CHAPTER 11.

FORECASTING WITH A SYSTEM OF EQUATIONS: VECTOR AUTOREGRESSION

SOLUTIONS
by
Wei Lin and Yingying Sun
(University of California, Riverside)

Exercise 1
We download the time series of the quarterly house price index for two MSAs (1975Q1 – 2012Q3): MSA 1 (San Francisco-Oakland-Fremont) and MSA 2 (San Jose-Sunnyvale-Santa Clara) in California from the Freddie Mac website:

`http://www.freddiemac.com/finance/fmhpi/`.

These two metropolitan statistical areas are about 50 miles apart and are connected by several freeways, including I-280, I-880 and CA101 etc. The economy of these two areas is mainly driven by the technology and financial sectors. In Figure 1, we plot both time series. The average quarterly growth rate is about 2% for both series with a standard deviation of approximately 2.5%. Though these two markets were not immune to the national collapse of the real estate market that started during 2006 (the largest quarterly drop was about 8 % in 2007-2008), they were among the least affected. Observe that both time series tend to move together; their correlation coefficient is 0.90, which is very high. To study the interaction between these two markets, we proceed to estimate a bivariate VAR system. We will reserve the last 20 observations (2007Q04 – 2012Q03) for the prediction exercise.

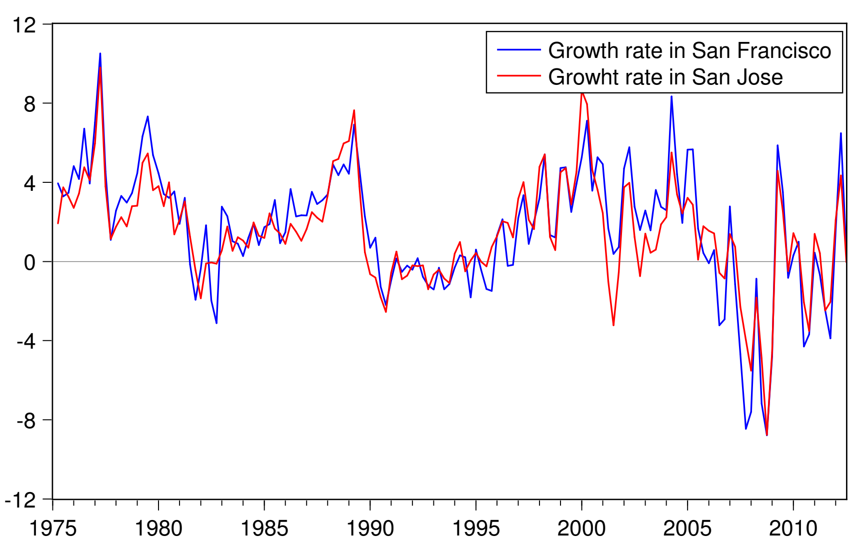


Figure 1: Growth Rate of House Prices in San Francisco and San Jose

In Table 1 we show the values of several information criteria as function of the number of lags in the VAR. The optimal lag is denoted by an asterisk. Most of the criteria select 3 lags as the optimal lag structure. It is interesting to note that the AIC and SIC agreed on the number of lags, which it is not always the case as SIC tends to choose more parsimonious models than AIC. In Table 2 we present the estimation results of a VAR(3). For both markets, the goodness of fit is very good with Adjusted R-squared of about 60%. If we look at the effect and statistical significance of the lags of GSJ on the GSF equation, and vice versa the lags of GSF on the GSJ equation, it is obvious that both markets are very much connected. However, the overall effect of one market on the other will be assessed when we compute the impulse response functions, as these functions capture the joint effect of all regressors in the system.

VAR Lag Order Selection Criteria
Endogenous variables: GSF GSJ
Exogenous variables: C
Sample: 1975Q1 2007Q3
Included observations: 120

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-470.78	NA	9.060428	7.87967	7.926128	7.898537
1	-410.191	118.1488	3.528188	6.936518	7.075892	6.993118
2	-406.095	7.851507	3.522788	6.93491	7.167201	7.029245
3	-386.78	36.37572	2.729665*	6.679668*	7.004875*	6.811736*
4	-384.243	4.693554	2.797876	6.70405	7.122174	6.873852
5	-380.614	6.593619	2.816523	6.710225	7.221265	6.917761
6	-378.824	3.191496	2.924191	6.747065	7.351021	6.992334
7	-371.433	12.93377*	2.766069	6.690553	7.387425	6.973556
8	-371.18	0.433979	2.947888	6.753006	7.542795	7.073743
9	-366.807	7.361442	2.934116	6.746787	7.629493	7.105258
10	-363.381	5.654033	2.96794	6.756342	7.731964	7.152547

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Table 1: Lag Order Selection in VAR

Vector Autoregression Estimates		
Sample (adjusted): 1976Q1 2007Q3		
Included observations: 127 after adjustments		
Standard errors in ( ) & t-statistics in [ ]		
	GSF	GSJ
GSF(-1)	0.278041	-0.06317
	-0.12489	-0.10844
	[2.22626]	[-0.58251]
GSF(-2)	-0.16813	-0.19186
	-0.13269	-0.11521
	[-1.26711]	[-1.66534]
GSF(-3)	0.576978	0.370858
	-0.12499	-0.10852
	[4.61619]	[3.41735]
GSJ(-1)	0.650732	1.002117
	-0.14912	-0.12947
	[4.36382]	[7.73998]
GSJ(-2)	-0.33706	-0.29834
	-0.1737	-0.15081
	[-1.94048]	[-1.97818]
GSJ(-3)	-0.08235	-0.01054
	-0.15007	-0.1303
	[-0.54871]	[-0.08090]
C	0.169282	0.286328
	-0.21193	-0.18401
	[0.79876]	[1.55606]
R-squared	0.624406	0.637044
Adj. R-squared	0.605626	0.618896
Sum sq. resids	327.9616	247.2342
S.E. equation	1.653183	1.435369
F-statistic	33.24896	35.10304
Log likelihood	-240.448	-222.506
Akaike AIC	3.896823	3.614262
Schwarz SC	4.053589	3.771029
Mean dependent	2.122466	1.866065
S.D. dependent	2.632491	2.325101
Determinant resid covariance (dof adj.)		2.782913
Determinant resid covariance		2.48459
Log likelihood		-418.202
Akaike information criterion		6.806334
Schwarz criterion		7.119867

Table 2: Vector Autoregression Estimates

Exercise 2
To test for Granger-causality, we choose the asymptotic chi-squared test:

$$F^* = \frac{T(SSR_0 - SSR_1)}{SSR_1} \xrightarrow{d} \chi^2_{k_0}$$

where SSR_0 is the sum of squared residuals from the restricted model (under the null hypothesis), SSR_1 the sum of those from the unrestricted model, and k_0 is the number of restrictions under the null. We report the results in Table 3. We choose a test size of 5%.

In the first panel, we are testing the effect of the San Jose market (GSJ) on the San Francisco (GSF) market. The value of the chi-squared test $F^* = 19.65$ is rather large with a p-value of almost 0%. Thus, we reject the null and conclude that the San Jose market Granger-causes San Francisco market, which means that San Jose market should have predictive ability for the San Francisco market. In the second panel, we test the effect of the San Francisco market on the San Jose market. The value of the chi-squared test $F^* = 12.04$ is still large with a p-value of 0.7% so that we reject the null hypothesis that the San Francisco market does not Granger-cause the San Jose market, and conclude that the San Francisco market has also predictive ability for the San Jose market. These tests are further indication that both markets are highly interdependent and positive or negative shocks to either market will be transmitted to the other.

