Home Price Dataset: Executive Report

The goal of this project is to analyze the Case-Shiller Home Price Index to assess the effects of the United States housing bubble. The bubble began in the first quarter of 1998 and continued to grow rapidly until it burst in late 2005. Both the beginning and end of the bubble can be seen clearly in Figure 5.1. At the start of 1998 the index began to increase much more quickly than it had for the past decade, and in the middle of 2005 the rapid growth slowed dramatically. Home prices continued to rise for a short period afterward, but by the middle of 2006 the precipitous fall had begun.

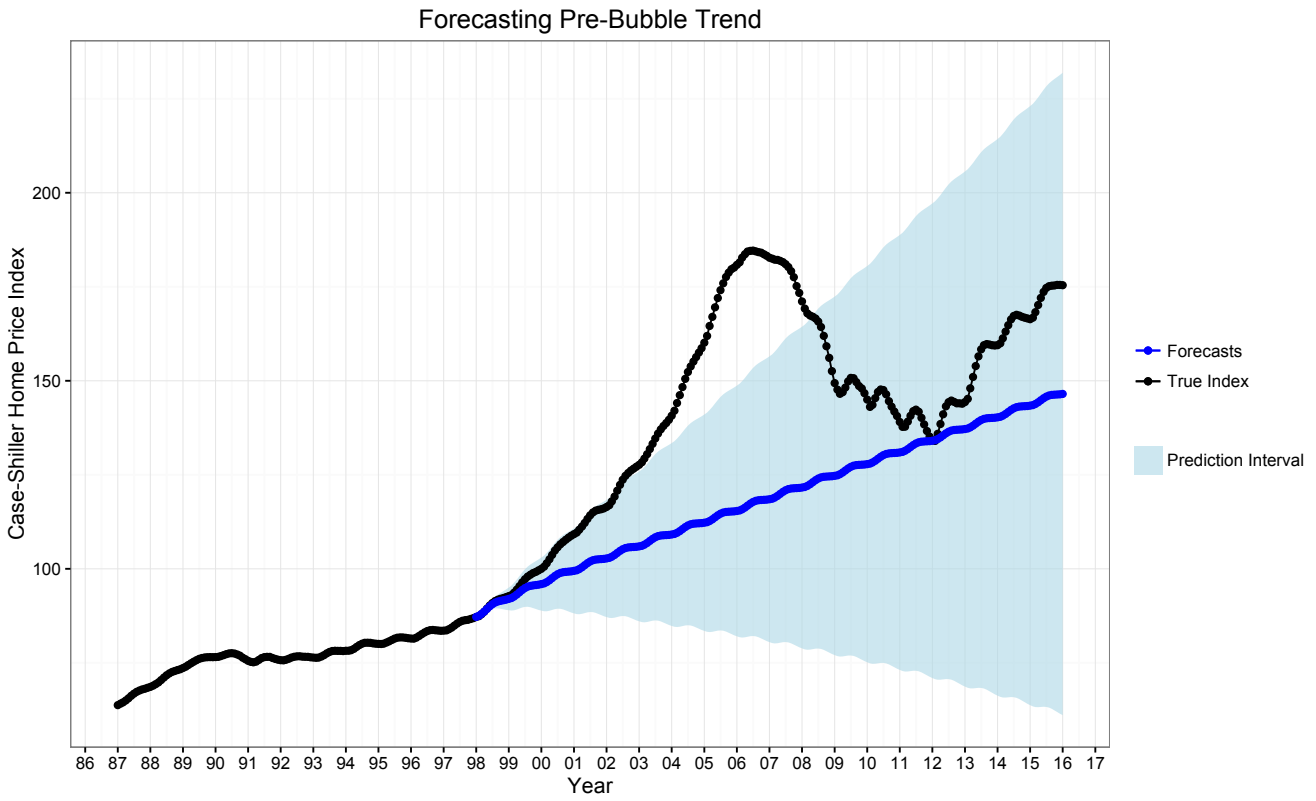

Figure 5.1: Case-Shiller Home Price Index from 1987-2016 along with forecasts for pre-bubble trends.

The bursting of the bubble has resulted in corrected housing prices, but the correction has not gone so far as to return to levels that would have been expected had there been no bubble. Had the trends from the 1990's continued without the bubble, housing prices would likely be lower than they are currently. The effects of the bubble are still lingering, and they can be seen mostly in the greater month-to-month volatility in home prices from 2009 to present.

