FORECASTING TRENDS IN THE HOUSING STOCK USING AGE-SPECIFIC DEMOGRAPHIC PROJECTIONS

Theodore M. Crone
Assistant Vice President and Economist

and

Leonard O. Mills
Senior Economist

Federal Reserve Bank of Philadelphia

February 1990
Revised September 1990

The views expressed here are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Philadelphia or of the Federal Reserve System. We would like to thank Jerry Carlino, Joe Gyourko, Dick Voith and three anonymous referees for comments on earlier drafts.

Abstract

Cointegration tests are used to examine the basic long-term relation between population and the housing stock. There is some weak evidence of a long-run relation between the constant-cost value of the housing stock and population-driven demand. Much stronger evidence exists for a long-term relation between owner-occupied housing units and the adult population. We generally cannot reject that the number of housing units intended for owner-occupancy has adjusted in proportion to the population 25 years of age and older. Using these results and current population projections, we produce trend forecasts through the year 2010 for the owner-occupied housing stock and single-family housing starts in the U.S.

The obvious connection between housing demand and population has not gone unnoticed by economists [see for example Gordon (1956) and Smith (1984)]. Few attempts, however, have been made to examine the empirical relation between the housing stock and the population group most relevant for housing demand. One exception is a recent article by Mankiw and Weil (1989), but they do not find a statistically significant relation between their housing demand variable and the value of the housing stock.[1] The modest goal of this article is to relate the historical trend in a single demographic variable, such as the number of persons over age 20 or over age 25, to the long-term movements in an appropriate housing stock measure. If a historical relation is found, long-term projections of the future housing stock based on simple population projections would seem to be reliable because projections of the population relevant to housing demand are fairly accurate at least 20 years into the future.

This article uses the recently developed time-series technique of cointegration to examine the long-term relations between various age-groups of the population and the housing stock. We find evidence in favor of a reliable long-term relation between housing and the relevant population. Using these results and population projections from the Bureau of the Census, we produce a trend forecast through the year 2010 for the total number of owner-occupied housing units in the U.S. We also produce forecasts for the trend in single-family housing starts from the historical relationship between starts and changes in the relevant population.

---

[1]They did find a statistically significant relationship between real housing prices and their measure of the population-driven demand for housing. On the basis of this relationship they forecast a 47 percent decline in the real price of housing between now and the year 2007. This result has been well publicized. See *The Wall Street Journal*, January 19, 1990, p. A10 for example.

I. THE BASIC RELATION

Since by definition a household is a group of related or unrelated individuals who occupy a housing unit, the number of housing units in any period t is determined by the following identity

$$S_t = HH_t + V_t \tag{1}$$

where

S_t = total housing stock in period t,

HH_t = total number of households in period t,

V_t = the number of vacancies in period t (defined as unoccupied units or units used as second residences).

We can express the identity in equation (1) slightly differently by introducing the vacancy rate, $VR_t = V_t/S_t$, so that

$$S_t = a_t \cdot HH_t \tag{1'}$$

where

$a_t = 1/(1-VR_t)$.

We can further refine the basic identity by noting that the total number of households is equal to the household headship rate times the population, or

$$HH_t = h_t \cdot P_t \tag{2}$$

where

h_t = the proportion of the population who are heads of households and

P_t = total population.

Substituting (2) into (1') and taking logs, we have

$$\ln S_t = \ln a_t h_t + \ln P_t \tag{3}$$

In this representation, $\ln a_t h_t$ is the gap between the log of the housing stock and the log of the population. As the t subscripts indicate, the vacancy and headship rates can vary over time. Both are likely to vary with the business cycle, but these cyclical fluctuations are of no concern for our purposes. Long-term changes in the vacancy rate and/or headship rate, however, would drive a long-term wedge between population and the housing stock. While the vacancy rate, and therefore a_t, has changed from year to year, there has been no discernable trend. The ratio of total vacancies (including second residences) to total housing units was about 13 percent in the mid-1960s, dropped to about 10 percent in the late 1970s, and rose again to about 14 percent in the late 1980s. Headship rates have also varied over time, but also without a pronounced trend. Headship rates for most age groups rose in the 1960s and 1970s but have declined since the late 1970s. The changing age distribution of the population has resulted in a general increase in the headship rate for all those 15 years of age and older, but the headship rate (h_t) for persons 25 and over has changed little in the past decade.[2]

