

DETERMINANTS OF HOUSING PRICES IN THE UK FOR 1978-2015

INTRODUCTION

In this project, factors affecting housing prices in the context of UK have been studied from 1978 to 2015. Housing market is one of the significant sectors which affect the UK economy. More specifically, housing prices inflation in the UK is particularly substantial which affects economic aspects of UK. HM Treasury, in reference to Barker Report, has described increasing housing price will have unhelpful and unwelcome consequences for the economic wellbeing of the UK. The report of PwC (PricewaterhouseCoopers), world's second largest professional services network headquartered in London, states that if current market demand and supply for houses continue, by 2025, more than 50% of the people aged between 20 and 29 years of age will start renting privately (PwC, 2017). Housing prices get affected by many factors including interest rate, number of houses built per year, and other demand and supply factors (Attanasio and Weber, 2014).

However, study of housing prices and factors affecting this variable was particularly interesting for me because, in reference to PwC report, I also lie in the age bracket of 20 and 29 years. So, the issue of rising housing prices is also a concern for me from a home ownership perspective. Apart from this motivation, empirical studies on housing prices also shown the importance of its study where number of factors has been identified to affect housing prices (Attanasio and Weber, 2014; Cheng and Fung, 2015). Major factors, based on empirical studies, of housing price for 1978-2015 periods have been analysed in context of the UK in this project. This project has considered several demand and supply variables and tested four main hypotheses. The hypotheses are:

H1: GDP Per Capita has positive association with Housing Price.

H2: Houses Build Per Year has negative association with Housing Price.

H3: Interest Rate has negative association with Housing Price.

H4: Unemployment Rate has negative association with Housing Price.

Housing price is also very important from economic perspective because number of factors get involved in the behaviour of housing prices (Cohen and Karpavičiūtė, 2017). The study of Meese and Wallace (2014) argued that interest rate is one of the significant factors that affect housing prices. Increasing interest rate during inflationary period has found to be associated with increasing housing prices while it was not true during deflationary period with lowering interest rate have no impact on decreasing housing prices (Meese and Wallace, 2014). Gross Domestic Product (GDP) per capita and its association with housing prices have been argued to be positive (Yun Joe Wong, Man Eddie Hui, and Seabrooke, 2013). Fiva and Kirkebøen (2011) argued that higher national income represents higher ability of people to buy houses and for this reason higher GDP may aid in higher housing prices. Number of houses built per year in the UK may be one of the significant factors as supply in tandem with demand may also aid in hike/decrease in housing prices (Belej and Kulesza, 2014). Employment within an economy is also an important determinant of housing prices because higher unemployment rate represents larger number of people is being unable to buy new houses (Cheng and Fung, 2015).

DATA DESCRIPTION

This project has studied housing prices and several other variables as its determinants. In order to serve this purpose, this project has used secondary data from archived sources of UK Government as national level statistics were essential. Collection of primary data on housing prices, interest rate, number of houses built per year, GDP per capita, and unemployment rate would have been irrelevant for this project. For instance, housing prices could not be accurately estimated if it were collected primarily through survey. All these variables have been expressed quarterly between 1978 and 2015. As all these data were collected from UK Government sources therefore this project did not require any ethical approval. However, following table provides brief description of variables:

Table 1: Definition of Variables

Name of Variable	Definition
Real Housing Price (RHP)	Average housing price in a given year adjusted for inflation (£)
GDP Per Capita (GDPPC)	Real GDP per capita (£)
Houses Built Per Year (HBPY)	Aggregate number of houses built per year in the UK
Interest Rate (IR)	Rate of interest according to the Bank of England (BoE)
Unemployment Rate (UR)	Rate of unemployed population in the UK

REAL HOUSING PRICE

Data on *real Housing Price* has been collected from *Nationwide* society though Helifax, Office for National Statistics (ONS), and Rightmove also provides housing prices data. Nationwide has been chosen over other sources for two reasons. Firstly, Nationwide is the only source which provides longest time series data which extends to 1975. It allowed this project to include more statistical observations in the analysis which, on the other hand, increased reliability. Secondly,

housing prices data provided by Nationwide were adjusted for retail price inflation. Inflation adjusted housing price provided this project more representative analysis of the issue. For this project, quarterly housing price data has been collected from Nationwide.

GDP PER CAPITA

GDP per capita data on the UK has been collected from the ONS because this dataset was compiled through three GDP measurement approaches which includes income, output, and expenditure. In the ONS' database for GDP, there are three separate estimations for this variable for each quarter which ensures accurate measurement of data. ONS' database for GDP also adjusts inflation which makes the data more reliable and aids in more representative analysis. Quarterly data on GDP per capita has been collected from ONS' UK economic accounts.

HOUSES BUILT PER YEAR

Houses built per year data has been collected from the archives of the UK Government. The archive includes data on *number of permanent dwellings completed* per year which incorporates properties built by local authorities, housing associations, and private sector. Quarterly data on houses built per year has been extracted from UK Government archive and being a government archive this data must be reliable.

INTEREST RATE

The online source of Bank of England (BoE) has been used to obtain interest rate time series data. The time series data on interest rate provides both in quarterly and annualised form for each calendar year. This project has chosen quarterly interest rate data. As this data has been collected from the BoE which sets interest rate thus this data should be highly reliable.

UNEMPLOYMENT RATE

Unemployment data has been sourced from the database of ONS. ONS database for unemployment data excludes economically inactive people (for example, sick, students, retirees, disabled, and people less than 16 years of age). Quarterly data on unemployment rate has been extracted from the ONS database.

METHODS

This project has used several statistical data analysis procedures to study housing prices and its determinants. EViews has been used to make all the analyses in this project because this software package allows effective and easier ways of analysing time series data. First of all, correlations among the variables have been tested. Secondly, this project has tested for stationarity of the time series data by applying unit root test. Thirdly, this project has tested for autocorrelation by applying ARMA model in EViews. Fourthly, Breusch-Pagan-Godfrey test for heteroskedasticity has been carried out. Finally, after adjusting for stationarity, autocorrelation, and heteroskedasticity ordinary least squares regression has been carried out for final outcomes.

Correlation among all the variables has been calculated based on the data collected from stated sources. Correlation has been carried out to provide information on nature of associations and strength of associations among the variables. This helped understanding the characteristic of relationship between the variables.

Time series data may suffer from non-stationarity because it is very possible that variables may be correlated with time. That is, a variable may change along with the passage of time. Non-stationarity in a time series data may indicate a systematic pattern that is unpredictable. Therefore, it was essential to run a unit root test to confirm there is no systematic pattern in the dataset that is unpredictable and affects the outcomes of the analysis.

Ordinary Least Squares (OLS) regression has been carried out for the time series data which helped estimating unknown parameters in the linear regression model. The goal of applying OLS was to minimise the sum of the squares of the differences between observed response and those predicted by the linear function of a set of explanatory variables. The consistency of OLS estimators is confirmed when regressors are optimal in class of linear unbiased estimators and exogenous especially when the errors are serially uncorrelated and homoscedastic.