Looking at the recent increase in housing prices from 2012 to 2014, the rate at which housing prices increased seems eerily similar to the rate of increase during the bubble's growth. This may have been a mini-bubble that formed, and it appears housing price increases have already begun to slow down as of late 2015. It seems likely that housing prices will be corrected again in the near future; therefore, now is not the optimal time for current renter's to purchase a home.

Home Price Dataset: Technical Appendix

A. Evaluating Beginning and End of the Bubble

Figure 5.2 shows the beginnings of the bubble in further detail. From 1990 to 1997 the index declined or flattened out seasonally from November to December; however, 1998 was the first year of the decade for which the index continued to increase over these months. Figure 5.2 also includes an index of real median household income measured annually. The increases in real median household income keep pace with housing price increases until 1997, after which the housing price increases begin to outpace real income growth. From 1999 to 2001 real median household income actually decreases while housing prices continue to increase more rapidly.

Figure 5.2: Pre-bubble trend and bubble beginning. Real Median Household Income Index, Jan 2000 = 100 (black line)

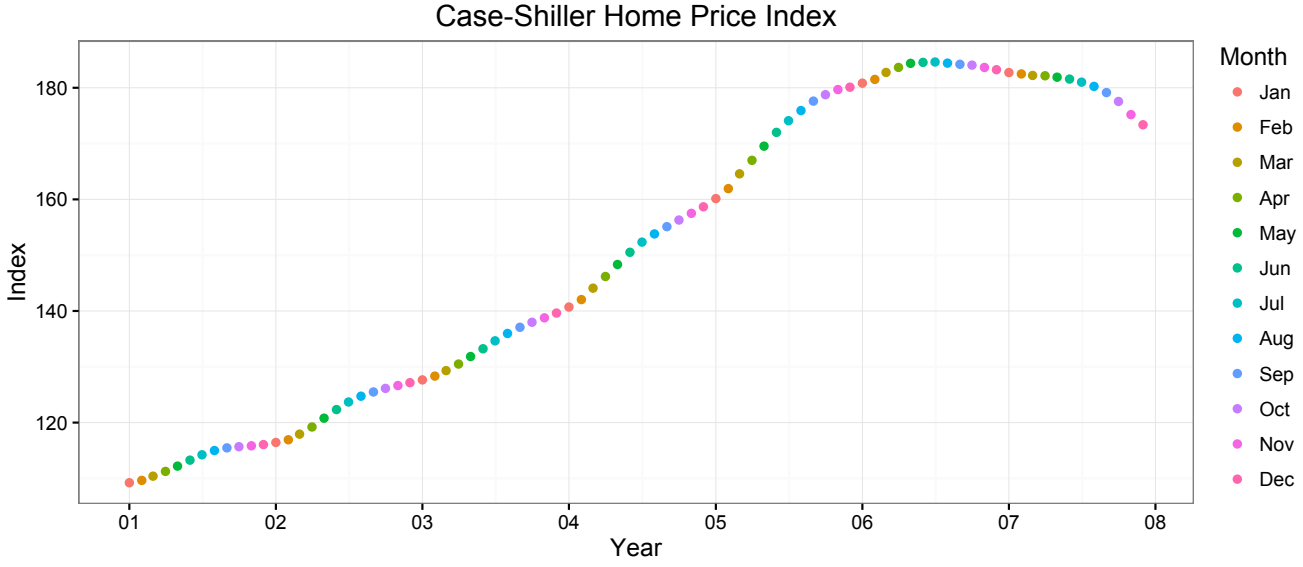

Figure 5.3: Bubble effects on home price index from 2001 to 2008. Effect of bubble bursting in 2005 can be seen in the decreased rate of growth and eventual decline in index value.

B. Forecasting Pre-Bubble Trends

The combination of trend and seasonality within the data suggest two differencing operations are required: 1) difference the data once at lag 12 to remove seasonality, then 2) difference the data once at lag 1 to remove trend. The sample ACF in the top row of Figure 5.2 shows relatively quick decay, which is characteristic of an AR component. The sample PACF in the bottom row of Figure 5.2 has a large spike at lag 1 which may provide further support for an AR component of order 1. The sample PACF also has values just above the significance threshold at lags 3, 12, and 13. The value at lag 3 may not be important, but the values at lag 12 and 13 indicate that a seasonal component may be appropriate.

