Title: HPI Forecast 3.0 Cookbook

Author: Kevin Gillen, Sr. Housing Economist

Date: 2/18/10

I. Overview

This "cookbook" is designed to allow the reader to replicate the monthly production of the FA HPI 3.0 Forecast.

The HPI Forecast currently produces a 12-month point forecast for a variety of geographies (National to Zip) and tiers (SFD, SFA, SFC and SFCX).

The following table lays out the process followed for the Forecast update, with each row in the table representing a step in the update process. For each step in the update process, the table gives:

- The names of the program(s) called in that step;
- The names of the input data files called in that step;
- The names of the output data files created in that step;
- The names(s) of the person(s) responsible for executing that step;
- The time allocated to completing that step;
- A description of that step.

A critical feature of the HPI forecast is its hierarchical structure. The forecast of the HPI for any given geography is not independent of the forecasts for other geographies. Rather, each forecast takes as an input the forecast of the larger geography it is embedded in. For example, the forecast for CBSA-level HPIs uses the forecasts of the State-level HPIs. The hierarchy that is followed is: National->State-CBSA->County->Zip. Hence, each step of the forecast has some dependency on the previous step.

Every month, the forecast is updated using the updated HPIs and the updated exogenous variables. The general procedure we follow is:

1) Update the National HPIs and Exogenous Variables
2) Generate forecasts of the Exogenous Variables
3) Generate forecasts of the National HPIs
4) Use output from 2) and 3) to generate forecasts of the States, and output for review

    a. Review State-level forecasts, and identify if any forecasts need to be withheld
5) Use output from 4) to generate forecasts of the CBSAs, and output for review
    a. Review CBSA-level forecasts, and identify if any forecasts need to be withheld
6) Use output from 5) to generate forecasts of the Counties, and output for review
    a. Review County-level forecasts, and identify if any forecasts need to be withheld
7) Use output from 6) to generate forecasts of the Zips, and output for review
    a. Review Zip-level forecasts, and identify if any forecasts need to be withheld
8) Send all output from 3)-7) to be loaded into ETL

II. SAS Programs used to generate the HPI forecast

The following table lists the SAS programs, in order of the sequence that they are run, to generate the updated forecast each month.

Step	Program	Input Data Files	Output Data Files	Responsibility	Time	Description
0	ExovarForecast.sas	Downloading data from 36 different sites	Forecast_exog_varsMMM.xls	FIC	4 days	Formatting the data, computation of price rent.
1	Forecast_Exog_Vars_MMMYY.sas	Forecast_exog_varsMMM.xls	ExogVarsFrcst_MMMYY_Export.xls	Kevin	1 day	This program takes a file of regularly updated time series variables and generates 12-month forecasts of these variables for input into the HPI Forecast model. FIC is currently responsible for updates the input data file.
2	National_Forecast_MMMYY.sas	Fixed_nat_final_00.sas7bdat	NationalHPIsExportMMMYY.xls	Kevin	1 day	Generate 12-month

		Fixed_nat_final_07.sas7bdat Fixed_nat_final_11.sas7bdat Fixed_nat_final_12.sas7bdat ExogVarsFrcst_MMMYY_Export.xls				forecasts of National HPIs for all four tiers.
3	State_tier0.sas – State_tier12.sas	NationalHPIsExportMMMYY.xls ExogVarsFrcst_MMMYY_Export.xls and SAS State input files from HPI index team.	State_HPI_3.0_tier0.xls State_HPI_3.0_tier7.xls State_HPI_3.0_tier11.xls State_HPI_3.0_tier12.xls State_SE_3.0_tier0.xls State_SE_3.0_tier7.xls State_SE_3.0_tier11.xls State_SE_3.0_tier12.xls	FIC	1 day	Generate 12-month forecasts of State HPIs for all four tiers and std errors.
4	ReviewForecast1.sas	State_HPI_3.0_tier0.xls State_HPI_3.0_tier7.xls State_HPI_3.0_tier11.xls State_HPI_3.0_tier12.xls	State_Censor.xls	Kevin	½ day	Read in forecast for all States and tiers, generate summary statistics and diagnostics, output file of States and tiers to be withheld from release (if any). Sign off on release to ETL
5	CBSA_tier0.sas – CBSA_tier12.sas	State_HPI_3.0_tier0.xls State_HPI_3.0_tier7.xls State_HPI_3.0_tier11.xls State_HPI_3.0_tier12.xls ExogVarsFrcst_MMMYY_Export.xls and SAS CBSA input files from HPI index team	CBSA_HPI_3.0_tier0.xls CBSA_HPI_3.0_tier7.xls CBSA_HPI_3.0_tier11.xls CBSA_HPI_3.0_tier12.xls CBSA_SE_3.0_tier0.xls CBSA_SE_3.0_tier7.xls CBSA_SE_3.0_tier11.xls CBSA_SE_3.0_tier12.xls	FIC	1 day	Generate 12-month forecasts of CBSA HPIs for all four tiers and std errors.
6	ReviewForecast1.sas	CBSA_HPI_3.0_tier0.xls CBSA_HPI_3.0_tier7.xls CBSA_HPI_3.0_tier11.xls CBSA_HPI_3.0_tier12.xls	CBSA_Censor.xls	Kevin	1 day	Read in forecast for all CBSAs and tiers, generate summary statistics and