Letting k equal the mean of $\ln a_t h_t$ and letting lower case s and p represent the logs of the respective variables, the basic relationship in (3)

---

[2]Some housing forecasts have applied age-specific headship rates to population projections to forecast net household formations. See Patric H. Hendershott (1988). The success of these forecasts depends on the accuracy of the headship rate forecasts.

can be written as

$$s_t = k + p_t + u_t \tag{4}$$

where u_t represents deviations from the mean value of ln $a_t h_t$.

This article addresses the question of whether the deviations u_t are permanent or whether they eventually die out. If the deviations do not eventually fade away, long-term projections of the housing stock should not be made solely on the basis of population. In contrast, if these deviations do fade away, such long-term projections seem valid. As discussed below, the recently developed tests for cointegrated time series are designed to answer this type of question.

There are many important questions about the housing market that cointegration tests are not designed to answer. While these tests help isolate trends in housing, they cannot by themselves describe business cycle fluctuations in housing demand or the short-term dynamics of the housing market. Moreover, as applied in this article, the tests tell us little about the relation of housing and other long-term economic variables such as permanent income.

II. USING COINTEGRATION TO EXAMINE LONG-RUN RELATIONS

We now formally present the concept of cointegration as applied to the relation between the housing stock and population. An individual time series may have a stochastic trend component which implies that the trend changes randomly each period rendering the level of the series nonstationary. If the series must be differenced only once to render it stationary, the series is

said to be integrated of order one, denoted I(1). Equivalently, the series is said to have a single unit root. Nelson and Plosser (1982) and others have shown that most economic time series appear to contain a unit root implying that they need to be differenced to induce stationarity.

Engle and Granger (1987) point out, however, that it is possible that a linear combination of two I(1) series does not contain a stochastic trend. The two nonstationary series are then said to be cointegrated. Although each of the series appears to wander aimlessly because each contains a stochastic trend, they eventually follow one another if they are cointegrated. In other words, the stochastic trend in each series is common to both, and there is a single source of the unit root behavior.

Consider a general bivariate autoregressive representation for the log of the housing stock, s_t, and the log of the population, p_t, for $t = 1,\ldots T$.

$$s_t - \alpha p_t = u_{1t} \; ; \; (1 - \rho_1 L)(1 - \phi_1 L)u_{1t} = \mu_1 + \epsilon_{1t}, \; \rho_1 \leq 1 \text{ and } \phi_1 < 1 \qquad (5)$$

$$-\beta s_t + p_t = u_{2t} \; ; \; (1 - \rho_2 L)(1 - \phi_2 L)u_{2t} = \mu_2 + \epsilon_{2t}, \; \rho_2 \leq 1 \text{ and } \phi_2 < 1 \qquad (6)$$

where u_{10} and u_{20} are fixed and ϵ_{1t} and ϵ_{2t} are mutually uncorrelated, i.i.d. processes.[3] The processes u_{1t} and u_{2t} are said to have unit roots if $\rho_1 = 1$ and $\rho_2 = 1$ respectively.