OLS required serially uncorrelated and homoscedastic dataset for further data analysis. Therefore, it was also essential to test for serial correlation (autocorrelation) and heteroskedasticity. This project has used Breusch-Godfrey LM and Durbin-Watson test to examine if the time series data of this project was auto correlated. To remove autocorrelation from the regression model this project has used ARMA (Autoregressive and Moving-average)

model. This project has used Breusch-Pagan-Godfrey test for heteroskedasticity which is actually performed to estimate if the variability of a variable is unequal across the range of values of a second variable that predicts it.

However, the OLS equation of this project is:

$$\text{LOG(RHP(-1))} = \text{C(1)} + \text{C(2)*LOG(GDPPC(-1))} + \text{C(3)*LOG(HBPY)} + \text{C(4)*LOG(IR(-1))} + \text{C(5)*LOG(UR(-2))} + \text{C(6)*@TREND}$$

In this OLS model;

Log(RHP(-1))= Log value of First Lag of Real Housing Price

C(1)= Constant of the OLS Regression

C(2)= Coefficient of First Lag of GDP Per Capita

GDPPC(-1)= Log value of First Lag of GDP Per Capita

C(3)= Coefficient of Houses Built Per Year

Log(HBPY)= Log value of Houses Built Per Year

C(4)= Coefficient of First Lag of Interest Rate

Log(IR(-1))= Log value of First Lag of Interest Rate

C(5)= Coefficient of Unemployment Rate

Log(UR(-2))= Log value of Second Lag of Unemployment Rate

C(6)= Coefficient of Trend

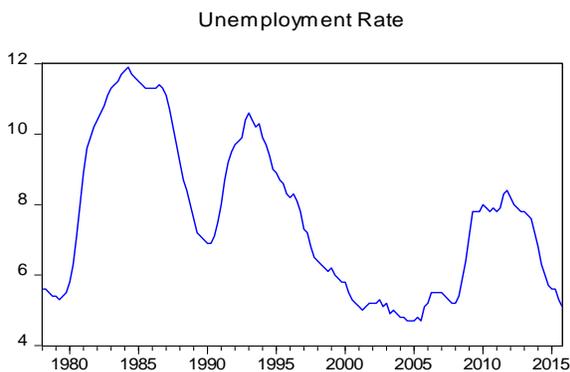
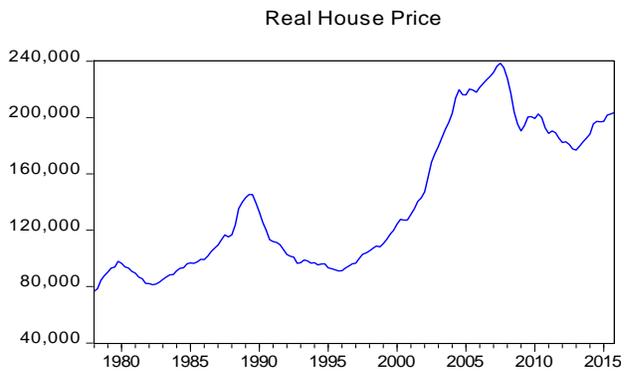
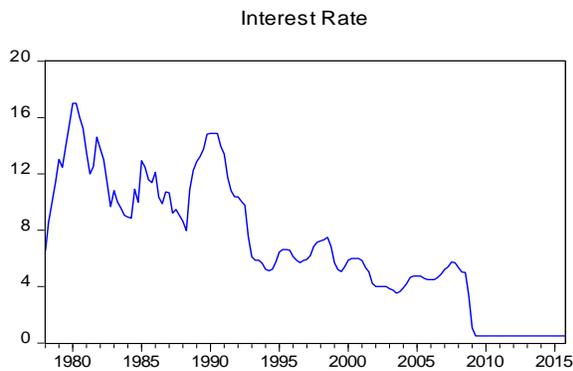
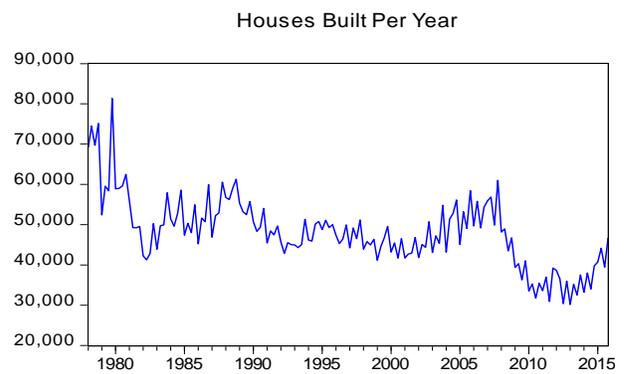
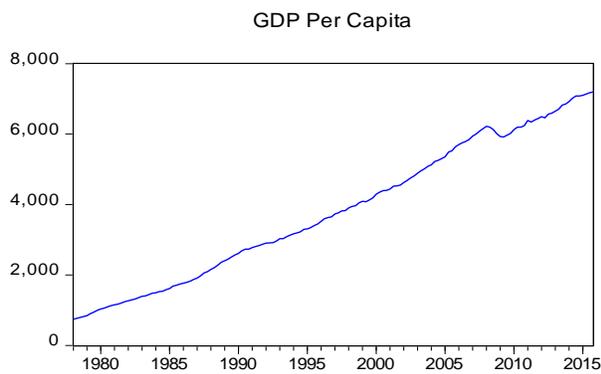
@Trend= Trend adjustment value for the OLS Model

This OLS model helped analysing the variables without any error as this project has removed non-stationarity, autocorrelation, and heteroskedasticity within the dataset alongside adjustment for lagged value and trend effect.

RESULTS

SIMPLE TIME SERIES REGRESSION

First of all, this project estimated a model that captured movements in housing prices based on GDP per capita, houses built per year, interest rate, and unemployment rate. From the graph presented below it can be observed that except for GDP Per Capita other variables have substantial variation across the time period observed which may be an indicator of non-stationarity.



However, this project has run simple time series regression on the raw data which returned following output:

Dependent Variable: LOG(RHP)
Method: Least Squares
Date: 04/28/17 Time: 01:23
Sample: 1978Q1 2015Q4
Included observations: 152

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.117958	1.605057	2.565615	0.0113
LOG(GDPPC)	0.361339	0.038273	9.441097	0.0000
LOG(HBPY)	0.502616	0.129701	3.875198	0.0002
LOG(IR)	-0.100928	0.020670	-4.882762	0.0000
LOG(UR)	-0.255714	0.060160	-4.250538	0.0000
R-squared	0.782475	Mean dependent var	11.78497	
Adjusted R-squared	0.776556	S.D. dependent var	0.346703	
S.E. of regression	0.163886	Akaike info criterion	-0.746950	
Sum squared resid	3.948216	Schwarz criterion	-0.647481	
Log likelihood	61.76822	Hannan-Quinn criter.	-0.706542	
F-statistic	132.1963	Durbin-Watson stat	0.151843	
Prob(F-statistic)	0.000000			

From this simple time series regression it can be observed that 78.24% of variations in housing prices have been explained by the variations in the dependent variables (i.e. GDP per capita, houses built per year, interest rate, and unemployment rate). All the variables were statistically significant while GDP per capita, interest rate, and unemployment rate having expected sign and houses built per year with unexpected sign. From this regression, housing prices get significantly affected by all the variables. As it has been said earlier that time series data may change along with time i.e. it may not be stationary. For this reason unit root test has been carried out to see if the dataset is stationary.

