2024 Northern European Stata Conference

Limitations and comparison of the DFA, PP and KPSS unit root tests: evidence for laboral market variables in Mexico

Presents:
Ricardo Rodolfo Retamoza Yocupicio

Oslo, Norway, September 2024

Abstract

Unit root tests have represented a great contribution to time series analysis by detecting when a variable is stationary or not. However, they present limitations, which, although known, are still used and it seems that these limitations go unnoticed when applied in time series studies. Examples of these limitations, mainly Dickey Fuller (DF) and Phillips-Perron (PP), are that they could be detecting the presence of a unit root, when the series does not have it. Consequently, this document includes some of the criticisms that have been made to the unit root tests, to consequently execute in the Stata program the three best-known unit root tests (DFA, PP and KPSS) for the main macroeconomic variables of Mexico, this with the intention of analyzing, both graphically and technically, whether the series are stationary or not.

Introduction

Literature in the field of time series.

Granger and Newbold (1974): They use the concept of spurious regression to explain the consequences of poor treatment of time series variables.

Dickey and Fuller (1979): They propose a test that identifies when a variable is stationary. The null hypothesis is that the series is non-stationary.

Phillips and Perron (1988): This is a generalization of DF, the null hypothesis is also that the variable is non-stationary.

Decomposition of a stationary variable

Where the following expression is estimated (considering the simplest case without drift):

$$Y_t = \varphi Y_{t-1} + u_t \qquad (1)$$

However, in (1) the function is shown in levels. Therefore, the test must be carried out on first differences of the variable. If Y_{t-1} is subtracted from both sides of the equality, we obtain:

$$Y_t - Y_{t-1} = \varphi Y_{t-1} - Y_{t-1} + u_t \quad (2)$$

Decomposition of a stationary variable

Reparametrizing, we know that on the left side of the equation $\Delta Y_t = Y_t - Y_{t-1}$, while on the right side we can factor Y_{t-1} into $\varphi Y_{t-1} - Y_{t-1}$, to finally obtain equation (3):

$$\Delta Y_t = (\varphi - 1)Y_{t-1} + u_t \qquad (3)$$

$$\Delta Y_t = \delta Y_{t-1} + u_t \qquad (4)$$

In (3) it is considered that δ= (ρ-1), so we arrive at the expression in (4), which represents a first-order autoregressive process. The theory of the test suggests that if φ = 1, it is equivalent to saying that the test has a unit root.

Limitations of unit root tests

One of these limitations is, as commented by Gujarati and Porter (2010), Dickey-Fuller type tests are biased in favor of not rejecting the null hypothesis that the series has a unit root, even if one does not exist.

Among the reasons, it is found that the power of the test is based more on the time span of the data than on the size of the sample, that is, if we had a sample of 40 annual observations it would be more powerful than 200 observations. daily.

In the case of Phillips and Perron (1988), one of the problems that arises in their test to detect unit root is that they propose to include a deterministic linear trend, which (and as will be detailed later) is not viable for the analysis of a time series in the economic field.

Limitations of unit root tests

In the field of econometric analysis there are two types of trends: deterministic and stochastic.

Stock and Watson (2012), showing evidence for different macroeconomic variables in the United States, point out that the trend is stochastic.

Nelson and Plosser (1982) study macroeconomic series for the United States, in which they do not reject the hypothesis that the variables analyzed are non-stationary stochastic processes without a trend.

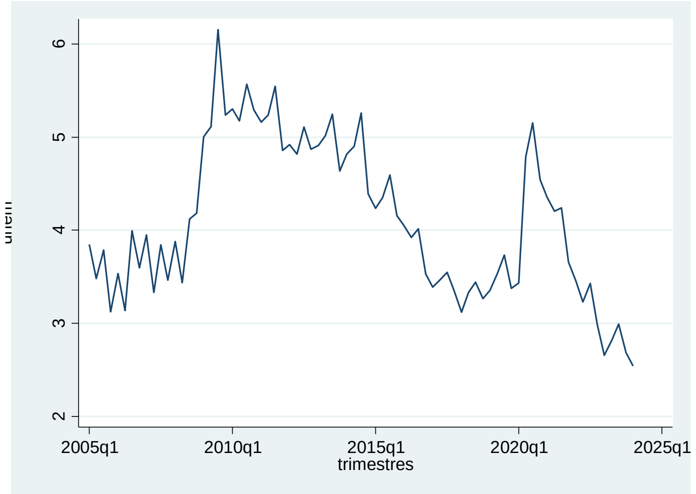
Econometric method

Table 1. Definition of variables

Laboral productivity	It is the gross domestic product obtained per hour worked with respect to a base year, in this case, 2018.
Gross Domestic Product	It is the sum of the value (in money) of all final-use goods and services generated by a country or federal entity during a period, based on 2018.
Unemployment rate	Percentage of the economically active population (EAP) that is not working, but is looking for work
Participation rate	Percentage that represents the economically active population (EAP) with respect to those aged 15 and over.
Laboral informality	Proportion of the employed population that includes the sum, without duplication, of those employed who are occupationally vulnerable due to the nature of the economic unit for which they work, with those whose labor link or dependence is not recognized by their source of work.

