Forecasting Sales and Price for Existing Single-Family Homes: A VAR Model with Error Correction

*Zhong-guo Zhou**

***Abstract*. In this paper we forecast demand for existing single-family housing in the United States. We first find that sales volume (sales) and median sales price (price) have unit roots. We then find that sales and price are cointegrated. We develop a vector autoregressive (VAR) model with error correction to further examine the causality between sales and price. We find that there exists a bidirectional causality relationship between sales and price. Price affects sales significantly and sales affects price weakly. With the VAR model we then forecast sales and price for existing single-family housing during the period 1991 to 1994 by using a recursive method. We find that our predictions for sales and price fit the actual data well.**

Introduction

In recent years researchers have begun exploring the relationships between the real estate market and other related markets. For example, Case and Shiller (1990) find that price changes and excess returns of single-family housing can be predicted by a number of information variables. Using the Granger equilibrium model Goebel and Ma (1993) find that mortgage rates and general interest rates are cointegrated after 1980. Schnitzel (1986), on the other hand, finds that deposit rates Granger-cause mortgage rates for Savings and Loans (S&L) during the period of 1970–78. Over the period 1978–84, however, he finds that it is mortgage rates that determine deposit rates. Less work has been done in exploring the fundamental relationship between sales and price for existing single-family homes. It is particularly interesting to investigate this housing market as it represents the biggest portion of home sales in the United States. The VAR model with error correction obtained here can help to analyze and predict the demand for existing single-family homes. Moreover, since residential investment has a timing lagged effect, forecasting the housing demand also appears to be important for policy makers.

In this study, we concentrate on the time-series behavior and relation between sales volume and median sales price. The sales and price data are for the existing single-family houses in the United States. We find that the levels of sales and price have unit roots. That is to say, the two real estate series are not stationary. Their first-order differences are, however, stationary. Further, we find that sales and price are cointegrated. That is, they tend to move together and converge in the long run. Following Engle and Granger (1987), therefore, we construct a VAR model with error correction to examine the Granger causality relationship between sales and price. We find that sales affect price

*Department of Finance, Real Estate, and Insurance, College of Business Administration and Economics, California State University, Northridge, California 91330.
Date Revised—July 1996; Accepted—October 1996.

significantly and price affects sales weakly. In addition, utilizing the VAR model we forecast sales and price of existing single-family homes in the whole nation by using a recursive method. We find that our predictions fit the actual data well. The goodness of fit of the model, measured by R^2s from the regression of the fitted values on the actual values, ranges from 0.77 to 0.86. Therefore, we conclude that the existing single-family housing market is not efficient. Sales volume and median sales price can be well predicted by our model.

The study is organized as follows. Section two describes the data set. The third section discusses the methodology used in this analysis, and section four provides the empirical results and their implications. Finally, section five concludes the paper.

Data Set

The data used in this study include monthly time series of the existing single-family housing market sales volume (sales) and median sales price (price) in the United States from January 1970 to December 1994. The data from January 1970 to December 1990 is used in the VAR modeling procedure and the data from January 1991 to December 1994 is used for testing the model's predictability of sales and price. The sources of housing data are provided by the National Association of Realtors in Washington, D. C.

The two time series are plotted in Appendix 1. The sales series has a strong seasonal pattern and a time trend. In particular, sales begin to increase in February and continue to increase until August. Starting from September, however, sales begin to fall and reach a bottom in January of the next year. This pattern is repeated year after year. In the long run, sales have a tendency to go up. The annual peak of sales typically exceeds that of the previous year, indicating a long-term upward trend. However, a major decline in sales occurred in 1980–81 when mortgage rates reached their peak. The price series indicates a clear time trend, suggesting strong autocorrelation in the time series and the possibility of the existence of nonstationarity. The first-order differences of these two series appear stationary.

Methodology

We first examine the possible existence of unit roots in our time-series data to ensure that the model constructed later is stationary in terms of the variables used. If a time series has a unit root, the first-order difference of the series is stationary and should be used. A series that is stationary after being differenced d times is said to be integrated of order d, or $I(d)$ (Granger, 1981). If two time series are both integrated of order d, a linear combination of these two series may result in a stationary time series, $I(0)$. In that case, we say that the two original series are cointegrated of order d (Granger, 1981). Following the stationarity tests, we then look at the cointegration of sales and price. If these two series are cointegrated, an error-correction term should be added to the modeling process as suggested by Engle and Granger (1987), and Phillips (1991). We further develop a VAR model with error correction terms to examine the Granger causality relationship between sales and price. Based on the VAR model, we forecast sales and price for existing single-family homes using a recursive method. Finally, we compare the forecasted sales and prices with the actual data to determine whether the VAR model with error correction provides a goodness of fit of the model.

