An Introduction to Testing for Unit Roots Using SAS®: The Case of U.S. National Health Expenditures

Donald McCarthy, Department of Research and Evaluation, Kaiser Permanente

ABSTRACT

Determining whether a dataset is stationary can be important for the economist who does empirical work and SAS® allows the user to test for a unit root using several different tests: the Augmented Dickey-Fuller, Phillips-Perron, and the Kwiatkowski-Schmidt-Shin tests can all be implemented using PROC AUTOREG. This paper presents a brief, non-technical, overview of unit roots and shows how to test for a unit root and stationarity in U.S. health expenditure data.

INTRODUCTION

A process y_t is covariance stationary if its mean, variance, and autocovariances are time independent. Such a series will be mean reverting with large deviations from its mean being rare. A covariance stationary process is said to be integrated of order 0 (I(0)).

A process is integrated of order1 (I(1)), if it is not stationary but its first difference is stationary (thus $\Delta y_t = y_t - y_{t-1}$ is I(0)). Such a series is not mean reverting and will tend to wander widely from its initial value. It is said to have a unit root.

Time series with unit roots are not mere curiosities and present problems of statistical inference for the empirical economist. Yule (1926) proved that the correlation between two unrelated I(1) series tends to plus or minus one; and termed this phenomenon "nonsense correlation". Yule's warning was repeated by Granger and Newbold (1974) who showed that a regression of one I(1) series on another I(1) series tended to produce an R^2 of close to one.

Such series are also not rare. Numerous series in economics are candidates for having unit roots. Examples include time series of output and employment in general (and by industry) as well as consumption (in aggregate and per capita). The frequency with which one may encounter non-stationary data combined with the problems that are presented by series with unit roots make it important for economists to think carefully about the dynamic properties of their data and to employ tests for unit roots (or stationarity) where appropriate.

DATA

The data used represent annual per capita national healthcare expenditures in the U.S. from 1975 through 2013 and are available from the Centers for Medicare and Medicaid Services. These data include all "consumption of health care goods and services … and … investment in a given year that is purchased or provided by [healthcare providers]" and are comprehensive and are consistent over time (Centers for Medicare and Medicaid Services, 2013). These data were selected for three reasons: the series is sufficiently long to include several business cycles (i.e., 1975, 1980, 1982, 1991, 2001, 2009); healthcare is a large and important sector of the U.S economy in terms of output (and employment); and healthcare spending is of great interest to public policy makers and private decision makers.

AUGMENTED DICKEY FULLER (ADF) TEST

The ADF test tests the hypothesis that a time series y_t is I(1) against the alternative that it is I(0) assuming that y_t is an ARMA process (and ARMA process has both autoregressive and moving average terms). To test for a unit root using the ADF test, one estimates the following model:

(1) $\quad y_t = \beta_o + \beta_1 y_{t-1} + \beta_2 t + \sum_{i=1}^{n} \alpha_i \Delta y_{t-1} + u_t$

Where the n lagged first differences approximate the ARMA dynamics of the time series, β0 is a constant, and t is a trend. If the series has a unit root, β1 = 0 and $\sum_{i=1}^{n} \alpha_i \Delta y_{t-1} = 1$. The ADF test is a test of the hypothesis that β1 = 0 given n lagged first differences.

The selection of n is made such that the residual term in the term model (1) is approximately white noise (e.g. has a mean of zero, a finite variance, and is not serially correlated). The choice of lag length is important. If too few lags are included, there will be remaining autocorrelation and size distortion, but if too many lags are included then the power of the test will suffer. In practice, there are several approaches to determining maximum lag length.

Often economic theory will suggest the correct lag length. For instance, one may be testing for a unit root in a series with known seasonality (e.g., construction employment) which might suggest 4 lags if the data are quarterly or 12 if they are monthly. If theory is no guide, one *could* choose to select the maximum lag length that minimizes an information criterion such as the Akaike Information Criterion. However, simulations (e.g., Schwert, 1989) have suggested that such an approach may lead to the inclusion of too few lags (and thus remaining autocorrelation and size distortion). Schwert suggests a rule of thumb: int{12(t/100)1/4} based on the number of observations (t) in the sample.

Ng and Perron (2001) present an approach where one sets a maximum lag length nmax (in practice often equal to Schwert's rule of thumb) and then estimates the ADF model (1) using n=nmax. If the resulting t-statistic for the last lagged first difference in (1) is greater than 1.6, then nmax may be used as the number of lags in the ADF test. If it is not greater than 1.6, then one can reduce the lag length by one and repeat the process.