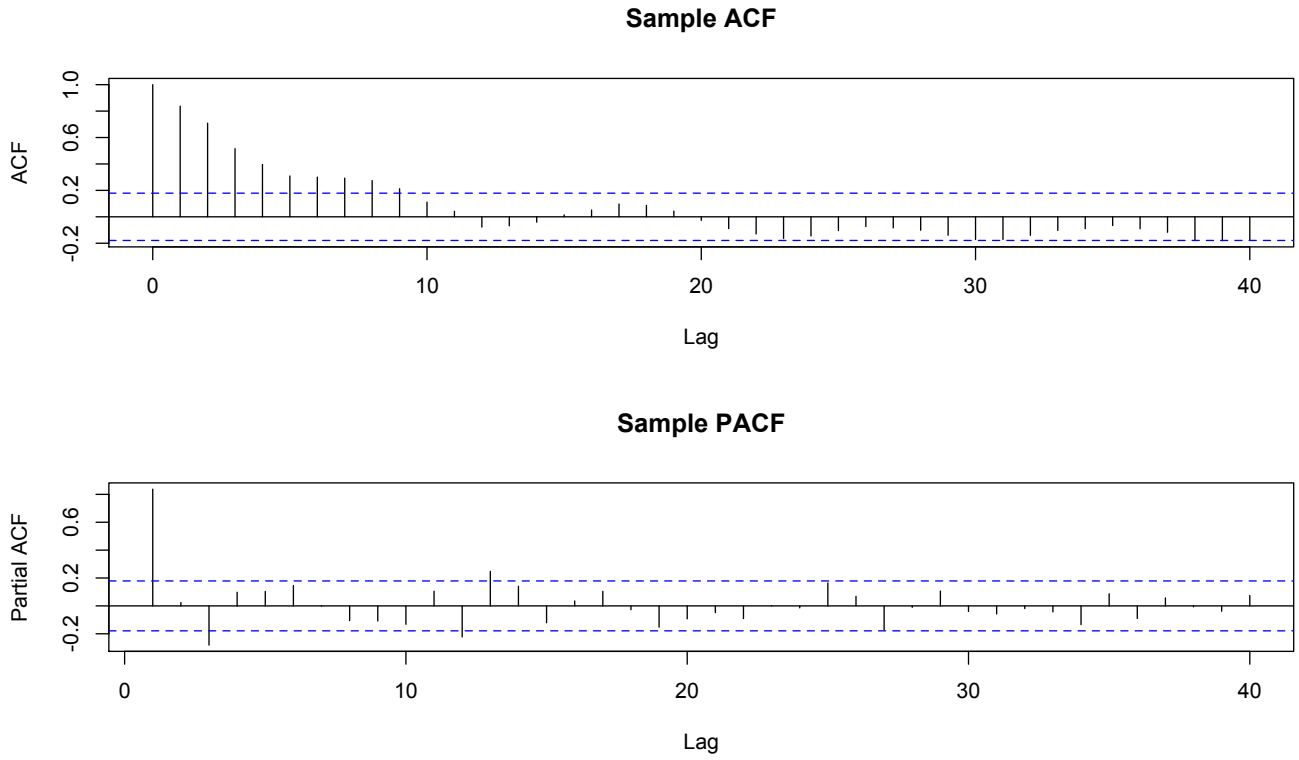

Figure 5.4: Sample ACF and PACF of home price data differenced once at lag 12 and again at lag 1.

The periodogram in Figure 5.5 does not exhibit any definitive peaks or dips, but it does appear to decrease slowly from the left to the right. Thus, the periodogram also seems to provide evidence in support of an order 1 AR component.

