}{\sqrt{Var\left(T^{-1}\sum_{t=1}^{T} z_t\right)}} = \frac{T^{-\frac{3}{2}}\sum_{t=1}^{T} z_t}{\sqrt{T^{-3}Var\left(\sum_{t=1}^{T} z_t\right)}} \rightsquigarrow$$

$$\rightsquigarrow \frac{\left(\left[\beta + \gamma(1-\lambda_\tau)^{\frac{3}{2}}\right]\omega_\epsilon(1) - \beta\int_0^1 \omega_\epsilon(r)dr - \gamma\int_{\lambda_\tau}^1 \omega_\epsilon(r)dr\right)}{\left[\frac{(\beta+\gamma)^2}{3} - \gamma\lambda_\tau(\gamma+\beta) + \gamma^2\lambda_\tau^2\right]} \sim N(0,1) \tag{2.1}$$

Proof: See Appendix D.

Chapter 3

Computational Statistics

3.1. Monte Carlo Simulations

We have now proved that data generated from any of the aforementioned data generating processes can be transformed in such a way that a very specific function of the data has a limiting standard gaussian distribution. One of the most recurring questions with respect to asymptotic results is the following: At what sample size can one say that the statistic in question *actually* begins to behave as theoretical limiting result says it is distributed? Of course, there's no definite and clear cut answer to this question but one way to provide an approximate answer is computational statistics, namely Monte Carlo simulations. What follows is a detailed description of the simulations that were done to test under which conditions the four asymptotic results begin to *show their true colors* and parameters affect the limiting distribution.

1. Simulate T realizations of the data generating process x_t

2. Multiply each of these data by an independent white noise ϵ_t and generate T new variables $z_t = x_t \epsilon_t$

3. Apply the transformation stated in each theorem to obtain the statistic $S_T = \frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z}$

4. Repeat steps 1-3 N times (we did 1000) to obtain a vector statistics S_T of length N

5. Apply a test of normality to this vector: Shapiro-Wilk, Kolmogorov-Smirnov, and/or Jarque-Bera and record whether the test rejects that the vector of statistics S_T has a standard normal distribution.

6. Repeat steps 1-5 M times (we used 1000) and calculate the proportion of times that the null hypothesis of normality was rejected

7. Repeat steps 1-6 for all desired parameter values: noise variances $(\sigma_u^2, \sigma_\epsilon^2)$, level breaks θ, trend breaks γ, break times λ_u, λ_τ, and sample sizes T

The Monte Carlo simulations were done for all possible combinations of samples sizes $T \in \{50, 200, 500, 1000, 10000\}$, noise variances (low, medium, high), level breaks (low, medium, high), trend breaks (low, medium, high), and break times (0.25, 0.5, 0.75). Note that we simulated the theoretical results for only 3 different values of each parameter due to computational limits. For the first data generating process, uncentered white noise, we did a total of 5x3=15 simulations, where doing step 1-6 counts as one simulation. That is, for the first process we obtained 45 null hypothesis rejection proportions (15 for each test of normality, 3 tests of normality). For the second process, uncentered white noise with level breaks, we did a total of 5x3x3x3=135 simulations for a total of 405 null hypothesis rejection proportions (15 for each test of normality, 3 tests of normality). For the third data generating process, trend stationary process, we did a total of 5x3=45 simulations for a total of 45 null hypothesis rejection proportions. Finally, for the fourth process, trend stationary with level and trend breaks, we did a total of 5x3x3x3x3=405 simulations for a total of 1215 null hypothesis rejection proportions. All this data was collected and analyzed via OLS regressions in order to asses how each parameter affects the proportion of times the null hypothesis of normality is rejected.

3.2. Regression Analysis of Simulated Data

Having realized the aforementioned simulations, we are now free to analyze the data as we wish. In particular, we wish to measure the magnitude and direction in which a variation in any of the parameters of the underlying data generating process x_t affects the null hypothesis that the test statistic defined as $S_T = \frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z}$ has a limiting standard gaussian distribution. With this objective in mind, we estimated the following OLS model:

$$R_i = \beta_0 + \beta_1\Lambda_i + \beta_2\Gamma_i + \beta_3\Theta_i + \beta_4\Sigma_i + \beta_5 T_i + \epsilon_i \quad (3.1)$$

where R_i is the dependent variable that measures the proportion of times (out of 1000) that the null hypothesis of normality of the test statistic was rejected, $\Lambda_i \in$\{0.25, 0.5, 0.75\} are the break times, $\Gamma_i \in$\{low, medium, high\} are the trend breaks, $\Theta \in$\{low, medium, high\} are the level breaks, $\Sigma_i \in$\{low, medium, high\} are the variances of the white noises, and $T \in \{50, 200, 500, 1000, 10000\}$ are the samples sizes. As reader will notice, we used only three different values for the break times, level/trend breaks, and white noise variances as well as only five different sample sizes. The reason for restricting our Monte Carlo simulations to such a small variety of parameter values is that we were limited in both time and computational capacity. Indeed, the simulations that generated our final data set took more than 75 hours to complete. Additionally, the decision of what constituted a "low", "medium" or "high" parameter value was quite arbitrary. One of the suggestions made in the *Future Work* section is surpassing the limitations that were just mentioned.

The results of this OLS regression are summarized in Table 3.1 and Table 3.2. Table 3.1 presents regression results for data from various data generating processes for different tests of normality: Shapiro-Wilk (S-W), Kolmogorov-Smirnov (K-S), and Jarque-Bera (J-B). The first column shows the regression results using data from the uncentered white noise and trend stationary process simulations for all three tests of normality. The second, third, and fourth column show the regression results for data simulated using all four time series processes using the null hypothesis rejection proportion of the Shapiro-Wilk, Kolmogorov-Smirnov, and Jarque-Bera tests, respectively. The fifth column shows the regression results using data from all data generating processes and all three tests of normality (hence the 1710 observations) and the final column shows the regression results for data simulated from an uncentered white noise with level breaks using all three tests of normality. Table 3.2 shows the regression results for the same model as in (3.1) for data simulated from a trend stationary process with trend and level breaks. The reason for separating the regression results into two tables is that the trend stationary process with trend and level breaks includes many more parameters than say the uncentered white noise or simply the trend stationary process. The first thing one should notice is that the intercept is correctly right around 0.05. This can be interpreted as follows. If all parameter values were equal to zero, in which case we are dealing with pure white noise as a data generating process, then the null hypothesis rejection rate should be exactly 0.05, theoretically. That is, the classical central limit theorem guarantees asymptotic normality for the test statistic S_T whenever the underlying data generating process is a sequence of independent and identically distributed random variables with finite first and second moments. The second row panel in Table 3.1 shows the regression coefficients for the white noise variance, where the omitted category is the "low" variance. A priori, one would expect that coefficient to be positive and significant. That is, the higher the variance of the white noise, the less the test statistic S_T behaves "normally". More intuitively, whenever a random variable has a high variance and one wishes to estimate its population mean using the sample mean, one would typically need a higher sample size for a better approximation compared to the case where the underlying random variable has a lower variance. This is because the variance of the sample variances depends on both the variances of the underlying random variable and the sample size. Another reason why one would expect a positive coefficient for the white noise variance has to do with the concept of signal to noise ratio. When the white noise variances rises, the signal to noise ratio decreases and hence the signal (or data) that one observes is comprised of more "undesired" random noise and less "desired" signal. On the other hand, the higher the variance of the white noise, the more the process x_t resembles pure white noise and by Theorem 3.1 the asymptotic normality of the test statistic follows directly. The third panel in Table 3.1 shows the regression coefficients for different sample sizes, being $T = 200$ the omitted reference category. As can be seen, a change from $T = 200$ to $T = 50$ causes an increase in the dependent variable. That is, increasing the sample size from 50 to 200 has the effect of lowering the rejection rate of the null hypothesis. Intuitively, this means that increasing the sample size makes the test statistic S_T *become more gaussian*, so to speak. On the other hand, the regression coefficients for larger sample sizes are not significant. This tells us that an increase in sample size from 200 to 500 or from 200 to 10,000 doesn't do much with regards to making the test statistic *more gaussian*. At first glance, one could interpret this as establishing the sample size at which the limiting distribution stated by any of the four central limit theorems begins to hold. Yet we will refrain making such authoritative statements as we prefer a more reserved stance: the data from our limited simulations *seem to suggest* that a sample size of 200 is enough for asymptotic normality. The fourth and fifth panel in Table 3.1 show the regression coefficients for the level break time and the level break size, respectively. A priori, one would expect a test statistic S_T, obtained from a data generating process with a break time $\lambda_u \in (0, 1)$ closer to the lower and upper bound, to be *more gaussian* than when the break occurs close to $\lambda_u = 0.5$. For example, consider data generated by an uncentered white noise with level breaks with the break occurring close to either extreme $\lambda_u \in \{0, 1\}$. Evidently, that specification corresponds to that of an uncentered white noise without level breaks for which the asymptotic normality of the test statistic follows directly from the classical central limit theorem. Hence, since the omitted category in the regression results is $\lambda_u = 0.5$, one would expect the regression coefficients to be negative. Yet, the results are partially contrary to that. The regression coefficients for the size of the level break are completely intuitive. The smaller the break size, the more the process represents uncentered white noise (i.e. a sequence of iid variables) and so the null hypothesis rejection rate drops. Finally, Table 3.2 shows exactly the same sign and significance associated with the regression coefficients for the data generating process' parameters. The only additional parameter used as an independent variable in this regression is the trend break, which is statistically insignificant in explaining the null hypothesis rejection rate.

