I. MODELING METHODOLOGY

The modeling of state-level HPI is divided into two phases:

1) Drift modeling
2) Volatility modeling.

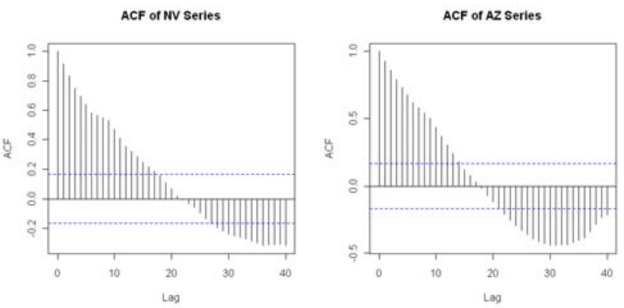
Both the phases are explained, with analysis and results for a set of states, as follows:

Phase#1: Drift Modeling

This stage consists of the following steps:

1) Data preparation
   The period selected for modeling is January 1998 to June 2008.
   a) Missing value treatment
   b) Exploratory data analysis
      i) Transformations
         Stationarity of the response and predictor variables was tested statistically using the Augmented Dickey-Fuller test and checked graphically by looking at the autocorrelations plot.
         After this initial analysis of the data, a simple returns transformation was carried out on the response and predictor variables for modeling & forecasting the HPI data. A simple returns transformation induced stationarity in the series. See Figure 1.b for results of the independent variables and the two response variables of Nevada and Arizona HPI.

	P-value	Lag
Median income	0.01	3
30-yr commitment rate	0.01	3
Mortgage originations	0.01	5
Unemployment rate	0.72	2
Population size	0.01	2
Foreclosures	0.27	6
Building permits	0.01	3
Nevada HPI	0.63	1
Arizona HPI	0.19	1



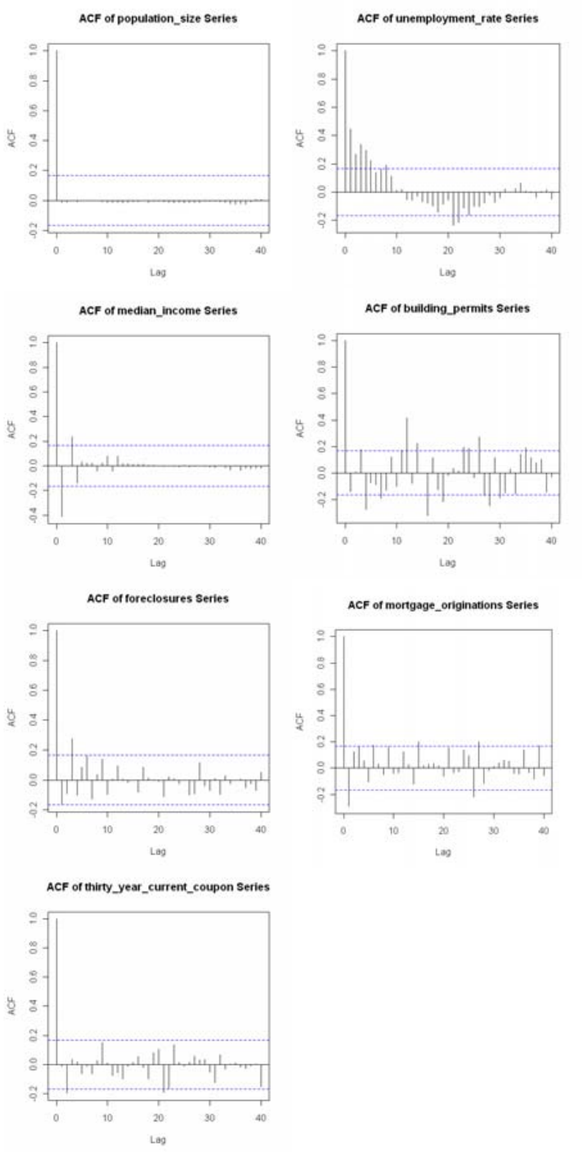
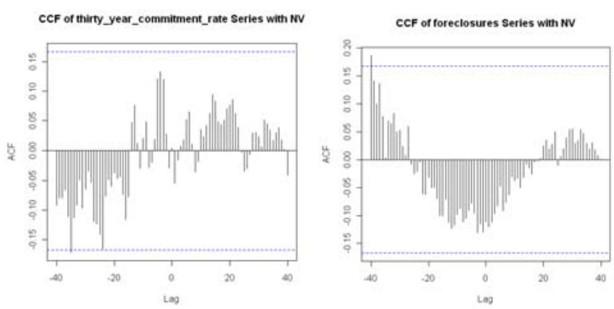


Figure 1.b: ADF results and ACF plots for the predictor variables and NV and AZ

   c) Identification of leading and lagging indicators
      i) Cross correlation analysis: This analysis was carried out to identify the leading and lagging indicators and the lags at which these indicators showed a high correlation with the HPI index. This step is crucial to capture true dependencies between the predictors and the response variable in the modeling stage, and develop a robust fit.
         Figure 1.c shows the lags showing significant correlations with the HPI simple returns data for the state of Nevada and Arizona.