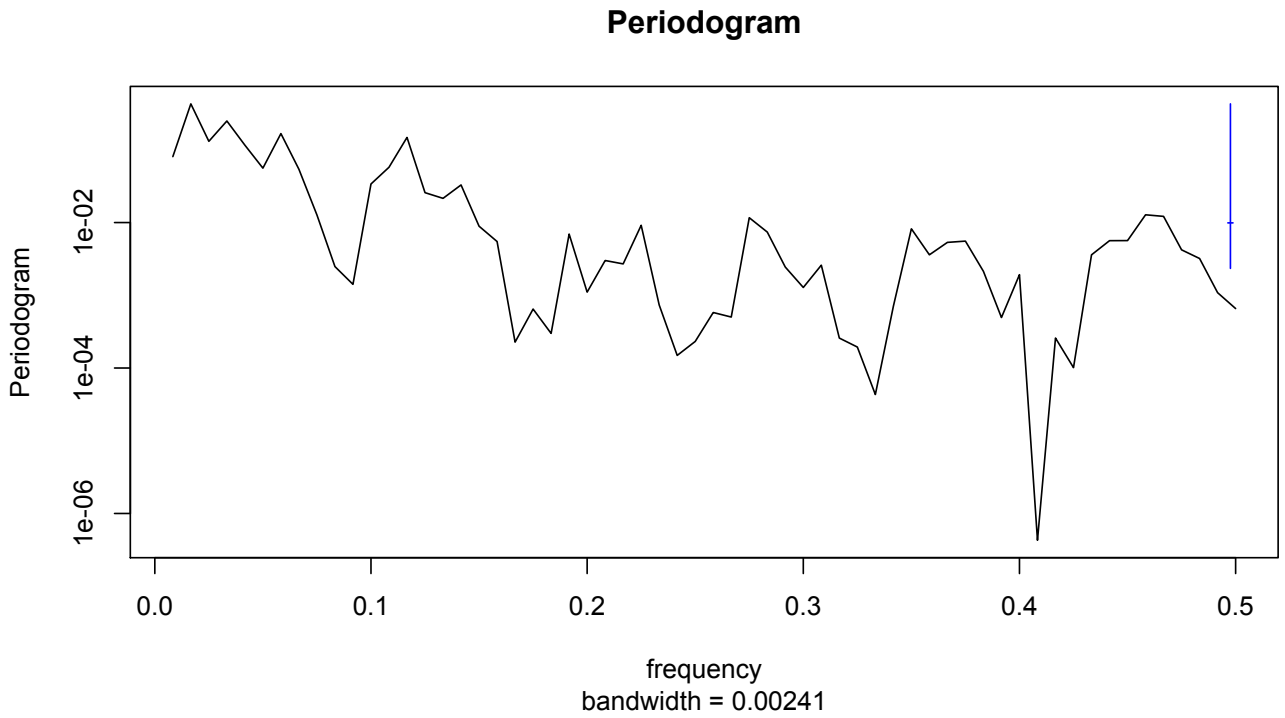

Figure 5.5: Periodogram of home price data differenced once at lag 12 and again at lag 1.

Initially, candidates were considered from SARIMA(p,1,q)$(0,1,0)_{s=12}$ for $0 \leq p \leq 3$ and $0 \leq q \leq 2$, but the residual diagnostics of all the candidates showed a significant residual ACF value at lag 12. As a result, new candidates from SARIMA(p,1,q)$(0,1,1)_{s=12}$ for $0 \leq p \leq 3$ and $0 \leq q \leq 2$ were evaluated. Table 5.1 lists the candidates as well as several evaluation criteria. The candidates were trained on data up to time January 1997 in order to evaluate prediction performance for the year leading up to the beginning of the bubble.

AR	MA	AIC	BIC	$Sigma^2$	Log Likelihood	Adjusted SS Residual	Significant Coefficients
1	2	-230.41273	-217.00208	0.005599858	120.20637	0.005731379	No
3	2	-228.4195	-209.64458	0.005209085	121.20975	0.005556155	No
1	0	-230.26452	-222.21813	0.005647525	118.13226	0.005580849	Yes
2	2	-228.5049	-212.41211	0.005592012	120.25245	0.005826191	No
3	0	-229.41736	-216.00671	0.005635111	119.70868	0.005764797	No
2	0	-228.40764	-217.67911	0.005657096	118.20382	0.005686025	No
1	1	-228.37062	-217.6421	0.005654026	118.18531	0.00568316	No
2	1	-228.30013	-214.88948	0.005652313	119.15007	0.005781308	No
3	1	-228.18722	-212.09443	0.005601627	120.09361	0.005835498	No
0	2	-174.51667	-163.78815	0.010454259	91.25834	0.010153366	No
0	1	-125.84382	-117.79743	0.016954673	65.92191	0.015930885	No
0	0	-58.67455	-53.31029	0.03271328	31.33728	0.029848162	No

Table 5.1: Candidate model evaluation. Greyed-out rows indicate candidate models that were not adequate.