Unit Root Test

Consider the autoregressive model:

$$y_t = c + \alpha t + \beta y_{t-1} + \varepsilon_t \, , \tag{1}$$

where we assume that $y_0=0$, β is a real number, and ε_t is a sequence of independent normal random variables with mean zero and a constant variance, σ^2, and $t=1, 2, \ldots, T$. The time-series y_t converges to a stationary time series if the absolute value of β is less than one. If the absolute value of β is one, the time series is not stationary and the variance of y_t is $t\sigma^2$. A time series with $\beta=1$ is said to have a unit root. Nelson and Plosser (1982) suggest that a unit root test should be imposed on most macroeconomic time series before any modeling procedure in order to ensure that the model constructed will be stationary. Granger and Newbold (1974) and Phillips (1986) also point out the serious problems associated with spurious regression models in which unit root time series are involved. To test for the existence of a unit root in a time series, the most unrestricted model by Dickey-Fuller (1979) is typically adopted although alternative tests (such as the Augmented Dickey-Fuller test) can also be used. The Dickey-Fuller test model is:

$$\Delta y_t = c + \alpha t + \beta y_{t-1} + \varepsilon_t \, , \tag{2}$$

where c is a drift term, αt is a time trend, and Δy_t is the first-order difference of the series y_t.

The null hypothesis for the test of the existence of a unit root is H_0: $\beta=0$ versus the alternative H_a: $\beta<0$.[1] In testing H_0, the statistic τ is used. It is defined as:

$$\tau = \frac{\beta - 0}{\mathrm{sd}(\beta)} . \tag{3}$$

Since τ is not distributed as the student's t, the tabulation from simulation provided by Dickey and Fuller (1981) is the correct reference to check for the existence of a unit root. However, the existence of a significant time trend and/or drift term will affect the distribution of τ. Specifically, if there exists a significant time trend and/or drift term, the usual Dickey-Fuller statistic is asymptotically standard normal. The Dickey-Fuller unit root test is performed on sales and price, respectively. The results are reported in the fourth section of this study.

Cointegration and Error Correction

Consider two time series x_t and y_t. Suppose that x_t is $I(1)$ and y_t is also $I(1)$. In general, we can find that a linear combination of x_t and y_t is still $I(1)$. It is, however, possible that a linear combination of two $I(1)$ series may result in a stationary time series of $I(0)$. If such a combination does exist then the two series are said to be cointegrated of order one. There are important implications if two series are cointegrated. As indicated earlier, if two series are cointegrated, there is a tendency for them to move together in the long run. To correctly specify the model with cointegrated variables, an error correction term should be added to the modeling procedure in order to capture the short-run dynamics.

To check for the existence of cointegration between the two $I(1)$ time series, x_t and y_t, we run a regression of x_t on y_t and check if the regression residual is stationary. To be more exact, we run the following regression:

$$y_t = \alpha + \beta x_t + z_t \, , \tag{4}$$

and test if the residual z_t is stationary. In testing the residual z_t we again use the Dickey-Fuller test of (2). If the regression residual is stationary we can conclude that x_t and y_t are cointegrated of order one. In this study, we check the cointegration between sales and price. The importance of checking for cointegration here is that if the two series are cointegrated of order one, then the first-order difference of each series plus a lagged regression residual, the error correction term, should be included in the modeling procedure. The model constructed can thus capture both long-term convergence between these two variables and the short-term dynamics. It is called an error correction model (Engle and Granger, 1987).

Granger Causality

In defining "causality" we follow Granger (1969): x "causes" y if and only if y is better predicted using the past history of x, together with the past history of y itself, than using just the past history of the y variable. Generally, the unidirectional Granger causality test is carried out by using an F-test on the coefficients of the lagged values of x's in the regression of y on its past values and the past values of x. If x and y are cointegrated of order one, then the first-order difference of each series plus an error correction term, the lagged residual from the regression of the level of x on the level of y, should be included in the Granger causality test. In this study we propose a more generalized VAR model with error correction terms to test for Granger causality.

