ECON 706 - Problem Set 1

André Victor D. Luduvice, Stefano Pietrosanti and Joseph Silverstein

February 19, 2015

Abstract

In this article we study the behavior of housing completions and starts in a time series environment. We conduct detailed correlagram analysis, identification and estimation of the series both for univariate and multivariate cases for forecasting purposes. After proceeding with residuals analysis, we select and estimate an ARIMA(1,1,2) for each separate series and a VAR(14) for the simultaneous approach. The forecasts obtained with the VAR(14) shows close fit to the "inside-sample" predictions and outperform the univariate models.

1 Introduction

We downloaded the data for housing starts and completions from the Federal Reserve Bank of St. Louis website (FRED, 2014), where the result of the US Census's collection are downloadable. Below, we plot the two series using data going from January 1968 to December 2014.

We observe a high level of correlation between the series, a high overall persistence and also a high persistence of oscillation in both series. One can also immediately notice the stark decline exhibited by both series levels starting in a period that coincides with the 2008 crisis.

The structure of the article is as follows: in the next section, we conduct the univariate analysis divided in specification, selection and residual analysis. In Section 3, we proceed with the multivariate analysis working through identification, Granger Causality and Impulse-Response functions. Section 4 shows the results of our forecasts and finally, Section 5 contains the conclusions of our study.

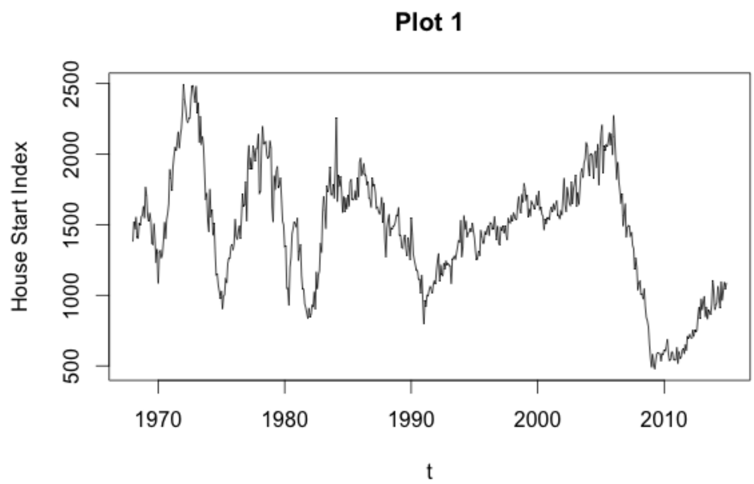

Figure 1

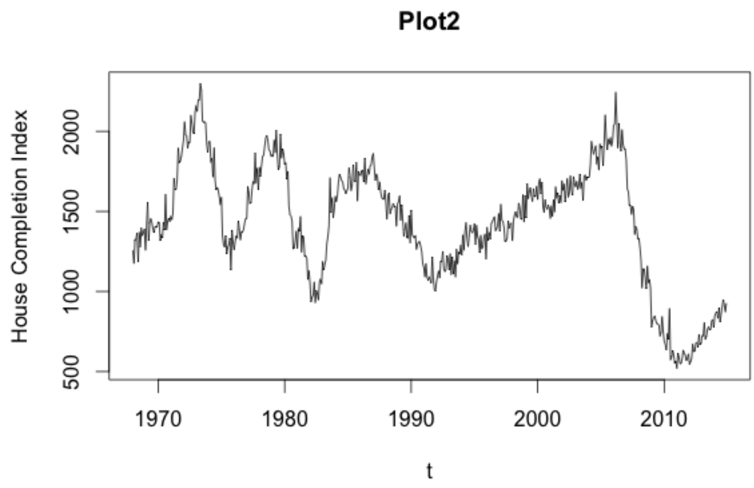

Figure 2

2 Univariate Analysis

2.1 Correlograms and Unit Root Detection

We first removed the last four observations so we can forecast them later. We then perform the autocorrelation (and partial autocorrelation) analysis to get insight into the structure of inter-temporal dependence of our series:

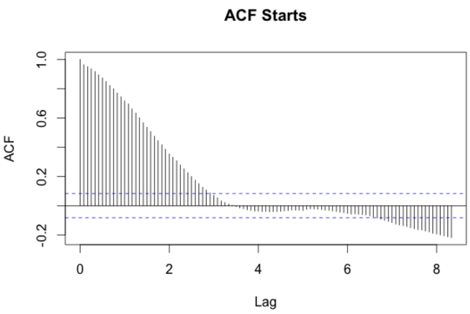

Figure 3

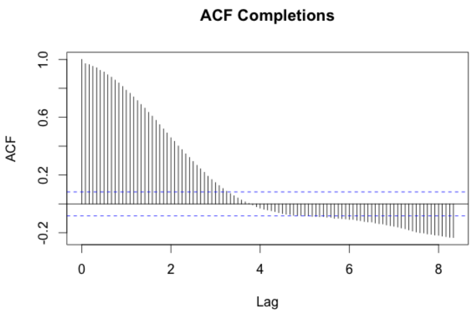

Figure 4

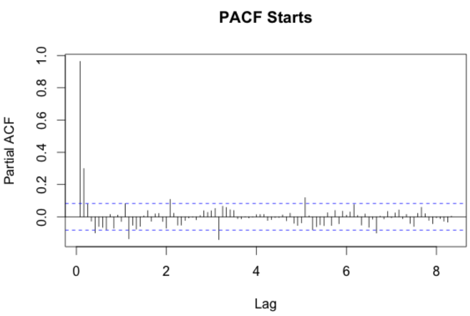

Figure 5

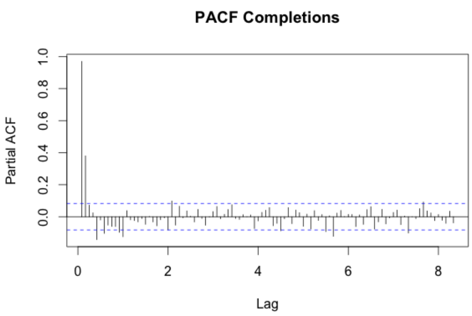

Figure 6

We can see how the autocorrelation plot is not a good evidence in favor of weak time stationarity of our series. Above we have plotted the values of the $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$ function - where the $\gamma(.)$ are the covariances of our processes at different lags. The dashed lines are the tolerance bands, here to represent the test of whether these $\rho(k)$ coefficients fade after a certain displacement k[1]. By this inspection, it might be that for our series this seems not to be the case, and so we suspect we are dealing with an integrated series[2].

