Predicting new housing starts using key economic indicators

Leo Hoare

43436021

Econ334

Data was collected from various key economic indicators to predict new housing starts. The data was collected between years 1974 and 2014.

The coefficients are defined as the following:

Hstart: New housing starts, monthly data at seasonally annual rate (thousands).

UN: seasonally adjusted civilian unemployment rate (%).

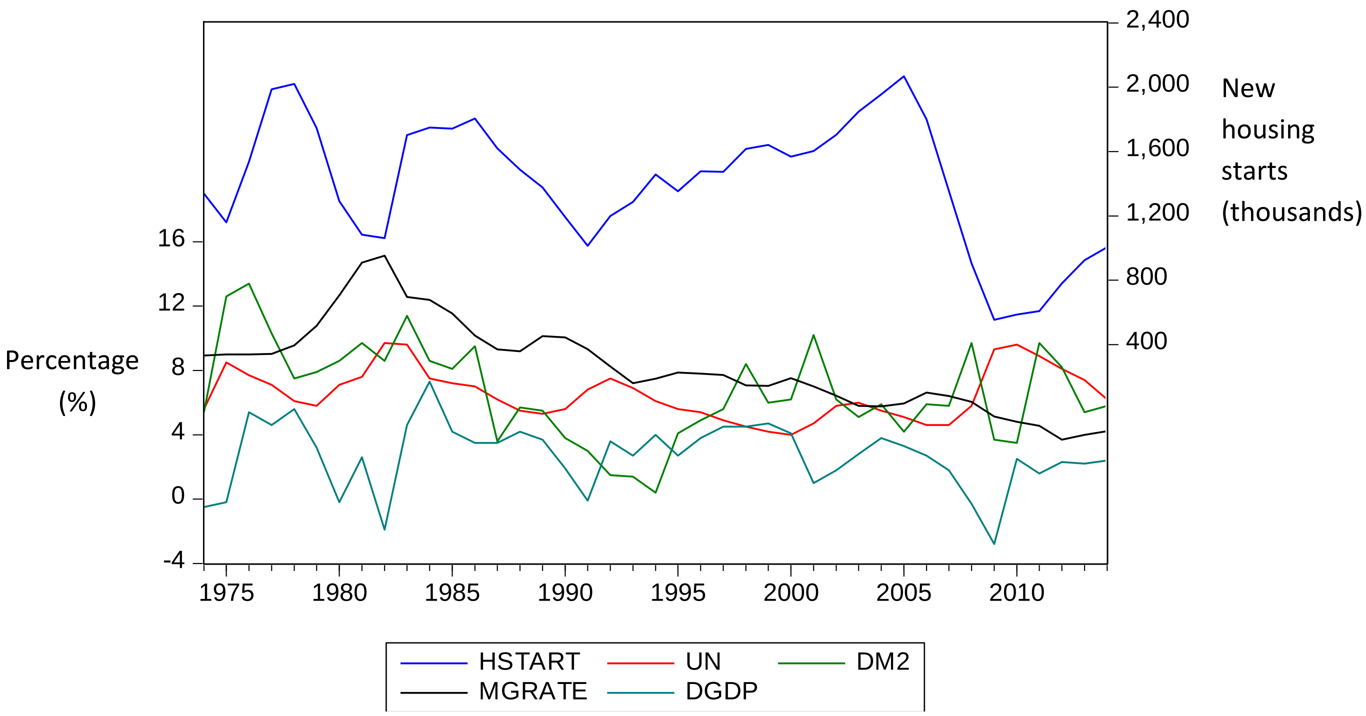
DM2: Percentage changes in (Seasonally adjusted) M2 money supply.

Mgrate: New home mortgage yield (%).

DGDP: Percentage changes in real gross domestic product.

（1） **[3 marks] Plot the data series using a line chart with Hstart on the right axis and all other variables on the left axis. Briefly comment on the dynamics (peaks and troughs) of the new housing starts. Hint on plotting data series on different axes: double click the graph, in the Graph Options window, click Axes & Scaling in the Option Pages box, assign left or right axis to each variable in the Series axis assignment box.**

New housing starts and other variables between years 1974 to 2014



New housing starts tend to fluctuate with a mean of 1407.200 (3 decimal places). There is an initial peak during 1977 to 1977 followed by a trough from 1981 to 1982. Housing starts then peak again at 1986 followed by a steady decline back to 1991. Housing starts then follow a steady peak until 2005 then rapidly decline to an all-time low during 2009 (554 housing starts). After 2009, housing starts begin to steadily increase until 2014. New housing starts follow distinct peaks and troughs throughout years 1974 to 2014.