						diagnostics, output file of CBSAs and tiers to be withheld from release (if any). Sign off on release to ETL.
7	Cnty_tier0.sas – Cnty_tier12.sas	CBSA_HPI_3.0_tier0.xls CBSA_HPI_3.0_tier7.xls CBSA_HPI_3.0_tier11.xls CBSA_HPI_3.0_tier12.xls ExogVarsFrcst_MMMYY_Export.xls and SAS County input files from HPI index	Cnty_HPI_3.0_tier0.xls Cnty_HPI_3.0_tier7.xls Cnty_HPI_3.0_tier11.xls Cnty_HPI_3.0_tier12.xls Cnty_SE_3.0_tier0.xls Cnty_SE_3.0_tier7.xls Cnty_SE_3.0_tier11.xls Cnty_SE_3.0_tier12.xls	FIC	1 day	Generate 12-month forecasts of County HPIs for all four tiers and std errors.
8	ReviewForecast1.sas	Cnty_HPI_3.0_tier0.xls Cnty_HPI_3.0_tier7.xls Cnty_HPI_3.0_tier11.xls Cnty_HPI_3.0_tier12.xls	county_Censor.xls	Kevin	1 day	Read in forecast for all counties and tiers, generate summary statistics and diagnostics, output file of counties and tiers to be withheld from release (if any). Sign off on release to ETL.
9	Zip_tier0.sas – Zip_tier12.sas	Cnty_HPI_3.0_tier0.xls Cnty_HPI_3.0_tier7.xls Cnty_HPI_3.0_tier11.xls Cnty_HPI_3.0_tier12.xls ExogVarsFrcst_MMMYY_Export.xls and SAS Zip input files from HPI index	Zip_HPI_3.0_tier0.xls Zip_HPI_3.0_tier7.xls Zip_HPI_3.0_tier11.xls Zip_HPI_3.0_tier12.xls Zip_SE_3.0_tier0.xls Zip_SE_3.0_tier7.xls Zip_SE_3.0_tier11.xls Zip_SE_3.0_tier12.xls	FIC	2 days	Generate 12-month forecasts of Zip HPIs for all four tiers and std errors.
10	ReviewForecast1.sas	Zip_HPI_3.0_tier0.xls	Zip_Censor.xls	Kevin	1 day	Read in forecast for

		Zip_HPI_3.0_tier7.xls Zip_HPI_3.0_tier11.xls Zip_HPI_3.0_tier12.xls				all Zips and tiers, generate summary statistics and diagnostics, output file of counties and tiers to be withheld from release (if any). Sign off on release to ETL.
11	State_tier0_bk.sas - State_tier12_bk.sas . CBSA_tier0_bk.sas - CBSA_tier12_bk.sas . Cnty_tier0_bk.sas - Cnty_tier12_bk.sas . Zip_tier0_bk.sas - Zip_tier12_bk.sas .	ExogVarsFrcst_MMMYY_Export.xls, SAS Input files provided by HPI index team for all geo's.	State_HPI_3.0_tier0.xls State_HPI_3.0_tier7.xls State_HPI_3.0_tier11.xls State_HPI_3.0_tier12.xls CBSA_HPI_3.0_tier0.xls CBSA_HPI_3.0_tier7.xls CBSA_HPI_3.0_tier11.xls CBSA_HPI_3.0_tier12.xls Cnty_HPI_3.0_tier0.xls Cnty_HPI_3.0_tier7.xls Cnty_HPI_3.0_tier11.xls Cnty_HPI_3.0_tier12.xls Zip_HPI_3.0_tier0.xls Zip_HPI_3.0_tier7.xls Zip_HPI_3.0_tier11.xls Zip_HPI_3.0_tier12.xls	FIC	4 days	Generate Back testing results by removing last 12 months index and rerunning the models for State, CBSA, County and Zip. Graphs are also generated.