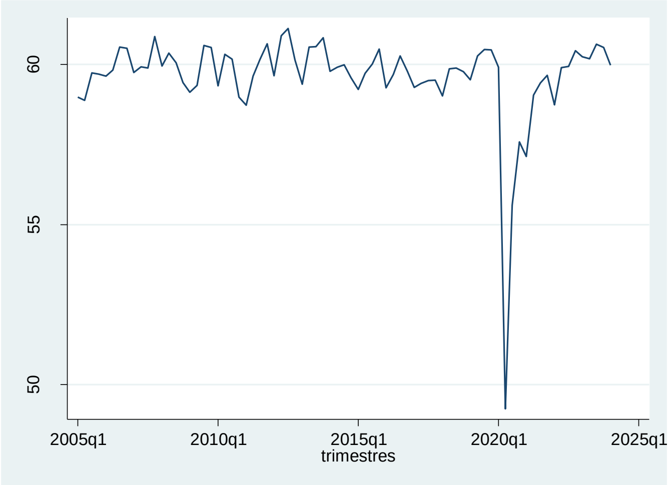
Econometric method

Graph 3. Unemployment rate in Mexico, 2005-2024



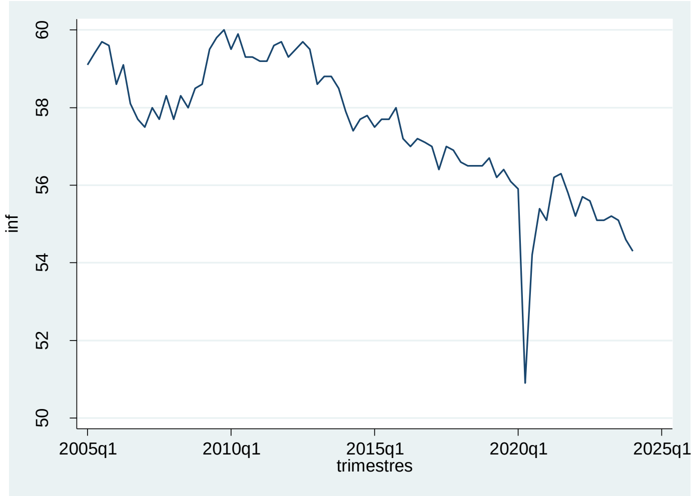
Source: Own elaboration in Stata

Graph 4. Participation rate in Mexico, 2005-2024



Source: Own elaboration in Stata

Graph 5. Laboral Informality rate in Mexico, 2005-2024



Source: Own elaboration in Stata

Unit root tests in levels

The theory of DFA and PP tests points to a null hypothesis that non-stationarity, in other words, the variable has a unit root. In the case of the KPSS test, the null hypothesis is that the series in question is stationary. Table 2 shows the performance of the tests at levels of the variables studied.

Table 2. Unit root tests in levels

		With drive and no trend			No drift	
		DFA	PP	KPSS	DFA	PP
lproductivity	Statistical	-1.414	-19.874	0.326	-0.688	-0.01
	Critical value	-3.54*	-19.368*	0.216*	-2.61*	-13.108
lgdp	Statistical	-1.016	-6.185	0.143	1.525	-0.018
	Critical value	-3.54*	-19.368*	0.216*	-2.61*	-13.108
unem	Statistical	-2.135	-6.849	0.277	-0.548	-0.485
	Critical value	-3.54*	-19.368*	0.216*	-2.61*	-13.108
part	Statistical	-3.175	-45.264	0.0747	-0.067	-0.008
	Critical value	-3.54*	-19.368*	0.216*	-2.61*	-13.108
inf	Statistical	-0.763	-4.564	0.264	-1.084	-0.086
	Critical value	-3.54*	-19.368*	0.216*	-2.61*	-13.108

Note: $p<0.1$*** $p<0.05$** $p<0.01$*
Source: Own elaboration in Stata

Unit root tests in first differences

Table 3. Unit root tests in first differences

		With drive and no trend			No drift	
		DFA	PP	KPSS	DFA	PP
dlproductivity	Statistical	-3.897	-110.945	0.0563	-3.852	-111.299
	Critical value	-3.55*	-19.35*	0.216*	-2.61*	-13.1*
dlgdp	Statistical	-4.691	-91.097	0.0405	-4.40*	-92.288
	Critical value	-3.55*	-19.35*	0.216*	-2.61*	-13.1*
dunem	Statistical	-3.423	-100.643	0.0614	-3.443	-100.633
	Critical value	-2.9**	-19.35*	0.216*	-2.61*	-13.1*
dpart	Statistical	-5.312	-80.48	0.0303	-5.351	-80.485
	Critical value	-3.55*	-19.35*	0.216*	-2.61*	-13.1*
dinf	Statistical	-4.946	-85.886	0.0335	-4.763	-86.653
	Critical value	-3.55*	-19.35*	0.216*	-2.61*	-13.1*