To illustrate, suppose we would like to examine the causality relationship between two time series, x_t and y_t. We examine the following VAR model:

$$\Delta y_t = c1 + \Sigma_{i=1}^{m} \beta_{1i} \, \Delta y_{t\text{-}i} + \Sigma_{i=1}^{n} \delta_{1i} \, \Delta x_{t-i} + \gamma_1 u_{1t-1} + \varepsilon_{1t} \, , \tag{5}$$

$$\Delta x_t = c2 + \Sigma_{i=1}^{p} \beta_{2i} \, \Delta x_{t\text{-}i} + \Sigma_{i=1}^{q} \delta_{2i} \, \Delta y_{t\text{-}i} + \gamma_2 \mu_{2t-1} + \varepsilon_{2t} \, , \tag{6}$$

where Δy_t and Δx_t are first-order differences of y_t and x_t, respectively, provided they are both $I(1)$, and $u1_{t-1}$ and $u2_{t-1}$ are the error correction terms obtained from regressions of x_t on y_t and y_r on x_t, respectively, assuming y_t and x_t are cointegrated, ε_{1t} and ε_{2t} are residuals in the VAR model that may be correlated with each other, and m, n, p, and q are numbers of lags. If y_t and x_t are not cointegrated but have unit roots then the error correction terms should be dropped. If y_t and x_t are not cointegrated and they do not have unit roots then Δy_t and Δx_t should be replaced by the levels of y_t and x_t. In that case, the model collapses to a traditional VAR model.

Our test procedure is as follows. First, we estimate equation (5) using ordinary least squares (OLS) by treating the VAR model as a system of 'seemingly unrelated regression equations' (SURE). By setting $n=0$, we then apply Akaike's Information Criterion to choose the optimal lag m^* in order to minimize Akaike's Final Prediction Error (FPE):

$$FPE(m) = \left(\frac{T+k}{T-k}\right)\left(\frac{SSR(m)}{T}\right), \tag{7}$$

where T is the sample size, $k=m+1$ if series y_t and x_t are not cointegrated (without the error correction term), and $k=m+2$ if y_t and x_t are cointegrated (with the error correction term), while $SSR(m)$ is the sum of the squared residuals given the lag m. By fixing m at its optimal lag m^* we further vary n to find the optimal value of n^* so as to minimize $FPE(m^*, n)$. Now the corresponding parameters in the $FPE(m^*, n)$ form are $k=m^*+n+1$ if y_t and x_t are not cointegrated and $k=m^*+n+2$ if y_t and x_t are cointegrated. This method has additional value in that it provides a double check on Granger causality. If $FPE(m^*, n^*) < FPE(m^*)$ then it suggests that x Granger causes y because the past history of series x helps to predict series y. If $FPE(m^*, n^*) > FPE(m^*)$ it implies that x does not Granger-cause y. After finding the optimal $FPE(m^*, n^*)$, we obtain the residual ε_{1t}. In the same manner we obtain the minimum $FPE(p^*, q^*)$ for equation (6) as well as the residual ε_{2t}. Finally, we test whether the two residuals are correlated or not. If uncorrelated then equations (5) and (6) can be estimated either together or separately; the result should not be significantly different. If the two residuals are correlated then we reestimate the VAR model jointly with the optimal lags m^*, n^*, p^*, q^* found previously, along with the adjustment for the correlation in residuals.

Another alternative test for Granger causality in equation (5) is to examine the following null hypothesis:

$$H_0\colon \delta_{1i}=0, \text{ for } i=1 \text{ to } n.$$

In testing H_0, the standard F-test is used. It is defined as:

$$F = \frac{\dfrac{SSR_R - SSR_U}{n}}{\dfrac{SSR_U}{T-(n+m)}},$$

where SSR_R is the restricted sum of squared residuals and SSR_U is the unrestricted sum of squared residuals, while n is the number of restrictions, and $n+m$ is the number of coefficients estimated (including the coefficient of the error correction term if it is present), and T is the number of observations. If H_0 is rejected for equation (5) we can say that x Granger-causes y. In the same way, if H_0 is rejected for equation (6) we say that y Granger-causes x. If H_0 is rejected for both (5) and (6) we can conclude that there exists a bidirectional causality between x and y. Of course, the conclusions reached here are dependent on the assumption that the residuals in (5) and (6) are uncorrelated. However, if the residuals are correlated the same test procedure is valid, along with an adjustment for the correlation in residuals.[2]

Following the procedures discussed above the Granger causality test with possible error correction terms is applied to sales and price to check the causality relationship between these two variables.

