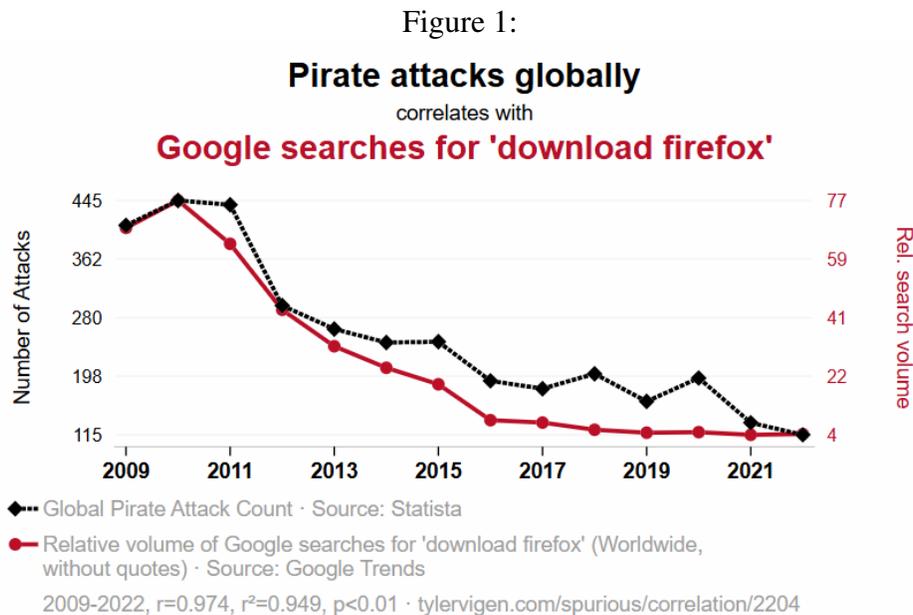# Non-Stationarity

We now turn our attention to a little time-series econometrics. Our first issue is to deal with a common problem, *non-stationarity*. A stationary variable is one where all moments are independent of time. A weakly stationary variable is one whose mean and variance are independent of time. For our purposes, weak stationarity is sufficient. If a variable has a time-trend, however, then it will typically be non-stationary. Non-stationarity violates the auss-Markov conditions and results in bias. Consider some stupid examples:[1]
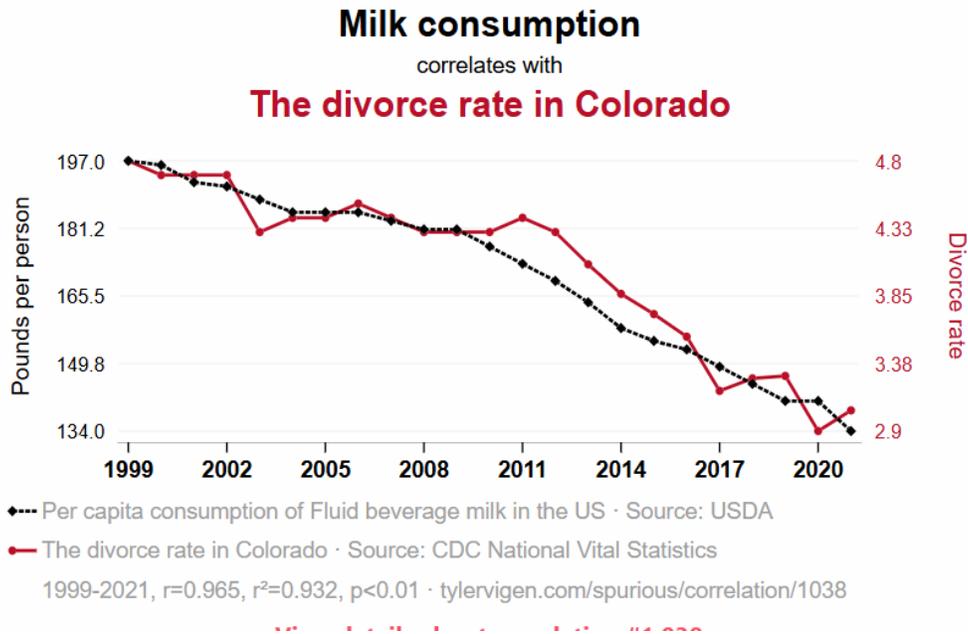
Figure 1:

In the previous example, it is obvious that neither milk causes divorce, nor that divorce causes milk. An OLS regression, would, however, suggest otherwise. This is because both variables are non-stationary. Milk consumption, for example, has been trending down. Its mean thus depends on the sample period selected, it has fallen over time. Many of the examples from that website are the result of non-stationarity.