Note: p<0.1*** p<0.05** p<0.01*
Source: Own elaboration in Stata

Conclusion

As a conclusion, it is important that the variables are stationary, in this way, the error of falling into a spurious relationship is avoided. In the context of time series it is important that a variable has a constant mean and variance over time so that it is easier to predict its behavior. Thus, we find more robust and reliable estimators.

Stata codes

```
*NORTHERN EUROPEAN STATA CONFERENCE
*ROUTINE IN STATA

*RENAMING VARIABLES
rename var1 productividad
rename var2 gdp
rename var3 unem
rename var4 part
rename var5 inf

*DECLARING THE TEMPORALITY OF THE SERIES
generate trimestres=tq(2005q1)+_n-1
format trimestres %tq
tsset trimestres

*GETTING LOGARITHMS
gen lproductividad=log(productividad)
gen lgdp= log(gdp)

*GRAPHING THE SERIES
tsline lproductividad
tsline lgdp
tsline unem
tsline part
tsline inf

*RUNNING UNIT ROOT TESTS AT LEVELS
*UNIT ROOT TEST: DICKEY-FULLER
dfuller lproductividad, regress
dfuller lgdp, regress
dfuller unem, regress
dfuller part, regress
dfuller inf, regress

*USING VARSOC TO DETERMINE THE OPTIMAL NUMBER OF LAGS
varsoc lproductividad
varsoc lgdp
varsoc unem
varsoc part
varsoc inf
```

Stata codes

```
*UNIT ROOT TEST: DICKEY-FULLER AUMENTED
dfuller lproductividad, lags(4) regress
dfuller lgdp, lags(4) regress
dfuller unem, lags(4) regress
dfuller part, lags(4) regress
dfuller inf, lags(4) regress

*No constant
dfuller lproductividad, lags(4) noconstant regress
dfuller lgdp, lags(4) noconstant regress
dfuller unem, lags(4) noconstant regress
dfuller part, lags(4) noconstant regress
dfuller inf, lags(4) noconstant regress

*UNIT ROOT TEST: PHILLIPS-PERRON
pperron lproductividad, lags(4) regress
pperron lgdp, lags(4) regress
pperron unem, lags(4) regress
pperron part, lags(4) regress
pperron inf, lags(4) regress

*No constant
pperron lproductividad, lags(4) noconstant regress
pperron lgdp, lags(4) noconstant regress
pperron unem, lags(4) noconstant regress
pperron part, lags(4) noconstant regress
pperron inf, lags(4) noconstant regress

*UNIT ROOT TEST: KPSS
kpss lproductividad
kpss lgdp
kpss unem
kpss part
kpss inf

*GENERATING THE FIRST DIFFERENCES OF THE SERIES.
gen dlproductividad=D.lproductividad
gen dlgdp=D.lgdp
gen dunem=D.unem
gen dpart=D.part
gen dinf=D.inf
```

Stata codes

```
*RUNNING UNIT ROOT TESTS IN FIRST DIFFERENCES
*UNIT ROOT TEST: DICKEY-FULLER AUMENTED
*With constant
dfuller dlproductividad, lags(4) regress
dfuller dlgdp, lags(4) regress
dfuller dunem, lags(4) regress
dfuller dpart, lags(4) regress
dfuller dinf, lags(4) regress

*Without constant
dfuller dlproductividad, lags(4) noconstant regress
dfuller dlgdp, lags(4) noconstant regress
dfuller dunem, lags(4) noconstant regress
dfuller dpart, lags(4) noconstant regress
dfuller dinf, lags(4) noconstant regress
```

```
*UNIT ROOT TEST: PHILLIPS-PERRON
*With constant
pperron dlproductividad, lags(4) regress
pperron dlgdp, lags(4) regress
pperron dunem, lags(4) regress
pperron dpart, lags(4) regress
pperron dinf, lags(4) regress

*Without constant
pperron dlproductividad, lags(4) noconstant regress
pperron dlgdp, lags(4) noconstant regress
pperron dunem, lags(4) noconstant regress
pperron dpart, lags(4) noconstant regress
pperron dinf, lags(4) noconstant regress
```

```
*UNIT ROOT TEST: KPSS
kpss dlproductividad
kpss dlgdp
kpss dunem
kpss dpart
kpss dinf
```