Forecasting Sales and Price with the VAR Model

The VAR model identified above provides us not only with the Granger causality relationship between sales and price but also the opportunity to forecast sales and price in the existing single-family housing market. Therefore, we forecast sales and price for the period of January 1991 to December 1994 using a recursive method. The recursive procedure works as follows. For example, if we would like to forecast sales and price in January 1991 we first estimate equations (5) and (6) to obtain all coefficients, using the data set from January 1970 to December 1990. We then forecast one-month-ahead sales and price (i.e., sales and price in January 1991). As time advances we reestimate (5) and (b), using the data set from January 1970 to January 1991 to forecast the sales and price in February 1991. Unlike most predictions, our forecast is an out-of-sample forecast because we separate the modeling data set from the testing data set. We obtain forty-eight predictions (from January 1991 to December 1994) for sales and price and run regressions of the fitted values on the actual values for both sales and price to test the goodness of fit of our forecasting model.

Results and Implications

In this section, we first report the results of the unit root tests and then provide results of the cointegration test. We then examine the Granger causality relationship between sales and price using the VAR model with error correction terms that were developed in the third section of this study. With the VAR model, we further forecast sales and price for existing single-family homes for the period 1991–94 by the recursive method. Finally, the implications of the results are discussed.

Results from Unit Root Test

First we observe from the plots of sales and price in Appendix 1 that these two time series are nonstationary. The plots indicate that both series move in a certain pattern over time. However, plots of the first-order difference, $y_t - y_{t-1} = \Delta y_t$, indicate stationary behavior. In order to test whether the nonstationarity arises from a type of unit root, we perform the unit root test on sales and price. The Dickey-Fuller test as discussed in section three is applied.[3] The results of the unit root test are summarized in Exhibit 1.

The critical value is 3.09 for a sample size of 250 if a drift term is included in the test. If a time trend is included, the critical value is 2.79 for a sample size of 250. The results in Exhibit 1 cannot reject the hypothesis that sales and price have unit roots. This implies that the first-order differences of sales and price are stationary and should be used in the late modeling and testing procedures.

Results of Cointegration Test and Error Correction Model

Earlier we identified the existence of unit roots for sales and price. Now, we examine whether the two variables are cointegrated in order to set up a VAR model to test for Granger causality and for forecasting. A regression of price on sales is run in order to test the stationarity of the regression residual. The results of the test procedure for cointegration as discussed in detail in section three are provided in Exhibit 2.

Exhibit 1
Results of Unit Root Test (January 1970–December 1990)

Series	τ-Statistic	H_0: Unit Root Exists
Sales	$\tau = -2.51$	H_0 cannot be rejected
Price	$\tau = -2.48$	H_0 cannot be rejected

Exhibit 2
Results of Cointegration Test (January 1970–December 1990)

Stationarity test on the residual from the regression of price on sales	
H_0: Sales and price are cointegrated of order one	
$\tau = -5.16$	H_0 cannot be rejected

From Exhibit 2 we find that the regression residual of price on sales is stationary. That is to say that sales and price are cointegrated of order one, implying that sales and price move together in the long run. In the short run, however, the error correction term captures the dynamics. In sum, in order to estimate the VAR model correctly, we must use the first-order differences of sales and price with the error correction term in order to examine Granger causality between these two variables. Based upon the correctly specified VAR model, we can forecast sales and price in the existing single-family housing market following the recursive method.

Results of Granger Causality Test

In testing for the Granger causality relationship between sales and price we follow the procedure as described previously. To be more exact, in testing the Granger causality between sales and price, we estimate the following pair of equations:

$$\Delta S_t = c1 + \Sigma_{i=1}^{m} \beta_{1i} \Delta S_{t-i} + \Sigma_{i=1}^{n} \delta_{1i} \Delta P_{t-i} + \gamma_1 \mu 1_{t-1} + \varepsilon_{1t}\ , \tag{8}$$

$$\Delta P_t = c2 + \Sigma_{i=1}^{P} \beta_{2i} \Delta P_{t-i} + \Sigma_{i=1}^{q} \delta_{2i} \Delta S_{t-i} + \gamma_2 \mu 2_{t-1} + \varepsilon_{2t}\ , \tag{9}$$

where all variables are defined earlier.

We first estimate equation (8) using an OLS regression by choosing optimal lag m^* to minimize the Final Prediction Error (FPE). We find that the optimal lag m^* is 12. In particular, we find that the first, second, fourth, and especially, the twelfth lag are significant. Sales tend to have strong autocorrelations and the regression coefficients switch signs with lags. It is consistent with previous findings that sales have a strong seasonal pattern. We then fix $m=m^*$ and choose the optimal lag n^* to minimize the $FPE(m^*, n)$. The optimal lag n^* happens to be 1. In the same way, we estimate equation (9) to get optimal lags p^* and q^*. The optimal p^* is also 12. Compared with sales, price tends to have negative short-run memories and positive long-run memories. The first, fourth and eighth lags are significant and negative. However, the eleventh and twelfth lags are significant and positive. The optimal lag q^* is also 1. Both error correction terms capture significant short-term dynamics. Finally, we test whether the residuals from equations (8) and (9) are correlated. If they are correlated, we go back and reestimate equations (8) and (9) jointly using the optimal lags m^*, n^*, p^*, q^*, along with a correction of the correlation of residuals in the model. If these two residuals are uncorrelated, the results from estimating equations (8) and (9) separately are valid.