An implication of the presence of a unit root is that an innovation in the process will permanently alter the level of that process. For example, suppose $\phi_1 = 0$ and $\rho_1 = 1$, then by recursive substitution,

[3]This illustration considers only a second order autoregressive process. Extensions to higher order processes, as in the quarterly results presented below, are straightforward.

$$u_{1t} = u_{10} + \mu_1 t + \Sigma_{i=1}^{t} \epsilon_{1t}$$

As seen by this equation, the effects of the first innovation, ϵ_{11}, as well as all subsequent innovations, never disappear when $\rho_1 = 1$. Consequently, u_{1t} can deviate permanently from its deterministic trend line, $u_{10} + \mu_1 t$. Thus, the trend effectively changes each period, i.e. the trend is stochastic, and u_{1t} is nonstationary. In contrast, if u_{1t} does not have a unit root, i.e. $\rho_1 < 1$, then

$$u_{1t} = \mu_1/(1-\rho_1) + [u_{10} - \mu_1/(1-\rho_1)]\rho_1^t + \Sigma_{i=1}^{t} \rho_1^{t-i}\epsilon_{1t}.$$

As t increases the effect of the first innovation, $\rho_1^{t-1}\epsilon_{11}$, gradually disappears because $\rho_1 < 1$. In this case, u_{1t} will eventually return to its deterministic path, i.e., u_{1t} is mean-reverting.

The presence of a unit root in s_t and p_t depends on the presence of unit roots in both u_{1t} and u_{2t}. This can be seen by examining the reduced form for s_t and p_t. For example,

$$s_t = (1 - \alpha\beta)^{-1}(u_{1t} + \alpha u_{2t}) \ .$$

If both $\rho_1 < 1$ and $\rho_2 < 1$, neither u_{1t} nor u_{2t} contains a unit root, and consequently there is no unit root in s_t. However, if either $\rho_1 = 1$ or $\rho_2 = 1$, then s_t will contain a unit root. Similarly, p_t will contain a unit root if either $\rho_1 = 1$ or $\rho_2 = 1$.

The basic equation relating the housing stock to population postulates a

cointegrating relation. The first equation in the bivariate system, equation (5), reduces to equation (4) under the assumption $\alpha = 1$. Now suppose $\rho_1 < 1$ and $\rho_2 = 1$. Because $\rho_2 = 1$, both s_t and p_t individually contain a unit root. However, the linear combination of the two given by $s_t - p_t = u_{1t}$ does not contain a unit root and is stationary. Thus, although both s_t and p_t are subject to permanent shocks, any difference between the two is temporary. When there is a permanent increase in the population, the housing stock permanently increases by a proportionate amount. In this case, the two series share a single, common unit root with the cointegrating vector being [1 -1].

In contrast, if $\rho_1 = 1$, then any deviation between the logs of the housing stock and population, $u_{1t} = s_t - p_t$, has a permanent component. That is, permanent gaps between the housing stock and population exist, and thus the two series would not be cointegrated. In our model, permanent changes in the vacancy and headship rates would lead to a permanent shift in the relation between the housing stock and population. In this case, there would be two unit roots in the bivariate system, and there would not be a cointegrating vector.

A full information maximum-likelihood procedure to test for cointegration and estimate the cointegrating vectors has recently been developed in a series of papers by Johansen (1988 and 1990) and Johansen and Juselius (1990). Here, we briefly describe the Johansen test procedure with the details and some of its advantages discussed in the Technical Appendix. The Johansen procedure is maximum likelihood, but under certain assumptions the procedure involves a series of ordinary least squares regressions. From these least squares regressions, two likelihood ratio test statistics for the number of cointegrating vectors in the bivariate system, which equals 2 minus

the number of unit roots, can be computed. The first statistic, called the trace statistic, tests whether the number of cointegrating vectors is a given number or less. The second statistic, called the maximum eigenvalue statistic, tests whether the number of cointegrating vectors is r under the maintained hypothesis that there are r+1 or fewer cointegrating vectors. The asymptotic distributions of these test statistics are found in Johansen (1990), and are not the usual χ^2 distributions. Johansen and Juselius (1990), however, provide simulated distributions. One advantage of the Johansen procedure over the more popular Engle and Granger (1987) two-step procedure is that the Johansen procedure allows hypothesis testing on the estimated cointegrating vector. In our application, we test whether the housing stock is proportional to the population as predicted by the basic relation.