Notation:

"MMM" is month; e.g. "Jan"

"YY" is year; e.g. 10

So, the output file of the National HPI for created in January 2010 would be "NationalHPIsExportJan10.xls".

III. Mathematical Specification of the forecast

This section outlines the mathematical specifications of the forecast model. The forecast for each geography is essentially a two-step process: First, the forecast of the evolution of the actual price-rent ratio towards its equilibrium price-rent ratio is generated from a vector of exogenous variables, and then that forecast is used to generate the forecast of the evolution of the HPI. A compact expression of the model's structure would be:

$$Ln\left({PR_{it}^{*}}\Big/{PR_{it-1}}\right) = f(Exogenous\ Variables) \quad (1)$$

$$Ln\left({HPI_{it}}\big/{HPI_{it-1}}\right) = f\left(Ln\left({\widehat{PR_{it}^{*}}}\Big/{PR_{it-1}}\right), Ln\left({\widehat{HPI_{jt}}}\Big/{HPI_{jt-1}}\right), \sum_{k} Ln\left({HPI_{it-k}}\big/{HPI_{it-k-1}}\right)\right) \quad (2)$$

Where:

PR_{it}^{*} = equilibrium price-rent ratio of geography i at time t

PR_{it-1} = price-rent ratio of geography i at time t-1

HPI_{it} = House Price Index of geography i at time t

HPI_{it-1} = House Price Index of geography i at time t-1

$\widehat{PR_{it}^{*}}$ = predicted equilibrium price-rent ratio of geography i at time t

HPI_{jt} = House Price Index of geography j at time t, where j indexes the next-largest geography that geography i is located in

HPI_{jt-1} = House Price Index of geography j at time t-1, where j indexes the next-largest geography that geography i is located in

$\widehat{HPI}_{jt}$ = predicted HPI of geography j at time t

HPI_{it-k} = House Price Index of geography i at time t-k, where k=1,2,3,4,12.

Both Equations (1) and (2) are linear regressions of the dependent variable on a vector of the variables in the parentheses, including an intercept. They are estimated via MLE in most cases, and via CLS if the SAS program encounters a non-convergence problem with MLE. The HPI forecast is performed sequentially for each equation, and hierarchically for each geography, by tier, as follows:

1) Beginning with the National HPIs, the log return of the price-rent ratio is regressed on the exogenous variables via Equation (1), and a 12-month forecast of the price-rent ratio's evolution is generated.

$$Ln\left({PR^*_{it}}/{PR_{it-1}} \right) = f(Exogenous\ Variables) \quad (3)$$

2) The 12-month forecast of the National HPI is then computed by regressing the log return in the HPI on the log return in price-rent (with forecast from step 1) appended) and lagged values of the log return in the National HPI. These variables are the first and third terms in Equation (2). Since LP does not estimate an HPI for any geography larger than the Nation, the second term in Equation (2) drops out:

$$Ln\left({HPI_{it}}/{HPI_{it-1}} \right) = f\left(Ln\left({\widehat{PR^*_{it}}}/{PR_{it-1}} \right), \sum_k Ln\left({HPI_{it-k}}/{HPI_{it-k-1}} \right) \right) \quad (4)$$

The forecast then proceeds hierarchically downwards by geography as follows:

3) For the States, the log return of the price-rent ratio is regressed on the exogenous variables via Equation (1), and a 12-month forecast of the price-rent ratio's evolution is generated.