[1]Indeed, under weak stationarity (under "fading after a certain displacement"), the ρs are ratios of sample averages which converge to well defined expected values. Hence, we can say that $\hat{\rho(k)} \sim^a N(0, \frac{1}{n})$, where n is the number of observation. Thus, we can derive the 95% confidence interval shown in the graphic.

[2]Series with a stochastic trend, so that shocks can have permanent effects.

We get further confirmation by repeating the autocorrelation plot for the differences of our original series[3]

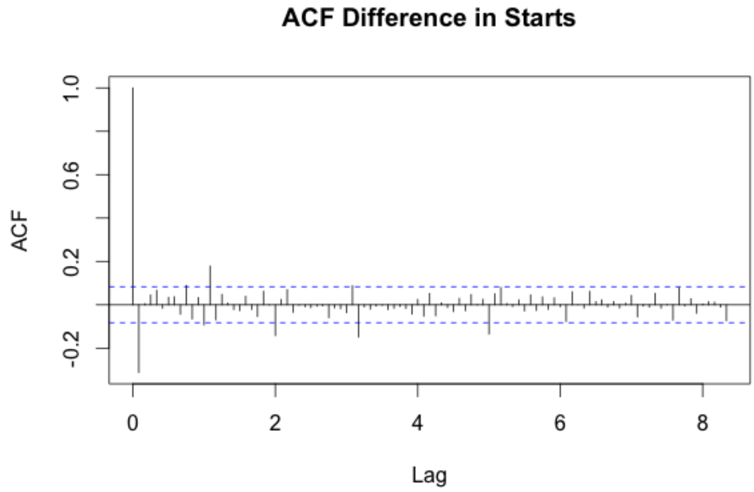

Figure 7

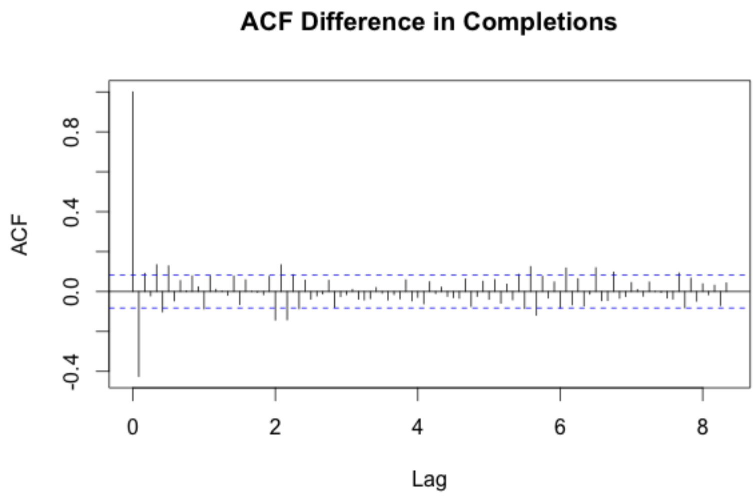

Figure 8

[3]A series $\{X_t\}_t$ is integrated of order d if $\left\{(1-L)^d X_t\right\}_t$ is weakly stationary.

This result is much nearer to the weak stationarity ideal. Looking at this plot, as at the partial autocorrelation functions in the previous group of plots, we can foresee how the first two lags may be important for the ARIMA modeling of the phenomenon under study.

We then include results for the Augmented Dickey-Fuller test with and without intercept, so as to get further confirmation of the integrated nature of the processes[4]. Below, the critical value to look at to understand the result of the test is the one related to τ, and the rule of rejection consists of rejecting whenever the test statistic is greater than the critical value. We avoided checking for the presence of a deterministic trend because, at first glance, the inspection of the initial time series plot does not suggest this; furthermore, to have both a deterministic trend and a unit root is very unrealistic for housing market related time series. The main rationale for our decision is that this would imply an ever-rising number of started/completed houses. For more in-depth treatment, we refer to Elder and Kennedy (2001).

```
COMPLETION SERIES, NO INTERCEPT, AIC SELECTION
```

[4]The Dickey Fuller test is based on the following intuition (I will report only a univariate, autoregressive of order 1 case, augmentation follows by insertion (e.g) of the differenced variable's lags):

$$\begin{aligned} X_t &= \rho X_{t-1} + \varepsilon_t \Rightarrow \\ \triangle X_t &= (\rho - 1) X_{t-1} + \varepsilon_t; \text{ equivalently} \\ \triangle X_t &= \gamma X_{t-1} + \varepsilon_t \end{aligned} \tag{1}$$

where we assume ϵ is a well behaved error term. Hence testing for $\gamma = 0$ means testing for the presence of a unit root. The only caveat is that, under the null, we cannot refer to the usual t-test critical value table. Indeed, the R package urca's "ur.df" test, which we use, refers to the critical values reported in Hamilton (1994). We repeat the test for both AIC and BIC lag selection in the model.

```
Value of test-statistic is: -0.6093

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

COMPLETION SERIES, NO INTERCEPT, BIC SELECTION

Value of test-statistic is: -0.6093

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

COMPLETION SERIES,  DRIFT, AIC SELECTION

Value of test-statistic is: -1.5435 1.2177

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78

COMPLETION SERIES,  DRIFT, BIC SELECTION

Value of test-statistic is: -1.5435 1.2177
```

```
Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78


START SERIES, NO INTERCEPT, AIC SELECTION

Value of test-statistic is: -0.8452

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

START SERIES, NO INTERCEPT, BIC SELECTION

Value of test-statistic is: -0.8452

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

START SERIES, DRIFT, AIC SELECTION

Value of test-statistic is: -2.1627 2.3702

Critical values for test statistics:
```

```
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