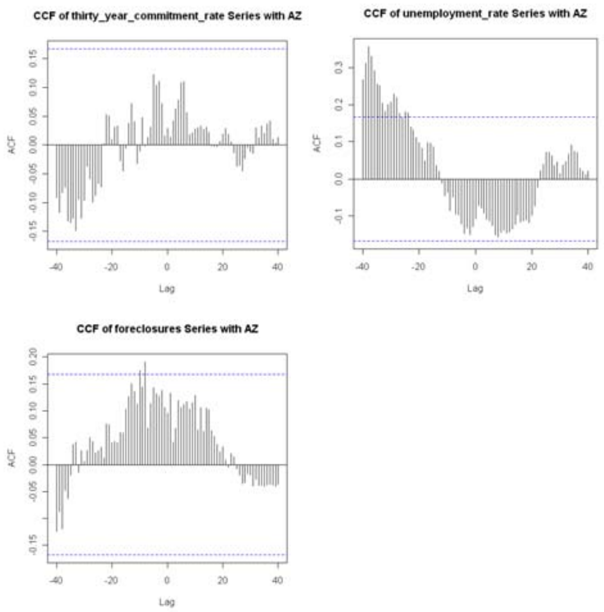


Figure 1.c: CCF plots of NV and AZ with lagged variables

2) Multiple regression analysis

a) Identification of the significant predictors: Once the leading and lagging indicators were identified, a multiple regression model was developed for each state separately. The modeling scheme employs a step-by-step inspection of p-values in the each regression output to identify the significant predictors and come up with a parsimonious model with highest explanatory power as measured by the R^2 values. At each iteration, the most insignificant variable, measured by the largest p-value, was dropped; then the model was refitted without the dropped parameter. Following results show the iterative and final models developed for Nevada and Arizona.

NEVADA - Iteration 1				
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.002	0.006	0.282	0.779
median_income	-0.071	0.185	-0.385	0.701
thirty_year_commitment_rate(-35)	-0.037	0.021	-1.785	0.078
mortgage_originations	0.004	0.002	2.146	0.035
unemployment_rate	-0.389	0.055	-7.098	0.000
population_size	7.217	7.891	0.915	0.363
foreclosures(-40)	0.004	0.002	1.675	0.097
building_permits	-0.001	0.003	-0.260	0.795
R Square	0.463			
Adjusted R Square	0.420			

Drop "building permits" and re-estimate

NEVADA - Iteration 5				
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.008	0.001	8.241	0.000
thirty_year_commitment_rate(-35)	-0.043	0.021	-2.105	0.038
mortgage_originations	0.004	0.002	2.152	0.034
unemployment_rate	-0.398	0.053	-7.483	0.000
R Square	0.440			
Adjusted R Square	0.422			

ARIZONA - Iteration 5				
	Coefficients	*Standard Error*	*t Stat*	*P-value*
(Intercept)	0.009	0.001	10.021	0.000
thirty_year_commitment_rate(-33)	-0.052	0.025	-2.066	0.042
unemployment_rate(-29)	0.126	0.049	2.563	0.012
foreclosures(-8)	0.002	0.001	2.666	0.009
Adjusted R Square	0.303			

3) Analysis of residuals:

a) Validation of assumptions of multiple regression: It is imperative to check that the final regression model satisfies the basic premises of the multiple regression framework to ensure that the regression is not spurious. The following checks are employed to test the validity of the drift models:

i) Homoskedasticity of residuals i.e. the residuals must have a constant variance.

ii) The final conclusion from the regression model must not be driven by influential outliers in the data. If outliers are detected then the data should be appropriately conditioned to get a robust fit. There are no outliers observed in the diagnostic plots.

iii) The residuals must be normally distributed. This is checked using the QQ plot as shown below for the modeled states. The points lie along the straight line, which clearly confirms normality of residuals for the fitted model.

