Chapter 5: Nonstationarity in Financial Time Series

Financial Econometric Modeling

Stan Hurn, Vance Martin, Peter C.B. Phillips and Jun Yu

Oxford University Press, 2020

Introduction

- An important property of asset prices is that they exhibit strong evidence of trends over long periods of time.
- Trend behaviour often manifests in a tendency for a time series to drift over time in such a way that no fixed mean value is revealed.
- But this long term growth is coupled with extended sub-periods in which prices wander above and below the growth line.
- Such time series are said to be nonstationary and may embody both a deterministic trend or a stochastic trend, which arises from the accumulation of random forces that drive prices to wander above and below the path of deterministic drift.
- Nonstationary behaviour needs to be respected in empirical work because of the importance of linkages between trending financial time series that are often the subject of investigation, because of the serious impact that trends can have on forecasting performance, and because of major changes in the econometric apparatus of inference when trends are present in the data.

Unit roots, Trends and the Order of Integration

The unit root hypothesis

- Identification of stochastic nonstationarity, typically hinges on testing evidence in support of a unit root restriction $\rho = 1$ in an autoregressive model of the form

$$y_t = \rho y_{t-1} + v_t \,,$$

  in which v_t is a stationary disturbance term.
- If the restriction $\rho = 1$ is satisfied and $v_t \sim iid\,(0, \sigma_v^2)$, the model is commonly known as a random walk.
- Tests of the restriction $\rho = 1$ are referred to as unit root tests and have very different characteristics from traditional regression tests in stationary time series models where $|\rho| < 1$. These distributions differ considerably from a normal distribution and lead to new procedures for testing.
- The classification of variables as either stationary or nonstationary has implications in both finance and econometrics. From a finance perspective, stochastic nonstationarity is important because the ubiquity of the random forces driving financial asset prices leads to the wandering price trajectories that are typically observed in practice and they are compatible with the efficient markets hypothesis.

- The return to a risky asset in an efficient market may be written as

$$r_t = p_t - p_{t-1} = \alpha + v_t \,, \qquad v_t \sim iid\,(0, \sigma_v^2)\,,$$

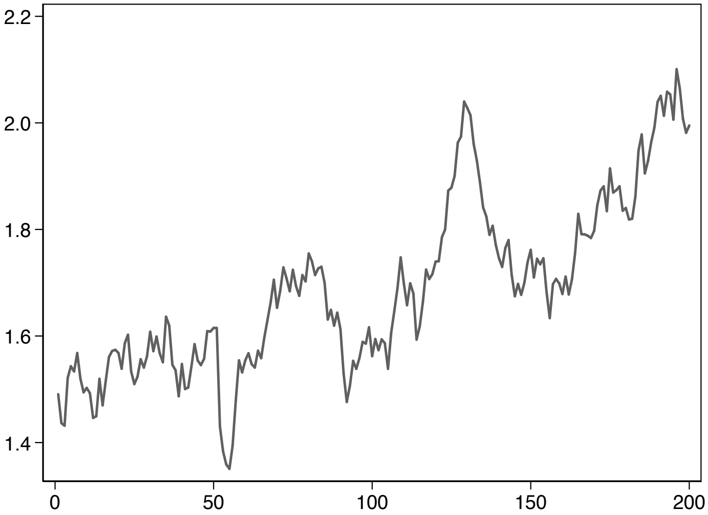
where p_t is the logarithm of the asset price. The parameter α represents the average return on the asset. From an efficient markets point of view, since $\alpha = 0$ and v_t is not autocorrelated, r_{t+1} cannot be predicted using information at time t.

- Another way of expressing the random walk equation is to write it in terms of p_t as

$$p_t = \alpha + p_{t-1} + v_t. \qquad (1)$$

The parameter α is the drift parameter with y_t now representing a random walk with drift.

Random walk with drift



Bond data

Consider fitting a simple AR(1) regression model

$$y_t = \alpha + \rho y_{t-1} + v_t \,,$$

to monthly data on United States zero coupon bonds for the period January 1947 to February 1987.

Table:

Ordinary least squares estimates of an AR(1) model estimated using monthly data on United States zero coupon bonds with maturities ranging from 2 months to 9 months for the period January 1947 to February 1987.