Unit root test for all the variables indicated that only houses built per year was stationary while all other variables were non stationary. The unit root test for real housing price indicated that RHP was non-stationary at level while it was stationary at its first difference or lag. Similarly, GDPPC and IR were also non-stationary at their level and stationary at their first lag. UR was non-stationary at level and its first lag but was stationary at its second difference. HBPY was stationary at its level. Therefore, for further econometric analysis this project has used first lag of RHP, GDPPC, and IR; second lag of UR, and level data of HBPY. Results of unit root test is

provided in Appendix. However, after correcting data from non-stationary to stationary the least squares regression provided following output:

Dependent Variable: LOG(RHP(-1))
 Method: Least Squares
 Date: 04/28/17 Time: 02:01
 Sample (adjusted): 1978Q3 2015Q4
 Included observations: 150 after adjustments

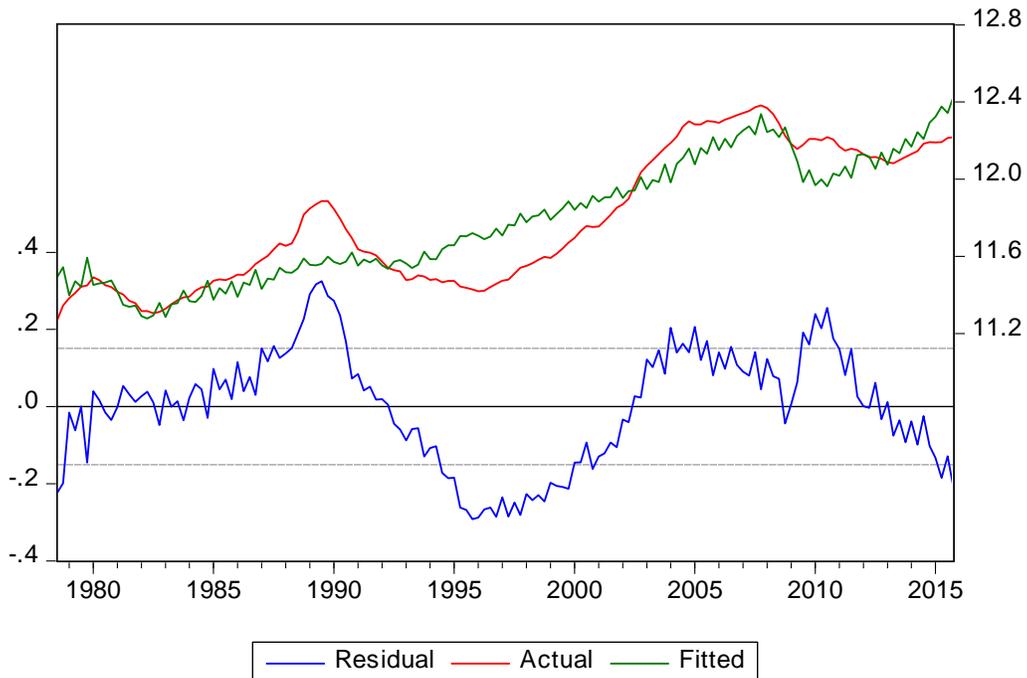
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.721966	1.526920	3.092478	0.0024
LOG(GDPPC(-1))	0.332455	0.038149	8.714693	0.0000
LOG(HBPY)	0.479226	0.126004	3.803260	0.0002
LOG(IR(-1))	-0.108252	0.021215	-5.102525	0.0000
LOG(UR(-2))	-0.307445	0.057796	-5.319530	0.0000
R-squared	0.783942	Mean dependent var		11.78563
Adjusted R-squared	0.777982	S.D. dependent var		0.344383
S.E. of regression	0.162269	Akaike info criterion		-0.766357
Sum squared resid	3.818030	Schwarz criterion		-0.666002
Log likelihood	62.47677	Hannan-Quinn criter.		-0.725586
F-statistic	131.5291	Durbin-Watson stat		0.148340
Prob(F-statistic)	0.000000			

Similar to the earlier regression, current regression also provides similar results although it has been adjusted for non-stationarity. As the results did not vary substantially this project considers adjusting for trend because economic time series data has a tendency to grow over time. Therefore, ignoring the fact of time trend can lead this project to falsely conclude regarding the impact of one variable on the other. Finding the relationship between two or more trending variable was essential because they grow over time which indicates spurious regression problem. For this reason, this project has added a time trend to this regression model. After adjusting for time trend following result is found:

Dependent Variable: LOG(RHP(-1))
 Method: Least Squares
 Date: 04/28/17 Time: 02:36
 Sample (adjusted): 1978Q3 2015Q4
 Included observations: 150 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.600316	1.735237	5.532569	0.0000
LOG(GDPPC(-1))	-0.480978	0.170239	-2.825314	0.0054
LOG(HBPY)	0.443705	0.117333	3.781576	0.0002
LOG(IR(-1))	0.096502	0.046318	2.083448	0.0390
LOG(UR(-2))	-0.058269	0.074073	-0.786644	0.4328
@TREND	0.016548	0.003387	4.885316	0.0000
R-squared	0.814660	Mean dependent var	11.78563	
Adjusted R-squared	0.808225	S.D. dependent var	0.344383	
S.E. of regression	0.150813	Akaike info criterion	-0.906378	
Sum squared resid	3.275204	Schwarz criterion	-0.785953	
Log likelihood	73.97836	Hannan-Quinn criter.	-0.857453	
F-statistic	126.5902	Durbin-Watson stat	0.154377	
Prob(F-statistic)	0.000000			

From the time trend adjusted regression output it can be seen that IR and UR became statistically insignificant and IR lost its expected sign. Furthermore, in previous regression coefficient of GDPPC was positive and statistically significant but this regression indicates that it has a negative coefficient and it is statistically insignificant.



Above figure is the actual, fitted, and residual plot of this regression where the behavior of residual appears to be correlated with their own lagged values which indicate serial correlation. Serial correlation is very common in time series data as data remains in ordered form over time. OLS regression assumes that error terms are uncorrelated and it is violated in above regression where error terms seem to be correlated. It was important to treat serial correlation because in the presence of lagged dependent variables the estimates provided by OLS will be inconsistent and biased. For this reason, this project has tested for serial correlation by using Durbin-Watson statistics and Breusch-Godfrey test for serial correlation.

EViews provides the value of Durbin-Watson statistics when a regression is run. From the regression table above it could be observed that Durbin-Watson statistics for this regression was 0.1543 which means that this project soundly rejects the null hypothesis of no serial correlation. If there were no serial correlation the Durbin-Watson statistics should have been around 2. In this case, the Durbin-Watson statistic falls below 2 which indicates positive serial correlation. On the other hand, the Breusch-Godfrey test for serial correlation indicates that the probability of chi-square is statistically significant. This implies that, the regression performed after adjustment for time trend is suffering from serial correlation.