Sample: 1975Q1 2007Q3			
Included observations: 127			
Dependent variable: GSF			
Excluded	Chi-sq	df	Prob.
GSJ	19.65114	3	0.0002
All	19.65114	3	0.0002
Dependent variable: GSJ			
Excluded	Chi-sq	df	Prob.
GSF	12.04215	3	0.0072
All	12.04215	3	0.0072

Table 3: Testing for Granger-Causality in VAR

Exercise 3

The correlation coefficient between the residuals from each equation in the VAR is 0.71, which implies that a shock in one market will produce a contemporaneous reaction in the other. Since we wish to track the transmission of a 'unit' impulse through the system, keeping everything else constant, we need to orthogonalize the shocks such that their correlation is zero. The standard orthogonalization is to use the Cholesky decomposition for the variance-covariance matrix of the residuals. However, on doing so, the order of the variables in the system matters, as we will see next. First, we choose the ordering (`GSF`, `GSJ`), which implies that a shock in the San Jose housing market does not have any *contemporaneous* effect in the San Francisco market (the effect may be coming in future periods) while a shock in the San Francisco market contemporaneously impacts both the San Jose and the San Francisco markets. In Figure 2, we show four impulse response functions for the order (`GSF`, `GSJ`). In each graph, there are 95% confidence bands (red dotted lines) around the values of the response. If zero is within the bands, we will consider that the response is statistically zero. We observe that a shock to the San Francisco market lives for about 9 or 10 quarters in San Francisco and about 6 quarters in San Jose. Similarly, a shock to the San Jose market lives for about 3 quarters in San Jose but it has a very small or no effect in the San Francisco market. Let us reverse the order of the impulses, i.e. (`GSJ`, `GSF`) and plot the resulting impulse response functions in Figure 3. Observe that the cross-responses (response of GSF to GSJ and response of GSJ to GSF) are now very different; San Francisco seems to respond to shocks in San Jose but San Jose does not respond to San Francisco shocks. Thus, we have somehow contradictory information and either we need to bring more information gathered from other sources to understand better the transmission of shocks or we could implement other methods such that the ordering of the variables does not matter. A 'generalized' impulse response is one such method. In Figure 4, we report the functions. Now, we observe that both markets feed each other and any shock in San Francisco is transmitted to San Jose and vice versa. It seems that a San Jose shock affects the San Francisco market for longer periods, about 8 or 9 quarters, while a San Francisco shock affects the San Jose market for about 2 quarters. This information is more in agreement with our findings regarding the estimation results of the VAR and the tests of Granger-causality, which all point towards a feedback mechanism between both markets.

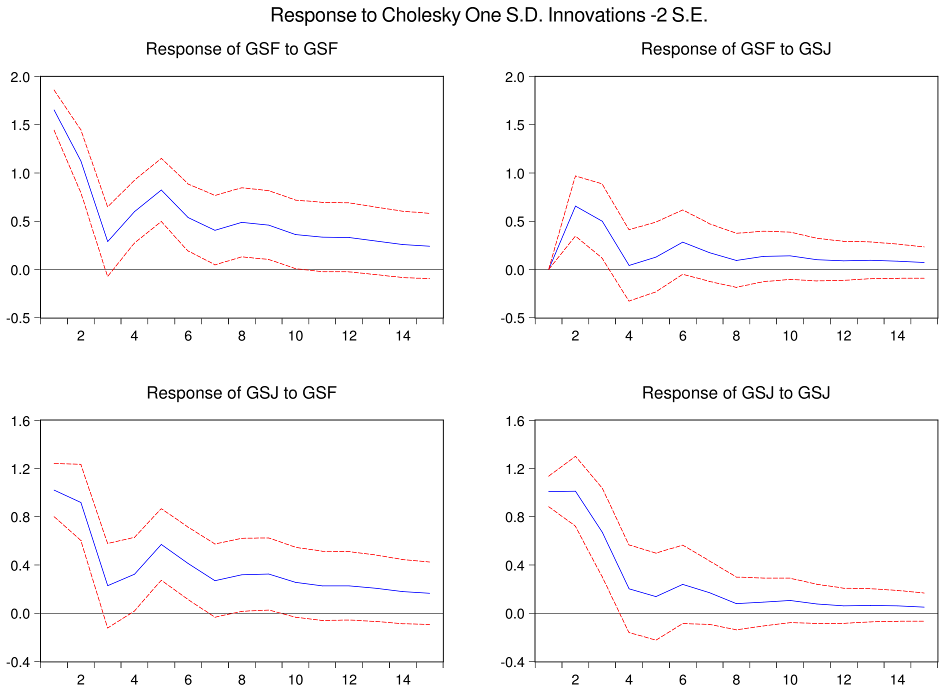


Figure 2: Impulse Response for San Francisco and San Jose Markets; ordering (GSF, GSJ)

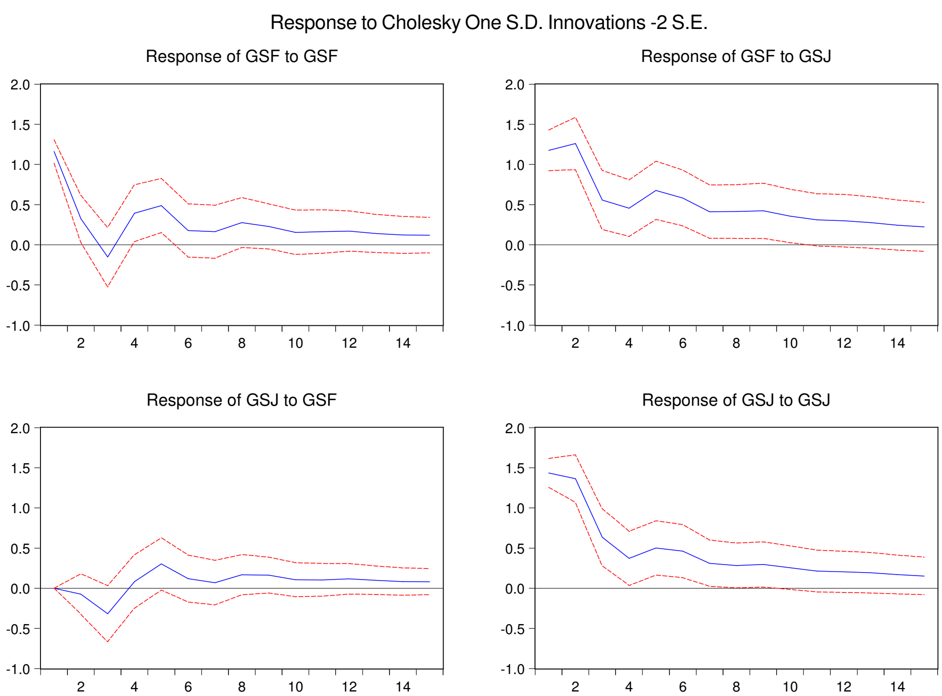


Figure 3: Impulse Response for San Francisco and San Jose Markets; ordering (GSJ, GSF)

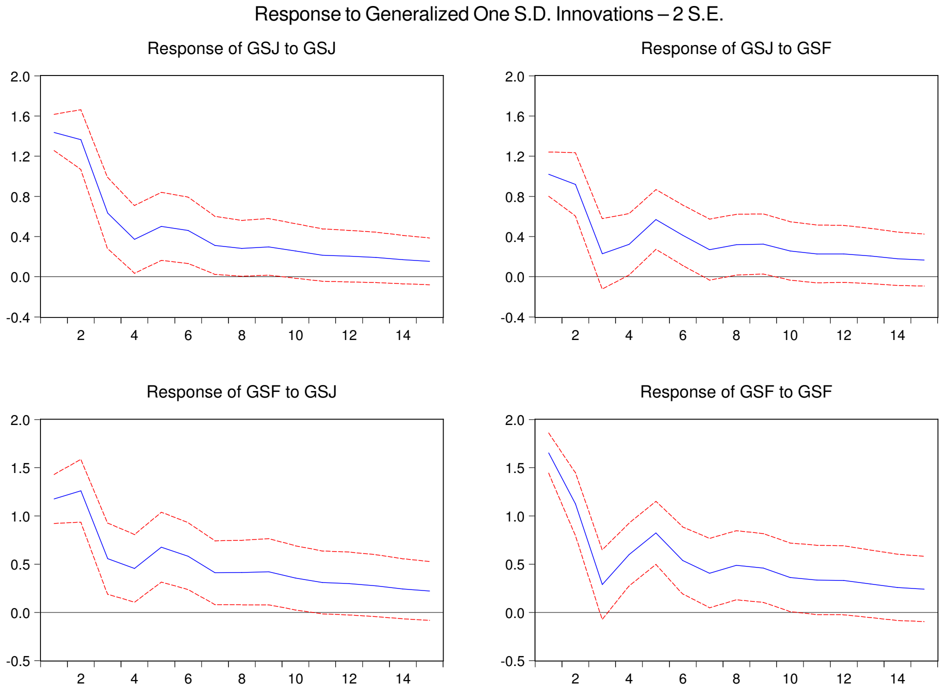
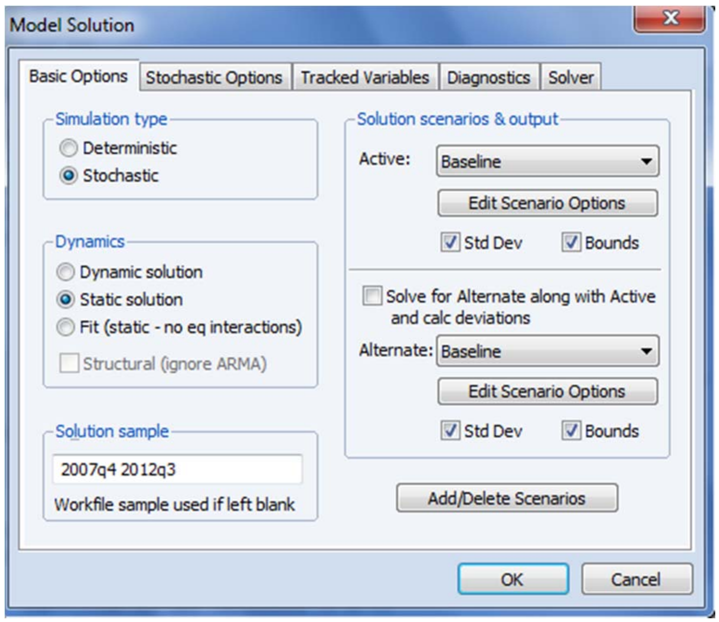