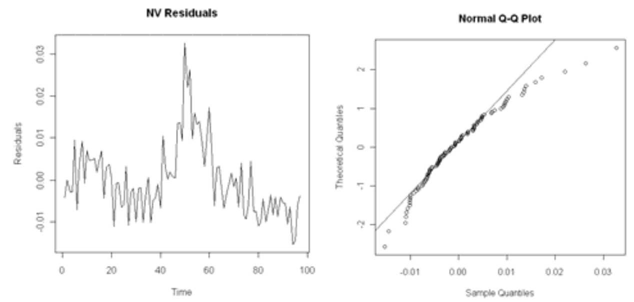


Figure 3.a.i: Nevada regression residual analysis

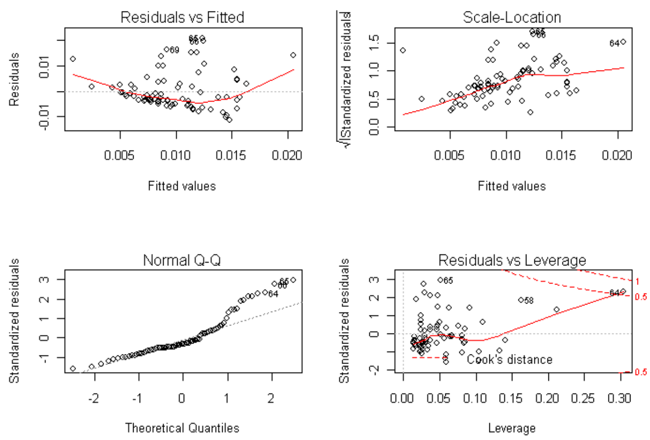


Figure 3.a.ii: Arizona regression residual analysis

iv) There should be no autocorrelation observed in the residuals. The ACF plots of the residuals are created to check for independence. If there are no significant autocorrelations observed, as seen in Figure 3.a.iv, then there is no need to model the innovations component for this series. However if there are significant autocorrelations, as seen in Figure 3.a.iii, then the residuals are correlated, which implies dependence amongst residuals. This calls for modeling the innovations, which takes us to phase#2 of the modeling methodology.