Table 3.1: Regression Results for Miscellaneous Data Generating Processes

H_0 Rejection Rate	UWN & TS	All S-W	All K-S	All J-B	All-All	UWN with LB
(Intercept)	0.0464***	0.0516***	0.0492***	0.0476***	0.0495***	0.0379***
	(0.0023)	(0.0029)	(0.0011)	(0.0027)	(0.0014)	(0.0050)
White Noise Variance						
Medium	0.0006	-0.0004	-0.0002	-0.0012	-0.0006	-0.0023
	(0.0022)	(0.0027)	(0.0010)	(0.0025)	(0.0013)	(0.0037)
High	-0.0027	-0.0018	0.0003	-0.0032	-0.0016	-0.0053
	(0.0022)	(0.0027)	(0.0010)	(0.0025)	(0.0013)	(0.0037)
Sample Size (Reference $T = 200$)						
$T = 50$	0.0111***	0.0280***	0.0164***	0.0254***	0.0233***	0.0369***
	(0.0028)	(0.0034)	(0.0013)	(0.0032)	(0.0016)	(0.0047)
$T = 500$	0.0024	0.0003	-0.0004	0.0020	0.0007	0.0005
	(0.0028)	(0.0034)	(0.0013)	(0.0032)	(0.0016)	(0.0047)
$T = 1000$	-0.0009	0.0012	$-0.0024^{\dagger}$	0.0027	0.0005	0.0007
	(0.0028)	(0.0034)	(0.0013)	(0.0032)	(0.0016)	(0.0047)
$T = 10000$	$0.0056^{\dagger}$	-0.0002	0.0005	0.0025	0.0009	0.0012
	(0.0028)	(0.0034)	(0.0013)	(0.0032)	(0.0016)	(0.0047)
Level Break Time						
$\lambda_u = 0.25$						0.0041
						(0.0037)
$\lambda_u = 0.75$						0.0167***
						(0.0037)
Level Break						
Medium Break						0.0076*
						(0.0037)
High Break						0.0113**
						(0.0037)
N	90	570	570	570	1710	405

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Source: Prepared by the author using simulated data not publicly accessible

Table 3.2: Regression Results for TS Process with Level and Trend Breaks

H_0 Rejection Rate	S-W	K-S	J-B	Combined
(Intercept)	0.0419***	0.0478***	0.0413***	0.0437***
	(0.0036)	(0.0016)	(0.0034)	(0.0018)
Break Time				
$\lambda_\tau = \lambda_u = 0.25$	0.0040	-0.0014	0.0057*	0.0028*
	(0.0025)	(0.0011)	(0.0023)	(0.0012)
$\lambda_\tau = \lambda_u = 0.75$	0.0117***	0.0039***	0.0077***	0.0078***
	(0.0025)	(0.0011)	(0.0023)	(0.0012)
Trend Break				
Medium	0.0002	0.0004	-0.0020	-0.0005
	(0.0025)	(0.0011)	(0.0023)	(0.0012)
High	-0.0008	0.0023*	-0.0029	-0.0005
	(0.0025)	(0.0011)	(0.0023)	(0.0012)
Level Break				
Medium	0.0019	-0.0006	-0.0005	0.0003
	(0.0025)	(0.0011)	(0.0023)	(0.0012)
High	0.0099***	0.0008	0.0088***	0.0065***
	(0.0025)	(0.0011)	(0.0023)	(0.0012)
White Noise Variance				
Medium	0.0005	-0.0008	-0.0003	-0.0002
	(0.0025)	(0.0011)	(0.0023)	(0.0012)
High	0.0000	0.0003	-0.0011	-0.0002
	(0.0025)	(0.0011)	(0.0023)	(0.0012)
Sample Size				
$T = 50$	0.0235***	0.0151***	0.0204***	0.0197***
	(0.0032)	(0.0014)	(0.0030)	(0.0015)
$T = 500$	0.0000	-0.0007	0.0024	0.0006
	(0.0032)	(0.0014)	(0.0030)	(0.0015)
$T = 1000$	0.0014	−0.0028*	0.0030	0.0005
	(0.0032)	(0.0014)	(0.0030)	(0.0015)
$T = 10000$	-0.0011	0.0009	0.0017	0.0005
	(0.0032)	(0.0014)	(0.0030)	(0.0015)
N	405	405	405	1215

Standard errors in parentheses

† significant at $p < .10$; $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$

Source: Prepared by the author using simulated data not publicly accessible

Chapter 4

Potential Applications

In this section, we will discuss two possible applications of the theoretical results presented in this paper. It is my conjecture that the central limit theorems presented here are also useful in the general context of instrumental variable regression for measuring the relationship between risk and return in finance, as in *The VIX, the Variance Premium, and Expected Returns.* However, given the high level of technicality of that paper, it is beyond the scope of the paper to endeavor into applications such at those. We now present an application of the central limit theorems which would allow us to derive the asymptotic distribution of regression coefficients between variables generated by any of the four time series processes considered in this work and another application where the central limit theorem is used to construct an exogeneity test for the independent variables in an OLS regression.

4.1. Asymptotic Distribution of Regression Coefficients

The intermediate results obtained in this thesis in the process of proving the final results are useful in the context of an OLS regression as in *Understanding Spurious Regression in Econometrics* (Phillips, 1986). Phillips obtains the asymptotic distribution of regression coefficients, test statistics, and the coefficient of correlation R^2 under the assumption that both the explanatory and dependent variable are independent AR(1) processes. Suppose one has a data set with two independent variables, x_t and y_t, both originating from any of the four data generating processes considered in this work. Suppose as well that you estimate the following simple OLS regression:

$$y_t = \alpha + \beta x_t + \epsilon_t \tag{4.1}$$

and obtain the residuals $\hat{e}_t = y_t - \hat{\alpha} - \hat{\beta} x_t$ from the resulting estimation of the model specified in (4.1). If the residuals have the desired properties of white noise (i.e. no autocorrelation, constant mean and variance) then

$$\sum_{t=1}^{T} x_t \hat{e}_t = \sum_{t=1}^{T} x_t (y_t - \hat{\alpha} - \hat{\beta} x_t) = \sum_{t=1}^{T} x_t y_t - \hat{\alpha} \sum_{t=1}^{T} x_t - \hat{\beta} \sum_{t=1}^{T} x_t \tag{4.2}$$

has a known limiting distribution and hence one derive the asymptotic behavior of $\hat{\alpha}$ and $\hat{\beta}$ since the asymptotic behavior of $\sum_{t=1}^{T} x_t$, $\sum_{t=1}^{T} y_t$, and $\sum_{t=1}^{T} x_t y_t$ can also be easily obtained. Just as in Phillips' paper, this would not only allow us to detect if and when we're in the presence of a spurious regression but more importantly *why*.

4.2. Exogeneity Test

A second possible application of the main results of this thesis is in the context of an OLS regression. Two of the many assumptions in OLS regression is that the explanatory variables are independent of and uncorrelated with the error term and that the covariance matrix of the error term is a diagonal matrix with a constant term in all entries of the diagonal. Also, recall that the theorems of this paper state that if x_t is independent from a white noise ϵ_t then it follows that $S_T = \frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z} \sim N(0,1)$ where $z_t = x_t\epsilon_t$. With this in mind, replace ϵ_t with the regression residuals $\hat{e}_t$ and we now have a test statistic. Under the null hypothesis that x_t is independent from $\hat{e}_t$, the test statistic S_T has a standard normal distribution, assuming of course that $\hat{e}_t$ has the properties of white noise. Hence, if the researcher can provide evidence that the residuals have constant first two moments and that they are not autocorrelated then S_T is an adequate test statistic for the hypothesis of independence between the independent variable x_t and the error term. Notice that this proposed test statistic exploits the contrapositive of the statement of the theorem

$$x_t \perp\!\!\!\perp \epsilon_t \Rightarrow \frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z} \sim N(0,1) \iff \frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z} \nsim N(0,1) \Rightarrow x_t \not\!\perp\!\!\!\perp \epsilon_t$$

We emphasize to the reader that this proposed test will only work when the assumption that the residuals have the property of white noise holds (i.e. is supported by the data) and when the underlying data generating process of the time series x_t is one of the four considered in this paper.

Conclusions and Future Work

In this thesis, we have derived central limit theorems for data generated from four different stochastic processes. We have shown that when this data multiplied by an independent white noise then the statistic used in the classical central limit theorem also has a standard normal distribution under relatively weak conditions. This is surprising since the sequence of random variables generated by the product of white noise and any of the trend stationary processes is not only heterogeneous in its distribution but has a variance that does not have an upper bound. In this sense, it does not meet the usual conditions for the central limit theorem and hence these data generating processes required asymptotic theory specific to their statistical properties.

We then used computational statistics in order to provide evidence regarding the sample size required for the limiting distribution to begin to emerge, the effect of break times and break sizes as well as the effect of the variances of the white noises on the asymptotic results. Some of the conclusions reached from this simulation exercise resulted counter intuitive while others did not. Yet, the regression results obtained from the simulated data must be taken with a grain of salt as limitations in computing power hindered us from constructing as rich and comprehensive a data set as we would have liked.

As future work, we propose to carry out much finer Monte Carlo simulations where a larger set of values for the white noise variances, break times, level/trend breaks, and sample sizes. This would allow us to reach far more robust and firm conclusions regarding the effect of the variations of data generating process' parameter values on the rejection rates of the null hypothesis of normality of the test statistic. Furthermore, we wish to develop similar asymptotic theory for a wider range of data generating process with the goal of building upon the ever growing literature on asymptotic theory for econometricians.

Appendix A - Proof of Theorem 3.1

Proof of Theorem 3.1

Theorem. *Let* $\{x_t\}_{t=1}^{T}$ *be a stochastic process defined as* $x_t = \mu + u_t$ *where* μ *is a constant and* $u_t \sim iid(0, \sigma_u^2)$ *is a white noise. Let* $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ *be another white noise independent from* u_t*. Define*

$$z_t = x_t \epsilon_t = \mu \epsilon_t + u_t \epsilon_t.$$

Let $\hat{\sigma}_z^2 = \frac{\sum_{t=1}^{T}(z_t - \bar{z}_T)^2}{T-1}$ *be the unbiased and consistent estimator of* $Var(z_t)$*, then*

$$\frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z} \rightsquigarrow N(0,1)$$

Proof. The fact that $u_t \perp\!\!\!\perp \epsilon_t$, where $\perp\!\!\!\perp$ denotes statistical independence, together with the properties of white noise imply that $z_t \perp\!\!\!\perp z_\tau \; \forall \, t \neq \tau$. Furthermore, the expected value of z_t is

$$\mathbb{E}[z_t] = \mathbb{E}[x_t \epsilon_t] = \mathbb{E}[x_t]\mathbb{E}[\epsilon_t] = 0 \; \forall \, t$$

where the second equality is an implication of the independence between x_t and ϵ_t. The following lemma is useful for the derivation of the variance of z_t.