Figure 4: Generalized Impulse Response for San Francisco and San Jose Markets

Exercise 4
Based on the VAR(3) system that we have estimated in Exercise 1, we construct the 1-step-ahead forecast for the housing price growth rate in San Francisco and San Jose. The prediction sample dates from 2007Q4 to 2012Q3. We use a fixed scheme for our forecasting environment.

In Tables 4a and 4b, we report the 1-step-ahead forecast values and their corresponding standard errors for San Francisco and San Jose housing price growth, respectively. To generate the one-step forecasts and corresponding standard errors in EViews, solve the model by choosing the 'static solution' and the 'stochastic' choice under simulation type. In Figure 5, we provide the EViews screens.

Observe that the standard error of the forecast is the the standard deviation of the error term in each equation: about 1.65% for the San Francisco equation and 1.44% for the San Jose equation. In Figures 6a and 6b, we plot the forecasts with their 95% confidence bands. Observe that with only a few exceptions, the forecasts follow very closely the actual values, and these tend to fall within the 95% confidence bands.



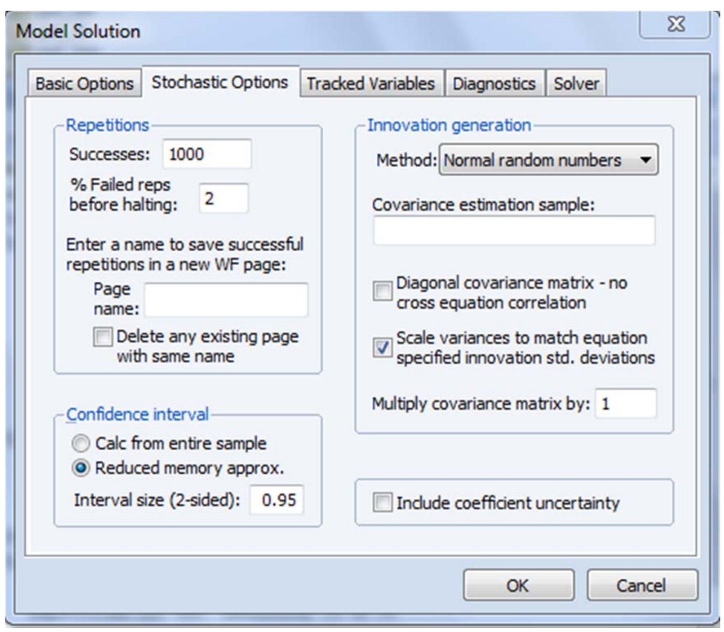


Figure 5: EViews Screens. Computation of One-step Forecasts

Date	Actual	Forecast	S.E.
2007Q4	-8.46469	-1.23783	1.625915
2008Q1	-7.60675	-3.95668	1.657396
2008Q2	-0.87833	-5.30805	1.576388
2008Q3	-7.18036	-2.66151	1.66342
2008Q4	-8.783	-8.19494	1.640984
2009Q1	-4.80468	-5.45296	1.664244
2009Q2	5.868813	-3.39386	1.629355
2009Q3	3.505712	2.811151	1.69506
2009Q4	-0.82108	-2.11604	1.627065
2010Q1	0.302694	1.174337	1.632801
2010Q2	0.990632	3.243704	1.657908
2010Q3	-4.29527	-0.10149	1.620383
2010Q4	-3.68378	-2.69743	1.676754
2011Q1	0.440821	-1.16471	1.638771
2011Q2	-0.68196	0.662075	1.638768
2011Q3	-2.45633	-2.13354	1.668716
2011Q4	-3.88592	-2.05067	1.650291
2012Q1	1.557425	-1.47599	1.626673
2012Q2	6.484462	2.089434	1.693259
2012Q3	0.221087	1.758209	1.686017

(a) San Francisco

Date	Actual	Forecast	S.E.
2007Q4	-3.96139	-0.73188	1.44291
2008Q1	-5.52328	-2.01518	1.404805
2008Q2	-1.83105	-3.66045	1.451764
2008Q3	-4.88968	-1.44767	1.454878
2008Q4	-8.76027	-6.20523	1.452298
2009Q1	-4.58544	-5.4235	1.467071
2009Q2	4.581605	-2.31416	1.404348
2009Q3	2.524675	3.585587	1.449422
2009Q4	-0.49201	-1.67692	1.449449
2010Q1	1.432683	0.519013	1.492353
2010Q2	0.730517	3.223456	1.440413
2010Q3	-2.06481	-0.00238	1.395111
2010Q4	-3.51032	-1.80993	1.440491
2011Q1	1.396968	-1.18229	1.4689
2011Q2	0.419369	1.841389	1.439334
2011Q3	-2.47261	-1.11306	1.459701
2011Q4	-2.0472	-1.92875	1.454267
2012Q1	2.095104	-0.58566	1.433906
2012Q2	4.351995	2.762011	1.474768
2012Q3	-0.04695	1.8781	1.478754

(b) San Jose

Table 4: **One-step-ahead Forecast** of Housing Price Growth in San Francisco and San Jose

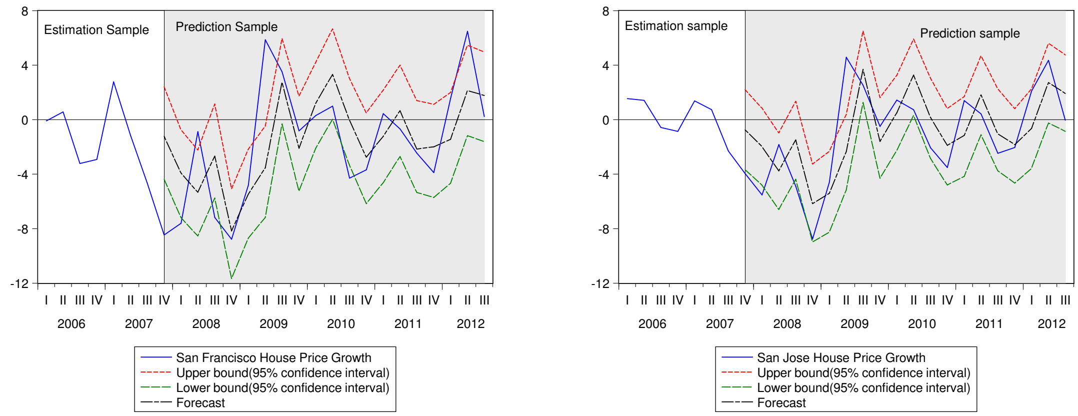
We will evaluate these 1-step-ahead forecasts by comparing them with the 1-step-ahead forecasts obtained from the best univariate model. We implement the unconditional predictability test, explained in Chapter 9. Based on the BIC, the best univariate models for both San Francisco and San Jose housing price growth rates are ARMA(3,2). We choose a quadratic loss function. For each period, we compute the loss associated with the forecast error from the VAR(3) and the ARMA(3,2) model, and compute the loss differential. Our goal is to test whether this loss differential is statistically significant. We proceed by running the following OLS regression:

$$L^{\text{VAR}}_{t+j,1} - L^{\text{ARMA}}_{t+j,1} \equiv \Delta L_{t+j,1} = \beta_0 + \varepsilon_{t+j}, \quad \text{for } j = 0, 1, 2, \cdots, T-t-1.$$