Breusch-Godfrey Serial Correlation LM Test

F-statistic	577.4182	Prob. F(2,142)	0.0000
Obs*R-squared	133.5754	Prob. Chi-Square(2)	0.0000

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 04/28/17 Time: 03:07

Sample: 1978Q3 2015Q4

Included observations: 150

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.285651	0.591821	3.862064	0.0002
LOG(GDPPC(-1))	0.047789	0.056995	0.838468	0.4032
LOG(HBPY)	-0.223393	0.040278	-5.546223	0.0000
LOG(IR(-1))	0.000196	0.015547	0.012600	0.9900
LOG(UR(-2))	-0.076591	0.024810	-3.087144	0.0024
@TREND	-0.001534	0.001134	-1.352696	0.1783
RESID(-1)	0.619206	0.071110	8.707680	0.0000
RESID(-2)	0.366841	0.070628	5.193979	0.0000

R-squared	0.890503	Mean dependent var	-2.77E-15
Adjusted R-squared	0.885105	S.D. dependent var	0.148261
S.E. of regression	0.050255	Akaike info criterion	-3.091568
Sum squared resid	0.358626	Schwarz criterion	-2.931000
Log likelihood	239.8676	Hannan-Quinn criter.	-3.026334
F-statistic	164.9766	Durbin-Watson stat	1.220907
Prob(F-statistic)	0.000000		

In addition to serial correlation, this project may also suffer from heteroskedasticity because in many financial and economic time series the conditional variance of the error term depends on past values of the error term. This may indicate autoregressive conditional heteroskedasticity. However, this project has tested for heteroskedasticity; if this regression did not suffer from heteroskedasticity only serial correlation will be corrected. To detect heteroskedasticity this project has employed White and ARCH LM test. Both White and ARCH LM test provided similar results where probability of chi-square was statistically significant. Therefore, this project has found that the regression performed after trend adjustment suffered from both serial correlation and heteroskedasticity. This project needed to adjust standard errors in case of the presence of both serial correlation and heteroskedasticity. Both these issues have been addressed by HAC (Newey-West) standard errors test. After adjusting for both autocorrelation and heteroskedasticity it has been found that only HBPY was statistically significant but the concern for low Durbin-Watson statistics remained (results in Appendix). This implies that even after

addressing autocorrelation and heteroskedasticity the problem of serial correlation remained in the regression model. To solve this issue ARMA model was applied. After applying ARMA model following result has been obtained:

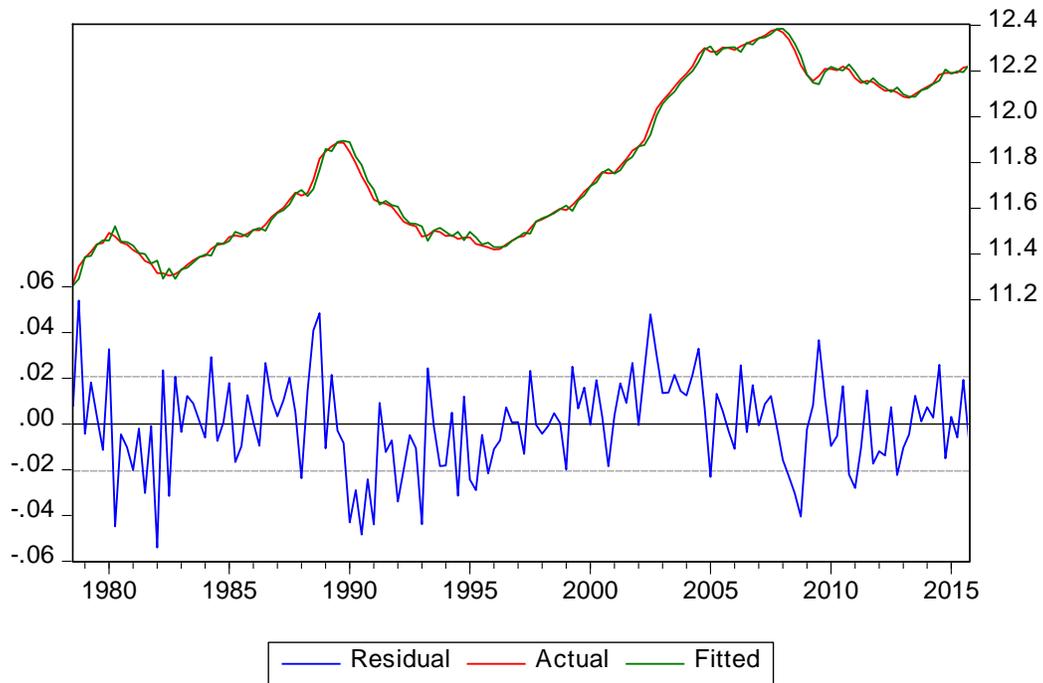
Dependent Variable: LOG(RHP(-1))
Method: ARMA Maximum Likelihood (BFGS)
Date: 04/28/17 Time: 03:28
Sample: 1978Q3 2015Q4
Included observations: 150
Convergence achieved after 11 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.392817	1.426556	6.584259	0.0000
LOG(GDPPC(-1))	0.288082	0.179835	1.601918	0.1114
LOG(HBPY)	-0.004987	0.008533	-0.584385	0.5599
LOG(IR(-1))	0.021291	0.015854	1.342971	0.1814
LOG(UR(-2))	-0.054152	0.059651	-0.907812	0.3655
@TREND	0.002711	0.003314	0.817845	0.4148
AR(1)	0.979854	0.018019	54.38041	0.0000
MA(1)	0.587802	0.073706	7.974982	0.0000
SIGMASQ	0.000401	4.72E-05	8.502465	0.0000

R-squared	0.996596	Mean dependent var	11.78563
Adjusted R-squared	0.996403	S.D. dependent var	0.344383
S.E. of regression	0.020654	Akaike info criterion	-4.833304
Sum squared resid	0.060150	Schwarz criterion	-4.652666
Log likelihood	371.4978	Hannan-Quinn criter.	-4.759916
F-statistic	5160.388	Durbin-Watson stat	1.654285
Prob(F-statistic)	0.000000		

Inverted AR Roots	.98
Inverted MA Roots	-.59

From this regression it can be seen that neither of the variables exhibit statistical significance which implies that housing prices do not get affected by these variables rather other unexplained variables. The Durbin-Watson statistics has also increased to 1.65 which is around 2 indicating reduced serial correlation. The residual plot of this regression (following graph) also indicates moderate behavior of residuals.