Lemma .1. *For $i = 1, ..., n$ let the random variables X_i be independent and consider (measurable) functions $g_i : \mathbb{R} \to \mathbb{R}$, so that $g_i(X_i), i = 1, ..., n$ are random variables. Then the random variables $g_i(X_i), i = 1, ..., n$ are also independent. Proof: See (Roussas, 1997).*

The variance of z_t is

$$Var(z_t) = \mathbb{E}[z_t^2] = \mathbb{E}[x_t^2 \epsilon_t^2] = \mathbb{E}[x_t^2]\mathbb{E}[\epsilon_t^2]$$

For the last equality we used the fact that $x_t^2 \perp\!\!\!\perp \epsilon_t^2$ by Lemma 1 and the fact that if two random variables X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[X]$. The expected value of ϵ_t^2 is $\mathbb{E}[\epsilon_t^2] = Var(\epsilon_t^2) = \sigma_\epsilon^2$. The expected value of x_t^2 is

$$\mathbb{E}[x_t^2] = \mathbb{E}[(\mu + u_t)^2] = \mathbb{E}[\mu^2 + 2\mu u_t + u_t^2] = \mu^2 + \sigma_u^2$$

We then have that

$$Var(z_t) = \mathbb{E}[x_t^2]\mathbb{E}[\epsilon_t^2] = (\mu^2 + \sigma_u^2)\sigma_\epsilon^2 = \sigma_z^2 \;\; \forall \;\; t.$$

Since ϵ_t and u_t are each respectively independent and and identically distributed, this implies that $z_t \sim iid(0, \sigma_z^2)$

with $\sigma_z^2 < \infty$. The stochastic process $\{z_t\}_{t=1}^T$ meets all conditions necessary for the classic Central Limit Theorem. Define $\bar{z}_T = T^{-1}\sum_{t=1}^T z_t$, then, by the Central Limit Theorem, as $T \to \infty$

$$\frac{T^{-1}\sum_{t=1}^T z_t}{\sqrt{Var(T^{-1}\sum_{t=1}^T z_t)}} = \frac{\sqrt{T}\bar{z}_T}{\sigma_z} \rightsquigarrow N(0,1) \tag{3}$$

where $\rightsquigarrow$ means weak convergence or convergence in distribution. Furthermore, since the z_t's are i.i.d., σ_z can be approximated by $\hat{\sigma}_z^2 = \frac{\sum_{t=1}^T (z_t - \bar{z}_T)^2}{T-1}$ and convergence is still guaranteed by Slutsky's theorem. That is,

$$\frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z} \rightsquigarrow N(0,1)$$

Here's another way to arrive at the same result. Take the sum over t of the realizations of the stochastic process $\{z_t\}_{t=1}^T$. Then

$$\sum_{t=1}^T z_t = \sum_{t=1}^T x_t \epsilon_t = \mu \underbrace{\sum_{t=1}^T \epsilon_t}_{O_p(T^{\frac{1}{2}})} + \underbrace{\sum_{t=1}^T u_t \epsilon_t}_{O_p(T^{\frac{1}{2}})}$$

Note that $u_t \epsilon_t$ is also white noise $\sim iid(0, \sigma_u^2 \sigma_\epsilon^2)$. By the Central Limit Theorem,

$$T^{-\frac{1}{2}}\sum_{t=1}^{T} z_t = \mu T^{-\frac{1}{2}}\sum_{t=1}^{T}\epsilon_t + T^{-\frac{1}{2}}\sum_{t=1}^{T} u_t\epsilon_t \rightsquigarrow Y_1 + Y_2$$

where $Y_1 \sim N(0, \mu^2\sigma_\epsilon^2)$ and $Y_2 \sim N(0, \sigma_u^2\sigma_\epsilon^2)$. Also notice that since

$$cov(\epsilon_t, \epsilon_t u_t) = \mathbb{E}[\epsilon_t(\epsilon_t u_t)] = \mathbb{E}[\epsilon_t^2 u_t] = \mathbb{E}[\epsilon_t^2]\mathbb{E}[u_t] = 0$$

then $cov(Y_1, Y_2) = 0$ which implies that

$$Y_1 + Y_3 \sim N(0, (\mu^2 + \sigma_u^2)\sigma_\epsilon^2)$$

which is the same result as (1).

■

Appendix B - Proof of Theorem 3.2

Theorem. *Let* $\{x_t\}_{t=1}^T$ *be a stochastic process defined as* $x_t = \mu + \theta DU_t + u_t$ *where* μ *is a constant,* $u_t \sim iid(0, \sigma_u^2)$ *is a white noise and* DU_t *is a level dummy variable defined as:*

$$DU_t = \begin{cases} 0 & if\, t \leq T_u \\ 1 & if\, t > T_u \end{cases}$$

where $T_u = \lfloor \lambda_u T \rfloor$ *is the break time (*$l\cdot\rfloor$ *is the floor function) and* $\lambda_u \in (0,1)$*. Let* $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ *be another white noise independent from* u_t *and assume that both* u_t *and* ϵ_t *have finite third moments. Define*

$$z_t = x_t \epsilon_t = \mu \epsilon_t + \theta DU_t \epsilon_t + u_t \epsilon_t.$$

Then

$$\underbrace{\frac{\sum_{t=1}^T z_t}{\sqrt{Var(\sum_{t=1}^T z_t)}}}_{\text{Lyapunov's Theorem}} = \underbrace{\frac{T^{-1}\sum_{t=1}^T z_t}{\sqrt{Var(T^{-1}\sum_{t=1}^T z_t)}} = \frac{\sqrt{T}\bar{z}}{\hat{\sigma}_z}}_{\hat{\sigma}_z \xrightarrow{p} \sigma_z \text{ and Slutsky}} \rightsquigarrow N(0,1)$$

where

$$Var(T^{-1}\sum_{t=1}^{T} z_t) = T^{-1}\left[\lambda_u \sigma^2_{z,B} + (1-\lambda_u)\sigma^2_{z,A}\right], \textit{ for large } T$$

Proof. By independence of x_t and ϵ_t, the expected value of z_t is

$$\mathbb{E}[z_t] = \mathbb{E}[x_t\epsilon_t] = \mathbb{E}[x_t]\mathbb{E}[\epsilon_t] = 0$$

Given that the expected value of z_t is zero, the variance of z_t is

$$Var(z_t) = \mathbb{E}[z_t^2] = \mathbb{E}[x_t^2\epsilon_t^2] = \mathbb{E}[x_t^2]\mathbb{E}[\epsilon_t^2] \tag{4}$$

where the last equality is due to the independence of x_t and ϵ_t (again by Lemma 1). The expected value of ϵ_t^2 is $\mathbb{E}[\epsilon_t^2] = Var(\epsilon_t^2) = \sigma^2_\epsilon$. The expected value of x_t^2 is

$$\mathbb{E}[x_t^2] = \mathbb{E}[(\mu + \theta DU_t + u_t)^2] =$$
$$\mathbb{E}[\mu^2 + \theta^2 DU_t + u_t^2 + 2\mu\theta DU_t + 2\mu u_t + 2\theta DU_t u_t]$$

$$\Rightarrow \mathbb{E}[x_t^2] = \begin{cases} \mu^2 + \sigma_u^2 & \text{if } t \leq T_u \\ \mu^2 + \sigma_u^2 + \theta^2 + 2\mu\theta & \text{if } t > T_u \end{cases} \tag{5}$$

Finally, from (2) and (3) we can deduce that

$$Var(z_t) = \begin{cases} (\mu^2 + \sigma_u^2)\sigma^2_\epsilon \equiv \sigma^2_{z,B} & \text{if } t \leq T_u \\ (\mu^2 + \sigma_u^2 + \theta^2 + 2\mu\theta)\sigma^2_\epsilon \equiv \sigma^2_{z,A} & \text{if } t > T_u \end{cases} \tag{6}$$

By a similar argument made in result 1, all the z_t's are independent with finite variance although **not identically distributed** since the variance is different before and after the level break time T_u. In this case, z_t does not meet the conditions for the standard Central Limit Theorem, reason for which we refer to the following theorem.

Theorem (Lyapunov)**.** *Let* $X_1, ..., X_T$ *have finite means* $\mathbb{E}[X_t]$*, variances* $Var(X_t) = \mathbb{E}[(X_t - \mathbb{E}[X_t])^2]$ *and absolute moments* $\mathbb{E}[|X_t - \mathbb{E}[X_t]|^{2+\delta}]$ *with* $\delta > 0$ *and suppose* $B_T = \sum_{t=1}^{T} Var(X_t)$ *is the variance of the sum* $\sum_{t=1}^{T} X_t$*. If for some* $\delta > 0$

$$\lim_{T\to\infty} \frac{\mathbb{E}[|X_t - \mathbb{E}[X_t]|^{2+\delta}]}{B_T^{1+\frac{\delta}{2}}} = 0$$

Then as $T \to \infty$

$$\frac{\sum_{t=1}^{T} X_t - \sum_{t=1}^{T} \mathbb{E}[X_t]}{\sqrt{B_T}} \rightsquigarrow N(0,1)$$