Table 5.1 shows that all candidate models with an at least one AR component performed similarly across all evaluation criteria; however, all models except SARIMA(1,1,0)$(0,1,1)_{s=12}$ contained statistically insignificant coefficients. Therefore, the chose model is SARIMA(1,1,0)$(0,1,1)_{s=12}$. The R output from fitting the model to the data up until January 1998 is included below.

```
Call:
arima(x = hp.98$index, order = c(1, 1, 0), seasonal = list(order = c(0, 1, 1),
    period = 12), include.mean = FALSE, method = "ML")

Coefficients:
         ar1     sma1
      0.9420  -0.8432
s.e.  0.0382   0.1222

sigma^2 estimated as 0.005537:  log likelihood = 133.64,  aic = -261.28
```

The model equation is:

$$X_t = X_{t-12} + 1.9420(X_{t-1} - X_{t-13}) + 0.9420(X_{t-2} - X_{t-14}) + Z_t - 0.8432Z_{t-12}$$

The residual diagnostics for the SARIMA(1,1,0)(0,1,1)$_{s=12}$ model do not show anything alarming. There is one residual ACF value at lag 5 that is nearly significant, but this should not be too concerning since lag 5 is not associated with the seasonality of the data.

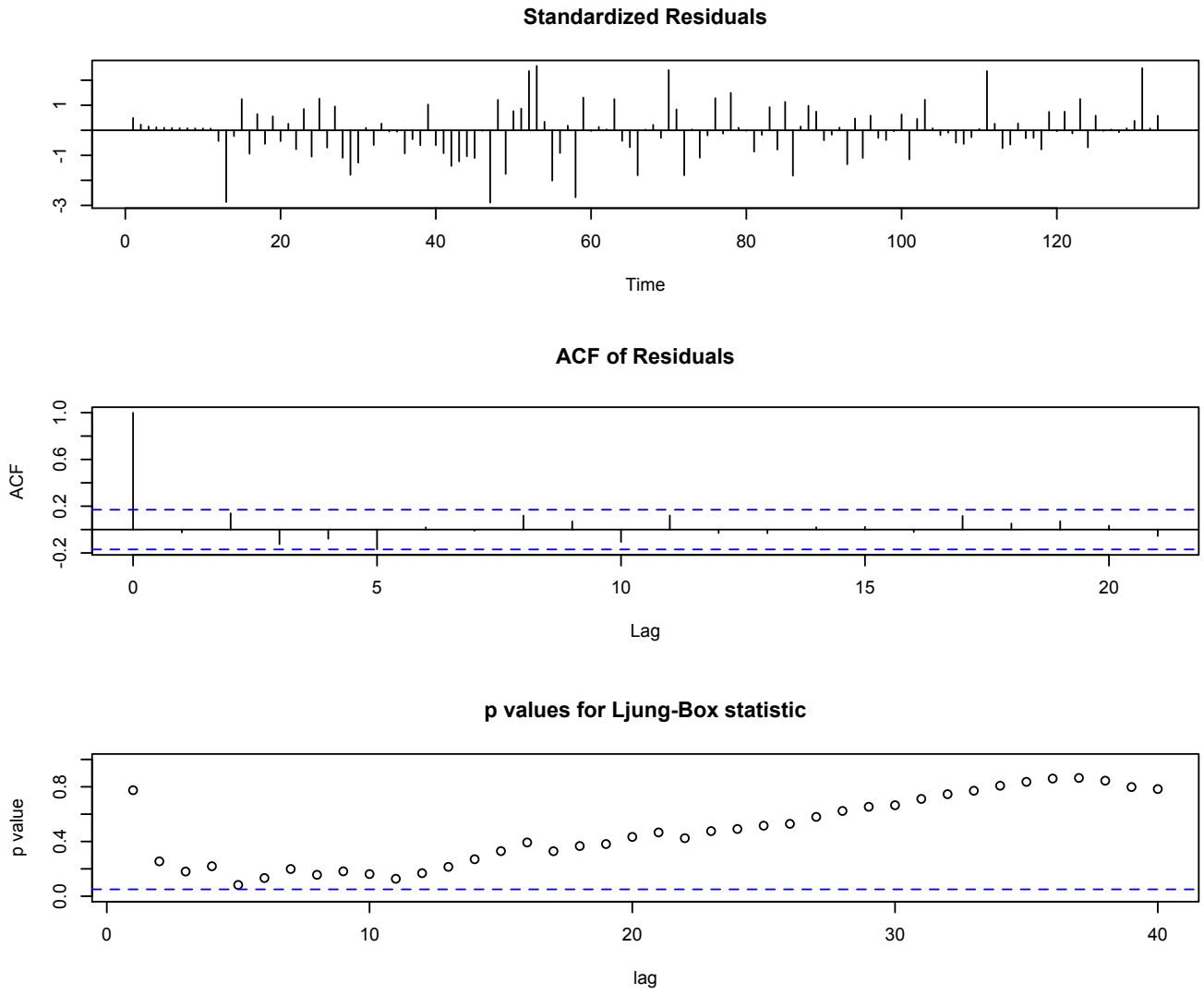

Figure 5.6: Residual diagnostics for SARIMA(1,1,0)(0,1,1)$_{s=12}$ model.