Maturity (mths)	Intercept $(\widehat{\alpha})$	Intercept $\mathrm{se}(\widehat{\alpha})$	Slope $(\widehat{\rho})$	Slope $\mathrm{se}(\widehat{\rho})$
2	0.090	0.046	0.983	0.008
3	0.087	0.045	0.984	0.008
4	0.085	0.044	0.985	0.007
5	0.085	0.045	0.985	0.007
6	0.087	0.045	0.985	0.007
9	0.088	0.046	0.985	0.007

Dependence on past shocks

Lat the random walk equation by one period

$$p_t = \alpha + p_{t-1} + v_t,$$
$$p_{t-1} = \alpha + p_{t-2} + v_{t-1},$$

Substituting for p_{t-1} gives

$$p_t = \alpha + \alpha + p_{t-2} + v_t + v_{t-1} \,.$$

Repeating this recursive substitution process for t-steps gives

$$p_t = p_0 + \alpha t + v_t + v_{t-1} + v_{t-2} + \cdots + v_1 = \alpha t + \sum_{j=1}^{t} v_j + p_0 \,,$$

in which p_t is fully determined by its initial value, p_0, a deterministic trend component and the accumulation of the complete history of shocks since initialisation of the process at $t = 0$.

The unit root mechanism is evident in the equation both in the unit coefficient of the lagged price variable p_{t-1} and in the accumulation process $\sum_{j=1}^{t} v_j$ whose weights are unity in all time periods. The drift parameter α now determines the extent of the deterministic drift measured by the linear time trend αt.

Properties

- Taking expectations and using the property that $\mathrm{E}(v_t) = \mathrm{E}(v_{t-1}) = \cdots = 0$, gives the mean of p_t

$$\mathrm{E}(p_t) = p_0 + \alpha t\,.$$

  Evidently when $\alpha > 0$, the mean price drifts upwards and increases over time at the same constant rate α. Even when the drift parameter α is small, over long periods of time the upward drift in the mean price becomes a prominent characteristic of the time series.

- The variance of p_t is given at each point in time by

$$\mathrm{var}(p_t) = \mathrm{E}\{[p_t - \mathrm{E}(p_t)]^2\} = t\sigma_v^2.$$

  Just as for the mean, the variance is also a linear increasing function over time. So the asset price p_t exhibits fluctuations with increasing amplitude as time passes.

- These properties reveal some of the implications of the efficient market hypothesis on the time series behaviour of financial asset prices. Specifically, in an efficient market asset prices may be expected to exhibit trending behaviour in levels and in long term fluctuations.

Order of integration

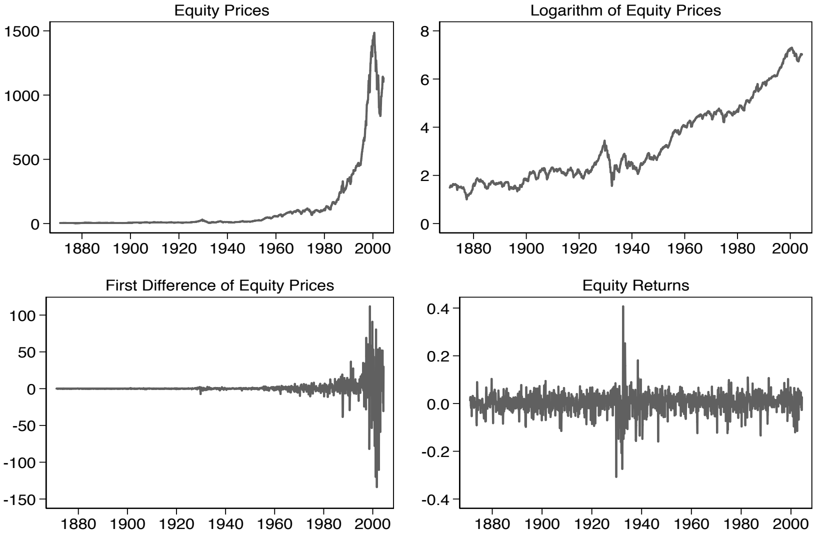
- The partial summation $\sum_{j=1}^{t} v_j$ aggregates (or integrates) up the component shocks v_j.
- Simply removing the deterministic trend, αt, from the process y_t will not be sufficient to obtain a stationary series because the stochastic trend component is still retained.
- This partial summation component is the origin of an important concept concerning nonstationarity, namely, *the order of integration* of a time series. A process is integrated of order d, denoted by $I(d)$, if it can be rendered stationary by differencing d times. Setting $d = 1$ to ensure differencing once, we have