```
START SERIES, DRIFT, BIC SELECTION

Value of test-statistic is: -2.1627 2.3702

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

As we can see, the "$\rho = 1$" hypothesis is never rejected. In an abundance of caution, we repeat the ADF tests on the differenced series[5].

```
DIFFERENCE IN STARTS, NO INTERCEPT, AIC SELECTION

Value of test-statistic is: -21.1396

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

DIFFERENCE IN COMPLETION, NO INTERCEPT, AIC SELECTION
```

[5]Indeed, in the event that a non-integrated series is differenced, a unit root will be inserted, causing the result of the ADF test to worsen. Since there is no relevant difference between using the AIC and BIC selection criterion, we will only report AIC results

```
Value of test-statistic is: -22.1988

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

DIFFERENCE IN STARTS, DRIFT, AIC SELECTION

Value of test-statistic is: -21.1237 223.1048

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78

DIFFERENCE IN COMPLETIONS, DRIFT, AIC SELECTION

Value of test-statistic is: -22.1833 246.0493

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

The hypothesis of ρ being equal to 1 is consistently rejected. Based on the results of this further testing, we feel safe in opting for an ARIMA$(p, 1, q)$ framework for modeling time dependency in our series.

2.2 Model Selection and Residual Analysis

We start model selection by checking the Bayesian Information Criterion (BIC henceforth) [6]. Below, we report the BIC for models with various combinations of autoregressive and moving average terms in ARIMA(p,1,q):

Table 1: BIC Array, Starts Series

	q=0	q=1	q=2	q=3
p=0	6891.962	6837.764	6842.062	6843.220
p=1	6840.843	6842.784	6839.404	6844.854
p=2	6841.334	6847.611	6844.825	6851.655
p=3	6847.513	6847.074	6851.157	6857.425

Table 2: BIC Array, Completions Series

	q=0	q=1	q=2	q=3
p=0	6603.746	6501.538	6495.167	6497.715
p=1	6498.350	6498.449	6481.824	6487.571
p=2	6497.927	6504.118	6487.508	6490.254
p=3	6503.734	6496.612	6508.821	6496.142

What we can see is that for the starts table, two values capture our attention: the one related to the ARIMA(0, 1, 1) model: (6837.764), and the one related to

[6] The BIC equals to $-2 * \log\left(\hat{L}\right) + \log(n) * k$, where $\hat{L}$ is the maximized value of the likelihood function, $\log(n) * k$ the total penalty for including k parameters, given that we have n observations. Hence it is an efficiency criterion, summarizing the tradeoff between likelihood gains and parameters insertion, and a low BIC value is “good”.

the ARIMA(1, 1, 2) model: (6839.404). For the completions table, the ARIMA(1, 1, 2) seems to be better than the rest of the models according to the BIC.

We then proceed to estimating the model and then checking the residuals to decide between the two possibilities for the starts series. We then verify the viability of the option the BIC suggests to us for the completions series.

The estimated models are below summarized:

Table 3: ARIMA(0,1,1) for Starts

	MA1
	-0.3195
s.e.	0.0365
AIC	6829.11

Table 4: ARIMA(1,1,2) for Starts

	AR1	MA1	MA2
	0.8250	-1.1912	0.3609
s.e.	0.1098	0.1069	0.0431
AIC	6822.09		

Table 5: ARIMA(1,1,2) for Completions

	AR1	MA1	MA2
	0.9058	-1.4031	0.5132
s.e.	0.0400	0.0496	0.0360
AIC	6464.51		

As a first check, we plot the ACF of the residuals to check whether the time dependency has been captured by our models:

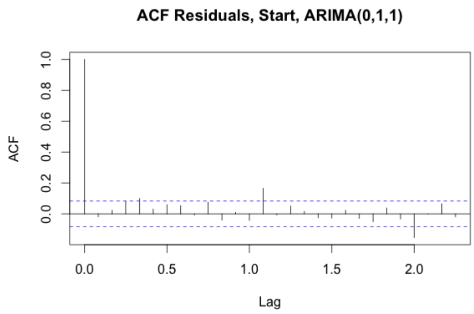

Figure 9

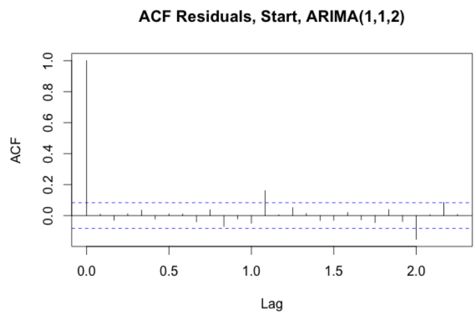

Figure 10

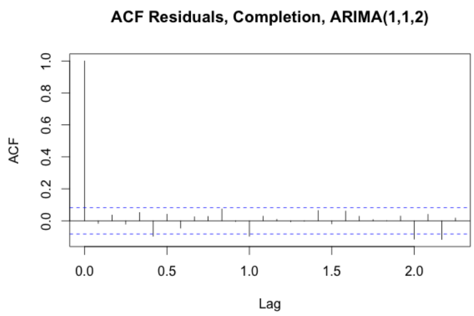

Figure 11

We can see how the ACF plots do not actually resolve our doubts: though arguably slightly better, the ACF of the ARIMA(1, 1, 2) specification on the starts series has as many significantly greater than zero coefficients as the the ACF of the ARIMA(0, 1, 1) specification. In search for a better answer, we perform the Box-Pierce test on our models' residuals[7].

[7]The Box-Pierce test is aimed at understanding whether a realization of some stochastic process hints for the presence of correlation in its generator. In particular, when dealing with time series the intuition goes as follows: as before asserted, if a time series shows no autocorrelation, $\hat{\rho(k)} \sim^a N(0, \frac{1}{n})$. Hence $n \sum_{\tau=1}^{m} \hat{\rho(m)}^2 \sim^a \chi^2(m)$, where m is the number of autocorrelations we

We present the p-values of the test in the tables below. In performing it, we choose the number of lags to include following the rule of thumb of equating it to the square root of the number of data points used in the estimation of our models. Obviously, low p-values stand for rejection with low probability of being wrong[8].