In the case of z_t defined in this section, supposing z_t meets the conditions for this theorem for $\delta = 1$ would require that,

$$\lim_{T\to\infty} \frac{\sum_{t=1}^{T} \mathbb{E}[|x_t|^3]\mathbb{E}[|\epsilon_t|^3]}{\underbrace{(\lambda_u T(\mu^2 + \sigma_u^2)\sigma_\epsilon^2 + (1-\lambda_u)T(\mu^2 + \sigma_u^2 + \theta^2 + 2\mu\theta)\sigma_\epsilon^2)^{\frac{3}{2}}}_{O_p(T^{\frac{3}{2}})}} = 0$$

which would clearly be satisfied if the numerator is

$o_p(T^{\frac{3}{2}})$. That is, the condition for Lyapunov's Theorem would be satisfied if the proper assumptions are made about the third moment of the white noises u_t and ϵ_t, which is not too strong of a requirement. Note that if we assume that u_t and ϵ_t have finite third moments, the numerator in the previous expression is $O_p(T)$ which means that the limit does indeed converge to zero. Applying Lyapunov's Theorem, we have that

$$\begin{aligned}\frac{\sum_{t=1}^{T} z_t}{\sqrt{Var(\sum_{t=1}^{T} z_t)}} &= \frac{T^{-1}\sum_{t=1}^{T} z_t}{\sqrt{Var(T^{-1}\sum_{t=1}^{T} z_t)}} \\ &= \frac{\sqrt{T}\bar{z}_T}{\sqrt{\left[\lambda_u \sigma_{z,B}^2 + (1-\lambda_u)\sigma_{z,A}^2\right]}} \rightsquigarrow N(0,1)\end{aligned} \tag{7}$$

where

$$\begin{aligned}Var(T^{-1}\sum_{t=1}^{T} z_t) &= T^{-2}\sum_{t=1}^{T} Var(z_t) \\ = T^{-2}&\left[\sum_{t=1}^{\lfloor \lambda_u T \rfloor} Var(z_t) + \sum_{t=\lfloor \lambda_u T \rfloor+1}^{T} Var(z_t)\right] \\ &\Rightarrow Var(T^{-1}\sum_{t=1}^{T} z_t) \\ \approx T^{-1}&\left[\lambda_u(\mu^2+\sigma_u^2)\sigma_\epsilon^2 + (1-\lambda_u)(\mu^2+\sigma_u^2+\theta^2+2\mu\theta)\sigma_\epsilon^2\right] \\ &= T^{-1}\left[\lambda_u \sigma_{z,B}^2 + (1-\lambda_u)\sigma_{z,A}^2\right], \text{ for large T}\end{aligned} \tag{8}$$

Note that it's not directly evident if the typical estimator for the variance of z_t, namely $\hat{\sigma}_z^2 = \frac{\sum_{t=1}^{T}(z_t-\bar{z}_T)^2}{T-1}$, will converge in probability to the true variance of z_t since the the z_t's are not identically distributed before and after the level break. As will be shown, the typical variance estimator converges to the weighted averaged of the variances of z_t before and after the level break.

Define $\hat{\sigma}_z^2 \equiv \frac{\sum_{t=1}^{T}(z_t - \bar{z}_T)^2}{T-1}$ as the estimator of $Var(z_t)$. First note that the numerator can be written as

$$\sum_{t=1}^{T}(z_t - \bar{z}_T)^2 = \sum_{t=1}^{T} z_t^2 - \frac{(\sum_{t=1}^{T} z_t)^2}{T}$$

by (E1) and that $\mathbb{E}[z_t z_\tau] = \mathbb{E}[z_t]\mathbb{E}[z_\tau] = 0$ for $t \neq \tau$. Then,

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}_z^2] = \mathbb{E}\left[\frac{\sum_{t=1}^{T}(z_t - \bar{z}_T)^2}{T-1}\right] &= \frac{1}{T-1}\mathbb{E}\left[\sum_{t=1}^{T} z_t^2 - \frac{(\sum_{t=1}^{T} z_t)^2}{T}\right] \\
&= \frac{1}{T-1}\left[\sum_{t=1}^{T}\mathbb{E}[z_t^2] - \frac{\mathbb{E}[(\sum_{t=1}^{T} z_t)^2]}{T}\right] \\
&= \frac{1}{T-1}\left[\sum_{t=1}^{T} Var(z_t) - \frac{\sum_{t=1}^{T} Var(z_t)}{T}\right] \\
&= \frac{1}{T-1}[\lambda_u T\sigma_{z,B}^2 + (1-\lambda_u)T\sigma_{z,A}^2 \\
&\quad -\frac{1}{T}(\lambda_u T\sigma_{z,B}^2 + (1-\lambda_u)T\sigma_{z,A}^2)]
\end{aligned} \tag{9}$$

$$\Rightarrow \mathbb{E}[\hat{\sigma}_z^2] = \frac{1}{T-1}\frac{T-1}{T}\left[\lambda_u T\sigma_{z,B}^2 + (1-\lambda_u)T\sigma_{z,A}^2\right] \\ = \lambda_u\sigma_{z,B}^2 + (1-\lambda_u)\sigma_{z,A}^2 \tag{10}$$

It should be clear from result (8) that the usual variance estimator is not an unbiased estimator of $Var(z_t)$ (because the z_t's are not identically distributed for all t) but rather its expected value is a convex combination of the true variances of z_t before and after the level break time, which is precisely what we need. In that sense, the usual variance estimator is unbiased with regards to the weighted mean of population variances we wish to approximate.

Now, what we want is a *consistent* estimator for

$$Var(T^{-1}\sum_{t=1}^{T} z_t).$$

The following results, whose proofs can be found in many books on probability, show why the unbiasedness of $\hat{\sigma}_{\bar{z}_t}^2$ is a useful property in proving that $\hat{\sigma}_{\bar{z}_t}^2 \xrightarrow{p} Var(T^{-1}\sum_{t=1}^{T} z_t)$.

Definition: Mean Squared Error (Wasserman, 2004): The mean squared error of an estimator $\hat{\theta}_n$ is

$$MSE(\theta_n) = \mathbb{E}[(\hat{\theta}_n - \theta)^2]$$

where θ is any constant.

Definition: Convergence in quadratic mean (Wasserman, 2004): Let $X_1, X_2, ...$ be a sequence of random variables and let X be another random variable. X_n converges to X in quadratic mean (also called convergence in L_2), written $X_n \overset{qm}{\to} X$, if

$$\mathbb{E}[(X_n - X)^2] \to 0$$

Note that this definition still holds even if X is a point mass at a c, for c constant. That is if $\mathbb{P}(X = c) = 1$ and if $X_n \overset{qm}{\to} X$ then we say $X_n \overset{qm}{\to} c$.

Theorem. *The mean squared error (MSE) can be written as*

$$MSE(\hat{\theta}_n) = bias^2(\hat{\theta}_n) + Var(\hat{\theta}_n)$$

Proof: See (Wasserman, 2004)

Theorem. *Convergence in quadratic mean implies convergence in probability. That is,* $X_n \overset{qm}{\to} X \Rightarrow \quad X_n \overset{p}{\to} X$.
Proof: See (Wasserman, 2004)

We have shown that $\hat{\sigma}_z^2 \equiv \frac{\sum_{t=1}^{T}(z_t - \bar{z}_T)^2}{T-1}$ is an unbiased estimator of $Var(T^{-1}\sum_{t=1}^{T} z_t)$. We now prove that $\hat{\sigma}_z^2 \overset{p}{\to} Var(T^{-1}\sum_{t=1}^{T} z_t)$. The following equalities with be useful in this task.

Equality 1 (E1)

$$\sum_{t=1}^{T}(z_t - \bar{z}_T)^2 = \sum_{t=1}^{T} z_t^2 - \frac{(\sum_{t=1}^{T} z_t)^2}{T} = \sum_{t=1}^{T} z_t^2 - T\bar{z}^2$$

Equality 2 (E2)

$$(\sum_{t=1}^{T} z_t)^2 = \sum_{t=1}^{T} z_t^2 + 2\sum_{t=1}^{T-1}\sum_{\tau>t}^{T} z_t z_\tau$$

By Equality 1,

$$\hat{\sigma}_z^2 = \frac{\sum_{t=1}^{T}(z_t - \bar{z})^2}{T-1}$$
$$= \frac{1}{T-1}\left[\sum_{t=1}^{T} z_t^2 - T\bar{z}^2\right] = \frac{T}{T-1}\frac{\sum_{t=1}^{T} z^2}{T} - \frac{T}{T-1}\bar{z}^2$$
$$= \underbrace{\frac{T}{T-1}}_{\to 1}\left[\underbrace{\frac{\lambda_u \sum_{t=1}^{\lambda_u T} z_t^2}{\lambda_u T}}_{\xrightarrow{p} \lambda_u \sigma_{z,B}^2} + \underbrace{\frac{(1-\lambda_u)\sum_{t=\lambda_u T+1}^{T} z_t^2}{(1-\lambda_u)T}}_{\xrightarrow{p} (1-\lambda_u)\sigma_{z,A}^2}\right]$$
$$- \underbrace{\frac{T}{T-1}}_{\to 1}\underbrace{\bar{z}^2}_{\xrightarrow{p} 0} \tag{11}$$

$$\Rightarrow \hat{\sigma}_z^2 = \frac{\sum_{t=1}^{T}(z_t - \bar{z})^2}{T-1} \tag{12}$$
$$\xrightarrow{p} \lambda_u \sigma_{z,B}^2 + (1-\lambda_u)\sigma_{z,A}^2 = Var(T^{-1}\sum_{t=1}^{T} z_t)$$

Result (12) together with result (5) imply that

$$\frac{\sqrt{T}\bar{z}}{\hat{\sigma}_z} \rightsquigarrow N(0,1) \tag{13}$$

■

Intermediate Results Concerning Stochastic Sums

We now present a set of results for stochastic sums that will themselves be useful for proving the following two results. These results are a collection gathered from three main sources: Phillips (1986), Hamilton (1994) and a formulary of asymptotics developed by my thesis advisor, Dr. Daniel Ventosa-Santaulària.