Equation (8) is used to test if price Granger-causes sales while equation (9) is used to examine if sales Granger-causes price. As indicated earlier, the Final Prediction Error (FPE) provides valuable information about the causality between sales and price. The detailed regression results are reported below, along with the adjusted R^2s, FPEs and Durbin-Watson statistics, using data from January 1970 to December 1990.[4] The t-values are in parentheses.

$$\underset{(-0.01)}{\Delta S_t=-9.028}\underset{(-2.71)}{-0.097\Delta S_{t-1}}\underset{(4.53)}{+0.168\Delta S_{t-2}}\underset{(-4.55)}{-0.179\Delta S_{t-4}}\underset{(19.53)}{+0.781\Delta S_{t-12}} \tag{8.1}$$

$$\bar{R}^2=0.70 \qquad \text{D-W}=2.14 \qquad \text{FPE}=292{,}880{,}000$$

$$\Delta S_t=\underset{(0.16)}{174.96}\underset{(-1.62)}{-0.061\Delta S_{t-1}}\underset{(5.16)}{+0.195\Delta S_{t-2}}\underset{(-3.55)}{-0.145\Delta S_{t-4}}\underset{(18.84)}{+0.758\Delta S_{t-12}}\underset{(-2.77)}{-0.066ul_{t-1}} \tag{8.2}$$

$$\bar{R}^2=0.71 \qquad \text{D-W}=2.15 \qquad \text{FPE}=285{,}720{,}000$$

$$\begin{aligned}\Delta S_t=&\underset{(-0.91)}{-1051}\underset{(-2.35)}{-0.088\Delta S_{t-1}}\underset{(5.19)}{+0.192\Delta S_{t-2}}\underset{(-3.85)}{-0.153\Delta S_{t-4}}+\\ &\underset{(18.67)}{0.739\Delta S_{t-12}}\underset{(-2.66)}{-0.061ul_{t-1}}\underset{(3.57)}{+4.298\Delta P_{t-1}}\end{aligned} \tag{8.3}$$

$$\bar{R}^2=0.73 \qquad \text{D-W}=2.15 \qquad \text{FPE}=273{,}000{,}000$$

$$\Delta P_t=\underset{(3.35)}{239.14}\underset{(-1.83)}{-0.107\Delta P_{t-1}}\underset{(-3.91)}{-0.239\Delta P_{t-4}}\underset{(-2.65)}{-0.172\Delta P_{t-8}}\underset{(5.52)}{+0.327\Delta S_{t-11}}\underset{(5.18)}{+0.351\Delta P_{t-12}} \tag{9.1}$$

$$\bar{R}^2=0.30 \qquad \text{D-W}=2.19 \qquad \text{FPE}=590{,}434$$

$$\begin{aligned}\Delta P_t=&\underset{(3.03)}{219.08}\underset{(-1.92)}{-0.112\Delta P_{t-1}}\underset{(-4.09)}{-0.252\Delta P_{t-4}}\underset{(-2.62)}{-0.169\Delta P_{t-8}}+\\ &\underset{(5.67)}{0.336\Delta P_{t-11}}\underset{(5.33)}{+0.362\Delta P_{t-12}}\underset{(-1.56)}{-0.004u2_{t-1}}\end{aligned} \tag{9.2}$$

$$\bar{R}^2=0.31 \qquad \text{D-W}=2.21 \qquad \text{FPE}=586{,}817$$

$$\begin{aligned}\Delta P_t=&\underset{(3.07)}{222}\underset{(-2.11)}{-0.126\Delta P_{t-1}}\underset{(-4.06)}{-0.250\Delta P_{t-4}}\underset{(-2.61)}{-0.168\Delta P_{t-8}}+\\ &\underset{(5.69)}{0.336\Delta P_{t-11}}\underset{(5.37)}{+0.3651\Delta P_{t-12}}\underset{(-1.34)}{-0.003u2_{t-1}}\underset{(1.55)}{+0.003\Delta S_{t-1}}\end{aligned} \tag{9.3}$$