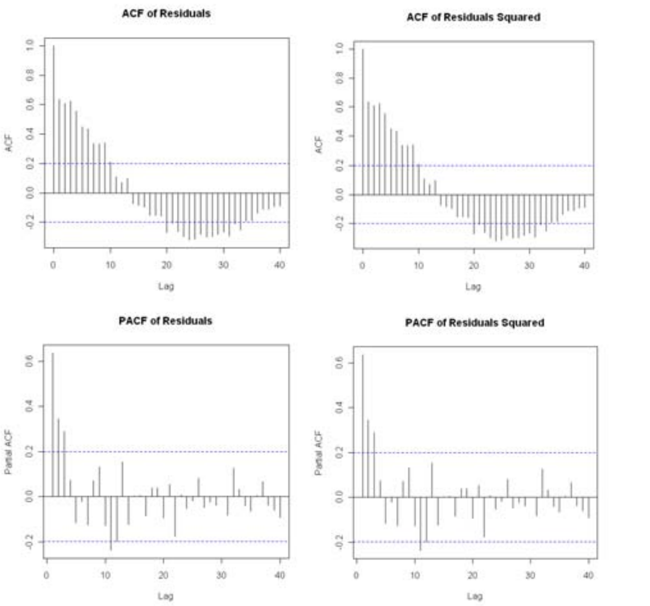


Figure 3.a.iii: ACF of residuals and squared residuals for NV

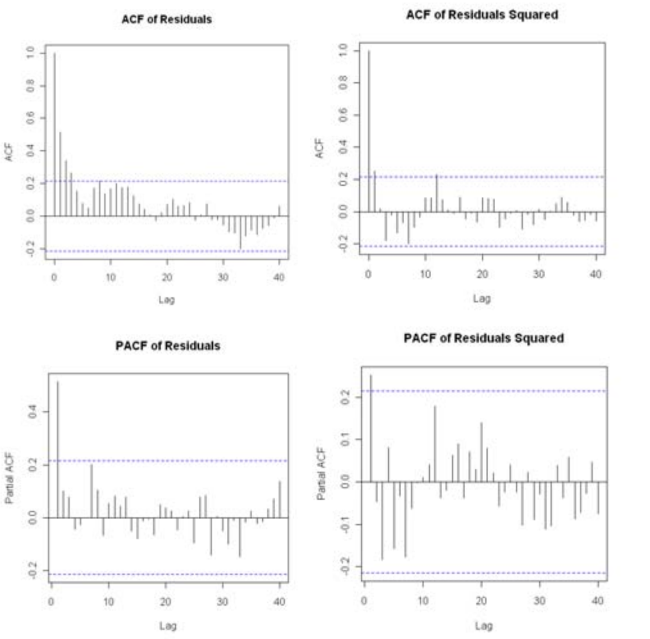


Figure 3.a.iv: ACF of residuals and squared residuals for AZ

Phase#2: Innovations (Volatility) Modeling

1) ARFIMA model fitting
   a) Determination of model order
      i) Time series, ACF, PACF and EACF plots are examined to evaluate test the series for unit root non-stationarity and to choose appropriate model order for the time series of the residuals series in question. Trial and error based approach is employed to determine the most parsimonious model. See Figure 4.b.i for the ARMA results of Nevada.
   b) Analysis of residuals
      i) Once the ARMA model is fitted the residuals are tested for autocorrelations. Box-Ljung test is carried out on the residuals and squared-residuals to test for autocorrelations and ARCH effects. The results of this test for Nevada are given in the Figure 4.b.ii. If autocorrelations are observed, then we re-fit the model using different orders. If ARCH effects are observed, based on the result of Box-Ljung test on the squared residuals, then we move to fit the ARFIMA-GARCH model to model the innovation component.

```
Series: residuals
ARIMA(0,1,1)

Call: auto.arima(x = residuals)

Coefficients:
          ma1
      -0.6094
s.e.   0.0708

sigma^2 estimated as 3.634e-05:
log likelihood = 354.24
AIC = -704.48   AICc = -704.35   BIC = -699.35
```

Figure 4.b.i: ARIMA results for Nevada residual analysis. An ARIMA(0,1,1) model was fit.

```
> Box.Ljung.test(resid.ts$residuals, lag = 12,
adj.DF = 12 - 1)
    Q(12)      P-value        df
 14.05752       0.2298        11

> Box.Ljung.test(resid.ts$residuals^2,lag=12,
adj.DF = 12)
    Q(12)      P-value        df
 20.80679       0.0533        12
```

Figure 4.b.ii: Box-Ljung test results on the residual and residual squared series after fitting an ARMA model. The large p-values indicate that the model is sufficient.

Once the model is finalized by carrying out the Phase#1 and Phase#2, we test the model for its validity and how well it forecasts the HPI returns from which we can back transform to get the actual HPI forecasts. We forecast the data held back as the validation set to assess the accuracy of the model and to document the forecast errors. The model prediction errors for the Nevada are documented in Table 5. If these prediction errors are within tolerable limits we can finalize on these models and use them to forecast long run HPI estimates.

30-yr Commitment Rate (-35)	Mortgage Originations	Unemploy-ment Rate	NV HPI	Predicted Regression	Predicted Residual	Predicted Value	Prediction Error
-0.011	0.031	0.000	-0.002	0.010	-0.010	0.000	120%
0.123	0.068	0.000	-0.002	0.004	-0.009	-0.005	-147%
0.038	-0.315	0.000	-0.005	0.008	-0.008	0.000	105%
-0.033	0.381	0.023	-0.005	0.006	-0.007	-0.001	85%
-0.024	-0.387	0.022	-0.005	0.005	-0.006	-0.001	90%
-0.047	-0.175	0.022	-0.008	0.006	-0.005	0.001	118%
0.011	-0.649	0.021	-0.008	0.004	-0.004	0.000	94%
-0.027	0.025	0.021	-0.008	0.006	-0.004	0.002	126%
0.034	-0.241	0.020	-0.007	0.003	-0.003	0.000	100%
-0.020	-0.493	0.020	-0.007	0.006	-0.003	0.003	138%
-0.018	-0.504	0.020	-0.007	0.006	-0.002	0.003	148%
0.047	-0.881	0.019	-0.016	0.003	-0.002	0.001	105%
0.027	-0.406	0.038	-0.017	-0.001	-0.002	-0.002	87%
-0.037	-0.842	0.018	-0.017	0.007	-0.001	0.005	130%
-0.031	0.000	0.036	-0.020	0.003	-0.001	0.001	107%
-0.006	-0.333	0.052	-0.020	-0.002	-0.001	-0.003	83%
0.062	-0.500	0.049	-0.021	-0.005	-0.001	-0.006	72%

Table 5: Backtest of combined regression and ARMA model for Nevada using forecasts and prediction error.