Proposition 17.1, (Hamilton, 1994)

Suppose that ξ_t follows a random walk without drift,

$$\xi_t = \xi_{t-1} + u_t = \xi_0 + \sum_{i=1}^{t} u_t$$

where $\xi_0 = 0$ and $u_t \sim iid\ WN(0, \sigma^2)$ is white noise. Let $\omega(r) \sim N(0, r)$ be standard brownian motion. Then as $T \rightarrow \infty$

(a) $T^{-\frac{1}{2}} \sum_{t=1}^{T} u_t \rightsquigarrow \sigma\omega(1)$

(b) $T^{-\frac{3}{2}} \sum_{t=1}^{T} \xi_t \rightsquigarrow \sigma \int_0^1 \omega(r)dr$

(c) $T^{-\frac{3}{2}} \sum_{t=1}^{T} tu_t \rightsquigarrow \sigma \left[\omega(1) - \int_0^1 \omega(r)dr\right]$

For subsamples where $\lambda \in (0,1)$ and $\lfloor \lambda T \rfloor$ denotes the largest integer less than or equal to λT, we have that

(d) $T^{-\frac{1}{2}} \sum_{t=\lfloor \lambda T \rfloor+1}^{T} u_t \rightsquigarrow \sigma\omega(1-\lambda)$

(e) $T^{-\frac{3}{2}} \sum_{t=\lfloor \lambda T \rfloor+1}^{T} \xi_t \rightsquigarrow \sigma \int_\lambda^1 \omega(r)dr$

(f) $T^{-\frac{3}{2}} \sum_{t=\lfloor \lambda T \rfloor+1}^{T} tu_t \rightsquigarrow \sigma \left[\omega(1-\lambda) - \int_\lambda^1 \omega(r)dr\right]$

Appendix C - Proof of Theorem 3.3

Theorem. *Let* $\{x_t\}_{t=1}^T$ *be a stochastic process defined as* $x_t = \mu + \beta t + u_t$ *where* μ *and* β *are constants,* $u_t \sim iid(0, \sigma_u^2)$ *is a white noise and* $t = 1, ..., T$*. Let* $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ *be another white noise independent from* u_t*. Define* $z_t = x_t \epsilon_t = (\mu + \beta t + u_t)\epsilon_t$*. Then*

$$\frac{T^{-1}\sum_{t=1}^T z_t}{\sqrt{Var\left(T^{-1}\sum_{t=1}^T z_t\right)}} \rightsquigarrow \sqrt{3}\left[\omega_\epsilon(1) - \int_0^1 \omega_\epsilon(r)dr\right] \sim N(0,1)$$

where $\omega_\epsilon(r)$ *is standard brownian motion (Wiener process).*

Proof. The sum of z_t's is then

$$\begin{aligned}\sum_{t=1}^T z_t = \sum_{t=1}^T x_t \epsilon_t = \sum_{t=1}^T (\mu + \beta t + u_t)\epsilon_t \\ = \mu \underbrace{\sum_{t=1}^T \epsilon_t}_{O_p(T^{\frac{1}{2}})} + \beta \underbrace{\sum_{t=1}^T t\epsilon_t}_{O_p(T^{\frac{3}{2}})} + \underbrace{\sum_{t=1}^T u_t \epsilon_t}_{O_p(T^{\frac{1}{2}})}\end{aligned} \tag{14}$$

$$
\begin{aligned}
\Rightarrow T^{-\frac{3}{2}} \sum_{t=1}^{T} z_t &= \beta T^{-\frac{3}{2}} \sum_{t=1}^{T} t\epsilon_t + o_p(1) \\
&\rightsquigarrow \beta\sigma_\epsilon \left[\omega_\epsilon(1) - \int_0^1 \omega_\epsilon(r)dr\right]
\end{aligned} \tag{15}
$$

where $\omega_\epsilon(r)$ is a standard brownian motion [result due to the functional central limit theorem and the continuous mapping theorem: see (Hamilton, 1994) & (Phillips, 1986)]. We now know that $\sum_{t=1}^{T} z_t = O_p(T^{\frac{3}{2}})$. By the independence of the z_t's,

$$
\begin{aligned}
Var\left(\sum_{t=1}^{T} z_t\right) &= \sum_{t=1}^{T} Var\left(z_t\right) = \sum_{t=1}^{T} \mathbb{E}[z_t^2] \\
&= \sum_{t=1}^{T} \mathbb{E}[x_t^2]\mathbb{E}[\epsilon_t^2] = \sum_{t=1}^{T} \mathbb{E}[x_t^2]\sigma_\epsilon^2
\end{aligned} \tag{16}
$$

The expected value of x_t^2 is:

$$
\begin{aligned}
\mathbb{E}[x_t^2] &= \mathbb{E}[\mu^2 + \beta^2 t^2 + u_t^2 + 2\mu\beta t + 2\mu u_t + 2\beta t u_t] \\
&= \mu^2 + \beta^2 t^2 + \sigma_u^2 + 2\mu\beta t
\end{aligned} \tag{17}
$$

Substituting (17) into (16) we obtain that:

$$
\begin{aligned}
Var\left(\sum_{t=1}^{T} z_t\right) &= \sum_{t=1}^{T} Var\left(z_t\right) \\
&= \underbrace{\sum_{t=1}^{T} \left(\mu^2 + \beta^2 t^2 + \sigma_u^2 + 2\mu\beta t\right) \sigma_\epsilon^2}_{O(T^3)}
\end{aligned} \tag{18}
$$

Finally, since

$$\sum_{t=1}^{T} z_t = O_p(T^{\frac{3}{2}}) \text{ and } Var\left(\sum_{t=1}^{T} z_t\right) = O(T^3)$$

we have that

$$\frac{T^{-1}\sum_{t=1}^{T} z_t}{\sqrt{Var\left(T^{-1}\sum_{t=1}^{T} z_t\right)}} = \frac{\overbrace{T^{-1}\sum_{t=1}^{T} z_t}^{O_p(T^{\frac{1}{2}})}}{\underbrace{T^{-1}\sqrt{Var\left(\sum_{t=1}^{T} z_t\right)}}_{O(T^{\frac{1}{2}})}} = O_p(1) \tag{19}$$

We conclude that the ratio in (19) does have a limiting distribution.

Notice that

$$T^{-3}Var\left(\sum_{t=1}^{T} z_t\right) \to \frac{\beta^2\sigma^2\epsilon}{3}$$

$$\Rightarrow \frac{T^{-1}\sum_{t=1}^{T} z_t}{\sqrt{Var\left(T^{-1}\sum_{t=1}^{T} z_t\right)}} = \frac{T^{-\frac{3}{2}}\sum_{t=1}^{T} z_t}{\sqrt{T^{-3}Var(\sum_{t=1}^{T} z_t)}} \tag{20}$$
$$\rightsquigarrow \sqrt{3}\left[\omega_\epsilon(1) - \int_0^1 \omega_\epsilon(r)dr\right]$$

It will later be shown that the asymptotic result in (20) actually has a standard Gaussian distribution. But first, we will show that (20) holds even when we substitute the usual variance estimator for the theoretical variance. That is, we want to show that the following holds:

$$\frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z} = \frac{\frac{\sum_{t=1}^T z_t}{\sqrt{T}}}{\sqrt{\hat{\sigma}_z^2}} = \frac{\frac{\sum_{t=1}^T z_t}{T^{\frac{3}{2}}}}{\frac{1}{T}\sqrt{\hat{\sigma}_z^2}} = \frac{T^{-\frac{3}{2}}\sum_{t=1}^T z_t}{\sqrt{T^{-2}\hat{\sigma}_z^2}}$$
$$= \frac{T^{-\frac{3}{2}}\sum_{t=1}^T z_t}{\sqrt{T^{-3}Var(\sum_{t=1}^T z_t)}} \rightsquigarrow \sqrt{3}\left[\omega_\epsilon(1) - \int_0^1 \omega_\epsilon(r)dr\right] \tag{21}$$

Notice that all that needs to be proved in this case is that as $T \to \infty$

$$T^{-2}\hat{\sigma}^2 \xrightarrow{p} \frac{\beta^2\sigma_\epsilon^2}{3} = \lim_{T\to\infty} T^{-3}Var\left(\sum_{t=1}^T z_t\right)$$

We prove this now. First, recall that $z_t = x_t\epsilon_t = (\mu + \beta t + u_t)\epsilon_t$. By definition, the variance estimator is

$$\begin{aligned}\hat{\sigma}_z^2 &= \frac{\sum_{t=1}^T (z_t - \bar{z}_T)^2}{T-1} \\ &= \frac{1}{T-1}\left[\sum_{t=1}^T z_t^2 - \left(\frac{\sum_{t=1}^T z_t}{T}\right)^2\right] \\ &= \frac{1}{T-1}\left[\frac{(T-1)\sum_{t=1}^T z_t^2}{T} - \frac{\sum_{t\neq\tau} z_t z_\tau}{T}\right] \\ &= \frac{\sum_{t=1}^T z_t^2}{T} - \frac{\sum_{t\neq\tau} z_t z_\tau}{T(T-1)}\end{aligned} \tag{22}$$

The second term in the last equality will approach the sum of all $(T-1)$ autocovariances of z_t divided by T. By independence of $z_t, z_\tau \; \forall \, t \neq \tau$ the second term will converge to zero and we can henceforth ignore it. We focus on $\frac{\sum_{t=1}^{T} z_t^2}{T}$, the first term of the last equality, from now on. First,

$$z_t^2 = (x_t \epsilon_t)^2 = (\mu + \beta t + u_t)^2 \epsilon_t^2$$
$$= (\mu^2 + \beta^2 t^2 + u_t^2 + 2\mu\beta t + 2\mu u_t + 2\beta t u_t)\epsilon_t^2$$

So that the sum over t of z_t^2 is

$$\begin{aligned} \sum_{t=1}^{T} z_t^2 = \mu^2 \underbrace{\sum_{t=1}^{T} \epsilon_t^2}_{Op(T)} + \beta^2 \underbrace{\sum_{t=1}^{T} t^2 \epsilon_t^2}_{Op(T^3)} + \underbrace{\sum_{t=1}^{T} u_t^2 \epsilon_t^2}_{Op(T)} \\ + 2\mu\beta \underbrace{\sum_{t=1}^{T} t \epsilon_t^2}_{Op(T^2)} + 2\mu \underbrace{\sum_{t=1}^{T} u_t \epsilon_t^2}_{Op(T)} + 2\beta \underbrace{\sum_{t=1}^{T} t u_t \epsilon_t^2}_{Op(T^{\frac{3}{2}})} \end{aligned} \tag{23}$$

We now make a quick side note in order to discuss the order of convergence of each individual sum in (23). The first term is clearly $Op(T)$ since scaling it by T^{-1} would make it converge to σ_ϵ^2. For the third term, notice that by independence of $u_t, \epsilon_t \; \forall \, t, \tau$, $w_t = u_t \epsilon_t$ has the properties of white noise and hence $w_t \sim WN(0, \sigma_u^2 \sigma^2 \epsilon)$, so that the third sum scaled by T^{-1} will converge to $\sigma_u^2 \sigma_\epsilon^2$. For the fifth sum, notice that $u_t \epsilon_t^2$ creates a new independent and identically distributed random variable with expected value equal to zero. It then suffices to scale that sum by T^{-1} to make it converge to zero. The order of convergence of the second, fourth and sixth term is discussed in further detail down below.