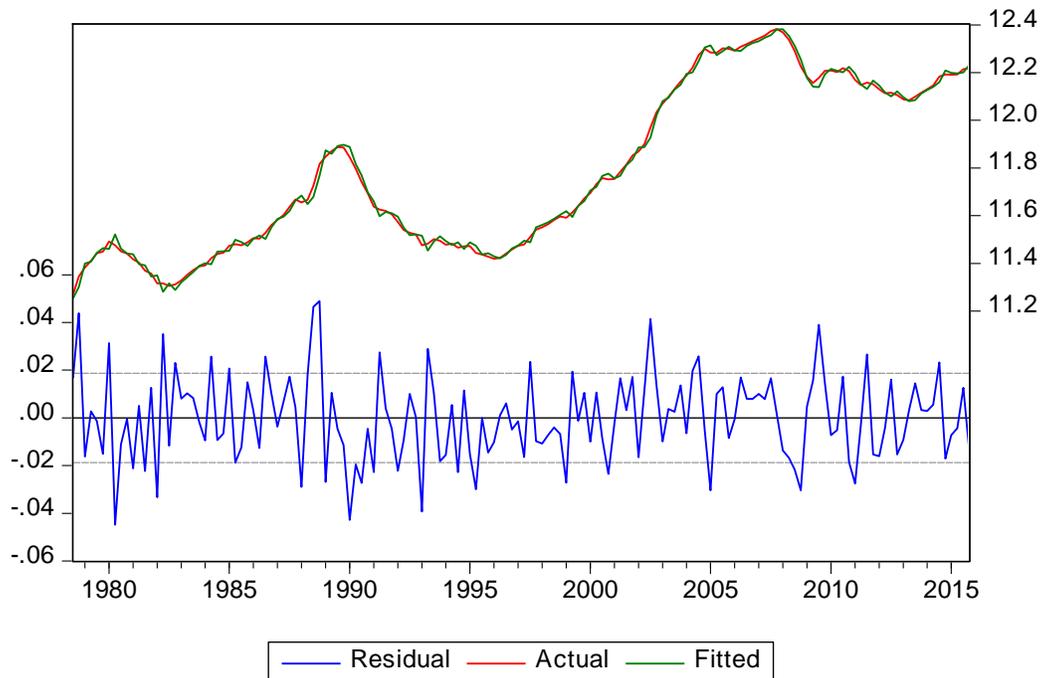
For confirmation of no persistent autocorrelation in this regression this project has run higher order ARMA model which provides following result:

Dependent Variable: LOG(RHP(-1))
 Method: ARMA Maximum Likelihood (BFGS)
 Date: 04/28/17 Time: 03:34
 Sample: 1978Q3 2015Q4
 Included observations: 150
 Convergence achieved after 29 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.538717	1.037757	9.191671	0.0000
LOG(GDPPC(-1))	0.212993	0.147327	1.445713	0.1505
LOG(HBPY)	-0.003371	0.008177	-0.412249	0.6808
LOG(IR(-1))	0.010943	0.012267	0.892061	0.3739
LOG(UR(-2))	0.085904	0.067814	1.266749	0.2074
@TREND	0.004752	0.002503	1.898412	0.0597
AR(1)	1.874983	0.058152	32.24289	0.0000
AR(2)	-0.889032	0.057770	-15.38911	0.0000
MA(1)	-0.187742	0.110789	-1.694596	0.0924
MA(2)	-0.254887	0.099151	-2.570694	0.0112
SIGMASQ	0.000325	4.07E-05	7.986031	0.0000

R-squared	0.997238	Mean dependent var	11.78563
Adjusted R-squared	0.997039	S.D. dependent var	0.344383
S.E. of regression	0.018740	Akaike info criterion	-5.008925
Sum squared resid	0.048814	Schwarz criterion	-4.788145
Log likelihood	386.6694	Hannan-Quinn criter.	-4.919229
F-statistic	5018.066	Durbin-Watson stat	2.017321
Prob(F-statistic)	0.000000		

Inverted AR Roots	.94-.10i	.94+.10i
Inverted MA Roots	.61	-.42



The result clearly demonstrates that this regression has no serial correlation and residuals are behaving more even. It has finally been found that GDPPC, HBPY, IR, and UR have no statistical influence over RHP. These variables had no statistical significance as well. Based on this regression model it can be concluded that all the hypotheses has been rejected for this project implying housing prices get affected by other unexplainable variable.

CONCLUSION

This project has studied the housing prices in the UK considering GDP per capita, interest rate, houses built per year, and unemployment rate as major determinants. It has been primarily observed that all these variables had statistically significant impact on housing prices but this was not true in the end after the regression model has been adjusted for non-stationarity, serial correlation, and heteroskedasticity. Finally, this project has found inconclusive evidence to support significant impact of GDP per capita, houses built per year, interest rate, and unemployment rate on housing prices in the UK for 1978-2015.

REFERENCE

- Attanasio, O. and Weber, G. (2014). The UK Consumption Boom of the Late 1980s: Aggregate Implications of Microeconomic Evidence. *The Economic Journal*, 104(427), p.1269.
- Belej, M. and Kulesza, S. (2014). The Influence of Financing On The Dynamics Of Housing Prices. *Folia Oeconomica Stetinensia*, 14(2).
- Cheng, A. and Fung, M. (2015). Determinants of Housing Prices. *Journal of Economics, Business and Management*, 3(3), pp.352-355.
- Cohen, V. and Karpavičiūtė, L. (2017). The analysis of the determinants of housing prices. *Independent Journal of Management & Production*, 8(1), pp.49-63.
- Fiva, J. and Kirkebøen, L. (2011). Information Shocks and the Dynamics of the Housing Market*. *Scandinavian Journal of Economics*, p.no-no.
- HM Treasury (2010). [ARCHIVED CONTENT] UK Government Web Archive – The National Archives. [online] Webarchive.nationalarchives.gov.uk. Available at: http://webarchive.nationalarchives.gov.uk/+http://www.hm-treasury.gov.uk/barker_review_of_housing_supply_recommendations.htm [Accessed 27 Apr. 2017].
- Meese, R. and Wallace, N. (2014). House Price Dynamics and Market Fundamentals: The Parisian Housing Market. *Urban Studies*, 40(5-6), pp.1027-1045.
- PwC (2017). *UK Economic Outlook July 2015: Housing market outlook moderates but rise of Generation Rent will continue*. [online] PwC. Available at: <http://www.pwc.co.uk/services/economics-policy/insights/uk-economic-outlook/ukeo-july2015-housing-market-outlook.html> [Accessed 27 Apr. 2017].
- Yun Joe Wong, T., Man Eddie Hui, C. and Seabrooke, W. (2013). The impact of interest rates upon housing prices: an empirical study of Hong Kong's market. *Property Management*, 21(2), pp.153-170.

RESEARCH LOG

1ST PHASE

- Choice of Topic between Housing Prices Determinants or Labour Productivity.
- Finally chosen to study Housing Prices because of data availability and personal interest.
- Selection of variables and collection of secondary data.
- Proposal for the project is submitted.

2ND PHASE

- Methods Planner is submitted.
- Assembled secondary data into MS Excel worksheet.
- Met supervisor for further guidelines.

3RD PHASE

- Extracted data into EViews workfare from MS Excel worksheet.
- Ran correlation analysis on the variables.
- Ran simple regression on the dataset.
- Unit Root test is performed for stationarity.
- Ordinary Least Square regression is performed.
- Tested the regression for Trend effect.
- Again, OLS is run.
- Tested regression for Autocorrelation and Heteroskedasticity.
- Adjusted for Autocorrelation and Heteroskedasticity.
- Finally, OLS is run after addressing for Autocorrelation and Heteroskedasticity.