These notes thus demonstrate how we can test for and fix non-stationary time series. We consider two cases, the first is that of a deterministic trend:

---

[1] Source: https://www.tylervigen.com/spurious-correlations

Figure 2:

Milk consumption

correlates with

The divorce rate in Colorado

- ···♦··· Per capita consumption of Fluid beverage milk in the US · Source: USDA
- ──●── The divorce rate in Colorado · Source: CDC National Vital Statistics

1999-2021, r=0.965, r²=0.932, p<0.01 · tylervigen.com/spurious/correlation/1038

Deterministic Trends:

Consider the following example:

$$y_t = \delta t + x_t \tag{1}$$

where x_t is a stationary variable. Note that the "t" is not a subscript here. The term δt captures a linear trend. The solution here is to simply de-trend by removing the trend:

$$\tilde{y}_t = y_t - \delta t = x_t \tag{2}$$

We can then work with the de-trended variable in the usual manner. Deterministic trends need not be linear. We can, for example, test for a quadratic trend.

Unit Roots

Our focus will be on a second type of non-stationary, known as a unit root or stochastic trend.

We will begin with the example of an AR(1) process:

2

$$y_t = \alpha y_{t-1} + u_t \tag{3}$$

where e_t is a white-noise error term. If we iterate forward from initial value, y_0 we see that:

$$y_t = \alpha^t y_0 + e_t + \alpha u_{t-1} + \alpha^2 u_{t-2}... \tag{4}$$

It then follows that if $|\alpha| > 1$, then y_t will behave explosively. This is the condition for a unit root.

The solution to dealing with unit roots is to difference. Subtract y_{t-1} from both sides of $\overset{\circ}{4}$:

$$\Delta y_t = (1 - \alpha)y_{t-1} + u_t \tag{5}$$

Most macroeconomic time series, such as prices and output, are non-stationary and must be differenced at least once. One cost of differencing is that the differenced series hasone fewer observation than the original.

Suppose that $\alpha > 2$. In this case, $\overset{\circ}{(5)}$ is also non-stationary. The solution is then to difference for a second time:

$$\Delta^2 y_t = (1 - \alpha)y_{t-1} + \Delta y_t + u_t \tag{6}$$

If $\overset{\circ}{(6)}$ is stationary, then the series is said to be I(2). In general, if we must difference d times to yield a stationary process, then the time series is said to be I(d).

Mistakes with Non-Stationary Time Series

Here, we consider the consequences of getting it wrong.
#1: Doing Nothing. As discussed earlier, if we fail to deal with non-stationarity then our results are biased and probably not useful. There are some signs that suggest that we may have erroneously included such a variable in our specification. These include highly autocorrelated error terms and suspiciously high R-squareds.

#2: De-trending when we should difference. Suppose that the data generating process is described by $\mathring{(4)}$, a single unit root. But we mistakenly assume that the variable includes a linear trend and we thus subtract off δt

$$y_t = y_t - 1 - \delta t + u_t \tag{7}$$

Notice that this does little to remedy the explosive dynamics of the time series. The transformed variable is thus still non-stationary and the consequences of this mistake are thus similar to #1.

#3. Differencing when you should de-trend. Now suppose that (3) is the true data generating process but we mistakenly first difference the time series instead:

$$\Delta y_t = \delta t + x_t - \delta(t - 1) - x_{t1} = \delta + x_t - x_{t-1} \tag{8}$$

If x_t is white noise, then we have transformed the time series into a stationary MA(1) process. We have thus fixed the bigger problem of non-stationarity, but added autocorrelation into the time series. This is problematic for the same reasons discussed in ECO 255. But because it does not bias the coefficients, just the standard errors, it it usually less serious than #2. For this reason, there is a tendency to difference rather than de-trend when in doubt.

#4 Overdifferencing: Now suppose that our time series is stationary but we mistakenly first difference it:

$$y_t = x_t \tag{9}$$

$$\Delta y_t = x_t - x_{t-1} \tag{10}$$

The result is similar to #3 (unsurprising given that we are differencing in both cases when we should not). The process remains stationary it now exhibits autocorrrelation. But because autocorrelation does not cause bias, there is a tendency to difference when in doubt. But it is important not to automatically difference your variables because overdifferencing is a source of misspecification. There are other ways to deal with non-stationarity. For example, when running a panel vector

4

autoregression (PVAR), ordinary differencing causes bias. We thus must use a more complicated technique. But for now, we will limit our attention to differencing and de-trending.