This process can be operationalized in SAS® as follows (using the same data as above):

```
data tstats;
set _null_;
run;
%macro np(indata, testvar);
%do i=1 %to &&nmax;
    %if (&i<2) %then %do;
    ods output ParameterEstimates=parms_&i.a;
        proc reg
                data = &indata;
                model d1&testvar= lag1&testvar lag&i.d1&testvar;
            run;quit;
    %end;
    %if  (&i>1) %then %do;
        ods output ParameterEstimates=parms_&i.a;
                proc reg
                data = &indata;
                model d1&testvar = lag1&testvar lag1d1&testvar --
                lag&i.d1&testvar;
            run;quit;
    %end;
    data parms_&i (keep=tvalue lag);
            set parms_&i.a (where=(variable="lag&i.d1&testvar"));
            lag=&i;
    run;
    data tstats;
            set tstats parms_&i;
            if abs(tvalue)>1.6 then flag=1;
    run;
    %end;
    proc sort data=tstats;
    by descending lag;
    run;
%mend;
%np(nhe_data,PCA);
```

The Schwert rule of thumb here suggests 9 lags and using this as the value for nmax in Ng and Perron's algorithm yields a lag length of 9 since the absolute value of t-statistic for the 9th lag >1.6. The syntax for the ADF test using a dataset (nhe_date) that includes per capita national health expenditures (PCA) and n=9 is:

```
proc autoreg
    data = nhe_data;
        model PCA = /stationarity=(adf=9);
run;
```

The ADF test generates the following output:

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	9	1.0524	0.9088	-0.5001	0.4894		
Single Mean	9	0.8627	0.9824	1.4134	0.9986	2.1559	0.5351
Trend	9	-0.3034	0.9921	-0.2734	0.9875	2.0307	0.7753

In interpreting the output, one starts from the most general model of a constant (single mean) and trend and continues to the most specific (zero mean and no trend) stopping when one can reject the null hypothesis of a unit root using the results of t and F type test, the results of which are provided in the table under the columns "tau", "F" (test statistics) , and "pr<tau", and "pr<F" (p-values). In the above output, the results for the trend, constant model are summarized in the third row. The F statistic is the test of the hypothesis that $\beta_1 = \beta_2 = 0$ and one cannot reject the null of a unit root. The next step is to examine the tau statistic which is the test that $\beta_1 = 0$. Again one cannot reject the null. One then proceeds to the constant, no trend model of row two. As with the trend, constant model, one cannot reject the null. Finally one considers the no constant, no trend model of row one. As with the more general models, one cannot reject the null of a unit root. The results of the test are consistent with the series having a unit root.

PHILLIPS PERRON (PP) TEST

To test for a unit root using the PP test, one estimates the following model:

(2) $y_t = \beta_o + \beta_1 y_{t-1} + \beta_2 t + u_t$

Where β_0 is a constant, and t is a trend. If the series has a unit root then $\beta_1 = 0$, and the PP test is a test of the hypothesis that $\beta_1 = 0$. The PP test differs from the ADF test insofar as the ADF test seeks to approximate the ARMA dynamics of the series at question through the use of lagged first differences in the test model (1) while the PP ignores autocorrelation in the test model (2) and instead calculates what can be thought of as the analogue of the ADF tau statistic that is robust to autocorrelation and heteroskedasticity.

The PP test does not require one to specify a model and to select the number of lags; however, SAS allows one to specify the number of autocovariance terms in the PP test (the lag truncation or bandwidth parameter). If no terms are specified, the bandwidth parameter is calculated automatically using the formula int(max(1,T1/2/5)) (SAS Institute, 2008).

The syntax for the PP test is as follows:

```
proc autoreg
    data = nhe_data;
        model PCA = /stationarity=(phillips);
run;
```

The PP test generates the following output:

Phillips-Perron Unit Root Test					
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	1	1.8270	0.9794	10.7581	1.0000
Single Mean	1	1.0289	0.9868	4.4964	1.0000
Trend	1	-1.7087	0.9732	-1.5202	0.8050

The columns "tau" and "pr<tau" can be interpreted in the same fashion as the columns "tau" and "pr<tau" in the ADF output and, as with the interpretation of the ADF output, one starts with the most general model and continues to the most specific stopping only when one can reject the null hypothesis. In the above output, the results for the trend, constant model are summarized in the third row. The tau statistic is the test that $\beta_1=0$ and one cannot reject the null of a unit root. One then proceeds to the constant, no trend model of row two. As with the trend, constant model, one cannot reject the null of a unit root. Finally one considers the no constant, no trend model. As with the more general models, one cannot reject the null of a unit root. The results of the test are consistent with the series having a unit root.