$$\Delta p_t = \alpha + v_t,$$

where the symbol Δ is known as the difference operator, leading to $\Delta p_t = p_t - p_{t-1}$. The series Δp_t is stationary with mean α, variance σ^2 and residual v_t is serially uncorrelated.
- A general time series y_t is said to be integrated of order one, denoted $I(1)$, if it is rendered stationary by differencing once: that is y_t is nonstationary, but $\Delta y_t = y_t - y_{t-1}$ is stationary
- If $d = 2$, then y_t is $I(2)$ and needs to be differenced twice to achieve stationarity as follows

$$\Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}.$$

Different filters



Deterministic trend

Nonrandom function of time.

Most common is linear trend

$$y_t = \alpha + \delta\, t + v_t$$

where v_t is stationary.

De-trending by OLS:

$$y_t - \hat{\alpha} - \hat{\delta}\, t$$

If de-trended y_t is stationary, then y_t is called trend-stationary.

Stochastic trend

Random forces drive the trending: autoregression with unit root

$$y_t = \alpha + y_{t-1} + v_t$$

where v_t is stationary. If $v_t \sim WN(0, \sigma_u^2)$, the above model is called a random walk with drift.

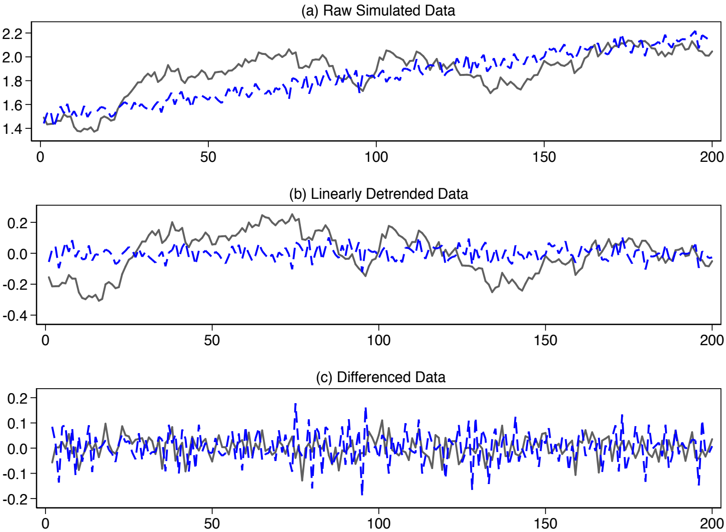
De-trending by differencing:

$$\Delta y_t = y_t - y_{t-1}$$

If differencing y_t produces a stationary series, then y_t is called difference-stationary.

Order of integration $I(\#)$: the number of times y_t needs to be differenced in order to reach a stationary series.

Trend and difference stationary



Testing for Unit Roots – the Dickey Fuller Framework

Testing

- Identification of stochastic nonstationarity, typically hinges on testing evidence in support of a unit root restriction $\rho = 1$ in an autoregressive model of the form

$$y_t = \rho y_{t-1} + v_t \ ,$$

  in which v_t is a stationary disturbance term.
- If the restriction $\rho = 1$ is satisfied and $v_t \sim iid\,(0, \sigma_v^2)$, the model is commonly known as a random walk.
- Tests of the restriction $\rho = 1$ are referred to as unit root tests and have very different characteristics from traditional regression tests in stationary time series models where $|\rho| < 1$. These distributions differ considerably from a normal distribution and lead to new procedures for testing.
- The classification of variables as either stationary or nonstationary has implications in both finance and econometrics. From a finance perspective, stochastic nonstationarity is important because the ubiquity of the random forces driving financial asset prices leads to the wandering price trajectories that are typically observed in practice and they are compatible with the efficient markets hypothesis.

Hypotheses

- Consider again the AR(1) regression equation

$$y_t = \alpha + \rho y_{t-1} + v_t$$

in which $v_t \sim N(0, \sigma_v^2)$.