III. APPLYING COINTEGRATION TESTS TO THE MANKIW AND WEIL DATA

Mankiw and Weil (1989) estimate a variant of equation (4). In their article, the housing stock variable is the value of all publicly and privately owned residential capital at constant cost, VSTOCK (mnemonics are ours). Corresponding to this constant-value-based measure of the housing stock, Mankiw and Weil develop a population-driven demand variable. They cross-sectionally estimate the demand for housing for every age group using 1970 Census data. Then they apply these estimates to the population profile in each year to obtain a time series for a population-driven housing demand variable, HD.

Using annual data, 1947-1985, and correcting for first-order serial correlation, Mankiw and Weil estimate the following regression:

$$\ln VSTOCK_t = \underset{(7.81)}{8.01} + \underset{(.0366)}{.0095}\ time + \underset{(.652)}{.01}\ \ln HD_t + \epsilon_t - \underset{(.035)}{.97}\epsilon_{t-1}$$

(standard errors)

The large standard error for the coefficient on $\ln HD_t$ leads to inconclusive results. They "cannot find a relation between [their] demographically driven housing demand variable and the stock of housing." Nor can they say that "there is no relation."

One possible reason for these results is the correction for serially correlated residuals. Specifically, their estimated first-order correlation coefficient is so close to one that the equation was effectively estimated in first-differenced form. By differencing the data, the long-run movements in the series are eliminated. Thus, the short-term movements in the housing stock and population are given primary emphasis in the Mankiw and Weil results. Because both of these stocks are essentially fixed for short horizons, it is not surprising that no significant relation appears in their regression. If a long-term relation is present, it may not be revealed by estimations using first-differenced data.

Cointegration tests preserve any long-term relation between two series because they do not require that the series be differenced. We begin by testing for cointegration of the same series used by Mankiw and Weil, $VSTOCK_t$ and HD_t. Specifically, we test for the number of cointegrating vectors in the bivariate error-correction representation of these two series (see Technical Appendix). The results are reported in the first two rows of Table 1. Using a .10 level of significance, the tests produce mixed results. The trace statistic does not reject the hypothesis that there are no cointegrating

vectors (second row). The maximum eigenvalue statistic, however, does reject that hypothesis at the 10 percent significance level. We conclude, therefore, that there is some weak evidence that the value of the housing stock does change with this demographically determined demand variable over the long-term.

The estimated coefficient on the demographic demand variable is 1.64. While this estimate is not significantly different from one at commonly accepted significance levels, the point estimate is much larger than one. One possible explanation for the large coefficient is that the HD_t series is computed from coefficients on the value of housing demanded by each age group in 1970. HD_t, therefore, does not capture the effect of any increase over time in permanent income on the value of housing demand. Thus HD_t may understate (overstate) housing demand after (before) 1970.[4]

We also tested for the cointegration of two other stock variables with the Mankiw-Weil demand variable -- the value of all non-farm privately owned housing stock (VOWNREN)[5] and the value of the non-farm owner-occupied stock (VOWN). The results as shown in Table 1 are similar to the results for the value of total residential capital (VSTOCK).

In a second set of regressions we tested for cointegration of the constant-cost-value measures of the housing stock with those population groups which have the largest influence on the number of households in the nation.

---

[4]In fact, Mankiw and Weil show that there was a large upward shift in the value of housing demand when the coefficients generating housing demand are based on the 1980 census rather than the 1970 census. As Mankiw and Weil point out, this shift may be at least partly attributable to permanent income growth.

[5]This includes both privately owned rental housing and owner occupied housing. Unlike VSTOCK, this series does not include residential farm property or government-owned housing.

The results are reported in Table 2. Mankiw and Weil indicate that the change in their population-driven housing demand variable is highly correlated with the growth of the population over 21. Therefore, we tested for the cointegration of the measures of the value of the residential housing stock and the population 20 and over (POP20).[6] For the value of the total housing stock (VSTOCK) and the value of all privately owned non-farm housing stock (VOWNREN) there is no evidence for cointegration with this population measure.