$$\bar{R}^2=0.31 \qquad \text{D-W}=2.20 \qquad \text{FPE}=586{,}723$$

These results show that sales and price are both autocorrelated and have long memories. The error correction terms capture short-run dynamics by reducing the Final Prediction Errors (FPE) in equations (8.2) and (9.2). The error correction terms turn out to be negative for both cases, consistent with the negative autocorrelation at lag 1 in equations (8.1) and (9.1). The results also indicate the existence of a bidirectional causality relationship between sales and price. To be specific, price affects sales significantly. The causality coefficient for price is 4.298 with a t-value of 3.57 (see equation (8.3)). By adding the first lagged price term in equation (8.3), the FPE is reduced significantly compared with that in equation (8.2). The adjusted R^2s range from 0.70 to 0.73, indicating the goodness of fit of the model. The Durbin-Watson (D-W) statistics indicate that the residuals are well behaved. On the other hand, sales also affects price. However, the effect is not as strong as that from price to sales. The causality coefficient is only 0.003 for sales with a t-value of 1.55 (see equation (9.3)). The FPE is reduced marginally. The adjusted R^2s are around 0.31. It appears that higher price stimulates sales, ceteris paribus, implying, as expected, that the supply curve for the existing single-family home market is upward sloping. A demand and supply model can provide a structural explanation. As demand rises, the market is fed from existing inventory because of timing lag in housing construction, so prices do not rise when sales do. Eventually, there is a bottleneck, and prices and sales start to rise together. Higher sales usually represents increased demand. Higher demand, in turn, drives the price up. In Exhibit 3, we summarize our previous results, along with the traditional F-tests on the causality coefficients.

The results in Exhibit 3 confirm the existence of strong Granger causality from price to sales. The $FPE(m^*, n^*)$ is significantly less than the $FPE(m^*)$ and the associated F-value on the causality coefficient is very significant (F-value=13.5). However, the $FPE(p^*, q^*)$ is only marginally less than the $FPE(p^*)$ and the corresponding F-value for the causality coefficient is 2.76. This indicates that sales weakly Granger-causes price. Note that the results from the FPE criterion are consistent with those from the traditional F-test and can be used to complement one another. Last, we check the correlation between the two residuals from equations (8.3) and (9.3). We find that the correlation is 0.138 and insignificant. Therefore, we conclude that our VAR model with error correction is correctly specified and can be used for forecasting as in the next section.

Forecasting of Sales and Price in the Existing Single-Family Housing Market

The VAR model constructed above is used to forecast sales and price in the existing single-family housing market. This model is superior to a traditional VAR model in that it includes an error correction term that can capture short-run dynamics and therefore

Exhibit 3
Results of Granger Causality Test (January 1970–December 1990)

Dependent Variable	Independent Variable	$FPE(m^*)$ $FPE(P^*)$	$FPE(m^*, n^*)$ $FPE(p^*, q^*)$	Residual Corr.	F-Stat.	$\bar{R}^2$
ΔSales	ΔPrice	285,720,000	273,000,000	0.138	13.5	0.73
ΔPrice	ΔSales	586,817	586,723	0.138	2.78	0.31

improve the forecast power. Since the two transformed series under investigation are stationary to forecast sales and price in January 1991 we first estimate equations (8.3) and (9.3) to obtain all coefficients, using the data set from January 1970 to December 1990. We then reestimate (8.3) and (9.3) using the data set from January 1970 to January 1991 to forecast the sales and price in February 1991. Not only does our model provide an out of the sample forecast, it also takes care of time-varying regression coefficients. In this way, we obtain forty-eight predictions (from January 1991 to December 1994) for sales and price. To check the accuracy of our forecasting model we compare the predicted values with the actual data. We do a fitness test by running the regression of the predicted values on the actual values for both sales and price as follows:

$$Y_t = a + b\hat{Y}_t \,, \tag{10}$$

where Y_t is the actual value and $\hat{Y}_t$ is the predicted value. We test the hypothesis H_0: $b=1$. The regression results are provided below.

$$\underset{}{S_t = \underset{(1.29)}{28511} + \underset{(12.66)}{0.903}\ \hat{S}_t} \tag{10.1}$$

$\bar{R}^2=0.77$ D-W$=2.50$ t-value for $H_0=-1.36$.

$$P_t = \underset{(3.27)}{16672} + \underset{(17.33)}{0.841}\ \hat{P}_t \tag{10.2}$$

$\bar{R}^2=0.86$ D-W$=2.50$ t-value for $H_0=-3.27$.