- We are now interested in testing the following hypotheses:

$$\begin{array}{llcll} H_0 : & \rho & = & 1 & \text{[Variable } y_t \text{ is nonstationary]} \\ H_1 : & |\rho| & < & 1 & \text{[Variable } y_t \text{ is stationary].} \end{array}$$

- To perform the test, the equation is estimated by ordinary least squares regression and a t statistic is constructed in the usual manner to test whether $\rho = 1$. This statistic has the conventional ratio form

$$t_\rho = \frac{\widehat{\rho} - 1}{\mathrm{se}(\widehat{\rho})}, \tag{2}$$

where $\mathrm{se}(\widehat{\rho})$ is the standard error of $\widehat{\rho}$.

- The difficulty in executing the test arises from the fact that under the null hypothesis, the time series y_t is nonstationary and nonstationarity affects both the finite sample and asymptotic distribution of the statistic t_ρ.

A useful transformation

- In practice, it is convenient to transform the AR(1) equation

$$y_t = \alpha + \rho y_{t-1} + v_t$$

  in a way that converts the t statistic to a test of a zero slope coefficient in the transformed equation.
- To achieve this transformation, simply subtract y_{t-1} from both sides of and collect terms to give

$$y_t - y_{t-1} = \alpha + (\rho - 1)y_{t-1} + v_t. \quad (3)$$

- Defining $\beta = \rho - 1$ gives the regression equation

$$y_t - y_{t-1} = \alpha + \beta y_{t-1} + v_t.$$

  These two equations are precisely the same with the connection between them being $\beta = \rho - 1$.
- This transformation has the great advantage that the t statistic commonly reported in standard regression packages directly yields the unit root test statistic.

Using the zero coupon bond data

To illustrate this equivalence in a practical application, consider the monthly data on United States zero coupon bonds with maturities ranging from 2 months to 9 months for the period January 1947 to February 1987.

- The AR(1) regression gives

$$y_t = \underset{(0.046)}{0.090} + \underset{(0.008)}{0.983}\, y_{t-1} + \widehat{v}_t$$

- The transformed equation yields

$$y_t - y_{t-1} = \underset{(0.046)}{0.090} - \underset{(0.008)}{0.017}\, y_{t-1} + \widehat{v}_t .$$

- The difference in the two slope estimates is easily reconciled

$$\widehat{\beta} = \widehat{\rho} - 1 = 0.983 - 1 = -0.017.$$

- To perform a statistical test of the null hypothesis $H_0 : \rho = 1$, the two relevant t statistics in these two regressions are

$$\begin{aligned} t_\rho &= \frac{\widehat{\rho} - 1}{\mathrm{se}(\widehat{\rho})} = \frac{0.983 - 1}{0.008} = -2.120\,, \\ t_\beta &= \frac{\widehat{\beta} - 0}{\mathrm{se}(\widehat{\beta})} = \frac{-0.017 - 0}{0.008} = -2.120\,, \end{aligned}$$

demonstrating that the two methods are indeed equivalent.

Dealing with a deterministic trend alternative

- These tests can be extended to deal with the possibility that under the alternative hypothesis the time series may be stationary around a deterministic trend.
- The form of the equation to be estimated is

$$\Delta y_t = y_t - y_{t-1} = \alpha + \delta t + \beta y_{t-1} + v_t \, .$$

  The Dickey-Fuller test still consists of testing $\beta = 0$. But under the alternative hypothesis, y_t is now a stationary process with a deterministic trend.
- Once again using the monthly data on United States zero coupon bonds, the estimated regression including the time trend gives the following results (with standard errors in parentheses)

$$\Delta y_t = \underset{(0.052)}{0.030} + \underset{(0.001)}{0.001}\ t - \underset{(0.014)}{0.046}\ y_{t-1} + \widehat{v}_t .$$

- The value of the Dickey-Fuller test statistic is

$$t_\beta = \frac{\widehat{\beta} - 0}{\mathrm{se}(\widehat{\beta})} = \frac{-0.046 - 0}{0.014} = -3.172.$$