Table 6: BP Test p-values for Starts Series

	MA=0	MA=1	MA=2	MA=3
AR=0	0	0.004232941	0.008307439	0.01437321
AR=1	3.566963e-03	0.005805429	0.054998835	0.03551411
AR=2	6.611755e-03	0.004446651	0.036461131	0.02319417
AR=3	4.702945e-03	0.021865523	0.024361715	0.01902152

Table 7: BP Test p-values for Completions Series

	MA=0	MA=1	MA=2	MA=3
AR=0	0	7.445260e-08	0.0001546419	0.0008136751
AR=1	3.757337e-05	1.354121e-05	0.1237698596	0.1309889773
AR=2	2.140424e-05	8.851688e-06	0.1355936977	0.2392724060
AR=3	4.392557e-06	1.505277e-02	0.0003873276	0.1903504619

As we can see, the ARIMA(1, 1, 2) specification is never rejected at 95% critical level, in both cases. Yet, we must say that the result for for the starts case

check. We will thus reject our null hypothesis of temporal independence if the value we get after computation of the test statistic are greater then the critical value of a chi-squared distribution of m degrees of freedom. The only caveat is that, given we are performing our estimate on the fitted errors retrieved from our model estimation, we need to correct the statistic for the number of estimated parameters.

[8]We do include also the p-values for model we did not consider in the first moment, so to be fully accountable about choosing good options, and allowing for a more thorough comparative evaluation.

is not very strong. Nonetheless, there is strong evidence against the integrated MA option. On the other hand, the result for the Completion case is quite strong[9], hence we choose for keeping the two ARIMA(1, 1, 2) specifications, keeping in mind the close and natural relationship between the two series.

3 Multivariate Analysis

In this section we conduct a multivariate analysis in order to identify a VAR model for housing starts and completions. Our goal is to estimate the coefficients $\Phi_1, \ldots, \Phi_p$ of the following system:

$$\boldsymbol{y_t} = \Phi_1 \boldsymbol{y_{t-1}} + \cdots + \Phi_p \boldsymbol{y_{t-p}} + \boldsymbol{\varepsilon_t} \tag{2}$$

where $\boldsymbol{y_t} = (starts_t, completions_t)$ or $\boldsymbol{y_t} = (completions_t, starts_t)$, to try all the possible Cholesky orderings, and p is the maximum lag of the system variables to be specified later.

3.1 Cross-correlations and Identification

To begin our analysis we plot the two cross-correlation for our vector time series:

[9]Here the ARIMA(1, 1, 2) is the most parsimonious (in the Box and Jenkins (1979) sense) among the specifications with arguably serially uncorrelated residuals.

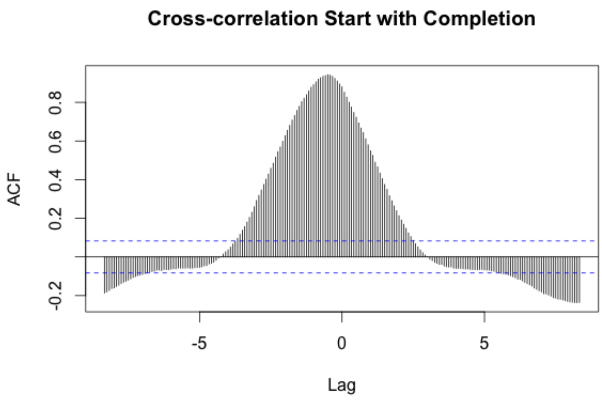

Figure 12

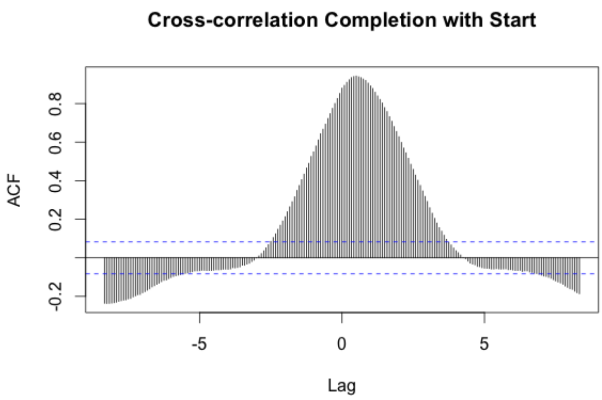

Figure 13

We can read the two plots - starting from the first - as follows: given the starts series, we can see at which lag the completions series covariates the most with it. The answer our graph gives to that question is “around five months after today”. Quite logically, the second graph is specular, showing us how much the Start series covariates with the Completion series: the peak is indeed around six months in the past.

This first confirms plain logic: starts comes before completions; then, it fosters our intuition that - in selecting models - the most plausible Cholesky ordering is from starts to completions, and not the other way round. However, the plots also suggest us that, given the general similarity in behavior, the Cholesky ordering will not produce major discrepancies in our results.

In order to proceed with our model selection we use the `VAR.select` command of the package `vars` in R. It uses many different selection criteria to choose the best VAR identification. We will inspect the ones suggested by the AIC and SC criteria. The first indicates we should estimate a VAR(14) and the latter indicates a VAR(3). We estimate both models obtaining relatively good significance of individual coefficients and high joint significance as well as a high R^2 for both of them.

Now, in order to select which of the VAR specifications is the one to stay with, we perform a residual analysis testing whether the residuals of these models are serially correlated. We use the command `serial.test`, also of the `vars` package, to perform the Portmanteau test. We apply both the asymptotic and small sample correction versions that yield us that only the VAR(14) does not reject the null hypothesis of no serial correlation. The outputs of such tests that confirm this are:

```
Portmanteau Test (asymptotic)


data:  Residuals of VAR object var3
Chi-squared = 155.6458, df = 80, p-value = 8.658e-07
------------------------------------
Portmanteau Test (asymptotic)


data:  Residuals of VAR object var14
Chi-squared = 36.1692, df = 36, p-value = 0.4607
```

Finally, to obtain the conclusion of whether we can proceed with our selected model to a forecasting exercise, we need to check for cointegration between the two series. From the previous univariate section, we know that both series are I(1) and thus it is possible that when both series are estimated simultaneously they become jointly stationary, namely cointegrated. To verify this we apply the Johansen procedure with the command `ca.jo` of package `urca` in R obtaining the output shown below.

```
######################
# Johansen-Procedure #
######################
Test type: trace statistic , with linear trend

Eigenvalues (lambda):
[1] 0.0950848 0.0135822

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  7.47  6.50  8.18 11.65
r = 0  | 62.02 15.66 17.95 23.52

Eigenvectors, normalised to first column:
(These are the cointegration relations)

             starttsf.l14 comptsf.l14
starttsf.l14    1.0000000    1.000000
comptsf.l14    -0.9929537   -1.785322
```

```
Weights W:
(This is the loading matrix)

            starttsf.l14 comptsf.l14
starttsf.d   -0.1998374  0.04319136
comptsf.d     0.2697194  0.01712912
```

The test was applied for the "trace" statistic. We have tested the null-hypothesis that there is no cointegration at all against the hypothesis that there might be one cointegration vector. Fortunately, we strongly reject the hypothesis of no cointegration while we do not reject the hypothesis of cointegration at a 5% level. Therefore we can accept that housing starts and completions cointegrate with the relations shown above and thus find forecasts for the VAR(14) with both series at their level values.

3.2 Granger Causality and Impulse-Response Functions

3.2.1 Granger Causality Test

We now investigate whether there is a particular relationship of Granger Causality - in one or the other direction - among our series. First, the intuition: what we are actually searching for is whether, say, the starts series is useful in forecasting the completions series, and vice-versa.

This means, assuming some autoregressive structure as

$$start_t = c_s + \phi_1 start_{t-1} + ... + \phi_p start_{t-p} + \rho_1 comp_{t-1} + ... + \rho_p comp_{t-p} + \varepsilon_t$$

to test whether $\rho_1, ...\rho_p$ are jointly equal to zero or not. Let us assume joint equality to zero is the null, in that case standing for "completions does not Granger

Cause starts", we can build both a test based on the F distribution, as a test based Chi-square distribution (as Hamilton (1994) shows in Chapter 11). Nevertheless, both test search to understand whether the Mean Square Error of similar regression is significantly different after the insertion of lags of the other time series as explanatory. We can hence go to the results, in our case for the Granger F-test[10]:

```
$Granger

Granger causality H0: starttsf do not Granger-cause comptsf

data:  VAR object bfitcol
F-Test = 18.2913, df1 = 14, df2 = 1034, p-value < 2.2e-16