Then, the t-statistic is

$$t = \frac{\hat{\beta}_0}{\hat{\sigma}_{\hat{\beta}_0}} \xrightarrow{A} N(0,1)$$

The null hypothesis is that the constant term is zero, implying that, on average, both forecasts are equally accurate over the entire prediction sample. In Tables 5 and 6, we show the regression output for San Francisco and San Jose loss differential series. The t-statistics are 1.598 and 1.294 for San Francisco and San Jose respectively. Observe that, because the t-statistics are positive, the loss associated with the VAR(3) model is larger than the loss associated with the univariate model ARMA(3,2). However, the respective two-sided (one-sided) p-values are 0.127 (0.0635) and 0.211 (0.1056), which are larger than the customary 5% significance level. Thus, we fail to reject the null hypothesis and conclude that the forecasts from the bivariate VAR model for both San Francisco and San Jose markets are statistically equivalent to those from the univariate ARMA models.



(a) San Francisco   (b) San Jose

Figure 6: **One-step-ahead Forecasts** of House Price Growth in San Francisco and San Jose

Dependent Variable: SFVLOSS-SFARMALOSS				
Method: Least Squares				
Sample: 2007Q4 2012Q3				
Included observations: 20				
HAC standard errors & covariance				
(Bartlett kernel, Newey-West fixed bandwidth = 3.0000)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.092604	4.43964	1.597563	0.1266
R-squared	0.000000	Mean dependent var		7.092604
Adjusted R-squared	0.000000	S.D. dependent var		23.29075
S.E. of regression	23.29075	Akaike info criterion		9.182696
Sum squared resid	10306.72	Schwarz criterion		9.232483
Log likelihood	-90.827	Hannan-Quinn criter.		9.192415
Durbin-Watson stat	2.028234			

Table 5: Unconditional Predictive Ability Test for San Francisco

Dependent Variable: SJVLOSS-SJARMALOSS				
Method: Least Squares				
Sample: 2007Q4 2012Q3				
Included observations: 20				
HAC standard errors & covariance				
(Bartlett kernel, Newey-West fixed bandwidth = 3.0000)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.658108	3.60034	1.293797	0.2112
R-squared	0.00000	Mean dependent var		4.658108
Adjusted R-squared	0.00000	S.D. dependent var		18.13434
S.E. of regression	18.13434	Akaike info criterion		8.682198
Sum squared resid	6248.229	Schwarz criterion		8.731985
Log likelihood	-85.822	Hannan-Quinn criter.		8.691917
Durbin-Watson stat	1.990828			

Table 6: Unconditional Predictive Ability Test for San Jose

Exercise 5

We construct multi-step-ahead forecasts with 95% confidence intervals for both San Francisco and San Jose house markets, which are reported in Tables 7a and 7b . The prediction period includes the great recession of 2008 and the deflation of the real estate bubble. By the nature of the multi-step forecast, we do not expect to track dramatic changes in the growth rates as those observed in 2008 and 2009. Observe in Figures 7a and 7b that the multi-step forecast is much smoother than the actual values. In addition, the 95% confidence bands are very wide, which means that there is much uncertainty in both markets. Compare the multi-step forecasts with the one-step forecasts in Exercise 4. In the one-step forecasts, the bands are tighter (uncertainty is lower, why?) and, since we are updating the information set one observation at the time, the forecast is better able to track the actual values of the the series.

To compute the multi-step forecasts and their corresponding standard errors follow the same EViews screens as in Exercise 4 but in the 'Basic Options' screen, choose the 'Dynamic solution'.

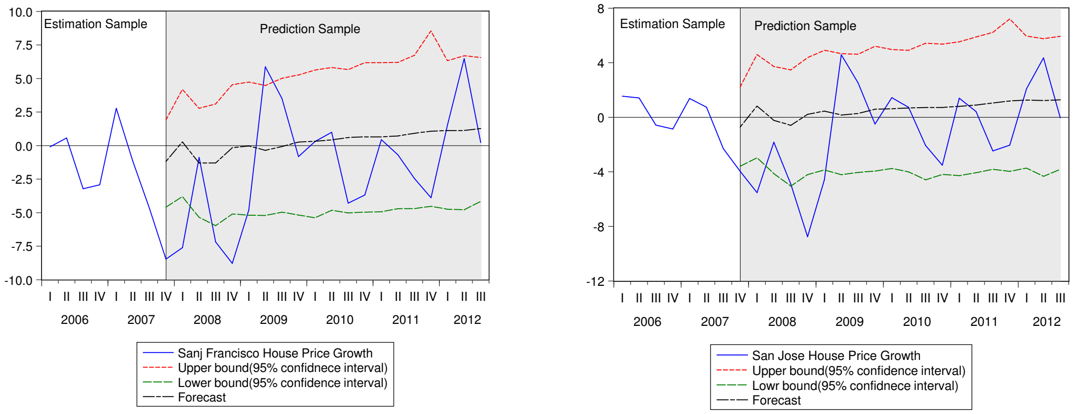
Date	Actual	Forecast	S.E.
2007Q4	-8.46469	-1.18501	1.681299
2008Q1	-7.60675	0.269525	2.055752
2008Q2	-0.87833	-1.29735	2.11206
2008Q3	-7.18036	-1.29214	2.244051
2008Q4	-8.783	-0.16133	2.457837
2009Q1	-4.80468	-0.03473	2.453821
2009Q2	5.868813	-0.35623	2.536526
2009Q3	3.505712	-0.0707	2.621605
2009Q4	-0.82108	0.262608	2.683101
2010Q1	0.302694	0.320318	2.656087
2010Q2	0.990632	0.416436	2.615725
2010Q3	-4.29527	0.598151	2.791921
2010Q4	-3.68378	0.643236	2.759838
2011Q1	0.440821	0.643245	2.745325
2011Q2	-0.68196	0.719891	2.817842
2011Q3	-2.45633	0.911988	2.785343
2011Q4	-3.88592	1.068153	2.824996
2012Q1	1.557425	1.117575	2.879059
2012Q2	6.484462	1.121851	2.853559
2012Q3	0.221087	1.269685	2.735504

(a) San Francisco

Date	Actual	Forecast	S.E.
2007Q4	-3.96139	-0.72076	1.450684
2008Q1	-5.52328	0.818759	1.973899
2008Q2	-1.83105	-0.22496	2.028436
2008Q3	-4.88968	-0.5944	2.13012
2008Q4	-8.76027	0.21293	2.247763
2009Q1	-4.58544	0.453822	2.255776
2009Q2	4.581605	0.160999	2.282395
2009Q3	2.524675	0.269248	2.317242
2009Q4	-0.49201	0.590378	2.352119
2010Q1	1.432683	0.61823	2.337871
2010Q2	0.730517	0.675278	2.31519
2010Q3	-2.06481	0.721586	2.460108
2010Q4	-3.51032	0.719536	2.436201
2011Q1	1.396968	0.795647	2.453503
2011Q2	0.419369	0.897814	2.422026
2011Q3	-2.47261	1.043506	2.451391
2011Q4	-2.0472	1.194167	2.483194
2012Q1	2.095104	1.260499	2.45427
2012Q2	4.351995	1.219532	2.469579
2012Q3	-0.04695	1.283982	2.44807

(b) San Jose

Table 7: **Multi-step-ahead Forecast** of Housing Price Growth in San Francisco and San Jose



(a) San Francisco

(b) San Jose

Figure 7: **Multi-step-ahead Forecasts** of House Price Growth in San Francisco and San Jose

Exercise 6

We choose MSA 3 as Albany-Schenectady-Troy in New York, which is in the east coast and geographically very distant from MSA 1 (San Francisco-Oakland-Fremont) and MSA 2 (San Jose-Sunnyvale-Santa Clara). In Figure 8, we plot the quarterly house price growth rates of the three MSAs. We find that Albany house price growth does not have strong co-movements with the house price growth of the other two MSAs in California; the correlation coefficient with San Francisco is 0.26 and with San Jose is 0.16. Nevertheless, we observe that the great recession of 2008, being a national crisis, affected all markets in similar fashion. We will reserve the last 20 observations (2007Q4 – 2012Q3) for the prediction exercise.