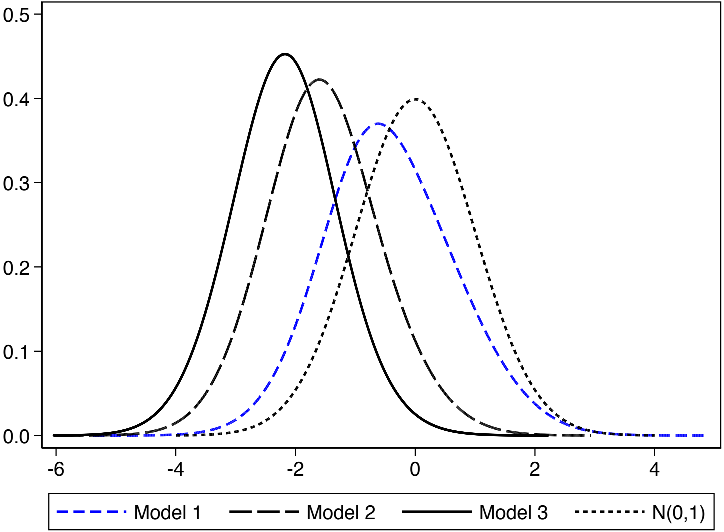
Dickey Fuller tests

- Three basic forms of the Dickey-Fuller unit root test are available, based on the following regression equations

$$\begin{aligned}
&\textbf{Model 1:} \quad && \Delta y_t = \beta y_{t-1} + v_t \\
&\textbf{Model 2:} \quad && \Delta y_t = \alpha + \beta y_{t-1} + v_t \\
&\textbf{Model 3:} \quad && \Delta y_t = \alpha + \delta t + \beta y_{t-1} + v_t.
\end{aligned}$$

- For each of these three models, the null hypothesis of the unit root test remains the same, namely, $H_0 : \beta = 0$.
- Unlike conventional statistical testing, however, the pertinent critical value for determining statistical significance in each case is different. The differences arise because the distribution of the unit root test statistic changes substantially depending on which model is used as the test regression. Thus, changing the regression equation by adding an intercept and/or a linear time trend not only affects the fitted regression coefficients and t statistics, it also changes their asymptotic distributions.

Dickey Fuller distributions



Augmented Dickey Fuller test (ADF)

- In estimating any of the previous test regressions, there is a real possibility that the disturbance term will exhibit autocorrelation.
- One common solution to correct for induced autocorrelation is to include lags of the dependent variable Δy_t in the test regressions. With adjustments for the extra lagged variables, the equations take the augmented form

$$
\begin{aligned}
\textbf{Model 1:} \quad & \Delta y_t = \beta y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + v_t \, , \\
\textbf{Model 2:} \quad & \Delta y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + v_t \, , \\
\textbf{Model 3:} \quad & \Delta y_t = \alpha + \delta t + \beta y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + v_t \, .
\end{aligned}
$$

- The unit root test procedures remain the same and involve the use of the same t statistic for testing $\beta = 0$ after taking into account the new specification of the various models.
- The lag length p in these specifications is an unknown parameter and may be chosen to ensure that the disturbances u_t do not exhibit autocorrelation.

Extensions – Beyond the Dickey Fuller Framework

Extensions

There is a huge literature on unit root testing, Extensions to the simple Dickey Fuller framework includes

- allowing for structural breaks;
- different approaches to detrending the data;
- testing with stationarity as the null hypothesis;
- testing for mildly explosive behaviour.

Structural Breaks

- When the timing of such structural breaks is known, it is straightforward to accommodate such shifts in the regression model.
- Level shift only

$$\Delta y_t = \beta y_{t-1} + \alpha + \delta t + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \gamma \mathrm{LBREAK}_t + v_t,$$

  where the structural break dummy variable is defined as

$$\mathrm{LBREAK}_t = \begin{cases} 0 & : \quad t \leq \tau \\ 1 & : \quad t > \tau, \end{cases}$$

  and τ is the observation (assumed to be known) where the break occurs.
- Additional shift in the time trend slope

$$\Delta y_t = \beta y_{t-1} + \alpha + \delta t + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + (\gamma_\alpha + \gamma_\delta t)\mathrm{LBREAK}_t + v_t.$$

- Unit root tests are now constructed and performed just as before by testing the hypothesis $\beta = 0$. However, because the regression equation has changed through the inclusion of the covariate LBREAK_t, the distribution of the ADF statistic under the null also changes to accommodate the presence of this covariate.