Take the fourth term in (23),

$$\sum_{t=1}^{T} t\epsilon_t^2 = \epsilon_1^2 + 2\epsilon_2^2 + ... + T\epsilon_T^2$$
$$= \sum_{t=1}^{T} \epsilon_t^2 + \sum_{t=2}^{T} \epsilon_t^2 + \sum_{t=3}^{T} \epsilon_t^2 + ... + \sum_{t=T-1}^{T} \epsilon_t^2 + \epsilon_T^2$$

Now scale the fourth term in (23) by T^{-1},

$$\frac{\sum_{t=1}^{T} t\epsilon_t^2}{T} = \frac{\sum_{t=1}^{T} \epsilon_t^2}{T} + \frac{\sum_{t=2}^{T} \epsilon_t^2}{T} + \frac{\sum_{t=3}^{T} \epsilon_t^2}{T} + ... + \frac{\epsilon_T^2}{T}$$
$$= \frac{T}{T}\frac{\sum_{t=1}^{T} \epsilon_t^2}{T} + \frac{T-1}{T}\frac{\sum_{t=2}^{T} \epsilon_t^2}{T-1} + \frac{T-2}{T}\frac{\sum_{t=3}^{T} \epsilon_t^2}{T-2} + ... + \frac{1}{T}\frac{\epsilon_T^2}{1}$$
$$= \hat{\sigma}_\epsilon^2 \left[\frac{T + (T-1) + (T-2) + ... + 1}{T}\right]$$
$$= \hat{\sigma}_\epsilon^2 \frac{T(T+1)}{2T} = \hat{\sigma}_\epsilon^2 \frac{(T+1)}{2} \tag{24}$$

where $\hat{\sigma}_\epsilon$ is a consistent estimator of the variance of ϵ_t. Clearly then, $\sum_{t=1}^{T} t\epsilon_t^2 = Op(T^2)$. For the sixth term, as was previously stated, $u_t\epsilon_t^2$ is another white noise. Hence, by result (c) of Proposition 17.1 it follows that $\sum_{t=1}^{T} tu_t\epsilon_t^2 = Op(T^{\frac{3}{2}})$. What remains to be shown is that the second term in (23) is $Op(T^3)$.

$$
\begin{aligned}
\sum_{t=1}^{T} t^2\epsilon_t^2 &= \epsilon_1^2 + 4\epsilon_2^2 + 9\epsilon_3^2 + 16\epsilon_4^2 \\
&+ ... + (T-1)^2\epsilon_{T-1}^2 + T^2\epsilon_T^2 \\
&= (1)\sum_{t=1}^{T}\epsilon_t^2 + (3)\sum_{t=2}^{T}\epsilon_t^2 + (5)\sum_{t=3}^{T}\epsilon_t^2 \\
&+ ... + (2[T-1]-1)\sum_{t=T-1}^{T}\epsilon_t^2 + (2T-1)\epsilon_T^2
\end{aligned} \tag{25}
$$

$$
\begin{aligned}
\Rightarrow \frac{1}{T}\sum_{t=1}^{T} t^2\epsilon_t^2 &= (1)\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 + (3)\frac{1}{T}\sum_{t=2}^{T}\epsilon_t^2 + (5)\frac{1}{T}\sum_{t=3}^{T}\epsilon_t^2 \\
&+ ... + (2[T-1]-1)\frac{1}{T}\sum_{t=T-1}^{T}\epsilon_t^2 + (2T-1)\frac{1}{T}\epsilon_T^2 \\
&= (1)\frac{T}{T}\frac{\sum_{t=1}^{T}\epsilon_t^2}{T} + (3)\frac{T-1}{T}\frac{\sum_{t=2}^{T}\epsilon_t^2}{T-1} + (5)\frac{T-2}{T}\frac{\sum_{t=3}^{T}\epsilon_t^2}{T-2} \\
&+ ... + (2[T-1]-1)\frac{2}{T}\frac{\sum_{t=T-1}^{T}\epsilon_t^2}{2} + (2T-1)\frac{1}{T}\frac{\epsilon_T^2}{1} \\
&= \frac{\hat{\sigma}_\epsilon^2}{T}[(1)T + (3)(T-1) + (5)(T-2) \\
&+ ... + [2(T-1)-1](2) + (2T-1)(1)T] \\
&= \hat{\sigma}_\epsilon^2 \frac{\sum_{t=1}^{T}(2t-1)(T+1-t)}{T} \\
&= \frac{\hat{\sigma}_\epsilon^2}{T}\left[2T\sum_{t=1}^{T}t + 2\sum_{t=1}^{T}t - 2\sum_{t=1}^{T}t^2 - T^2 - T + \sum_{t=1}^{T}t\right] \\
&= \underbrace{\frac{\hat{\sigma}_\epsilon^2}{T}\left[\overbrace{\frac{2T^2(T+1)}{2}}^{O(T^3)} + \overbrace{\frac{2T(T+1)}{2}}^{O(T^2)} - \overbrace{\frac{2T(T+1)(2T+1)}{6}}^{O(T^3)} - \overbrace{T^2}^{O(T^2)} - \overbrace{T}^{O(T)} + \overbrace{\frac{t(T+1)}{2}}^{O(T^2}\right]}_{O(T^2)}
\end{aligned} \tag{26}
$$

where $\hat{\sigma}_\epsilon^2$ is a consistent estimator of σ_ϵ^2. This implies that

$$T^{-3}\sum_{t=1}^{T} t^2\epsilon_t^2 = \hat{\sigma}_\epsilon^2 \frac{T^3 - \frac{2T^3}{3}}{T^3} + o(1) \xrightarrow{p} \frac{\sigma^2\epsilon}{3} \tag{27}$$

where all the terms that converge or collapse to zero have been grouped in the term $o(1)$. Finally, equations (22), (23) and (27) together imply that

$$\begin{aligned} \frac{\hat{\sigma}_z^2}{T^2} &= \frac{\sum_{t=1}^T z_t^2}{T^3} + o_p(1) = \beta^2 \frac{\sum_{t=1}^T t^2\epsilon_t^2}{T^3} + o_p(1) \\ &\xrightarrow{p} \frac{\beta^2\sigma_\epsilon^2}{3} = \lim_{T\to\infty} T^{-3} Var(\sum_{t=1}^{T} z_t) \end{aligned} \tag{28}$$

Result (28) proves that (21) holds.

Useful Results for Functionals of Brownian Motion

Definition (Beran, 1994): A standard Wiener process (also called Brownian motion) is a stochastic process $\{w_t\}_{t\geq 0}$ indexed by nonnegative real numbers t with the following properties:

(P1) $w_0 = 0$

(P2) With probability 1, the function $t \to w_t$ is continuous in t

(P3) The process $\{w_t\}_{t\geq 0}$ has stationary, independent increments

(P4) The increment $w_{t+s} - w_s$ has a normal(0,t) distribution

The following results, which are proved below, also hold and will prove useful for deriving the asymptotic distribution of results 3 and 4.

(i) $w(r) \sim N(0, r)$

(ii) $Cov\left(w(r), w(s)\right) = \mathbb{E}\left[w(r)w(s)\right] = min\{r, s\}$

(iii) $\int_l^u w(r)dr \sim N(0, \frac{u^3}{3} + \frac{2l^3}{3} - l^2 u)$
In particular:

$$\int_0^t w(r)dr \sim N(0, \frac{t^3}{3}) \text{ and } \int_t^1 w(r)dr \sim N(0, \frac{1}{3}+\frac{2t^3}{3}-t^2)$$

(iv)
$$Cov\left(w(t), \int_l^u w(r)dr\right) = \min\{t, u\}u - \frac{\min\{t, u\}^2}{2} - \min\{l, t\}l + \frac{\min\{l, t\}^2}{2}$$

(v) $Cov\left(\int_0^c w(r)dr, \int_0^d w(s)ds\right) = \frac{\min\{c,d\}^2 \max\{c,d\}}{2} - \frac{\min\{c,d\}^3}{6}$

(vi) $Cov\left(\int_0^1 w(r)dr, \int_t^1 w(s)ds\right) = \frac{1}{3} + \frac{t^3}{6} - \frac{t^2}{2}$

Proof: For (i), since $w(0) = 0$ write $w(r) = w(r) - w(0)$ and by property (4) we conclude that $w(r) \sim N(0, r)$. For (ii), suppose $r < s$. Then,

$$\begin{aligned}Cov\,(w(r)w(s)) &= \mathbb{E}[w(r)w(s)] - \mathbb{E}[w(r)]\mathbb{E}[w(s)] \\ &= \mathbb{E}[w(r)w(s)] = \mathbb{E}[w(r)(w(s) - w(r) + w(r))] \\ &= \mathbb{E}[w(r)(w(s) - w(r))] + \mathbb{E}[w^2(r)] \\ &\underbrace{=}_{(P3)} \mathbb{E}[w(r)]\mathbb{E}[w(s) - w(r)] + r = r\end{aligned} \tag{29}$$

For (iii), since the integral is a linear functional of (Gaussian) Brownian motion, then the integral in (iii) is Gaussian. The formal proof uses a Riemann sum approximation of the integral of Brownian motion but will not be done here as it's a ubiquitous result. We are interested only in the expected value and variance of the functional. The expected value is

$$\mathbb{E}\left[\int_l^u w(r)dr\right] = \int_l^u \mathbb{E}[w(r)]dr = \int_l^u 0dr = 0$$