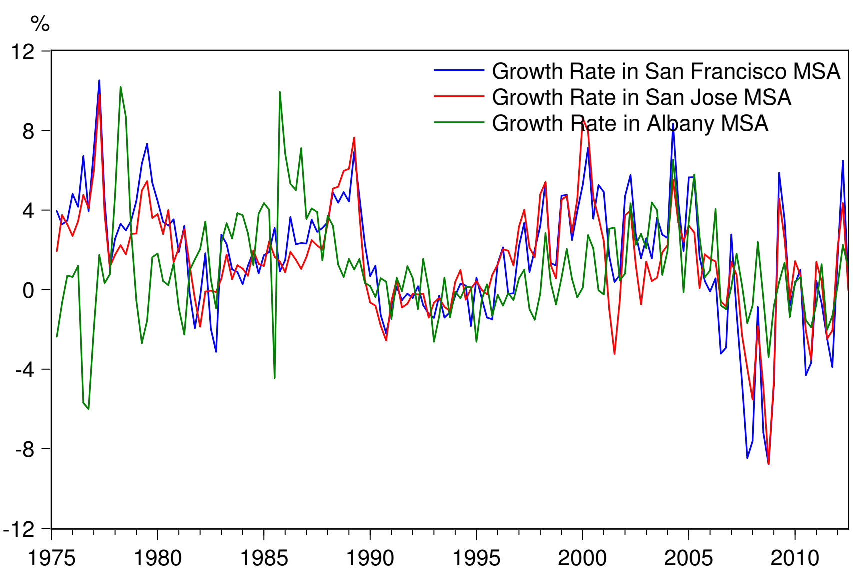


Figure 8: House Price Growth in San Francisco, San Jose, and Albany MSAs

We proceed to estimate a bivariate VAR for the price growth of MSA 1 (San Francisco, denoted as `GSF`) and MSA 3 (Albany, denoted as `GAL`). In Table 8, we report the optimal choices of the lag order based on different criteria. Most of the information criteria point to towards an optimal length of 3 lags.

VAR Lag Order Selection Criteria
Endogenous variables: GSF GAL
Exogenous variables: C
Sample: 1975Q1 2007Q3
Included observations: 122

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-567.6740	NA	38.98454	9.338919	9.384886	9.357589
1	-514.8728	103.0057	17.51686	8.538898	8.676801	8.594910
2	-512.5212	4.510485	17.99796	8.565921	8.795758	8.659274
3	-492.1243	38.45312*	13.75796*	8.297119*	8.618892*	8.427814*
4	-488.8890	5.993239	13.93549	8.309656	8.723364	8.477691
5	-486.4796	4.384316	14.31001	8.335731	8.841374	8.541108
6	-483.8325	4.730082	14.64033	8.357910	8.955488	8.600627
7	-480.4335	5.962099	14.79826	8.367763	9.057276	8.647822
8	-479.0201	2.432993	15.45699	8.410165	9.191614	8.727565

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Table 8: Lag Order Selection in VAR

In Table 9, we show the estimation results of the VAR(3) model. The San Francisco regression offers a stronger fit with a high adjusted R-squared of 59% while the adjusted R-squared for the Albany regression is only about 30%. Overall, the most significant terms in each regression are its own lagged regressors; the cross-lagged terms are less significant, statistically speaking, indicating that the relation between both markets is weaker than the relation we found between the San Francisco and San Jose markets. The Granger-causality tests seem to provide similar message. As in Exercise 2, we use the asymptotic chi-squared test statistic,

$$F^* = \frac{T(SSR_0 - SSR_1)}{SSR_1} \xrightarrow{d} \chi^2_{k_0}$$

where SSR_0 and SSR_1 are the sum of squared residuals for the restricted model under the null and unrestricted model, and k_0 is the number of restrictions under the null. In Table 10, we report the test statistics and their corresponding p-values. For the `GSF` equation, the p-value is about 5%, thus we will be able to reject the null hypothesis that the Albany market does not help to predict the San Francisco only marginally. For the `GAL` equation, the p-value is about 6% so that, at the 5% significance level, we fail to reject the null hypothesis that the San Francisco market does not Granger-cause the Albany market. Taking both results together, we conclude that both markets are not strongly linked, and so there is no much predictive power from San Francisco to Albany and vice versa.

Vector Autoregression Estimates Sample (adjusted): 1976Q1 2007Q3 Included observations: 127 after adjustments Standard errors in ( ) & t-statistics in [ ]		
	GSF	GAL
GSF(-1)	0.692497	0.161959
	(-0.08284)	(-0.10433)
	[8.35977]	[1.55233]
GSF(-2)	-0.263536	-0.307784
	(-0.10115)	(-0.12739)
	[-2.60550]	[-2.41602]
GSF(-3)	0.403678	0.227374
	(-0.08481)	(-0.10681)
	[4.75998]	[2.12869]
GAL(-1)	-0.045699	0.550064
	(-0.07034)	(-0.0886)
	[-0.64967]	[6.20868]
GAL(-2)	-0.083842	-0.133635
	(-0.07942)	(-0.10003)
	[-1.05569]	[-1.33597]
GAL(-3)	0.18768	0.175369
	(-0.06956)	(-0.08762)
	[2.69791]	[2.00154]
C	0.221307	0.367012
	(-0.23401)	(-0.29474)
	[0.94571]	[1.24522]
R-squared	0.589644	0.335547
Adj. R-squared	0.569126	0.302325
Sum sq. resids	358.3152	568.408
S.E. equation	1.727993	2.176404
F-statistic	28.73814	10.09996
Log likelihood	-246.069	-275.3696
Akaike AIC	3.985339	4.446766
Schwarz SC	4.142106	4.603532
Mean dependent	2.122466	1.329584
S.D. dependent	2.632491	2.60563
Determinant resid covariance (dof adj.)		14.13765
Determinant resid covariance		12.62212
Log likelihood		-521.4115
Akaike information criterion		8.431677
Schwarz criterion		8.74521

Table 9: Vector Autoregression Estimates

VAR Granger Causality/Block Exogeneity Wald Tests Sample: 1975Q1 2007Q3 Included observations: 127			
Dependent variable: GSF			
Excluded	Chi-sq	df	Prob.
GAL	7.821011	3	0.0499
All	7.821011	3	0.0499
Dependent variable: GAL			
Excluded	Chi-sq	df	Prob.
GSF	7.429494	3	0.0594
All	7.429494	3	0.0594

Table 10: Testing for Granger-Causality in VAR

Exercise 7
We calculate the impulse response functions for the bivariate VAR(3) in Exercise 6. First, we follow the ordering (`GSF`, `GAL`), which implies that a shock in the Albany housing market does not have any *contemporaneous* effect in the San Francisco market (the effect may be coming in future periods) while a shock in the San Francisco market contemporaneously impacts both the Albany and the San Francisco markets.

In Figure 9, we show the four impulse response functions for the order (`GSF`, `GAL`). In each graph, there are 95% confidence bands (red dotted lines) around the values of the response in blue solid lines. The top-left and bottom-right graphs show the responses of San Francisco and Albany house markets to their *own* shocks. A shock to the San Francisco market lives for about 8 to 9 quarters within the market; a shock to the Albany market is much short-lived, and it is absorbed within 2 to 3 quarters. The top-right (bottom-left) graph shows the impulse responses of San Francisco (Albany) house market to the shock in the Albany (San Francisco) house market. Both graphs show that San Francisco does not respond to Albany and neither Albany responds to shocks in the San Francisco market. Observe that the confidence bands include zero.

Now we reverse the ordering. In Figure 10, we plot the four impulse response functions for the *reversed* order (`GAL`, `GSF`). The four impulse response functions in Figure 10 are all very similar to their corresponding ones in Figure 9. Therefore, the ordering of the series does not affect the results of impulse response functions for the house price growth rates in San Francisco and Albany, thus both markets are rather independent from each other. This information is in agreement with our previous results, i.e. estimates from the VAR(3) and Granger-causality tests.