Ordinary Least Squares Detrending

- Consider

$$y_t = \alpha + \delta t + u_t ,$$
$$u_t = \phi u_{t-1} + v_t ,$$

- Instead of proceeding in the manner described previously and using Model 3, an alternative approach is to use a two-step procedure.

  Step 1: **Detrending** Estimate the parameters of the equation by ordinary least squares and then construct a detrended version of y_t given by

$$y_t^* = y_t - \widehat{\alpha} - \widehat{\delta} t .$$

  Step 2: **Testing** Test for a unit root using the deterministically detrended data, y_t^*, from the first step, using the Dickey-Fuller or augmented Dickey-Fuller test. Model 1 will be the appropriate model to use because, by construction, y_t^* will have zero mean and no deterministic trend.

  This procedure is equivalent to the single-step approach based on Model 3 .

GLS Detrending

- An alternative approach to defines the constant ϕ^*

$$\phi^* = 1 + \overline{c}/T \qquad \text{where } \overline{c} = \begin{cases} -7 & \text{[Constant } (\alpha \neq 0, \delta = 0) \\ -13.5. & \text{[Trend } (\alpha \neq 0, \delta \neq 0), \end{cases}$$

and use it to construct the following variables

$$y_t^* = y_t - \phi^* y_{t-1}\,, \qquad \alpha^* = 1 - \phi^*\,, \qquad t^* = t - \phi^*(t-1).$$

- The starting values for each of these variables at $t = 1$ are $y_1^* = y_1$ and $\alpha_1^* = 1$ and $t_1^* = 1$, respectively. The starting values are important because if $\overline{c} = -T$ then this differencing procedure has no effect. If, on the other hand, $\overline{c} = 0$ then the procedure reverts to a simple first difference.
- Using the newly defined variables run the regression

$$y_t^* = \pi_0 \alpha^* + \pi_1 t^* + u_t^*\,,$$

in which u_t^* is a composite disturbance term. Once the ordinary least squares estimates $\widehat{\pi}_0$ and $\widehat{\pi}_1$ are available, detrended data

$$\widehat{u}_t^* = y_t^* - \widehat{\pi}_0 \alpha^* - \widehat{\pi}_1 t^*\,,$$

can be constructed and tested for a unit root using Model 1 of the Dickey-Fuller framework. New critical values are required.

Nonparametric Adjustment for Autocorrelation

- The test is based on estimating the Dickey-Fuller regression equation by ordinary least squares and then correcting for the autocorrelation in a nonparametric way.
- The Phillips-Perron statistic is

$$\widetilde{t}_\beta = t_\beta \left(\frac{\widehat{\gamma}_0}{\widehat{f}_0}\right)^{1/2} - \frac{T(\widehat{f}_0 - \widehat{\gamma}_0)\mathrm{se}(\widehat{\beta})}{2\widehat{f}_0^{1/2} s},$$

  where t_β is the ADF statistic, s is the standard error of the Dickey-Fuller test regression, and $\widehat{f}_0$ is known as the long-run variance.
- The long-run variance is computed as

$$\widehat{f}_0 = \widehat{\gamma}_0 + 2\sum_{j=1}^{p}(1 - \frac{j}{p})\widehat{\gamma}_j,$$

  where p is the length of the lag, and $\widehat{\gamma}_j$ is the j^{th} estimated autocovariance function of the ordinary least squares residuals

$$\widehat{\gamma}_j = \frac{1}{T-j}\sum_{t=j+1}^{T} \widehat{u}_t \widehat{u}_{t-j}.$$

- The critical values are the same as the Dickey-Fuller critical values when the sample size is large.

KPSS test

- The Dickey-Fuller testing framework, including the GLS detrending and Phillips-Perron variants, are designed for testing the null hypothesis that a time series y_t is nonstationary or $I(1)$.
- There is also a popular test commonly known as the KPSS test, after Kwiatkowski, Phillips, Schmidt and Shin that is often reported in the empirical literature which has a null hypothesis of stationarity or $I(0)$.
- Consider the regression model