$Granger

Granger causality H0: comptsf do not Granger-cause starttsf

data:  VAR object bfitcol
F-Test = 2.6457, df1 = 14, df2 = 1034, p-value = 0.0008471
```

It indeed seems that we have good evidences in favour of the conclusion that both completions are relevant for starts forecasting, as starts are relevant for completions forecasting. This is pretty logical, as a question similar to H0 - given a dataset that does not start with the start of a particular building market - is

[10]Note that we do not repeat the test for the alternative Cholesky ordering. This since the Granger test works on the reduced form of the model (searching whether the coefficient matrix of the lagged vectors are lower triangular), hence flipping Cholesky -though searching no more for a lower triangular, yet for an upper triangular matrix - does not change the final result.

a close statistical equivalent to the “which came first, the chicken or the egg?” dilemma.

3.2.2 Impulse Response Functions

The last analysis of the multivariate case, and perhaps the more significant in economic insight and intuition is the plot of the impulse-response functions. As our vector consists of two variables we will have two possible Cholesky orderings for the orthogonalization of the residuals. The first one will assume no contemporaneous effect of completions on starts and the second, the converse.

We start commenting the Cholesky ordering from starts to completions. First we interpret the cross-effects of shocking our variables. Shocking starts, coherently with our identification strategy, results in a contemporaneous increasing effect in completion that, after a few lags, accumulates, peaks and then fades away. A possible interpretation of the shock is a one period surge in market demand for houses, or in anticipation of the latter by construction companies, while the effect on completion follows the common sense.

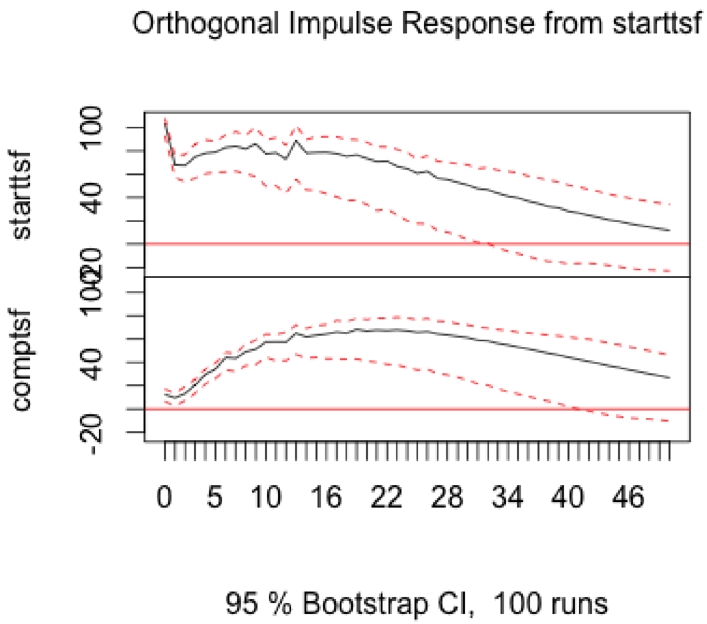

Figure 14

On the other hand, let us consider a shock in completions - which may be interpreted as a momentary raise in productivity, leading to a sudden increase in housing completion. This - given Cholesky - will have no contemporaneous effect on housing starts. Going along the lines of this interpretation, after some time (around 4 months) in which we see a slight decrease in housing starts, the series moderately increasing, perhaps after the "over-supply" has been absorbed by the market demand.