Different results are obtained for the value of owner-occupied housing (VOWN) and the population 20 and over. We reject the hypothesis that there are no cointegrating vectors at the .10 significance level using the trace statistic and at the .05 significance level using the maximum eigenvalue statistic. Moreover, the coefficient on the population variable is close to one, and we cannot reject the hypothesis that it is one. Evidence for cointegration is also found between the population 25 and over (POP25) and the value of non-farm owner-occupied housing (VOWN).[7] The estimated coefficient on the demographic variable in this case was very low, but again we cannot reject the hypothesis that it is one.

IV. HOUSING UNITS AND POPULATION

The usual measure of the housing stock in equations (1) through (4) is the total number of housing units (HSTOCK) rather than the value of the housing stock, and the number of units is more likely to be related to a simple population variable. This basic relation focuses on the fact that

---

[6]This age cohort is useful for the results below.

[7]Some weak evidence was found for cointegration between the population 25 and over the more inclusive housing stock variables (VSTOCK and VOWNREN).

households must live somewhere without considering the size or quality of their residence. The value of the housing stock increases with larger or higher quality housing units which can depend on such economic variables as permanent income. Thus, the long-run relationship between housing and population is not as simple in value units as it is in physical units.

In terms of the total population, only those who become heads of households effectively increase the demand for housing units, and for all practical purposes the headship rate of those below 20 is zero. Therefore, the relevant population to consider is age 20 and over (POP20). Because few householders become homeowners before age 25, we consider separately the group 25 and over (POP25). The homeownership rate for households headed by persons less than 25 years old is currently only about 16 percent. It rises to more than 35 percent for those 25 to 29 and is more than 66 percent for all those over 25. We obtained a quarterly series of total housing units from the Bureau of the Census from 1965:I through 1989:II. Using annual Bureau of the Census estimates of the population by age and interpolating between quarters, we tested for cointegration in quarterly data of housing units and population.[8]

The results in Table 3 indicate that HSTOCK and POP20 are not cointegrated. Nor is this housing stock variable cointegrated with POP25. One possible explanation for the lack of cointegration is that the proportion of housing units which are vacant but not for sale or for rent (e.g. second residences) has been changing over time. We excluded these types of units

[8]We interpolate a quarterly population series by assigning one-fourth of the annual population growth to each quarter within the year. Because we are now examining quarterly data, we included four lags of the first differences in the error-correction representation to correct for serial correlation.

from our stock variable and included only those housing units occupied by owners or renters or for sale or rent (HOWNREN). Even though the difference between these housing stock measures is small, HOWNREN is cointegrated with POP20 and POP25. The relationships, however, do not seem to be proportional as hypothesized. In both cases we reject the hypothesis that the coefficient on the population variable is equal to one.

Our housing stock variable (HOWNREN) still contained the volatile rental market where vacancy rates have varied from 5 percent to 8 percent over the last 20 years. The rental housing market is also important for people in transition, e.g. those recently divorced. Therefore, we confined our housing stock measure to the more stable owner-occupied market which we define as housing units which are owner-occupied or for sale (HOWN).[9] We tested for the cointegration of the population variables with this measure of the housing stock thus eliminating any effect of long-term shifts affecting the rental market. In the case of owner-occupied housing and the population 20 and over, the tests are anomalous because they suggest that neither series contains a unit root, i.e., rejections are in both rows.

Our best results were obtained in tests of cointegration between HOWN and POP25. At the 5 percent significance level, both the tests indicated cointegration between the these two series. Moreover, we cannot reject the hypothesis that the relationship is proportional, that is, that the coefficient on the population variable is equal to one. The constant of proportionality between units intended for owner occupancy and population over

[9]We include vacant units for sale on the assumption that the great majority of them are intended for owner occupancy. Most rental units for sale are not vacant.

25 is simply the mean value of HOWN/POP25 over the sample, .3817. The proportionality indicates that when the baby-boom entered its first-time house-buying years the number of owner-occupied housing units eventually increased by the same percentage as the population.