(2) [3 marks] Estimate the regression model $Hstart_t = \beta_1 + \beta_2 UN_t + \beta_3 DM2_t + \beta_4 Mgrate_t + \beta_5 DGDP_t + \varepsilon_t$ by OLS and write down the fitted equation.

The regression above was estimated using ordinary least squares (OLS) method. The coefficients are stated in the OLS are defined at the beginning of the report. All estimators were rounded to three decimal places.

Original OLS:

$Hstart_t = \beta_1 + \beta_2 UN_t + \beta_3 DM2_t + \beta_4 Mgrate_t + \beta_5 DGDP_t + \varepsilon_t$

Fitted equation:

$\widehat{Hstart}_t$	= 1528.991 – 122.784(UN_t) + 18.321($DM2_t$) + 39.112 ($Mgrate_t$) + 86.078 ($DGDP_t$)				
s.ê.	= (223.212)	(29.707)	(15.136)	(16.166)	(21.439)
t.stat	= (6.850)	(-4.133)	(1.210)	(2.419)	(4.015)

$R^2 = 0.606$

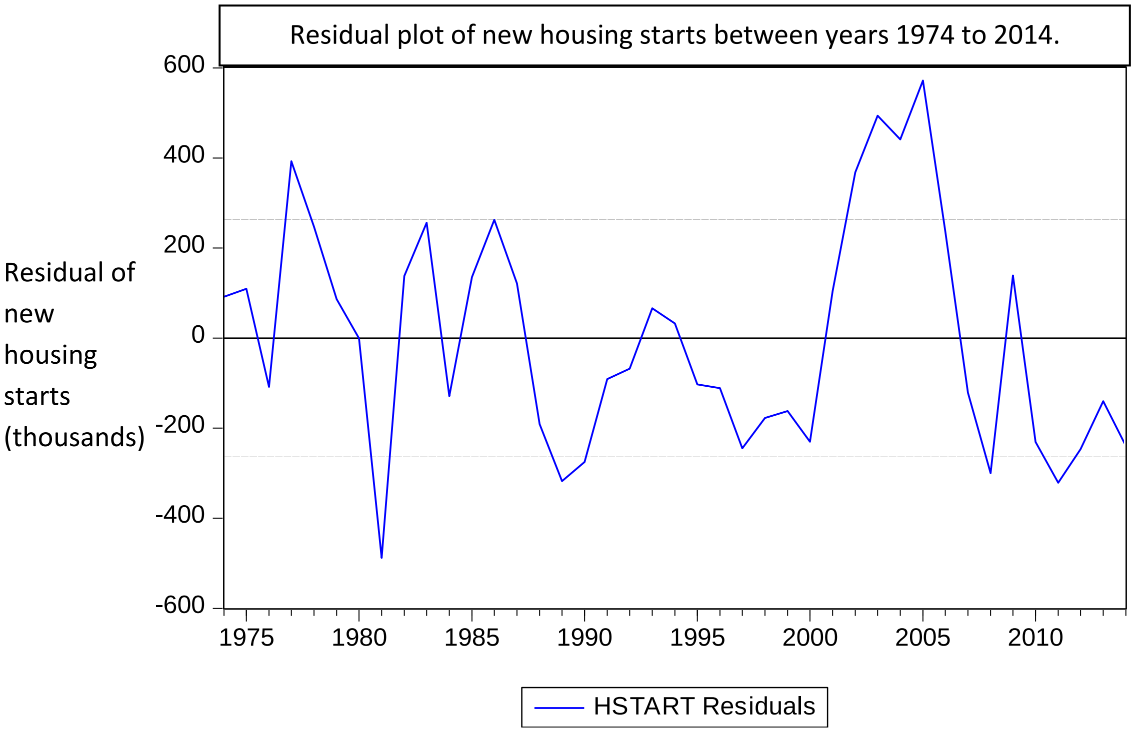
Where

s.ê. : the standard error of regression coefficients

t.stat : the test statistic $t.stat = \frac{\beta(a)}{s.e.(a)}$

R^2 : Coefficient of determination

(3) [3 marks] Examine the plot of the residuals from part (2) vs. time, i.e. generate the residual series and plot the line chart of the residual. Comment on the dynamics of the residual (i.e. showing pattern of serial correlation or not).



The residuals do not appear to be random but display patterns of peaks and troughs. Thus, the residual plot shows a general pattern of positive serial correlation. This is because if a positive or negative residual occurs the next score is similarly more likely to be the same. Due to this pattern, it is likely positive serial correlation is present. For example during 2001 to 2006 the residual graph is continuously positive in residuals. As the model doesn't continually go from positive to negative it is unlikely the serial correlation is negative. This shows a trend that the residual of the last score (t-1) has an effect on the next score (t), therefore, serial correlation may be present. Further analysis must be conducted to determine if serial correlation is present.