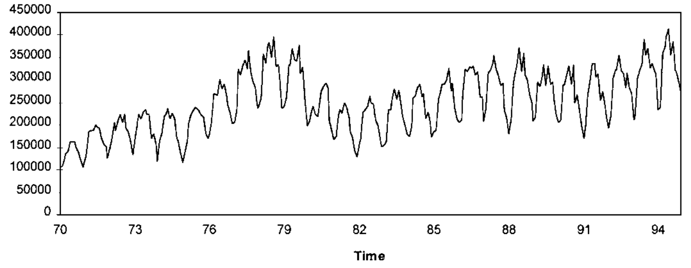
The above results show that the predicted values for both sales and price fit the actual values well with adjusted R^2s of 0.77 and 0.86, respectively, and t-values of -1.36 and -3.27. The null hypothesis cannot be rejected in equation (10.1) at the 95% confidence level. Although the null hypothesis is rejected for equation (10.2), the model still seems to provide an excellent model fit as shown by the R^2. The predicted values and the actual values are plotted in Appendix 2. The plots clearly show that the VAR model we developed with error correction can accurately predict sales and price in the existing single-family housing market. The predicted values, especially the predicted prices, appear to be leading the actual values. They are good forecasts for policy making.

Conclusions

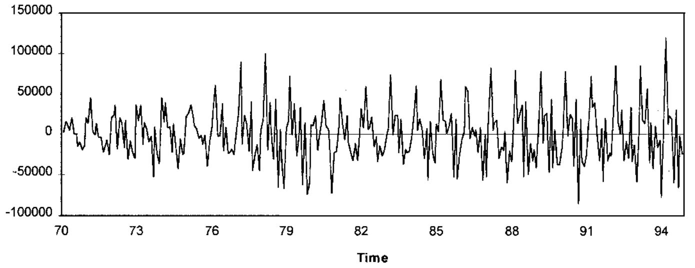
In this paper we examine demand in the existing single-family housing market and the causality relationship between sales volume and median price using a nationwide data set. We find that sales and price have unit roots and are not stationary, but are cointegrated of order one. A VAR model with error correction is developed to examine the causality relationship between sales and price. We find that price significantly Granger-causes sales and sales weakly Granger-causes price. Using the VAR model we then forecast sales and price for existing single-family homes. We find that our VAR model provides a good predictive model as the predictions for sales and price fit the actual data well. Our model is useful for policy makers in planning the residential investments.

Appendix 1
Plots of the Original Time Series and the Transformed Time Series

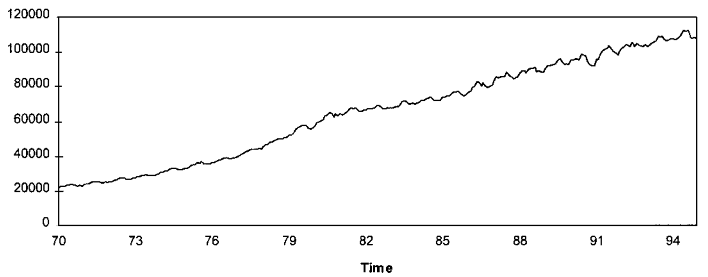
Existing Single-Family Home Sales for the US



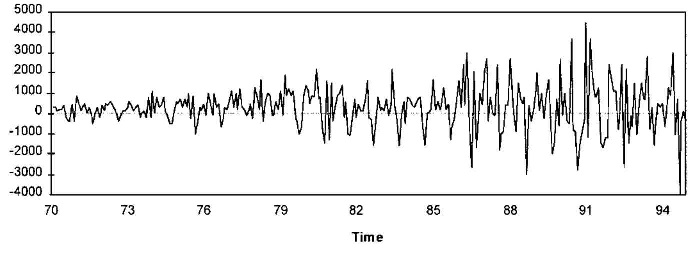
First-Order Difference of Existing Single-Family Home Sales for the US