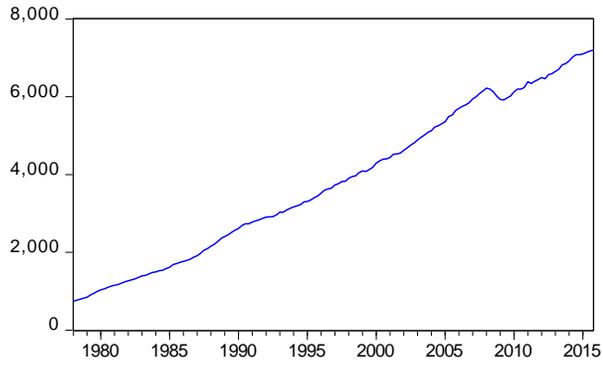
4TH PHASE

- Started writing report based on analyses performed.
- Created reference list and appendix.
- Formatted the report and finalized for submission.

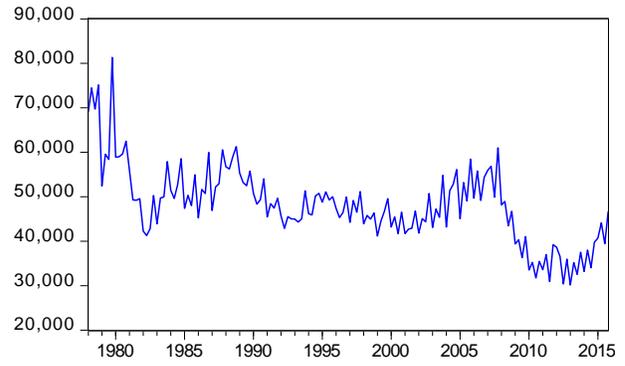
APPENDICES

GRAPH

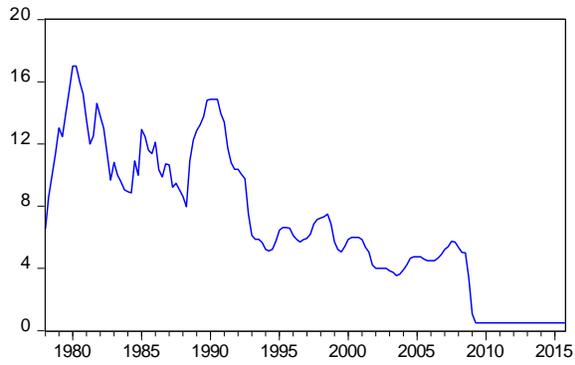
GDP Per Capita



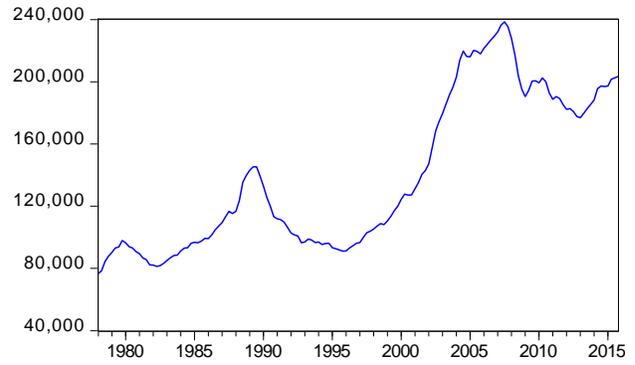
Houses Built Per Year



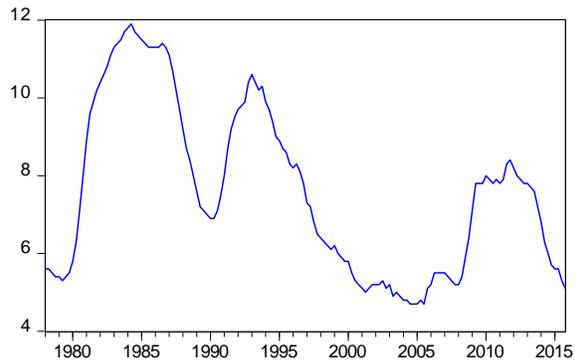
Interest Rate



Real House Price



Unemployment Rate



CORRELATION

Table 2: Correlation of Original Data

	GDPPC	HBPY	IR	RHP	UR
GDPPC	1	-0.6244346...	-0.8865111...	0.86783222...	-0.5249707...
HBPY	-0.6244346...	1	0.62249754...	-0.3382037...	-0.0489138...
IR	-0.8865111...	0.62249754...	1	-0.6760091...	0.34677704...
RHP	0.86783222...	-0.3382037...	-0.6760091...	1	-0.5802455...
UR	-0.5249707...	-0.0489138...	0.34677704...	-0.5802455...	1

Table 3: Correlation of Lagged Data

	DRHP	DIR	DGDPPC	D2UR	HBPY
DRHP	1	-0.6765518...	0.86488382...	-0.5801746...	-0.3203483...
DIR	-0.6765518...	1	-0.8934154...	0.32389651...	0.61484825...
DGDPPC	0.86488382...	-0.8934154...	1	-0.5242098...	-0.6095439...
D2UR	-0.5801746...	0.32389651...	-0.5242098...	1	-0.0286966...
HBPY	-0.3203483...	0.61484825...	-0.6095439...	-0.0286966...	1

UNIT ROOT TEST

Null Hypothesis: GDPPC has a unit root
 Exogenous: Constant
 Lag Length: 1 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.616002	0.9898
Test critical values:		
1% level	-3.474265	
5% level	-2.880722	
10% level	-2.577077	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(GDPPC)
 Method: Least Squares
 Date: 04/26/17 Time: 16:34
 Sample (adjusted): 1978Q3 2015Q4
 Included observations: 150 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDPPC(-1)	0.000894	0.001451	0.616002	0.5388
D(GDPPC(-1))	0.252090	0.080202	3.143196	0.0020
C	28.58415	6.804908	4.200519	0.0000
R-squared	0.068468	Mean dependent var		42.82000
Adjusted R-squared	0.055794	S.D. dependent var		35.21936
S.E. of regression	34.22275	Akaike info criterion		9.923456
Sum squared resid	172165.9	Schwarz criterion		9.983669
Log likelihood	-741.2592	Hannan-Quinn criter.		9.947918
F-statistic	5.402248	Durbin-Watson stat		2.069263
Prob(F-statistic)	0.005446			

GDPPC has a unit root with t-statistic being lower than all critical values. This clearly demonstrates the non-stationarity characteristic of GDPPC. So we have tested if first difference of GDPPC has unit root or not.

Null Hypothesis: D(GDPPC) has a unit root
 Exogenous: Constant
 Lag Length: 0 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-9.332958	0.0000
Test critical values:		
1% level	-3.474265	
5% level	-2.880722	
10% level	-2.577077	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(GDPPC,2)
 Method: Least Squares
 Date: 04/26/17 Time: 16:36
 Sample (adjusted): 1978Q3 2015Q4
 Included observations: 150 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(GDPPC(-1))	-0.742566	0.079564	-9.332958	0.0000
C	31.77263	4.408289	7.207474	0.0000
R-squared	0.370492	Mean dependent var		-0.093333
Adjusted R-squared	0.366238	S.D. dependent var		42.89825
S.E. of regression	34.15093	Akaike info criterion		9.912701
Sum squared resid	172610.4	Schwarz criterion		9.952842
Log likelihood	-741.4525	Hannan-Quinn criter.		9.929009
F-statistic	87.10411	Durbin-Watson stat		2.074382
Prob(F-statistic)	0.000000			

We have found that the first difference of GDPPC has no unit root with t-statistic being higher than all critical values which implies the first difference of GDPPC is stationary. For this reason, we have used the first difference of GDPPC for further econometric analysis.