（4） **[8 marks] Test that the errors in model (1) are not serially correlated against the alternative that they are autoregressive of order 1 using the Breusch-Godfrey test at the 5% significance level. Your answer should state the null and alternative in terms of parameters in the model, your test statistic, what distribution this test statistic should follow under the null, and what decision rule you should use (including the exact critical value obtained from the statistical tables assuming that you are testing at the 5% level of significance).**

The Breusch-Godfrey LM test statistic is obtained from the auxiliary regression through regressing the residuals from the OLS against all regressors as well as the lagged residual to the first order. The test is testing for 1^{st} order correlation in the model.

Original OLS:

$Hstart_t = \beta_1 + \beta_2 UN_t + \beta_3 DM2_t + \beta_4 Mgrate_t + \beta_5 DGDP_t + \varepsilon_t$

Axillary regression:

$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \mu_t$

Where:

ρ1: coefficient for lagged residual to the first order

ε_t : Residuals from original OLS

ε_{t-1} : lagged residuals to one degree

H0: $\rho_1 = 0$

H1: $\rho_1 \neq 0$

The test statistic under the null

$LM = T.R^2_{aux} \sim \chi^2_1$

R^2: Coefficient of determination obtained from the auxiliary regression

T : number of observations from auxiliary regression

LM = 0.332163*40 (calculated via manual method)

LM = 13.287 (3 decimal places)

χ^2_1 (5% significance) = 393214 x 10^{-8}

Where χ^2_1: chi-squared distribution to one degree of freedom.

if LM<393214 x10^{-8} (χ^2_1) then retain H_0

As 13.287>393214 x 10^{-8}

As LM > χ^2_1 similarly P<0.000 (3 decimal places)

Reject H0 to 5% significance level.

According to the Breusch-Godfrey LM test to a 5% significance level, the model shows evidence of serial correlation. Therefore, despite the estimators still being unbiased and consistent, the standard errors will be incorrect. The model must be adjusted before correct statistical analysis can take place. Thus, the current estimators are not the best linear unbiased estimators (BLUE).

(5) [2 marks] How do serial correlated errors affect OLS estimation results?

The OLS will still be considered unbiased and consistent; however, the efficiency of the model is affected. This will result in the estimators of the OLS model not being the best linear unbiased estimators (BLUE). If the model is not efficient, the variance of estimators will not be minimised. The standard errors will be incorrect as well as statistical tests are invalid. The coefficients are still unbiased and consistent. The model must be adjusted to make statistical tests valid. This can be done using different methods that will be explored throughout the text.

(6) [5 marks] Estimate model (1) using Newey-West error correction method and write down the fitted equation. Comment on the estimation results in relation to your answer in part (5).

OLS regression is estimated with Newey-West error correction method. All estimators are rounded to three decimal places. The Newey-West error correction method is a method of calculating standard errors that will correct for serial correlation present in the model.

$$Hstart_t = \beta_1 + \beta_2 UN_t + \beta_3 DM2_t + \beta_4 Mgrate_t + \beta_5 DGDP_t + \varepsilon_t$$

Fitted equation (using Newey-West correction method):

$\widehat{Hstart}_t$	= 1528.991 – 122.784 (UN_t) + 18.321($DM2_t$) + 39.112($Mgrate_t$) + 86.078($DGDP_t$)				
s.ê.	= (265.349)	(31.158)	(10.341)	(22.607)	(13.150)
T.stat	= (5.762)	(-3.941)	(1.772)	(1.730)	(6.545)

$R^2 = 0.606$

As stated earlier, the coefficients from the Newey-West error correction method will have the same estimators that will be unbiased and consistent. The standard errors were previously incorrect, this was due to serial correlation being present. This will change the test statistics, hence why previous statistical analysis is invalid. The main difference in standard errors is between DGDP in both models; with the corrected standard errors DGDP is much more significant. Statistical tests and standard error interpretation is now correct with the error correction method. The Newey-West model can be considered BLUE due to the efficiency issue will be fixed.

(7) [3 marks] Re-estimate your model allowing for autoregressive errors and write down the fitted equation.

The following regression is estimated using autoregressive model. The residuals were lagged to the first order to account for positive serial correlation stated in the model. All regression estimators were rounded off to three decimal places.

The following autoregressive model will be estimated.

$Hstart_t = \beta_1 + \beta_2 UN_t + \beta_3 DM2_t + \beta_4 Mgrate_t + \beta_5 DGDP_t + \varepsilon_t$

$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \mu_t$

Where:

μ_t: residuals not accounted for by past residuals from the original OLS equation.