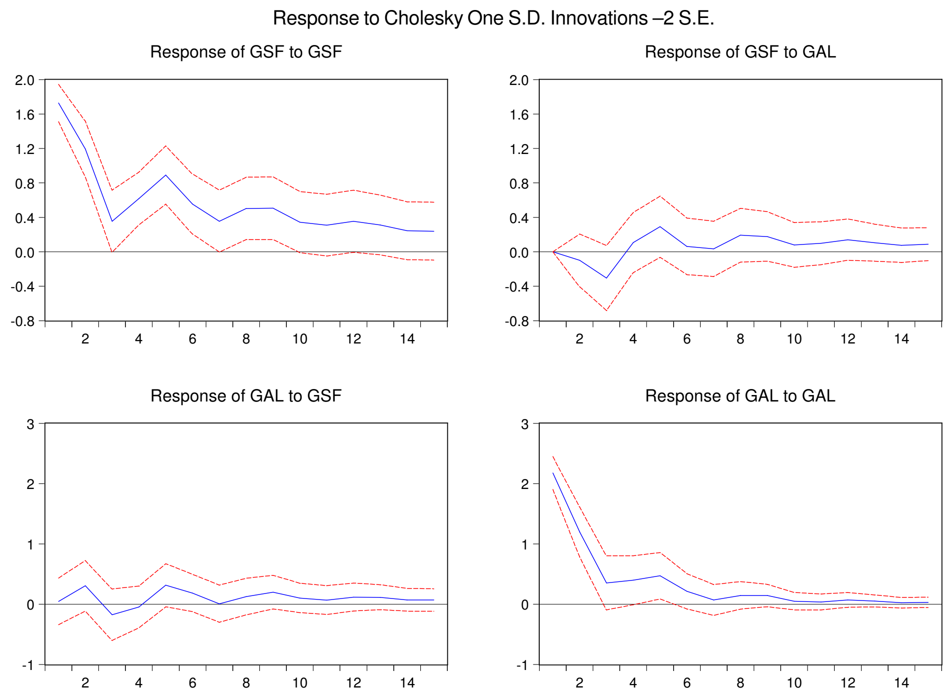


Figure 9: Impulse Response Functions for San Francisco and Albany House Markets (`GSF`, `GAL`)

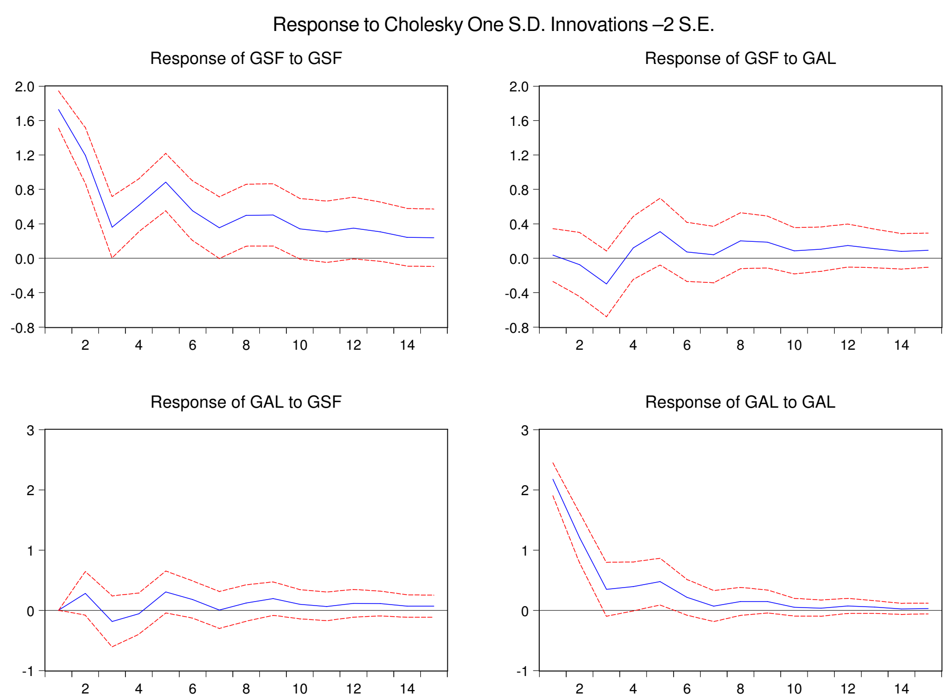


Figure 10: Impulse Response Functions for Albany and San Francisco House Markets (`GAL`, `GSF`)

Exercise 8

Following the results from Exercise 6 and 7, we construct the 1-step-ahead forecast for the quarterly house price growth rate in San Francisco and Albany based on the VAR(3) model that we estimated previously. The prediction sample dates from 2007Q4 to 2012Q3.

In Tables 11a and 11b, we report the 1-step-ahead forecast values and their corresponding standard errors for San Francisco and Albany growth rates, respectively. In Figures 11a and Figure 11b we plot the forecasts with 95% confidence bands. Observe that the values of the forecast for the San Francisco area are slightly different from those provided in Exercise 4. This is expected because we have already seen that the San Jose market affects the San Francisco market. In this new VAR, we have already concluded that Albany is not influential in San Francisco, so the model for the the San Francisco market is slightly different from the model we have estimated in Exercise 4. Nevertheless, looking at the plot in Figure 11a, we see that the overall performance of the one-step-ahead forecast is very similar to that in Exercise 4. The forecast for the Albany market is very good as the predicted values follow very closely the realized values.

Date	Actual	Forecast	S.E.
2007Q4	-8.464693	-1.787516	1.755675
2008Q1	-7.606754	-4.507933	1.727364
2008Q2	-0.878331	-4.553730	1.772140
2008Q3	-7.180357	-2.218489	1.744155
2008Q4	-8.782996	-7.893525	1.743471
2009Q1	-4.804676	-3.664810	1.750027
2009Q2	5.868813	-3.500153	1.775015
2009Q3	3.505712	1.455590	1.752002
2009Q4	-0.821082	-1.015413	1.748780
2010Q1	0.302694	1.109341	1.767939
2010Q2	0.990632	2.528051	1.716763
2010Q3	-4.295270	0.166776	1.715989
2010Q4	-3.683776	-2.881894	1.719407
2011Q1	0.440821	-0.415604	1.728513
2011Q2	-0.681955	-0.313240	1.708284
2011Q3	-2.456326	-2.229018	1.736950
2011Q4	-3.885916	-1.311470	1.777117
2012Q1	1.557425	-1.689183	1.743101
2012Q2	6.484462	1.015987	1.749516
2012Q3	0.221087	2.438120	1.725830

(a) San Francisco

Date	Actual	Forecast	S.E.
2007Q4	-1.683196	0.566084	2.091851
2008Q1	-0.795481	-0.478162	2.212506
2008Q2	2.388146	0.407276	2.205246
2008Q3	-0.436979	1.681833	2.121946
2008Q4	-3.380073	-2.951479	2.269350
2009Q1	-0.831809	-0.395150	2.173591
2009Q2	0.374777	0.539525	2.207159
2009Q3	1.357198	0.499460	2.151417
2009Q4	-1.374108	-1.380475	2.097125
2010Q1	0.397225	-0.368914	2.065166
2010Q2	0.620054	2.059960	2.150389
2010Q3	-1.537649	0.312627	2.163746
2010Q4	-1.883531	-1.361907	2.158429
2011Q1	-0.741146	0.549809	2.097079
2011Q2	1.273028	0.156689	2.127596
2011Q3	-2.012449	-0.284839	2.124768
2011Q4	-1.319847	-1.197729	2.185842
2012Q1	0.187500	0.117998	2.307886
2012Q2	2.239551	1.202126	2.186498
2012Q3	1.116603	1.072701	2.178629

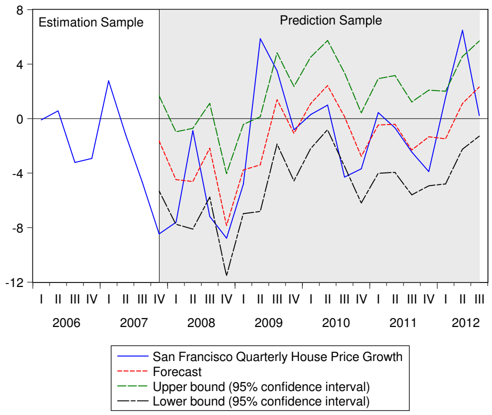
(b) Albany

Table 11: **One-step-ahead Forecast** of Housing Price Growth in San Francisco and Albany