Null Hypothesis: HBPY has a unit root
 Exogenous: Constant
 Lag Length: 7 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.695491	0.0051
Test critical values: 1% level	-3.476143	
5% level	-2.881541	
10% level	-2.577514	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(HBPY)
 Method: Least Squares
 Date: 04/26/17 Time: 16:52
 Sample (adjusted): 1980Q1 2015Q4
 Included observations: 144 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
HBPY(-1)	-0.141860	0.038387	-3.695491	0.0003
D(HBPY(-1))	-0.425294	0.076638	-5.549393	0.0000
D(HBPY(-2))	-0.045851	0.083075	-0.551921	0.5819
D(HBPY(-3))	0.052606	0.082906	0.634525	0.5268
D(HBPY(-4))	0.469121	0.074193	6.322988	0.0000
D(HBPY(-5))	0.194146	0.077536	2.503957	0.0135
D(HBPY(-6))	-0.040114	0.076607	-0.523632	0.6014
D(HBPY(-7))	-0.181279	0.068565	-2.643911	0.0092
C	6522.275	1850.671	3.524276	0.0006
R-squared	0.690546	Mean dependent var		-240.1389
Adjusted R-squared	0.672208	S.D. dependent var		5450.974
S.E. of regression	3120.854	Akaike info criterion		18.99006
Sum squared resid	1.31E+09	Schwarz criterion		19.17568
Log likelihood	-1358.285	Hannan-Quinn criter.		19.06549
F-statistic	37.65647	Durbin-Watson stat		1.993321
Prob(F-statistic)	0.000000			

Null Hypothesis: IR has a unit root
 Exogenous: Constant
 Lag Length: 1 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.171655	0.6859
Test critical values:		
1% level	-3.474265	
5% level	-2.880722	
10% level	-2.577077	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(IR)
 Method: Least Squares
 Date: 04/26/17 Time: 16:53
 Sample (adjusted): 1978Q3 2015Q4
 Included observations: 150 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
IR(-1)	-0.016402	0.013999	-1.171655	0.2432
D(IR(-1))	0.306424	0.077211	3.968644	0.0001
C	0.073511	0.116730	0.629752	0.5298
R-squared	0.099222	Mean dependent var		-0.053783
Adjusted R-squared	0.086967	S.D. dependent var		0.801337
S.E. of regression	0.765700	Akaike info criterion		2.323744
Sum squared resid	86.18550	Schwarz criterion		2.383956
Log likelihood	-171.2808	Hannan-Quinn criter.		2.348206
F-statistic	8.096149	Durbin-Watson stat		2.087797
Prob(F-statistic)	0.000462			

IR is non-stationary.

Null Hypothesis: D(IR) has a unit root
 Exogenous: Constant
 Lag Length: 0 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-9.199565	0.0000
Test critical values:		
1% level	-3.474265	
5% level	-2.880722	
10% level	-2.577077	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(IR,2)
 Method: Least Squares
 Date: 04/26/17 Time: 16:53
 Sample (adjusted): 1978Q3 2015Q4
 Included observations: 150 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(IR(-1))	-0.705253	0.076662	-9.199565	0.0000
C	-0.041930	0.062674	-0.669021	0.5045
R-squared	0.363802	Mean dependent var		-0.013568
Adjusted R-squared	0.359503	S.D. dependent var		0.957958
S.E. of regression	0.766663	Akaike info criterion		2.319706
Sum squared resid	86.99036	Schwarz criterion		2.359847
Log likelihood	-171.9779	Hannan-Quinn criter.		2.336014
F-statistic	84.63199	Durbin-Watson stat		2.076191
Prob(F-statistic)	0.000000			

First difference of IR is stationary.

Null Hypothesis: RHP has a unit root
 Exogenous: Constant
 Lag Length: 6 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.918051	0.7802
Test critical values:		
1% level	-3.475819	
5% level	-2.881400	
10% level	-2.577439	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(RHP)
 Method: Least Squares
 Date: 04/26/17 Time: 16:55
 Sample (adjusted): 1979Q4 2015Q4
 Included observations: 145 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RHP(-1)	-0.004079	0.004443	-0.918051	0.3602
D(RHP(-1))	0.896736	0.084163	10.65476	0.0000
D(RHP(-2))	-0.336150	0.110286	-3.047980	0.0028
D(RHP(-3))	0.198140	0.114497	1.730534	0.0858
D(RHP(-4))	0.171978	0.114818	1.497823	0.1365
D(RHP(-5))	-0.351300	0.110379	-3.182668	0.0018
D(RHP(-6))	0.143842	0.084254	1.707238	0.0900
C	801.0514	654.6548	1.223624	0.2232
R-squared	0.585961	Mean dependent var		755.5151
Adjusted R-squared	0.564806	S.D. dependent var		3885.817
S.E. of regression	2563.444	Akaike info criterion		18.58968
Sum squared resid	9.00E+08	Schwarz criterion		18.75392
Log likelihood	-1339.752	Hannan-Quinn criter.		18.65642
F-statistic	27.69812	Durbin-Watson stat		1.980926
Prob(F-statistic)	0.000000			

RHP is non-stationary.

Null Hypothesis: D(RHP) has a unit root
 Exogenous: Constant
 Lag Length: 5 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.758298	0.0042
Test critical values:		
1% level	-3.475819	
5% level	-2.881400	
10% level	-2.577439	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(RHP,2)
 Method: Least Squares
 Date: 04/26/17 Time: 16:56
 Sample (adjusted): 1979Q4 2015Q4
 Included observations: 145 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(RHP(-1))	-0.290162	0.077206	-3.758298	0.0003
D(RHP(-1),2)	0.187749	0.090631	2.071575	0.0402
D(RHP(-2),2)	-0.151167	0.092137	-1.640670	0.1031
D(RHP(-3),2)	0.046954	0.091209	0.514797	0.6075
D(RHP(-4),2)	0.214741	0.082940	2.589105	0.0107
D(RHP(-5),2)	-0.137781	0.083947	-1.641283	0.1030
C	235.4163	221.1467	1.064526	0.2889
R-squared	0.280732	Mean dependent var		1.621854
Adjusted R-squared	0.249459	S.D. dependent var		2957.258
S.E. of regression	2561.983	Akaike info criterion		18.58202
Sum squared resid	9.06E+08	Schwarz criterion		18.72573
Log likelihood	-1340.197	Hannan-Quinn criter.		18.64041
F-statistic	8.976944	Durbin-Watson stat		1.976761
Prob(F-statistic)	0.000000			

First difference of RHP is stationary.