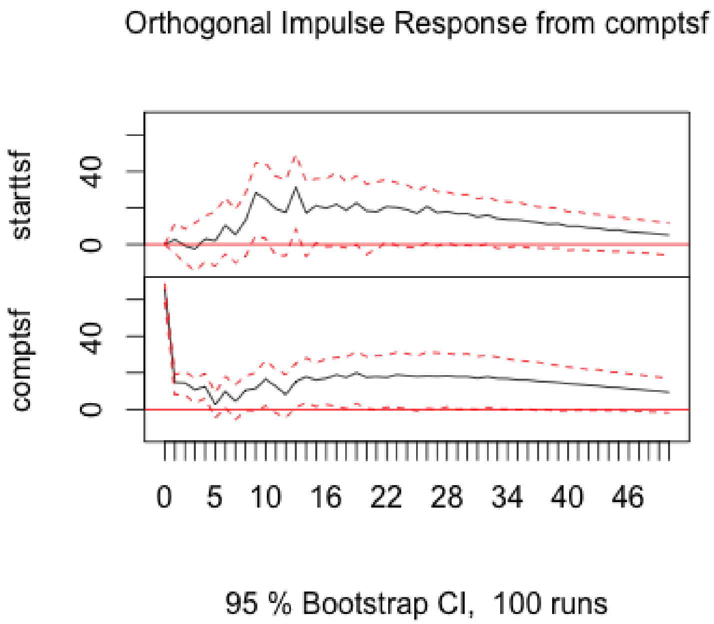

Figure 15

Going to the alternative Cholesky ordering, we can notice how - apart from the straightforward flip in contemporaneous and lagged effect of shocks - the behavior is rather similar, coherently with the close relationship among the two series, previously evidenced in the Granger Causality and cross-correlation analysis.

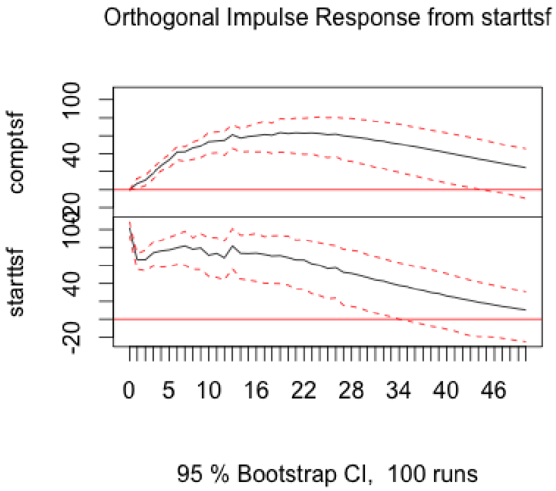

Figure 16

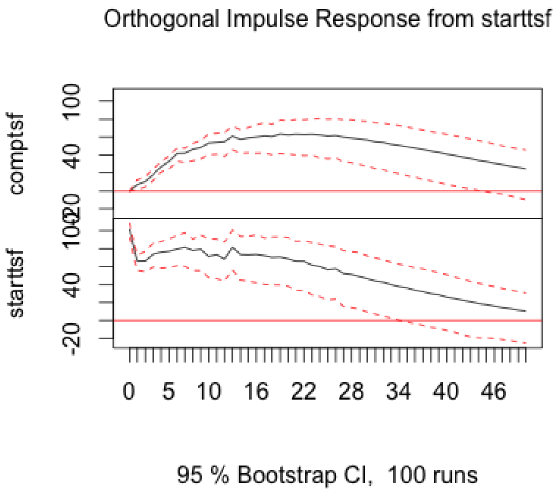

Figure 17

4 Forecasting

4.1 Univariate Forecasting

We generate forecasts for the four quarters of hold-out housing starts data using the ARIMA(1,1,2) model by using the `predict` command in R. The results are presented below, along with the realized values:

Table 8: Univariate Housing Starts Forecasts

	Forecast	Std. Error	Realized Value
Sept. 2014	997.3625	107.3158	1028
Oct. 2014	998.1446	127.0512	1092
Nov. 2014	998.7898	147.1895	1043
Dec. 2014	999.3221	167.2952	1089

We also include a graph of the last two years of the housing starts data (2013 and 2014), including our forecasts for the hold-out months. We show the forecasted data with a dashed line, and we show 95% confidence bands using dotted lines around the forecasted data:

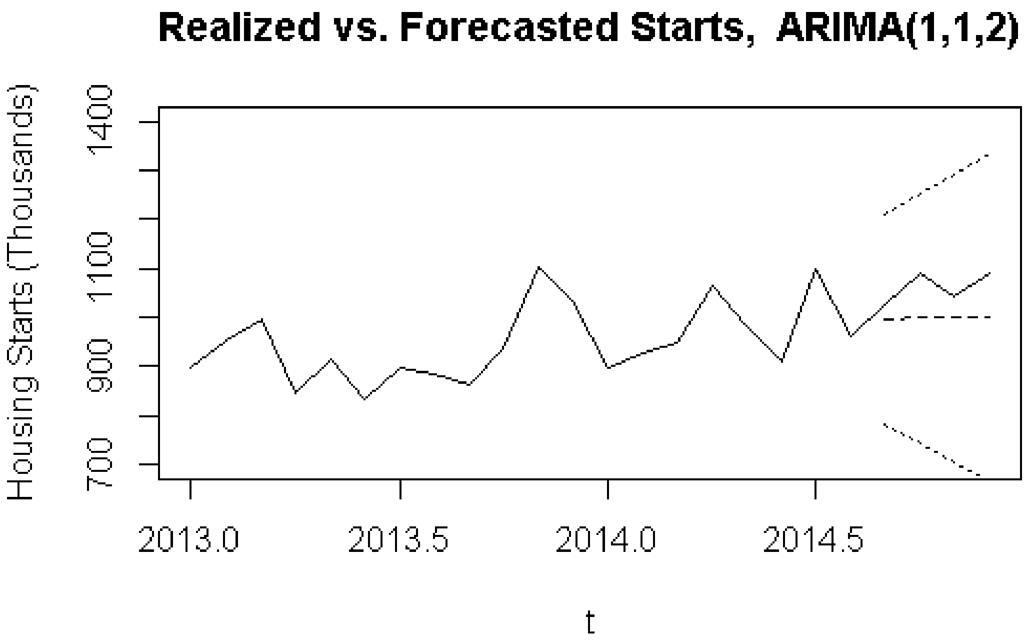

Figure 18

We also generate forecasts using the `predict` command for the four quarters of hold-out housing completions data using the ARIMA(1,1,2) model, with results presented below:

Table 9: Univariate Housing Completions Forecasts

	Forecast	Std. Error	Realized Value
Sept. 2014	885.1978	77.91895	950
Oct. 2014	889.9915	87.21017	915
Nov. 2014	894.3338	97.70534	872
Dec. 2014	898.2672	109.07076	927

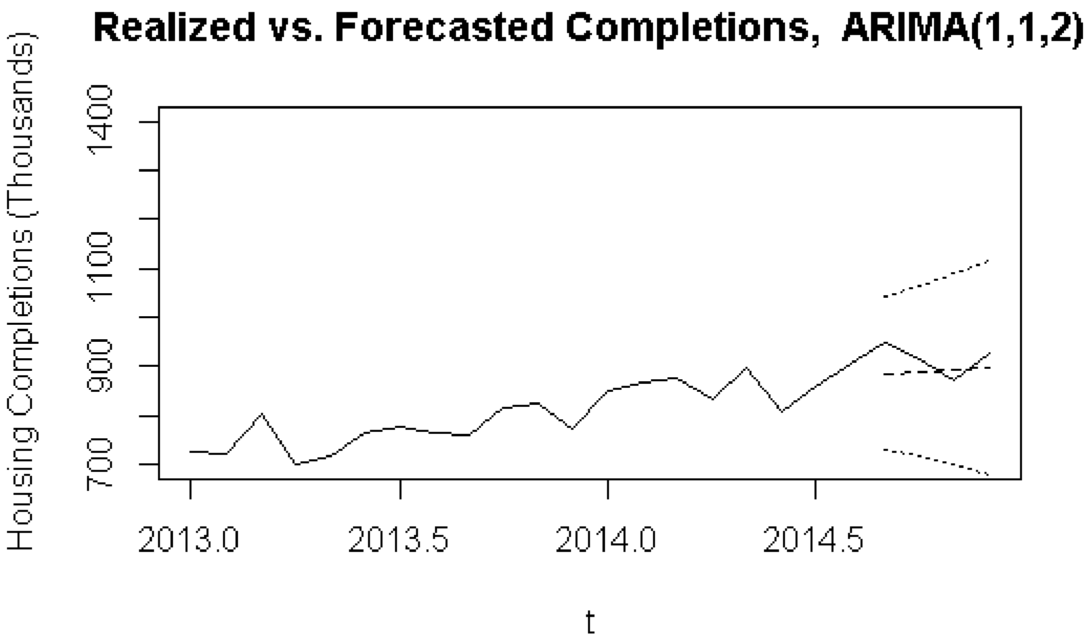

Figure 19

These forecasts perform poorly versus the realized data, and this leads us to believe that perhaps a multivariate forecast would perform better for both housing starts and housing completions.

4.2 Multivariate Forecasting

We generate forecasts for the four quarters of hold-out housing data using the VAR(14) model by using the `predict` command in R. The results are presented below, along with the realized values:

Table 10: Multivariate Housing Starts Forecasts

	Forecast	95% Confidence Interval	Realized Value
Sept. 2014	984.7744	203.6795	1028
Oct. 2014	1000.0370	243.4221	1092
Nov. 2014	1045.0702	277.0708	1043
Dec. 2014	1058.7864	313.5252	1089

Table 11: Multivariate Housing Completions Forecasts