This paper has concentrated on the long-run relationship between the housing stock and the adult population. Similar results on the cointegration of population and housing are reported by Holland (1990) using the Engle and Granger (1987) two-step tests. In addition, Holland found no long-term relationship between population growth and the price of housing. This is consistent with earlier empirical evidence that the long-run supply of housing is perfectly elastic (see Follain (1979)).

The long-run relation between the housing stock and population tells us nothing about the short-run dynamics governing the housing market. We can begin to look at these dynamics by examining the error-corrections equation for the housing stock. The ordinary least squares estimate of this equation is:

$$\begin{aligned}\Delta HOWN_t = & -\underset{(.0088)}{.0095} - \underset{(.10)}{.37}*\Delta HOWN_{t-1} - \underset{(.11)}{.21}*\Delta HOWN_{t-2} + \underset{(.11)}{.046}*\Delta HOWN_{t-3} + \underset{(.10)}{.28}*\Delta HOWN_{t-4} \\ & + \underset{(.62)}{1.56}*\Delta POP25_{t-1} - \underset{(.79)}{1.31}*\Delta POP25_{t-2} + \underset{(.80)}{.67}*\Delta POP25_{t-3} + \underset{(.63)}{.013}*\Delta POP25_{t-4} \\ & - \underset{(.056)}{.12}*\{HOWN_{t-1} - [\ln(.3817) + POP25_{t-1}]\}\end{aligned}$$

(standard errors)

The last term in this regression is the error-corrections term, and it implies that 12 percent of the "gap" between HOWN and [ln(.3817) + POP25] will be

closed in the next quarter, ceteris paribus. Thus, 88 percent of any gap between the housing stock and the (adjusted) population remains after one quarter. The remaining gap in subsequent quarters depends upon the coefficients on the autoregressive terms as well as the coefficient on the error-corrections term. For example, two quarters after a cyclical gap opens,

$$\Delta HOWN = (-.37)*(-.12) + (-.12)*(.88) = -.06$$

which leaves 82 percent (88 percent minus 6 percent) of the gap remaining. Continuing in this manner we can trace out the dynamics of the adjustment in the housing stock to any cyclical gap. Some results are shown in Figure 1 which plots the percent of a disturbance to the gap that remains over time. For example, about 30 percent of a gap is eliminated after one year so that about 70 percent remains. Similarly, less than one-half of the gap remains after two years and less than 1/3 of the gap remains after 3 years. Thus, the adjustment to a gap may be considered fairly slow, and the relative price of housing may be affected during this adjustment process.

The R-square from the error-corrections regression is about .31 implying that the majority of the short-run fluctuations in housing are not explained by the regression. In particular, the regression does not include other factors such as interest rates and income growth which are also important in explaining the dynamics of housing. Moreover, these other factors may also affect the speed of adjustment to a gap. Nevertheless, the cointegration and error-correction results presented here suggest that the trend in the housing stock can be successfully modelled by a simple population variable.

V. LONG-TERM FORECASTS

On the strength of our results from these cointegration tests we have plotted in Figure 2 the predicted housing stock intended for owner occupancy through the year 2010. The predicted line is computed as .3817*POP25 where the future values of POP25 are from the Bureau of the Census projections. On the same graph we have plotted the actual housing stock intended for owner occupancy (HOWN) through the second quarter 1989. There is some difference between the two series over the historical 25 year period, but the actual stock does return to the predicted trend line as indicated by the formal cointegration tests. It is obvious from Figure 2 that in the 1990s the trend growth of housing units intended for owner occupancy is predicted to slow.

In the next step in our analysis, we examined the usefulness of our simple demographic variable, POP25, for predicting single-family housing starts. We use single-family starts because about 83 percent of occupied single-family units in 1985 were owner-occupied, and few single-family units are built for rental purposes.[10] We regressed the four-quarter moving average of single-family starts on the year-over-year change in the population over age 25.[11]

$$SFS = \underset{(69.64)}{764.67} + \underset{(.080)}{.31}*(.3817*[POP25_t - POP25_{t-4}])$$

(standard errors)

---

[10]Single-family starts are an approximation of the gross additions to the owner occupancy housing stock but not net additions because there are always some removals.