The variance is

$$\begin{aligned}Var\left(\int_l^u w(r)dr\right) &= \mathbb{E}\left[\int_l^u w(r)dr \int_l^u w(s)ds\right] \\ &= \int_l^u \int_l^u \mathbb{E}[w(r)w(s)]drds = \int_l^u \int_l^u \min\{r, s\}drds \\ &= \int_l^u \left[\int_l^s rdr + \int_s^u sdr\right] ds = \int_l^u \left(\frac{r^2}{2}\bigg|_{r=l}^{s} + sr\bigg|_{r=s}^{u}\right) ds \\ &= \int_l^u \left[\frac{-s^2}{2} - \frac{l^2}{2} + su\right] ds = \left[\frac{-s^3}{6} - \frac{l^2 s}{2} + \frac{us^2}{2}\right]\bigg|_{s=l}^{u} \\ &= \ldots = \frac{u^3}{3} + \frac{2l^3}{3} - l^2 u\end{aligned} \tag{30}$$

For the particular results of (iii) substitute $u = t, l = 0$ and $u = 1, l = t$, respectively. For (iv), we have that

$$Cov\left(w(t), \int_l^u w(r)dr\right)$$
$$= \mathbb{E}\left[\int_l^u w(t)w(r)dr\right] = \int_l^u \min\{t, r\}dr$$

There are three cases. Suppose $t \leq l < u$. Then

$$\begin{aligned} &\Rightarrow Cov\left(w(t), \int_l^u w(r)dr\right) \\ &= \int_l^u tdr = tr\Big|_{r=l}^{u} = ut - lt \end{aligned} \tag{31}$$

Now suppose $l < t \leq u$. Then

$$\begin{aligned} &\Rightarrow Cov\left(w(t), \int_l^u w(r)dr\right) = \int_l^t rdr + \int_t^u tdr \\ &= \frac{r^2}{2}\Big|_{r=l}^{t} + tr\Big|_{r=t}^{u} = ... = tu - \frac{t^2}{2} - \frac{l^2}{2} \end{aligned} \tag{32}$$

Finally, suppose that $l < u < t$. Then

$$\begin{aligned} &\Rightarrow Cov\left(w(t), \int_l^u w(r)dr\right) \\ &= \int_l^u rdr = \frac{r^2}{2}\Big|_{r=l}^{u} = \frac{u^2}{2} - \frac{l^2}{2} \end{aligned} \tag{33}$$

So that results (31), (32) and (33) can be summarized in one general result (iv).

For (v), suppose $c < d$. Then

$$
\begin{aligned}
&Cov\left(\int_0^c w(r)dr, \int_0^d w(s)ds\right) \\
&= \mathbb{E}\left[\int_0^c w(r)dr \int_0^d w(s)ds\right] - \mathbb{E}\left[\int_0^c w(r)dr\right]\mathbb{E}\left[\int_0^d w(s)ds\right] \\
&= \int_0^c \int_0^d \mathbb{E}[w(r)w(s)]drds = \int_0^c \int_0^d \min\{r,s\}drds \\
&= \int_0^c \left[\int_0^r sds + \int_r^d rds\right] dr = \int_0^c \left[\frac{s^2}{2}\Big|_{s=0}^{r} + rs\Big|_{s=r}^{d}\right] dr \\
&= \int_0^c \left[\frac{r^2}{2} + rd - r^2\right] dr = \int_0^c \left[\frac{-r^2}{2} + rd\right] dr \\
&= \left[\frac{-r^3}{6} + \frac{dr^2}{2}\right]\Big|_{r=0}^{c} = \frac{-c^3}{6} + \frac{dc^2}{2}
\end{aligned} \tag{34}
$$

Now suppose $d < c$. Then

$$
\begin{aligned}
&\int_0^c \int_0^d \min\{r,s\}drds \\
&= \int_0^d \int_0^d \min\{r,s\}dsdr + \int_d^c \int_0^d \min\{r,s\}dsdr \\
&= \int_0^d \left[\int_0^r sds + \int_r^d rds\right] dr + \int_d^c \int_0^d sdsdr \\
&= \int_0^d \left[\frac{s^2}{2}\Big|_{s=0}^{r} + rs\Big|_{s=r}^{d}\right] dr + \int_d^c \frac{s^2}{2}\Big|_{s=0}^{d} dr \\
&= \int_0^d \left[\frac{r^2}{2} + rd - r^2\right] dr + \int_d^c \frac{d^2}{2} dr \\
&= \left[\frac{-r^3}{6} + \frac{dr^2}{2}\right]\Big|_{r=0}^{d} + \frac{d^2 r}{2}\Big|_{r=d}^{c} \\
&= \frac{-d^3}{6} + \frac{d^3}{2} + \frac{d^2 c}{2} - \frac{d^3}{2} = \frac{-d^3}{6} + \frac{cd^2}{2}
\end{aligned} \tag{35}
$$

Taken together, (34) and (35) imply that

$$Cov\left(\int_0^c w(r)dr, \int_0^d w(s)ds\right) = \frac{1}{2}\min\{c,d\} - \frac{1}{6}\min\{c,d\}^3 \tag{36}$$

For (vi), we have

$$\begin{aligned}
Cov\left(\int_0^1 w(r)dr, \int_t^1 w(s)ds\right) &= \mathbb{E}\left[\int_0^1 \int_t^1 w(r)w(s)dsdr\right] \\
&= \int_0^1 \int_t^1 \min\{r,s\}dsdr \\
&= \int_0^1 \int_t^1 rdsdr + \int_t^1 \int_t^1 \min\{r,s\}dsdr \\
&= \int_0^t rs\Big|_{s=t}^1 dr + \int_t^1 \left[\int_t^r sds + \int_r^1 rds\right] dr \\
&= \int_0^t r(1-t)dr + \int_t^1 \left[\frac{s^2}{2}\Big|_{s=t}^r + rs\Big|_{s=r}^1\right] dr \\
&= ... = \frac{(1-t)t^2}{2} + \left[\frac{r^2}{2} - \frac{r^3}{6} - \frac{t^2 r}{2}\right]\Big|_{r=t}^1 \\
&= \frac{1}{3} + \frac{t^3}{6} - \frac{t^2}{2}
\end{aligned} \tag{37}$$

We now prove that $\sqrt{3}\left[\omega_\epsilon(1) - \int_0^1 \omega_\epsilon(r)dr\right] \sim N(0,1)$. Because this expression is a sum of two normally distributed random variables, it itself is also normally distributed. By results (i) and (iii) for brownian motion, it can easily be shown that the expected value of (21) is equal to zero. We now show that the variance of (21) is equal to one.

$$
\begin{aligned}
&Var\left(w(1) - \int_0^1 w(r)dr\right) \\
&= Var\left(w(1)\right) + Var\left(\int_0^1 w(r)dr\right) \\
&-2Cov\left(w(1), \int_0^1 w(r)dr\right) \\
&= 1 + \frac{1}{3} - 2(1 - \frac{1}{2}) = \frac{1}{3}
\end{aligned}
\tag{38}
$$

The second to last line follows from results (i), (iii) and (iv) for brownian motion. Finally, result (21) follows directly from (38). ■

Appendix D - Proof of Theorem 3.4

Theorem. *Let* $\{x_t\}_{t=1}^{T}$ *be a stochastic process defined as* $x_t = \mu + \theta DU_t + \beta t + \gamma DT_t + u_t$ *where* μ, θ, β *and* γ *are constants and* $u_t \sim iid(0, \sigma_u^2)$ *is a white noise. The level dummy and trend dummy variables,* DU_t *and* DT_t, *respectively, are defined as:*

$$DU_t = \begin{cases} 0 & \text{if } t \leq T_u \\ 1 & \text{if } t > T_u \end{cases}$$

where $T_u = \lfloor \lambda_u T \rfloor$ *is the level break time and* $\lambda_u \in (0, 1)$. *The trend dummy variable is*

$$DT_t = \begin{cases} 0 & \text{if } t \leq T_\tau \\ t - T_\tau & \text{if } t > T_\tau \end{cases}$$

where $T_\tau = \lfloor \lambda_\tau T \rfloor$ *is the trend break time and* $\lambda_\tau \in (0, 1)$. *Let* $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ *be another white noise* *independent from u_t. Define*

$$z_t = x_t \epsilon_t = (\mu + \theta DU_t + \beta t + \gamma DT_t + u_t)\epsilon_t.$$

Then

$$\frac{T^{-1}\sum_{t=1}^{T} z_t}{\sqrt{Var\left(T^{-1}\sum_{t=1}^{T} z_t\right)}} = \frac{T^{-\frac{3}{2}}\sum_{t=1}^{T} z_t}{\sqrt{T^{-3}Var\left(\sum_{t=1}^{T} z_t\right)}} \rightsquigarrow$$

$$\rightsquigarrow \frac{\left(\left[\beta + \gamma(1-\lambda_\tau)^{\frac{3}{2}}\right]\omega_\epsilon(1) - \beta\int_0^1 \omega_\epsilon(r)dr - \gamma\int_{\lambda_\tau}^1 \omega_\epsilon(r)dr\right)}{\left[\frac{(\beta+\gamma)^2}{3} - \gamma\lambda_\tau(\gamma+\beta) + \gamma^2\lambda_\tau^2\right]}$$

$$\sim N(0,1)$$

Proof. The sum of z_t's is

$$\sum_{t=1}^{T} z_t = \sum_{t=1}^{T} x_t\epsilon_t = \sum_{t=1}^{T}(\mu + \theta DU_t + \beta t + \gamma DT_t + u_t)\epsilon_t$$
$$= \mu \underbrace{\sum_{t=1}^{T} \epsilon_t}_{O_p(T^{\frac{1}{2}})} + \theta \underbrace{\sum_{t=1}^{T} DU_t\epsilon_t}_{Op(T^{\frac{1}{2}})} + \beta \underbrace{\sum_{t=1}^{T} t\epsilon_t}_{O_p(T^{\frac{3}{2}})} + \gamma \underbrace{\sum_{t=1}^{T} DT_t\epsilon_t}_{O_p(T^{\frac{3}{2}})} + \underbrace{\sum_{t=1}^{T} u_t\epsilon_t}_{O_p(T^{\frac{1}{2}})} \tag{39}$$