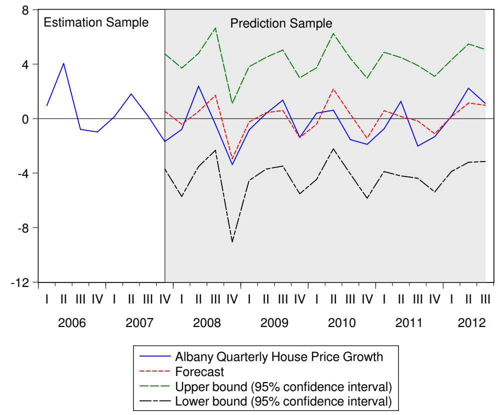
To evaluate these 1-step-ahead forecasts derived from the VAR(3) model, we compare them with those 1-step-ahead forecasts obtained from the best univariate model and we will implement a test of unconditional predictability. Based on the BIC, the best univariate models for San Francisco and Albany housing price growth rates are both ARMA(3,2), from which we obtain the 1-step-ahead forecasts[1]. We choose the mean squared error loss function. As in Exercise 4, we calculate the loss differences between the bivariate VAR and the univariate ARMA model, and regress the loss differences on a constant as follows

$$L_{t+j,1}^{\text{VAR}} - L_{t+j,1}^{\text{ARMA}} \equiv \Delta L_{t+j,1} = \beta_0 + \varepsilon_{t+j}, \quad \text{for } j = 0, 1, 2, \cdots, T-t-1.$$

[1] In order to avoid overfitting, we set the maximal lag for AR and MA terms to be 5. Therefore, the model candidates include different ARMA(p, q) specifications with $0 \leq p \leq 5$, and $0 \leq q \leq 5$.



(a) San Francisco



(b) Albany

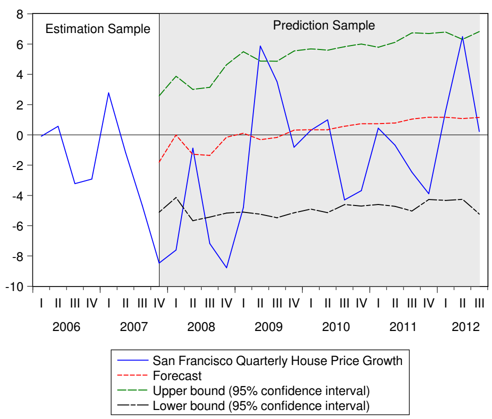
Figure 11: **One-step-ahead Forecasts** of House Price Growth in San Francisco and Albany

Then, we perform a t-statistic on the constant of the regression:

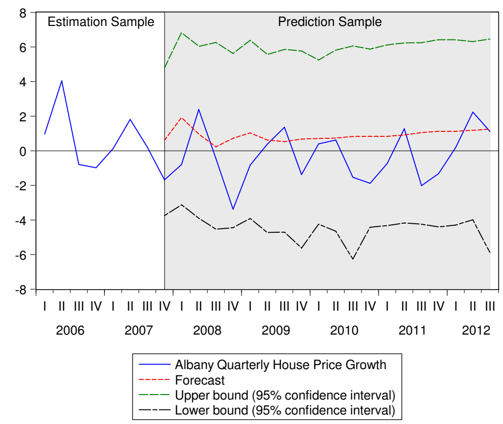
$$t = \frac{\hat{\beta}_0}{\hat{\sigma}_{\hat{\beta}_0}} \xrightarrow{A} N(0,1)$$

using a robust estimator for the variance of the test. The null hypothesis is that the constant term is zero, implying that on average both forecasts are equally accurate over the entire prediction sample. The test statistics for San Francisco and Albany series are 0.183423 and -2.037057 with associated two-sided p-values as 0.9218 and 0.0061, respectively. A positive t-test indicates that the loss from the VAR model is larger than the loss from the univariate ARMA model. Then, at the 5% significance level, we could conclude that, on average, the univariate ARMA model for the San Francisco market performs as well as the bivariate VAR model. There is no advantage to bring the Albany market into the model to forecast the San Francisco market. However, for the Albany market, the test indicates that the bivariate VAR model provides more accurate forecasts than the univariate ARMA model, indicating that the San Francisco market may have some effect in the Albany market. Given that the Granger-causality tests and the impulse-response functions point towards lack of interaction between these two markets, the preference for the bivariate VAR, in terms of predictability ability, should be interpreted with care. It is possible that the San Francisco market may be a proxy for a common national component in house prices.

In addition, we construct multi-step-ahead forecasts with 95% confidence intervals for the two house price growth rates, which are reported in Tables 12a, and 12b, and plotted in Figure 12a, and 12b. As we have commented in Exercise 5, the multi-step forecast is much smoother than the one-step-ahead forecast, and for a long forecasting horizon, it will converge towards the unconditional mean of the series. The wide confidence intervals also suggest that there is much uncertainty in both markets.



(a) San Francisco



(b) Albany

Figure 12: **Multi-step-ahead Forecasts** of House Price Growth in San Francisco and Albany

Date	Actual	Forecast	S.E.
2007Q4	-8.464693	-1.802864	1.680924
2008Q1	-7.606754	-0.007907	2.036892
2008Q2	-0.878331	-1.289379	2.149609
2008Q3	-7.180357	-1.361531	2.169138
2008Q4	-8.782996	-0.160856	2.468456
2009Q1	-4.804676	0.104304	2.559846
2009Q2	5.868813	-0.322635	2.523047
2009Q3	3.505712	-0.166978	2.618648
2009Q4	-0.821082	0.307984	2.719140
2010Q1	0.302694	0.332936	2.668168
2010Q2	0.990632	0.335621	2.701840
2010Q3	-4.295270	0.567177	2.719795
2010Q4	-3.683776	0.733152	2.721712
2011Q1	0.440821	0.732595	2.750514
2011Q2	-0.681955	0.780021	2.658232
2011Q3	-2.456326	1.034361	2.765615
2011Q4	-3.885916	1.163280	2.858362
2012Q1	1.557425	1.158318	2.811665
2012Q2	6.484462	1.071383	2.723702
2012Q3	0.221087	1.154010	2.773614

(a) San Francisco

Date	Actual	Forecast	S.E.
2007Q4	-1.683196	0.604649	2.173132
2008Q1	-0.795481	1.909514	2.473732
2008Q2	2.388146	0.961734	2.540341
2008Q3	-0.436979	0.221520	2.465711
2008Q4	-3.380073	0.724351	2.573075
2009Q1	-0.831809	1.032167	2.627469
2009Q2	0.374777	0.622930	2.676097
2009Q3	1.357198	0.527256	2.636630
2009Q4	-1.374108	0.671370	2.641923
2010Q1	0.397225	0.708234	2.552238
2010Q2	0.620054	0.730283	2.674573
2010Q3	-1.537649	0.821433	2.697683
2010Q4	-1.883531	0.832463	2.677257
2011Q1	-0.741146	0.828844	2.633735
2011Q2	1.273028	0.907758	2.603420
2011Q3	-2.012449	1.048054	2.708103
2011Q4	-1.319847	1.122429	2.734947
2012Q1	0.187500	1.121581	2.733531
2012Q2	2.239551	1.177017	2.586307
2012Q3	1.116603	1.250933	2.652500

(b) Albany

Table 12: **Multi-step-ahead Forecast** of Housing Price Growth in San Francisco and Albany

Exercise 9

We gather the three series of house price growth for the three MSAs (San Francisco, San Jose, and Albany; denoted as `GSF`, `GSJ` and `GAL`) and estimate the best trivariate VAR. We select a lag order of 3 lags according to the results in Table 13.[2]

VAR Lag Order Selection Criteria
Endogenous variables: GSF GSJ GAL
Exogenous variables: C
Sample: 1975Q1 2007Q3
Included observations: 417