Fitted equation:

$\widehat{Hstart}_t$	= 1542.210 – 58.761(UN_t)	- 1.802($DM2_t$)	- 95.398($Mgrate_t$)	+ 38.895($DGDP_t$)	+ 0.978(ε_{t-1})	
s.ê.	= (2870.561)	(27.815)	(10.124)	(35.167)	(13.059)	(0.060)
T.stat	= (0.537)	(-2.113)	(-0.178)	(-2.713)	(2.978)	(16.289)

$R^2 = 0.839$

（8） [5 marks] Carefully interpret the autocorrelation-corrected OLS estimates of β_1 β_2 β_3 β_4 and β_5 based on results in (7).

The estimator β_1 is the intercept term, thus, when all coefficients are equal to zero, the intercept term will still be present. β_1 states that if there is zero unemployment, no change in the money supply, the new home mortgage yield is zero, and GDP does not change, then there will be 1,542,210 new houses built.

The estimator β_2 states that for each increase 1% of unemployment rate there will be 58,761 fewer houses built during the current month. Due to the regression model having lagged residuals changes in unemployment will also affect future outcomes.

The estimator β_3 claims that for each 1% change increase in the money supply there will be 1,802 less houses built during the current month. Similar to other regresses, this will affect future time periods due to the lagged residual in the autoregressive model.

The estimator β_4 explores the effect of the new home mortgage yield. For each 1% increase in the mortgage yield rate, new house starts will decrease by 95,398 during the current month. This will similarly have an effect on future months due to the nature of the autoregressive model.

The estimator β_5 states that for each 1% change increase in gross domestic product new housing starts will increase by 38,895 during the current month. This will also affect future new housing starts because of the lagged residual.

(9) [5 marks] What does economic theory suggest about the impact of the various regressors on new housing starts? Which set of regression results, part (6) or part (7), is closer to your prior expectation?

Economic theory would suggest that, in terms of unemployment, if unemployment was to increase then due to less disposable income and confidence, then new housing starts would decrease. This is shown in both models Newey-West and autoregressive. Although the Newey-West model has unemployment as a lot stronger effect, for every 1% rise in the unemployment rate, new housing starts will decrease by 122,784 compared to a decrease of 58,761 in the autoregressive model.

As money supply increase, the cash rate will decrease. This will cause other rates to decrease such as mortgage rate. A lower mortgage rate would promote an increase in housing starts according to economic theory. Therefore, as money supply increases, new housing starts to rise would be predicted to rise, thus, a positive relationship should be present. This is shown in the Newey-West estimators where for every 1% change increase in money supply new houses will increase by 18,321 that month. Alternatively, the autoregressive model shows that for a 1% change increase in money supply housing starts will decrease by 1,802 that month. Despite this going against economic theory, in practice individuals may not respond in the current time period to changes in money supply. A low response to money supply changes may also be due to individuals respond to economic indicators like gross domestic product (GDP) and unemployment that more directly affect them. If money supply changes, often it may be due to unemployment being high or GDP being low which is why new housing starts may not change very significantly. Further investigation would be needed on reactions to money supply and delay periods to determine which model would be more correct.

When mortgage yield increases economic theory would suggest that housing starts would decrease. This is due to the cost of maintaining a mortgage would increase promoting less houses to be built. This negative relationship is present in the autoregressive model and for every 1% rise in mortgage yield

95,398 less houses will be built. Alternatively in the Newey-West model for every 1% rise in mortgage yield 39,112 more houses will be built. The Newey-West model goes against economic theory in this model.

If GDP increases, this will lead to more income, thus promoting new building projects and houses. Therefore, as GDP increases, new housing starts should increase according to economic theory. This is shown in both models where a 1% change increase in GDP will cause an increase of 86,073 new houses in the Newey-West model and 38,895 increase in the autoregressive model. To determine which model is closer to economic theory, further econometric analysis of GDP on new housing starts must be conducted.

According to economic theory, both models are not completely sufficient. However, the autoregressive model follows economic theory closer to expectation. The primary issue is how individuals respond to changes to money supply, however, this could be due to many reasons as discussed. Further econometric analysis must be conducted on mortgage rates and new housing starts to evaluate the true effect of money supply on new housing starts.

Final proposed model to predict new housing starts:

Autoregressive model:

$$Hstart_t = \beta_1 + \beta_2 UN_t + \beta_3 DM2_t + \beta_4 Mgrate_t + \beta_5 DGDP_t + \varepsilon_t$$

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \mu_t$$

Fitted equation:

$$\widehat{Hstart}_t = 1542.210 - 58.761(UN_t) - 1.802(DM2_t) - 95.398(Mgrate_t) + 38.895(DGDP_t) + 0.978(\varepsilon_{t-1})$$

s.ê.	= (2870.561)	(27.815)	(10.124)	(35.167)	(13.059)	(0.060)
T.stat	= (0.537)	(-2.113)	(-0.178)	(-2.713)	(2.978)	(16.289)

$R^2 = 0.839$