	Forecast	95% Confidence Interval	Realized Value
Sept. 2014	909.7207	132.0820	950
Oct. 2014	902.7373	136.4896	915
Nov. 2014	925.4044	141.8086	872
Dec. 2014	935.0333	148.9400	927

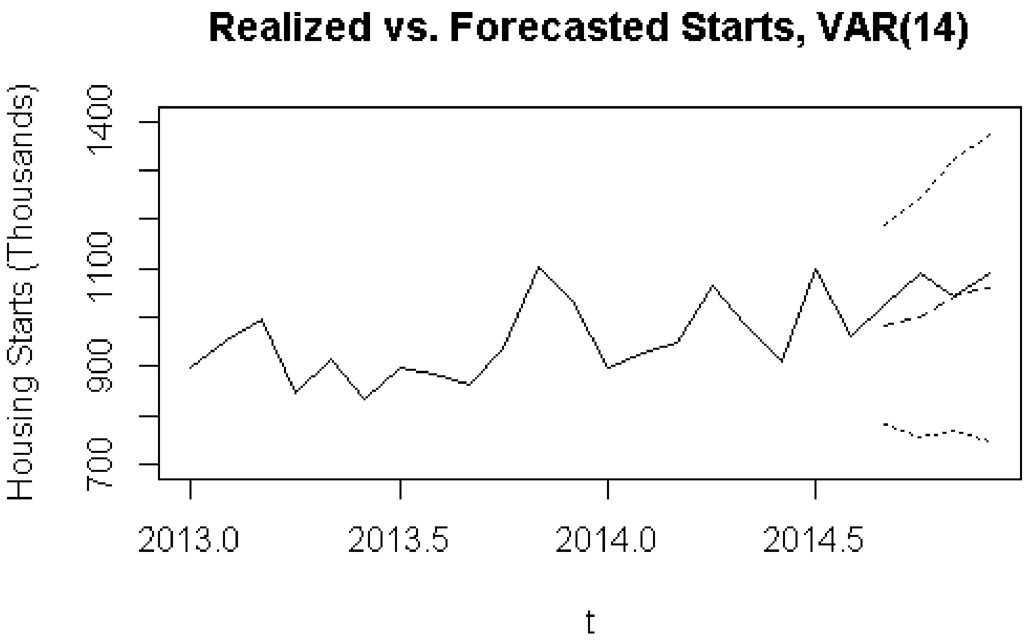
Realized vs. Forecasted Starts, VAR(14)
Housing Starts (Thousands)
700
900
1100
1400
2013.0
2013.5
2014.0
2014.5
t

Figure 20

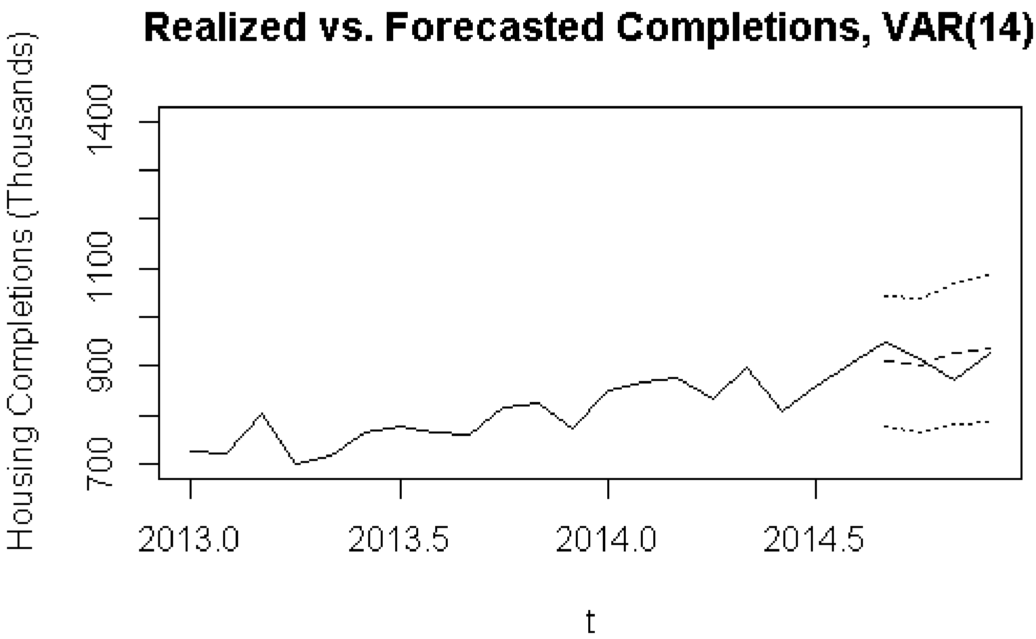

Figure 21

These predictions are much closer to the realized values than the univariate predictions. This is the case because the VAR model takes into account cross-correlations between starts and completions when estimating the coefficients of the model. Due to this, we find that the VAR(14) model is the better model for forecasting both housing starts and housing completions.

5 Conclusion

In our article we analyse the series for housing starts and completions in order to realize an “inside-sample” forecasting exercise. We select and estimate, through information criterion calculation and residual analysis, an ARIMA(1,1,2) for each univariate series. We proceed in the same way for the multivariate case reaching the VAR(14) specification.

The multivariate analysis also yields us that both series are highly and similarly correlated and also that both"Granger-cause" each other. The impulse-response function analysis also provides further insight of how shocks in starts and shocks in completions affect, separately, the dynamic system. The first one yields a relatively short term increase in the latter while the converse is more persistent but also with positive reaction.

Finally, our forecasting exercise shows us that the VAR(14) outperform both of each ARIMA(1,1,2) in predicting the levels of the series 4 periods ahead. We understand that this result is due to the fact that the multivariate specification is able to account for the close correlation between series in a larger span of lags and thus improve its accuracy.

Bibliography

Box, Jenkins and Gwilym M. Jenkins. 1979. *Time Series Analysis: Forecasting and Control*. Holden-Day.

Elder, John and Peter E. Kennedy. 2001. "Testing for unit roots: what should students be taught?" *The Journal of Economic Education* 32 (2):137–146.

FRED, Federal Reserve Bank of St. Louis from US Bureau of Census. 2014. "US. Bureau of the Census, New Privately-Owned Housing Units, Start and Completion."

Hamilton, James Douglas. 1994. *Time series analysis*. Princeton University Press.