Median Sales Price of Existing Single-Family Homes for the US

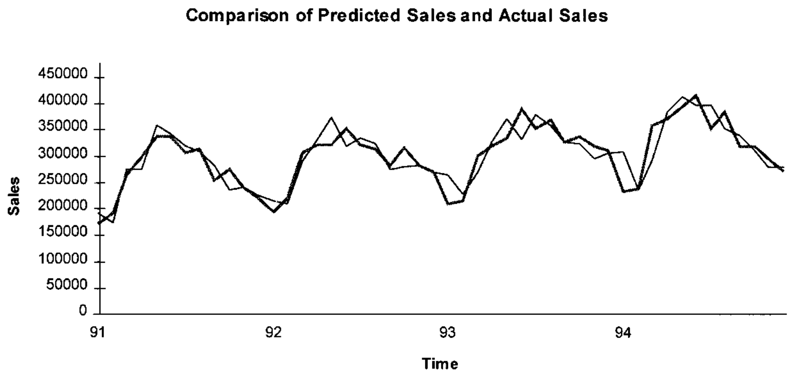


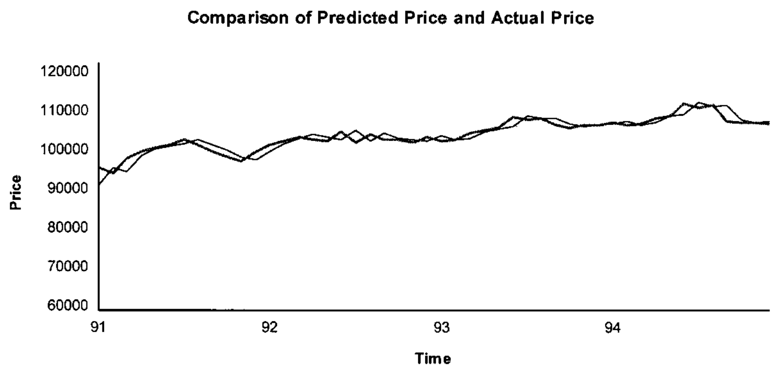
First-Order Difference of Median Sales Price of Existing Single-Family Homes for the US



Appendix 2
Comparison of the Predictions with the Actual Values

The dotted lines below represent predicted sales and price while the solid lines represent actual sales and price for existing single-family homes in the United States for the period of 1990–94.





Notes

[1]If $\beta > 0$ the series will not be stationary. That is why we test $\beta < 0$ for stationarity.
[2]If the two residuals are correlated, we estimate equations (5) and (6) simultaneously with optimal lags m^*, n^*, p^*, and q^*, along with the adjustment for correlated residuals. The F-test is still the correct way to test for causality.
[3]We not only apply the Dickey-Fuller test on sales and price but also other test procedures, for example, the Augmented Dickey-Fuller test. The results are similar. Thus, we report only the results from the general Dickey-Fuller test.
[4]One can argue that in a large structure prices and sales can be both endogenous variables. To address this issue, we add several other explanatory variables, such as the FHA/VA thirty-year mortgage rates and the New York Stock Exchange (NYSE) value-weighted monthly stock returns in our model. We find that although the mortgage rates have a significant negative relation with sales the overall fitness of the model remains almost the same as the model with only sales and

prices. We cannot find a significant relationship between the mortgage rates and prices. Even though the stock market returns seem to affect sales and price, they do not contribute additional predictability to sales and price. Therefore, we only report the results from the VAR model with only sales and price variables. Results from a more generalized VAR model with sales, price, mortgage rates, and stock index returns are available upon request.

References

Case, K. and R. Shiller, Forecasting Prices and Excess Returns in the Housing Market, *AREUEA Journal*, 1990, 18, 253–73.

Dickey, D. A. and W. A. Fuller, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, 1979, 74, 427–31.

——, Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, 1981, 49, 1057–73.

Engle, R. R. and C. W. J. Granger, Co-integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, 1987, 55, 251–76.

Goebel, P. R. and C. K. Ma, The Integration of Mortgage Markets and Capital Markets, *Journal of the American Real Estate and Urban Economics Association*, 1993, 21, 511–38.

Granger, C. W. J., Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, 1969, 37, 424–38.

——, Some Properties of Time Series Data and Their Use in Econometric Model Specification, *Journal of Econometrics*, 1981, 28, 121–30.

—— and P. Newbold, Spurious Regressions in Econometrics, *Journal of Econometrics*, 1974, 2, 111–20.

Nelson, C. R. and C. I. Plosser, Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications, *Journal of Monetary Economics*, 1982, 10, 139–62.

Phillips, P. C. B., Understanding Spurious Regression in Econometrics, *Journal of Econometrics*, 1986, 33, 311–40.

——, 1991, Optimal Inference in Cointegrated Systems, *Econometrica*, 1991, 59, 283–306.

Schnitzel, P., 1986, Do Deposit Rates Cause Mortgage Loan Rates? *AREUEA Journal*, 1986, 14, 448–64.

The author would like to thank Peter Chung, Judy Posnikoff, Mark Tengesdal, Aman Ullah, Ko Wang, and an anonymous referee for helpful comments and suggestions on this project. He would also like to thank Heeyoung Kim for providing him with part of the data set.