KWIATKOWSKI-PHILLIPS-SCHMIDT (KPSS) SHIN TEST

Unlike the ADF and PP tests, the KPSS test is a test of stationarity with the null being that the series is stationary (i.e. I(0)). To this extent the KPSS *might* serve as a complement to unit root tests where the null hypothesis – and thus the "benefit of the doubt" – is that the series is I(1). A rejection of the null hypothesis of stationarity in the KPSS test would then tend to corroborate a failure to reject the null hypothesis of a unit root in a ADF or PP test. However, power concerns means that one needs to be cautious in this interpretation. To test for stationarity using the KPSS test, one estimates the following model:

(3a) $\quad y_t = \beta_o + \beta_1 t + \mu_t + u_t$

(3b) $\quad \mu_t = \mu_{t-1} + \epsilon_t \sim iid(0, \sigma_\epsilon^2)$

Where β_0 is a constant, t is a trend, and μt is a random walk. The null hypothesis is specified as the variance of the error term in the random walk being equal to zero. Thus the KPSS test tests the hypothesis that $\sigma_\epsilon^2 = 0$.

There are several options available in the implementation of the KPSS test. To determine the bandwidth, one can set the bandwidth parameter manually (based on theory, if it is a guide), one can use a rule of thumb (given in SAS as floor{c(T/100)1/4} where c is a choice variable) or one can choose to allow SAS to implement an automatic bandwidth selection algorithm. In addition, one can select between the use of a Newey-West estimator and the Quadratic Spectral (QS) kernel in calculating the estimator of long run variance of the process.

Holbijn, Franses, and Ooms (1998) find in Monte Carlo simulations that the use of the QS kernel with automatic bandwidth selection leads to higher power in small samples and as with the ADF and PP tests, the power of the KPSS test is a concern. The syntax for the PP test using a QS kernel and automatic bandwidth selection is:

```
proc autoreg
    data = nhe_data;
        model PCA = /stationarity=(kpss=(kernel=qs auto));
    run;
```

To use the Schwert rule of thumb for the lag truncation parameter, one specifies schw=c in place of auto, to set the number of lags one specifies lag=c (where c is a choice parameter) and to substitute the Newey-West estimator for the Quadratic Spectral kernel one specifies kernel=nw.

The KPSS test generates the following output for the case of the QS kernel and automatic bandwidth selection:

KPSS Stationarity Test			
Type	Lags	Eta	Pr > Eta
Single Mean	2	1.5189	0.0002
Trend	2	0.3723	0.0004

In reading the output, one starts with the most general model (constant and trend) and continues to the most specific (no constant and no trend) stopping when one can reject the null hypothesis of stationarity. In the above output, the results for the trend, constant model are summarized in the second row. The eta statistic is the test that $\beta_1=0$ (in 3a) and one can reject the null of stationarity. The results of the test are consistent with the series having a unit root.

TEST COMPARISONS

The choice of lag length is very important in the ADF test. If one incorrectly includes too few lagged terms, the test will have size distortions while if one incorrectly includes too many lags, the power of the test will suffer. Additionally, the test has low power when testing for a unit root in a stationary process with a "near" unit root (such as a series with a trend).

The PP test has two advantages over the ADF test. The first is that one does not need to specify a lag length and can avoid the problems that mis-specifying the ADF regression model presents. The second is that the test is robust to heteroskedasticity unlike the ADF test which implicitly assumes heteroskedastic error terms. However, if the ADF regression model is correctly specified, the ADF test has higher power than the PP test. In addition, the PP involves the choice of a bandwidth parameter, and in finite samples, this can lead to size problems similar to those in the ADF test.

As with the ADF and PP tests, there is still the potentiality for size and power problems with the KPSS test. These relate (as with the PP test) to the selection of the bandwidth parameter. If too large a bandwidth parameter is chosen then the test has low power, while if a bandwidth parameter is chosen that is too small then the test is oversized. Nonetheless, the KPSS test can be a useful test for the applied economist; particularly in conjunction with one of the above unit root tests.

REFERENCES

Centers for Medicare and Medicaid Services. 2013. *National Health Expenditure Accounts: Methodology Paper*. Washington, D.C.: Centers for Medicare and Medicaid Services.

Granger, C.W.J. and P. Newbold. 1974. "Spurious regressions in econometrics." *Journal of Econometrics*. Volume 2:111-120.

Hobijn, B., P. H. Franses, M.Ooms. 1998. "Generalizations of the KPSS-test for Stationarity." *Econometric Institute Report*, no. 9802/A

Ng, S and P. Perron. 2001. "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power." *Econometrica*, Vol. 69, No. 6: 1519–1554.

SAS Institute Inc. 2008. *SAS/ETS® 9.3 User's Guide*. Cary, NC: SAS Institute Inc.

Schwert, G. W. 1989. "Tests for Unit Roots: A Monte Carlo Investigation." Journal of Business & Economic Statistics. Vol. 7, No. 2: 147-159.

Yule, G. U. 1926. "Why do we Sometimes get Nonsense-Correlations between Time-Series?--A Study in Sampling and the Nature of Time-Series." Journal of the Royal Statistical Society, Vol. 89, No. 1:1-63.

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Donald McCarthy
Department of Research and Evaluation, Kaiser Permanente
donald.p.mccarthy@kp.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.