The fourth term from the previous expression can be rewritten using the definition of DT_t as:

$$
\begin{aligned}
\gamma \sum_{t=1}^{T} DT_t \epsilon_t = \gamma \sum_{t=T_\tau+1}^{T} (t - T_\tau)\epsilon_t = \\
\gamma \sum_{t=T_\tau+1}^{T} t\epsilon_t - \gamma T_\tau^T \sum_{t=T_\tau^T+1}^{T} \epsilon_t \\
= \gamma \underbrace{\sum_{t=\lfloor \lambda_\tau T \rfloor+1}^{T} t\epsilon_t}_{O_p(T^{\frac{3}{2}})} - \gamma \underbrace{\lfloor \lambda_\tau T \rfloor \sum_{t=\lfloor \lambda_\tau T \rfloor+1}^{T} \epsilon_t}_{O_p(T^{\frac{3}{2}})}
\end{aligned} \tag{40}
$$

Substituting (40) into (39) and scaling by $T^{-\frac{3}{2}}$ we obtain

$$
\begin{aligned}
T^{-\frac{3}{2}} \sum_{t=1}^{T} z_t = \beta T^{-\frac{3}{2}} \sum_{t=1}^{T} t\epsilon_t + \gamma T^{-\frac{3}{2}} \sum_{t=\lfloor \lambda_\tau T \rfloor+1}^{T} t\epsilon_t \\
-\gamma T^{-\frac{3}{2}} \lambda_\tau T \sum_{t=\lfloor \lambda_\tau T \rfloor+1}^{T} \epsilon_t + o_p(1)
\end{aligned} \tag{41}
$$

Finally, taking the limit as $T \to \infty$

$$
\begin{aligned}
T^{-\frac{3}{2}} \sum_{t=1}^{T} z_t \rightsquigarrow \beta\sigma_\epsilon \left[\omega_\epsilon(1) - \int_0^1 \omega_\epsilon(r)dr\right] \\
+\gamma\sigma_\epsilon \left[(1-\lambda_\tau)^{\frac{1}{2}}\omega_\epsilon(1) - \int_{\lambda_\tau}^1 \omega_\epsilon(r)dr\right] - \gamma\lambda_\tau(1-\lambda_\tau)^{\frac{1}{2}}\sigma_\epsilon\omega_\epsilon(1) \\
= \sigma_\epsilon \left(\beta \left[\omega_\epsilon(1) - \int_0^1 \omega_\epsilon(r)dr\right] + \gamma \left[(1-\lambda_\tau)^{\frac{1}{2}}\omega_\epsilon(1) - \int_{\lambda_\tau}^1 \omega_\epsilon(r)dr\right] - \gamma\lambda_\tau(1-\lambda_\tau)^{\frac{1}{2}}\omega_\epsilon(1)\right) \\
= \sigma_\epsilon \left(\left[\beta + \gamma(1-\lambda_\tau)^{\frac{3}{2}}\right]\omega_\epsilon(1) - \beta\int_0^1 \omega_\epsilon(r)dr - \gamma\int_{\lambda_\tau}^1 \omega_\epsilon(r)dr\right)
\end{aligned} \tag{42}
$$

We now know that $\sum_{t=1}^{T} z_t = O_p(T^{\frac{3}{2}})$. Now, from the independence of the z_t's we have that

$$\begin{aligned} Var\left(\sum_{t=1}^{T} z_t\right) &= \sum_{t=1}^{T} Var\left(z_t\right) = \sum_{t=1}^{T} \mathbb{E}[z_t^2] \\ &= \sum_{t=1}^{T} \mathbb{E}[x_t^2]\mathbb{E}[\epsilon_t^2] = \sum_{t=1}^{T} \mathbb{E}[x_t^2]\sigma_\epsilon^2 \end{aligned} \tag{43}$$

The expected value of x_t^2 is

$$\begin{aligned} \mathbb{E}[x_t^2] &= \mathbb{E}[\mu^2 + \theta^2 DU_t^2 + \beta^2 t^2 + \gamma^2 DT_t^2 + u_t^2 \\ &+ 2\mu\theta DU_t + 2\mu\beta t + 2\mu\gamma DT_t + 2\mu u_t + 2\theta\beta DU_t t \\ &+ 2\theta\gamma DU_t DT_t + 2\theta DU_t u_t + 2\beta\gamma DT_t t + 2\beta t u_t + 2\gamma DT_t u_t] \\ &= \mu^2 + \theta^2 DU_t^2 + \beta^2 t^2 + \gamma^2 DT_t^2 + \sigma_u^2 + 2\mu\theta DU_t + 2\mu\beta t \\ &+ 2\mu\gamma DT_t + 2\theta\beta DU_t t + 2\theta\gamma DU_t DT_t + 2\beta\gamma DT_t t \end{aligned} \tag{44}$$

Note that (43) and (44) imply that

$$Var\left(\sum_{t=1}^{T} z_t\right) = \sum_{t=1}^{T} Var\left(z_t\right) = O(T^3)$$

since the terms

$$\sum_{t=1}^{T} t^2, \qquad \sum_{t=1}^{T} DT_t^2 = \sum_{t=1}^{T} (t - T_\tau)^2$$
$$\text{and } \sum_{t=1}^{T} DT_\tau t = \sum_{t=1}^{T} (t - T_\tau) t$$

dominate since they are all $0(T^3)$.

Finally, since

$$\sum_{t=1}^{T} z_t = O_p(T^{\frac{3}{2}}) \text{ and } Var\left(\sum_{t=1}^{T} z_t\right) = O(T^3)$$

we have that

$$\frac{T^{-1}\sum_{t=1}^{T} z_t}{\sqrt{Var\left(T^{-1}\sum_{t=1}^{T} z_t\right)}} = \frac{\overbrace{T^{-1}\sum_{t=1}^{T} z_t}^{O_p(T^{\frac{1}{2}})}}{\underbrace{T^{-1}\sqrt{Var\left(\sum_{t=1}^{T} z_t\right)}}_{O(T^{\frac{1}{2}})}} = O_p(1) \quad (45)$$

We conclude that the ratio in (45) does have a limiting distribution.

Note that

$$T^{-3}Var(\sum_{t=1}^{T} z_t) \to \left[\frac{(\beta+\gamma)^2}{3} - \gamma\lambda_\tau(\gamma+\beta) + \gamma^2\lambda_\tau^2\right]\sigma_\epsilon^2$$

$$\Rightarrow \frac{T^{-1}\sum_{t=1}^{T} z_t}{\sqrt{Var\left(T^{-1}\sum_{t=1}^{T} z_t\right)}} = \frac{T^{-\frac{3}{2}}\sum_{t=1}^{T} z_t}{\sqrt{T^{-3}Var\left(\sum_{t=1}^{T} z_t\right)}} \rightsquigarrow$$

$$\rightsquigarrow \frac{\left(\left[\beta+\gamma(1-\lambda_\tau)^{\frac{3}{2}}\right]\omega_\epsilon(1)-\beta\int_0^1\omega_\epsilon(r)dr-\gamma\int_{\lambda_\tau}^1\omega_\epsilon(r)dr\right)}{\left[\frac{(\beta+\gamma)^2}{3}-\gamma\lambda_\tau(\gamma+\beta)+\gamma^2\lambda_\tau^2\right]} \sim N(0,1) \tag{46}$$

so that the asymptotic distribution depends on unknown parameters. Notice that the numerator of (46) is a linear combination of normally distributed random variables (since both the brownian motion and its integral are gaussian), hence (46) also has a normal distribution. It should be clear from the results for brownian motion that (46) has expected value equal to zero. Then, it follows that (46) has variance equal to one since we are dividing the numerator by its theoretical standard deviation. We do not prove following for this data generating process and hence we present here are a conjecture

$$\begin{aligned} \frac{\sqrt{T}\bar{z}_T}{\hat{\sigma}_z} &= \frac{T^{-\frac{3}{2}}\sum_{t=1}^T z_t}{\sqrt{\frac{\hat{\sigma}^2}{T^2}}} \\ \xrightarrow{p} \frac{T^{-1}\sum_{t=1}^T z_t}{\sqrt{Var\left(T^{-1}\sum_{t=1}^T z_t\right)}} &= \frac{T^{-\frac{3}{2}}\sum_{t=1}^T z_t}{\sqrt{T^{-3}Var\left(\sum_{t=1}^T z_t\right)}} \end{aligned} \tag{47}$$

Another equivalent way to express the conjecture is that

$$\frac{\hat{\sigma}^2}{T^2} \xrightarrow{p} T^{-3}Var\left(\sum_{t=1}^T z_t\right) \tag{48}$$

which is what was proved for the previous data generating process. However, it has been shown through simulation that (46) holds and indeed has a standard gaussian distribution.

■

Bibliography

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman & Hall.

Box and Newbold (1971). Some comments on a paper of Coen, Gomme and Kendall. *Journal of the Royal Statistical Society. Series A*, 134(2):229–240.

Granger and Newbold (1974). Spurious regression in econometrics. *Journal of Econometrics*, 2(2):111–120.

Hamilton, J. D. (1994). *Time series analysis.* Princeton University Press, USA.

Osterrieder, D., Ventosa-Santaulària, D., and Vera-Valdés, J. E. (2018). The VIX, the variance premium, and expected returns. *Journal of the Financial Econometrics.*

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33(3):311–340.

Roussas, G. G. (1997). *A Course in Mathematical Statistics.* Academic Press.

Tsay and Chung (2000). The spurious regression of fractionally integrated processes. *Journal of Econometrics*, 96(1):155–182.

Vigen, T. (2015). *Spurious Correlations.* Hachette Books.

Wasserman, L. (2004). *All of Statistics - A Concise Course in Statistical Inference.* Springer.

White, H. (2001). *Asymptotic theory for econometricians.* Academic Press.