Null Hypothesis: UR has a unit root
 Exogenous: Constant
 Lag Length: 2 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.547902	0.1064
Test critical values:		
1% level	-3.474567	
5% level	-2.880853	
10% level	-2.577147	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(UR)
 Method: Least Squares
 Date: 04/26/17 Time: 17:31
 Sample (adjusted): 1978Q4 2015Q4
 Included observations: 149 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
UR(-1)	-0.017288	0.006785	-2.547902	0.0119
D(UR(-1))	0.637246	0.080534	7.912714	0.0000
D(UR(-2))	0.189195	0.081750	2.314297	0.0221
C	0.131081	0.053979	2.428390	0.0164
R-squared	0.622351	Mean dependent var		-0.002685
Adjusted R-squared	0.614537	S.D. dependent var		0.285919
S.E. of regression	0.177514	Akaike info criterion		-0.593050
Sum squared resid	4.569151	Schwarz criterion		-0.512408
Log likelihood	48.18226	Hannan-Quinn criter.		-0.560287
F-statistic	79.65136	Durbin-Watson stat		1.993992
Prob(F-statistic)	0.000000			

UR is non-stationary.

Null Hypothesis: D(UR) has a unit root
 Exogenous: Constant
 Lag Length: 1 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.455662	0.0106
Test critical values:		
1% level	-3.474567	
5% level	-2.880853	
10% level	-2.577147	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(UR,2)
 Method: Least Squares
 Date: 04/26/17 Time: 17:31
 Sample (adjusted): 1978Q4 2015Q4
 Included observations: 149 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(UR(-1))	-0.191500	0.055416	-3.455662	0.0007
D(UR(-1),2)	-0.153214	0.082022	-1.867965	0.0638
C	-0.001365	0.014814	-0.092153	0.9267
R-squared	0.133616	Mean dependent var		-0.000671
Adjusted R-squared	0.121748	S.D. dependent var		0.192949
S.E. of regression	0.180822	Akaike info criterion		-0.562676
Sum squared resid	4.773717	Schwarz criterion		-0.502193
Log likelihood	44.91933	Hannan-Quinn criter.		-0.538103
F-statistic	11.25826	Durbin-Watson stat		1.981007
Prob(F-statistic)	0.000028			

First difference of UR is also non-stationary.

Null Hypothesis: D(UR,2) has a unit root
 Exogenous: Constant
 Lag Length: 0 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
<u>Augmented Dickey-Fuller test statistic</u>	-15.66031	0.0000
Test critical values: 1% level	-3.474567	
5% level	-2.880853	
10% level	-2.577147	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(UR,3)
 Method: Least Squares
 Date: 04/26/17 Time: 17:32
 Sample (adjusted): 1978Q4 2015Q4
 Included observations: 149 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(UR(-1),2)	-1.250518	0.079853	-15.66031	0.0000
C	-0.001176	0.015356	-0.076553	0.9391
R-squared	0.625234	Mean dependent var		0.001342
Adjusted R-squared	0.622685	S.D. dependent var		0.305133
S.E. of regression	0.187431	Akaike info criterion		-0.497480
Sum squared resid	5.164168	Schwarz criterion		-0.457158
Log likelihood	39.06223	Hannan-Quinn criter.		-0.481098
F-statistic	245.2453	Durbin-Watson stat		2.008259
Prob(F-statistic)	0.000000			

Second difference of UR is stationary.

AUTOCORRELATION AND HETEROSKEDASTICITY

Heteroskedasticity: White

Heteroskedasticity Test: White

F-statistic	5.028243	Prob. F(5,144)	0.0003
Obs*R-squared	22.29606	Prob. Chi-Square(5)	0.0005
Scaled explained SS	13.94783	Prob. Chi-Square(5)	0.0159

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 04/28/17 Time: 03:14

Sample: 1978Q3 2015Q4

Included observations: 150

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.147156	0.113128	-1.300796	0.1954
LOG(GDPPC(-1))^2	0.002185	0.000524	4.172498	0.0001
LOG(HBPY)^2	0.000520	0.000815	0.638090	0.5244
LOG(IR(-1))^2	-0.001081	0.002216	-0.488006	0.6263
LOG(UR(-2))^2	-0.001383	0.002297	-0.602040	0.5481
@TREND^2	-3.28E-06	8.52E-07	-3.846028	0.0002

R-squared	0.148640	Mean dependent var	0.021835
Adjusted R-squared	0.119079	S.D. dependent var	0.025526
S.E. of regression	0.023958	Akaike info criterion	-4.585849
Sum squared resid	0.082654	Schwarz criterion	-4.465424
Log likelihood	349.9387	Hannan-Quinn criter.	-4.536924
F-statistic	5.028243	Durbin-Watson stat	0.349643
Prob(F-statistic)	0.000282		

Heteroskedasticity: ARCH LM

Heteroskedasticity Test: ARCH

F-statistic	161.1595	Prob. F(4,141)	0.0000
Obs*R-squared	119.7971	Prob. Chi-Square(4)	0.0000

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 04/28/17 Time: 03:16

Sample (adjusted): 1979Q3 2015Q4

Included observations: 146 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003011	0.001246	2.416930	0.0169
RESID^2(-1)	0.740939	0.084849	8.732413	0.0000
RESID^2(-2)	0.569721	0.095181	5.985669	0.0000
RESID^2(-3)	-0.499403	0.096996	-5.148720	0.0000
RESID^2(-4)	0.062848	0.082031	0.766144	0.4449

R-squared	0.820528	Mean dependent var	0.021794
Adjusted R-squared	0.815437	S.D. dependent var	0.025623
S.E. of regression	0.011008	Akaike info criterion	-6.146762
Sum squared resid	0.017085	Schwarz criterion	-6.044584
Log likelihood	453.7136	Hannan-Quinn criter.	-6.105245
F-statistic	161.1595	Durbin-Watson stat	1.942291
Prob(F-statistic)	0.000000		

Addressing Autocorrelation and Heteroskedasticity

Dependent Variable: LOG(RHP(-1))
 Method: Least Squares
 Date: 04/28/17 Time: 03:20
 Sample (adjusted): 1978Q3 2015Q4
 Included observations: 150 after adjustments
 HAC standard errors & covariance (Bartlett kernel, Newey-West fixed
 bandwidth = 5.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.600316	2.734059	3.511379	0.0006
LOG(GDPPC(-1))	-0.480978	0.315099	-1.526433	0.1291
LOG(HBPY)	0.443705	0.167423	2.650201	0.0089
LOG(IR(-1))	0.096502	0.076552	1.260607	0.2095
LOG(UR(-2))	-0.058269	0.108999	-0.534586	0.5938
@TREND	0.016548	0.006054	2.733581	0.0071
R-squared	0.814660	Mean dependent var	11.78563	
Adjusted R-squared	0.808225	S.D. dependent var	0.344383	
S.E. of regression	0.150813	Akaike info criterion	-0.906378	
Sum squared resid	3.275204	Schwarz criterion	-0.785953	
Log likelihood	73.97836	Hannan-Quinn criter.	-0.857453	
F-statistic	126.5902	Durbin-Watson stat	0.154377	
Prob(F-statistic)	0.000000	Wald F-statistic	70.97805	
Prob(Wald F-statistic)	0.000000			