$$y_t = \alpha + \delta t + w_t + u_t,$$

where w_t is given by

$$w_t = w_{t-1} + v_t\,, \qquad v_t \sim iid\ N(0, \sigma_v^2).$$

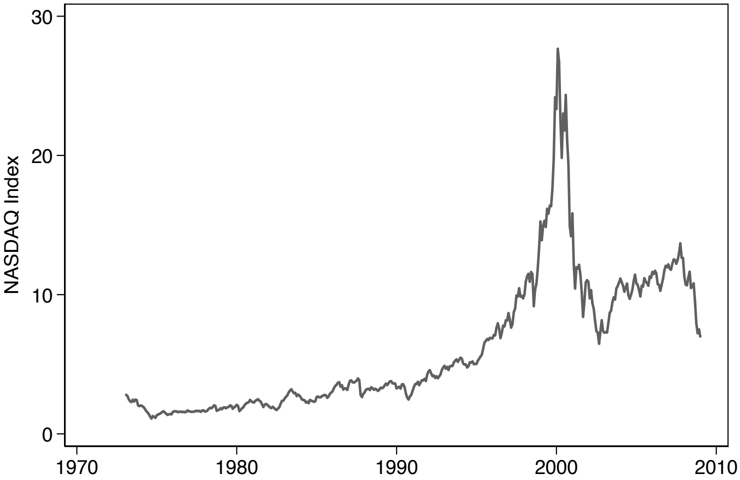
- The null hypothesis that y_t is a stationary $I(0)$ process is tested in terms of the null hypothesis $H_0 : \sigma_v^2 = 0$ in which case w_t is simply a constant equal to zero. Define $\{\widehat{u}_1, \cdots, \widehat{u}_T\}$ as the ordinary least squares residuals from a regression of y_t on a constant and a deterministic trend. Now define the standardised test statistic

$$\mathrm{S} = \frac{1}{T^2 \widehat{f}_0} \sum_{t=1}^{T} \Big(\sum_{j=1}^{t} \widehat{u}_j\Big)^2,$$

in which $\widehat{f}_0$ is a consistent estimator of the long-run variance of u_t.

Testing for Bubbles

Log Real Monthly NASDAQ Index



Bubbles and Unit Root Tests

- Consider the model
$$P_t = \frac{1}{1+R}\mathrm{E}_t\left[P_{t+1} + D_{t+1}\right],$$
where P_t is the price of an asset, R is the risk-free rate of interest assumed to be constant for simplicity, D_t is the dividend payment.
- By recursive substitution,
$$P_t = \sum_{j=1}^{N} \beta^j E_t\left[D_{t+j}\right] + \beta^N E_t\left[P_{t+N}\right] = \sum_{j=1}^{N} \beta^j E_t\left[D_{t+j}\right] + B_t,$$
where $\beta = (1+R)^{-1}$ and $B_t = \beta^N E_t\left[P_{t+N}\right]$. The process B_t is known as the rational bubble component.
- In the absence of bubbles ($B_t = 0$), the degree of nonstationarity of the asset price is controlled by the character of the dividend series.
- Asset prices will be explosive in the presence of bubbles because $1 + R > 1$.
- One way of testing for the presence of B_t in asset prices is to apply a right-sided unit root test to P_t and also to D_t. If one rejects the null in P_t and cannot reject the null in D_t, then there is evidence of explosive behavior in P_t. SInce this explosive behavior is not due to D_t it must be due to a bubble.

Right-Tailed Unit Root Tests

- If the ADF test were applied to the full sample where there is a bubble and the bubble collapses in the sample, the unit root test would not reject the null hypothesis $H_0 : \rho = 1$ in favour of the right-tailed alternative $H_1 : \rho > 1$. Evans (1991) shows that standard unit root tests have difficulties in detecting periodically collapsing bubbles.
- Phillips, Wu and Yu (2011) and Phillips and Yu (2011) suggest implementing recursive a unit root test based on an expanding window of observations, starting with a minimum window size and eventually using the entire sample. The test statistic is the maximum of the *t* statistics obtained in from the recursive regression. Critical values for the recursive test are obtained by simulation.
- Recursive testing of this kind also has the potential to deliver a date-stamping procedure for the origination and termination of bubbles. By matching the recursive *t* statistics with the path of the right tailed critical values an estimate of the origination of bubble behaviour in the data is obtained by determining the first observation for which the test statistic crosses the critical value path. This is known as the first crossing time principle.
- An estimate of the termination of the bubble is obtained by noting the observation for which the recursive test statistic crosses back over and falls below the critical value path.

Phillips, Wu and Yu (2011) Test for Price Bubbles