$$Ln\left({PR^*_{it}}/{PR_{it-1}} \right) = f(Exogenous\ Variables) \quad (5)$$

4) The 12-month forecast of the State HPIs are then computed by regressing the log return in the HPI on the log return in price-rent (with forecast from Step 3) appended), the forecast of the National HPI from Step 2) and lagged values of the log return in the State HPI. These variables are the three terms in Equation (2):

$$Ln\left({HPI_{it}}/{HPI_{it-1}}\right) = f\left(Ln\left({\widehat{PR^*_{it}}}/{PR_{it-1}}\right), Ln\left({\widehat{HPI_{jt}}}/{HPI_{jt-1}}\right), \sum_k Ln\left({HPI_{it-k}}/{HPI_{it-k-1}}\right)\right) \quad (6)$$

5) Steps 3) and 4) are then repeated for the CBSAs[1], Counties and Zips, with the forecast of the next higher-level geography entered as an independent variables in Equation (6).

IV. Estimation of the forecast

The forecast regressions specified in Equations (1) and (2) are estimated in SAS using Proc Arima. The method of Maximum Likelihood Estimation (MLE) is the default estimation algorithm used for the regression, with the method of Conditional Least Squares (CLS) used in the event that the MLE estimates fail to converge.

In time series regressions such as the HPI Forecast, the method of Ordinary Least Squares (OLS) is not typically used because the structure of the errors do not conform to the traditional assumptions of OLS. In regression analysis, a critical assumption is that the error terms are independent of each other; i.e. distributed i.i.d. If the error terms are not independent (autocorrelated), the efficiency of the OLS parameter estimates is adversely affected and the standard error estimates are biased. In addition, OLS also assumes that the error variance is the same for all observations. When the error variance is not constant (heteroscedastic), the OLS estimates are inefficient.

[1] Since CBSAs can include multiple States, the HPI forecasts for all of the States covered by a CBSA is used. For example, the forecast for the New York City CBSA includes the forecasts for New York, New Jersey and Connecticut.

Both of these assumptions are often violated with time series data. Errors are autocorrelated, and their variance can be non-constant with respect to time. Indeed, merely by estimating and ARMA or ARCH model, we are explicitly assuming these assumptions are violated by imposing an autocorrelation structure on the residuals and/or their variance.