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-742.4701	NA	49.94247	12.4245	12.49419	12.4528
1	-667.5729	144.8013	16.65412	11.32621	11.60496*	11.43942
2	-658.2265	17.60242	16.5636	11.32044	11.80825	11.51854
3	-633.1	46.06522	12.66950*	11.05167*	11.74854	11.33467*
4	-627.8166	9.422071	13.49797	11.11361	12.01954	11.48151
5	-623.5286	7.432544	14.63336	11.19214	12.30714	11.64495
6	-618.034	9.249227	15.56454	11.25057	12.57462	11.78827
7	-606.6226	18.63856*	15.01913	11.21038	12.7435	11.83298
8	-605.7115	1.44267	17.29003	11.34519	13.08737	12.0527
9	-598.2196	11.48744	17.86721	11.37033	13.32157	12.16274
10	-591.0383	10.65239	18.59582	11.40064	13.56094	12.27795

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Table 13: Lag Order Selection in Trivariate VAR

In Table 14, we show the estimation results for the trivariate VAR(3) model. Overall, these estimates point towards interdependence between San Francisco and San Jose markets, while the Albany market seems to be less connected to the West markets. We leave as an exercise the implementation of the Granger-causality tests and the impulse response functions. Based on this model, we obtain the one-step-ahead forecasts for the three series. With a mean squared error loss function, we implement the tests of unconditional predictability to compare the forecasts from the trivariate VAR(3) with those from the bivariate VAR(3) in Exercise 4. We present the test results in Table 15. The loss difference $\Delta L_{t+j,1}$ is defined as $\Delta L_{t+j,1} \equiv L^{\text{TriVAR}}_{t+j,1} - L^{\text{BiVAR}}_{t+j,1}$. Since the tests are positive, it indicates that the loss from the trivariate system is larger than that from the bivariate system. At the 5% significance level, we reject the null hypothesis of equal predictability and we choose the bivariate VAR as the preferred model because it produces a smaller loss. In order words, the Albany market is not helpful to improve upon the forecasts for San Francisco and San Jose markets based on a bivariate VAR.

[2]Note that although BIC prefers 1 lag, most others prefer 3 lags. We still choose 3 for the lags.

Vector Autoregression Estimates			
Sample (adjusted): 1976Q1 2007Q3			
Included observations: 127 after adjustments			
Standard errors in ( ) & t-statistics in [ ]			
	GSF	GSJ	GAL
GSF(-1)	0.289462	-0.026089	0.182116
	(0.12649)	(0.10965)	(0.16864)
	[2.28841]	[-0.23793]	[1.07993]
GSF(-2)	-0.180653	-0.192844	-0.40665
	(0.13337)	(0.11561)	(0.1778)
	[-1.35457]	[-1.66805]	[-2.28707]
GSF(-3)	0.520574	0.310472	0.410779
	(0.12701)	(0.11011)	(0.16934)
	[4.09854]	[2.81977]	[2.42582]
GSJ(-1)	0.615389	0.962589	-0.077542
	(0.1517)	(0.13151)	(0.20225)
	[4.05656]	[7.31972]	[-0.38340]
GSJ(-2)	-0.263381	-0.256437	0.243552
	(0.17487)	(0.15159)	(0.23314)
	[-1.50615]	[-1.69165]	[1.04467]
GSJ(-3)	-0.062861	0.013045	-0.303206
	(0.14917)	(0.12931)	(0.19887)
	[-0.42141]	[0.10088]	[-1.52462]
GAL(-1)	-0.01271	-0.098499	0.555023
	(0.06726)	(0.05831)	(0.08968)
	[-0.18895]	[-1.68924]	[6.18908]
GAL(-2)	-0.044306	0.029398	-0.145653
	(0.07607)	(0.06595)	(0.10142)
	[-0.58242]	[0.44579]	[-1.43612]
GAL(-3)	0.154916	0.094955	0.178077
	(0.06751)	(0.05852)	(0.09)
	[2.29482]	[1.62261]	[1.97863]
C	0.059864	0.260739	0.403861
	(0.22771)	(0.19739)	(0.30358)
	[0.26290]	[1.32091]	[1.33032]
R-squared	0.641406	0.654568	0.349414
Adj. R-squared	0.613822	0.627996	0.299369
Sum sq. resids	313.1172	235.2972	556.5458
S.E. equation	1.635914	1.418128	2.181009
F-statistic	23.25273	24.63404	6.981977
Log likelihood	-237.507	-219.3632	-274.0304
Akaike AIC	3.897748	3.61202	4.47292
Schwarz SC	4.1217	3.835971	4.696872
Mean dependent	2.122466	1.866065	1.329584
S.D. dependent	2.632491	2.325101	2.60563
Determinant resid covariance (dof adj.)			12.58646
Determinant resid covariance			9.841241
Log likelihood			-685.8135
Akaike information criterion			11.27265
Schwarz criterion			11.94451

Table 14: Vector Autoregression Estimates

		Trivariate VAR(3)		
		GSF	GSJ	GAL
Bivariate VAR(3)	GSF	2.290141 (2.907085) [0.0090]		
	GSJ		0.584851 (3.561469) [0.0021]	
t-statistics in ( ) & p-values in [ ]				

Table 15: Tests of Unconditional Predictive Ability between Trivariate and Bivariate VAR Models

Exercise 10

There is no standard or golden rules on how to write a report. There are many styles and formats but, in general, a good empirical report should contain a well-defined question, careful data description and appropriate econometric methodology, and conclusions derived from your data analysis. It is always very helpful to write one-page executive summary that condense the main findings of the report. Concise and clear writing is a must, avoid verbose statements and write in short sentences. Here we provide some general guidelines on the sections that a good empirical report should contain.

1. **Summary.** The objective should be to present a clear and to-the-point overview of the relevant findings in the report. Do not exceed two pages. This section is also known as 'executive summary'.

2. **Introduction.** It should state the objective of the report, what question(s) are addressed and why. It is always a good idea to write some context for the questions that you would like to analyze. The context could be a literature review of what has been done so far and what issues the current literature has not resolved yet. Sometimes it is a political debate or discussion that offers background, sometimes it is a request from a firm or from a unit within a firm that needs your predictions as an intermediate input for a larger project. You could bring some summary of the most important findings to this section as a way to keep the reader interested in what is coming up. At the end of this section, offer some guidelines on the organization of your writing to facilitate the reading of the report.

3. **Data and econometric models.** Your conclusions will be as good as the data that you work with and the models that you propose. Critics of your report will focus intensively on this part of your report. You need to provide the sources of your data, the measurement units, methods of collection when they are known, and your judgment on the overall quality of the data sets. If there are several sources providing similar data, exploring the robustness of your results is a valuable exercise that will speak well about your professional acumen. Generally, the econometric model(s) that you propose should have a conceptual framework as a starting point. It is a good idea to have some micro or/and macro economic theories that support the hypothesis/questions that you wish to analyze. Since here you are interested in a forecasting question, you will bring dynamics to the model and only the data will guide you on the specification of the best dynamics. Be explicit and specific about your model choices, state any assumptions that you will entertain, and the estimation methods that are suitable for the problem at hand. Inform the reader when you are using your scientific judgment. If you encounter econometric problems, be honest about the remedies that you implement and explain your confidence on why a remedy is a total or partial solution to the problem.

4. **Results.** You will present estimation results and forecasting results. First, present the esti-

mation results concerning the models that you propose above. It is likely that you estimate more than one model. Do not crowd the report with unnecessary tables, these can go directly to the Appendix. Comment on the main findings of your preferred model(s), e.g. the estimates and their statistical significance, the overall fit of the model to the data, the behavior of residuals, etc. Secondly, present the construction of your forecast. Remember that the loss function guides the forecast, thus you need to be explicit about the function you choose or the function that is given to you. Be also explicit about your information set and whether you are interested in short or long term forecasts. Competing models will produce competing forecasts. Explain which ones you would prefer and why. Figures and plots are the most attractive ways to present a forecast. It is extremely important that you quantify the uncertainty of the forecasts. As a validation of your forecasting techniques, you may reserve some observations from your sample and write a short paragraph comparing the actual realizations with your forecast(s). If there are other additional reports analyzing the same or similar question, you may want to compare your results with those, and explain your agreement or disagreement.

5. **Conclusions.** This section should be short. It is summary of the previous sections. You may want to underscore what you think is the main contribution and findings of your analysis. You may discuss availability of new data, when to update the forecast, and further methodological improvements.

6. **Appendix.** This section is for intermediate results that the interested reader may need either for replication or for a more detailed understanding of your estimation and forecasting methods. It is always a good idea to provide your data or to provide a website as a reference, if it is not proprietary.