Dicky-Fuller

There are a great many statistical tests for unit roots and trends. We will focus on the test is that typically appropriate for most economic time series, the (sometimes Augmented) Dickey Fuller Test.

Suppose that we are willing to impose that the time series follows an AR(1) process. This is known as imposing structure on the regression specification and may be reasonable in the face of convincing theoretical support. Write the process as:

$$y_t = \alpha + \beta t + \rho y_{t-1} + u_t \tag{11}$$

To run a Dickey Fuller test, we simply obtain the OLS regression coefficients for $\overset{\circ}{(12)}$. It can also be shown, however, that the standard errors and autoreggressive coefficient are biased downward. Intuitively, suppose that $\rho = 1$ so that the time series is non stationary. We know that non stationarity leads to bias. So we can't use ordinary standard errors to reject a hypothesis that implies bias.

The null hypothesis is $H_0 : \rho = 1$, implying a unit root. The alternative is $H_a : |\rho| < 1$, implying stationarity. Dickey and Fuller (1979, 1981) derive unbiased standard errors for this. The Dickey Fuller critical values depend on the exact specification. Recall that for a one sided test that $\rho < 1$ at the 95% confidence level, the critical value is 1.65. Consider three cases:

1. If, as in $\overset{\circ}{1}2$, we include an intercept and trend. In this case, the critical value is 3.46. If we use unadjusted standard errors we will thus be too likely to reject unit roots that do exist and we will thus be too likely to include a non-stationary variable in our specification.

2. Suppose, as is often the case in macroeconomics, we are willing to impose no trend so that $\beta = 0$. In this case, the critical value is 2.89.

5

3. Finally, suppose that are willing to impose both that $\beta = 0$ and $\alpha = 0$. In this case, the critical value is 1.94.

As long as we are willing to impose an AR(1) process then making the process stationary is straightforward. We run a Dickey Fuller test and then difference or de-trend as needed. If differencing, we repeat until the order of integration, d is determined and we must difference that many times.

If we are interested in testing for unit roots for a more general AR(p) process, then we can instead run an Augmented Dickey Fuller test. The specification for this test is:

$$\Delta y_t = \alpha + \beta t + (\rho - 1)y_{t1} + \theta_1 \Delta y_{t-1} + \theta_2 \Delta y_{t-2} \theta_k \Delta y_{t-k} = u_t \tag{12}$$

The first order of business is to choose k. This is done using an information criteria. Recall that an information criteria is a measure of fit (such as R^2) less a penalty for including additional variables. In general, this penalty is ad-hoc. Along literature examines which information criteria do best in different circumstances.

Conducting the test is otherwise similar to the Unaugmented Dickey Fuller test. We obtain the OLS regression coefficients. We then test for $H_0 : \rho = 1$ using Dickey Fuller critical values that depend on whether the specification includes a trend or constant. We then difference until we are able to reject a unit root.

Stata

We will use employment, excluding health care and education as an example. I begin by creating a logged version of thsi variable:

gen jerbs=ln(employ)-ln(hced);

The first choice we must make is whether or not we are concerned that the variable may include both a trend and a unit root(s). In some cases, there may be established results or related work

that convinces us that we need only test for either a trend or a unit root. Here, however, we will consider the possibility of both.

The first step is to choose the number of lags we should include in our Augmented Dickey Fuller tests. Stata's dfgls command is a good way to do this:

dfgls jerbs

This command includes a trend unless we specifically include the notrend option. Stata will use a linear trend by default. This shows another case for logging variables. A constant trend growth rate is a linear trend when we work with logs. Stata will choose the maximum number of lags to include, or we can impose a value using the maxlag() option. Below the output, dfgls reports the optimal lag length.

Stata's dfuller command can provide more detail:

dfuller pchs, lags(4) trend regress

Notice that the results are a little different. This is because, the dfgls automatically tests for a unit root around de-trended data. To test for a trend itself, we should use dfuller. We must include the trend option if we want to allow for a linear trend. We must also include the regress option if we want to see more than the test statistic and critical values.

A variable may have more than one unit root. We must thus check if the first difference of logged GDP also has a unit root.

dfgls $d.jerbs$

Each time we difference, we lose an observation. Because Dickey Fuller tests require that we reject a null hypothesis, they are known for finding unit roots that are nit really there. We can find ourselves in a trap where by continually differencing, we continually increase our standard errors, and we thus continually find unit roots. WE thus cannot ignore intuition. I know housing starts are not I(65), even if this process yields such a result.