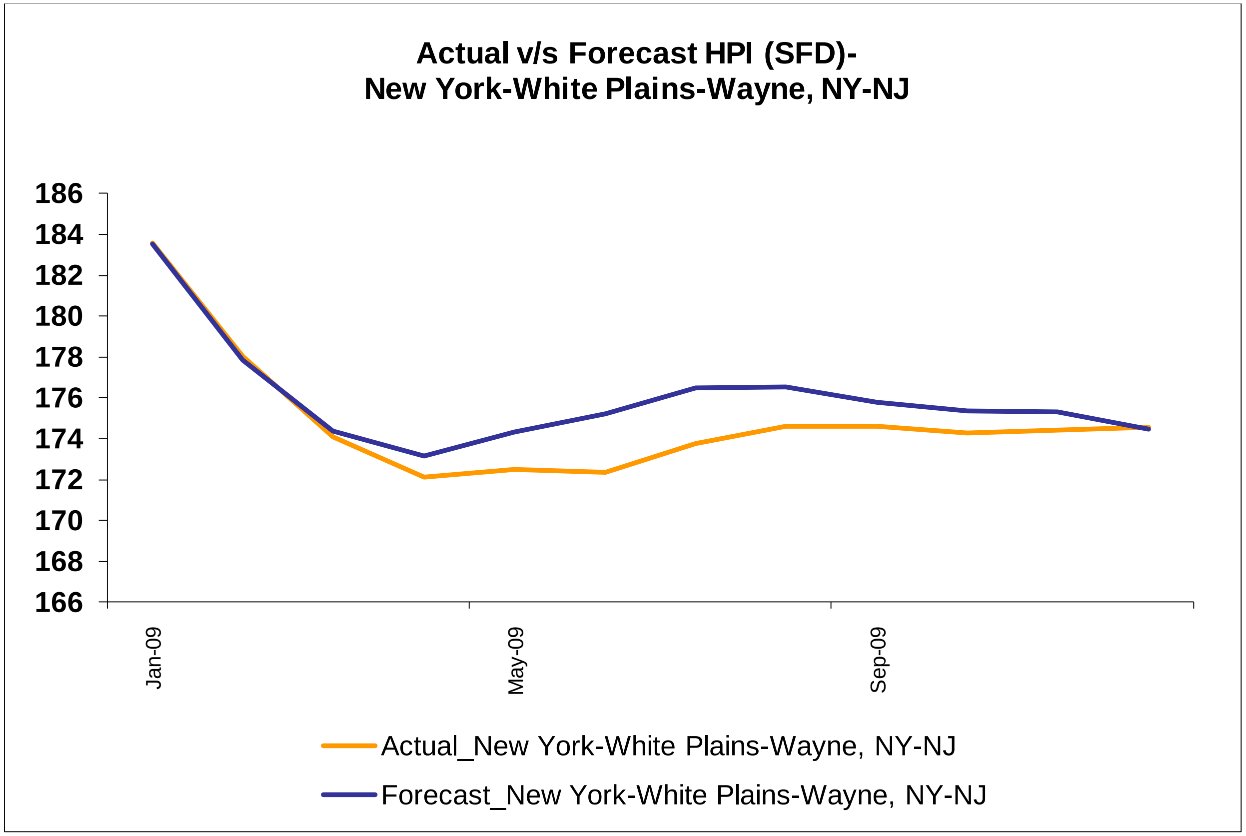
V. Validation of the Forecast

As part of each month's update to the forecast, a backtest routine is performed by simulating an out-of-sample forecast. Since the current forecast horizon is 12 months, the model's performance over a horizon of this length is the primary performance metric of interest. The out-of-sample backtest is done by dropping the most recent 12 months of the HPI, re-running the forecast for these 12 months, and then comparing this forecast to the actual HPI over this period. If the most recent time period of the observed HPI is t, then the last t-12 months of data are dropped from the HPIs and exogenous variables, and the model is re-estimated according to Equations (1) and (2). There are two outputs from the validation, for each set of geographies and tiers:

- An Excel file plotting the Forecasted HPI against the Actual HPI, for each backtested geography and tier.
- A SAS dataset of the values of the Forecasted and Actual HPIs, with the total 12-month percent change in each.

The Excel file is used to not only provide visual examination of the forecast model's performance, but also in presentations to clients to provide evidence of the model's performance. An example is shown below. The SAS dataset can be used to perform more formal tests of the model's performance; for example, regressing the (percent change in the) forecasted HPI against (the percent change in the) actual HPI can yield a number of useful metrics characterizing the forecast's accuracy: R-squared, level of heteroscedasticity, what percent of forecasts are within X% of the actual, etc.

Example of 12-month Out-of-Sample Backtest: Forecasted HPI v. Actual HPI



VI. Exogenous Variables in the Forecast

The following table lists the exogenous variables that are used in the forecast, per Equation (1)

Exogenous Variables in the Forecast		
Variable	**Definition**	**Source**
10-Year Treasury Yield	Yield (%) on the 10-Year Treasury Note	U.S. Federal Reserve
Slope of Yield Curve	10-Year treasury Yield minus 1-Year Treasury Yield	U.S. Federal Reserve
Unemployment rate	Percent of U.S. working population that is unemployed	U.S. Bureau of Labor Statistics (BLS)
Months' Supply of Homes	# of months it would take to sell current inventory of listed homes	National Association of Realtors
Foreclosure Inventories	# of homes in foreclosure	Mortgage Bankers' Association
REO Inventories	# of homes in REO	TrueStandings
Homebuilder Sentiment	An index of homebuilder sentiment	National Association of Homebuilders
Probabililty of Recession	Probability that the U.S. economy is in recession	J. Piger, University of Oregon

The set of exogenous variables used in the forecast are the same regardless of the geography being forecast. However, their measurement may be local, where appropriate. For example, the unemployment rate is always in the set of exogenous variables, but we will use the national unemployment rate for the national HPI forecast, the California unemployment rate for the California HPI forecast, etc. As of the time of this writing, the build-out of the exogenous variables to the local level is still an ongoing effort. By default, the forecast will use variables measured at the National level if the local level is either unavailable or not appropriate (e.g. interest rates). Lastly, the exogenous variables are observed through time period t-1, and forecast for periods t=1,2,…,12.