C. Assessing Post-Bubble Effects

Figure 5.7 shows the home price index differenced once at lag 1. This shows how much greater the volatility in month-to-month housing prices is now than before the bubble. Thus, Figure 5.7 provides further evidence that the effects of the bubble are still present in January 2016.

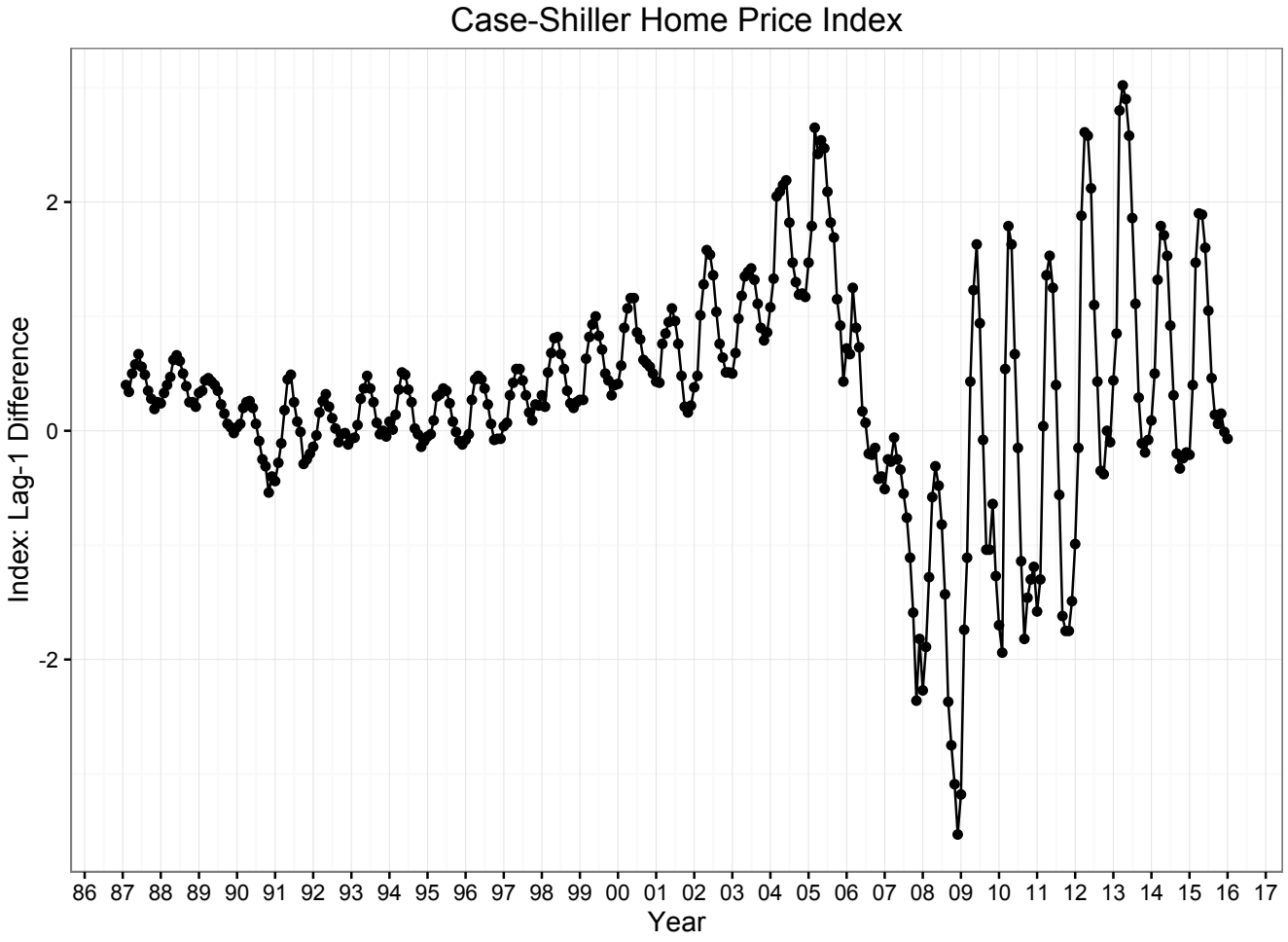

Figure 5.7: Case-Shiller Home Price Index differenced once at lag 1.