[11]POP25 refers to the level of population over age 25 in this regression, rather than the log-level in the previous regressions.

where SFS is the four-quarter moving-average of single-family starts. The R-square from this regression is about .13 indicating that, as with growth in the housing stock, a large amount of quarterly variation in single-family starts is unexplained by this simple demographic variable. This lack of explanatory power is not surprising since cyclical influences such as interest rates and income have been omitted.

Our cointegration results suggest that the coefficient on the adjusted change in population should be unity instead of the estimated .31. The t-statistic on the hypothesis that this coefficient equals one is -8.62 indicating that the unity restriction should be rejected. If we ignore this statistical rejection and constrain the coefficient to be one, we give greater weight to the population variable and we have the following estimated relation:

$$SFS = \underset{(25.68)}{187.96} + 1.00*(.3817*[POP25_t - POP25_{t-4}])$$

(standard errors)

In figure 3 we have plotted the four-quarter moving-average of single-family starts as well as the predicted values from both the unrestricted and restricted regressions. As expected, single-family starts are more variable than the predictions based on population, but starts do seem to follow the trend in population. The most notable gap occurs during the back to back recessions in 1980 to 1982. It should also be noted that this period follows the late 1970s when the single family starts were as high as 1.4 million units annually, far above the predicted change. As with the housing stock, we can use this regression and the Census projections for the population over 25 to

forecast the future level of single-family starts. The unrestricted regression predicts that single-family starts will fall to an annual rate around 900 thousand by the end of the 1990s. The restricted regression, which gives more weight to population, predicts a bigger decline to around 700 thousand units by the end of the decade.

VI. CONCLUSIONS

The tests presented in this paper provide evidence that over the long term population-driven demand translates directly into total housing stock or at least the portion of the housing stock intended for owner-occupancy. A reexamination of Mankiw and Weil's results using cointegration techniques provided some weak evidence that the value of the housing stock does follow their computed housing demand series over the long term. Stronger evidence was found for a long-term relation between value measures of the owner-occupied housing stock and simple population measures such as the population 20 and over or 25 and over. The best evidence for cointegration was found for the number of housing units intended for owner-occupancy and the population 25 and older. Over the long run, changes in these two series seem to have been proportional.

Applying the historic relation of owner-occupied housing units to the population 25 and older to current population projections results in a prediction of a sharp slowdown in the growth of the owner-occupied housing stock in the 1990s. A similar but less sharp decline in single-family housing starts is indicated by projecting into the next decade the historic relation between such housing starts and the change in the population 25 and older.

REFERENCES

Engle, Robert F. and C.W.J. Granger, "Co-integration and Error Correction: Representation, Estimation and Testing," Econometrica 55 (1987) 251-276.

Follain, James R. "The Price Elasticity of the Long-Run Supply of New Housing Construction," Land Economics 55 (1979) 190-199.

Gordon, Robert A., "Population Growth, Housing, and the Capital Coefficient," American Economic Review 46 (1956) 307-322.

Hendershott, Patric H. "Household Formation and Homeownership: Impacts of Demographic, Sociological, and Economic Factors," Housing Finance Review 7 (1988) 201-224.

Holland, A. Steven, "The Baby Boom and the Housing Market: Another Look at the Evidence," manuscript, University of Kentucky (1990).

Johansen, Soren, "Statistical Analysis of Cointegration Vectors," Journal of Economic Dynamics and Control 12 (1988) 231-254.

_____, "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," Econometrica, forthcoming (1990).

_____, and Katarina Juselius, "Maximum Likelihood Estimation and Independence on Cointegration - With Applications to the Demand for Money," Oxford Bulletin of Economics and Statistics 52 (1990) 169-210.

Mankiw, Gregory N. and David N. Weil, "The Baby Boom, The Baby Bust, and the Housing Market," Journal of Regional Science and Urban Economics 19 (1989) 235-258.

Nelson, Charles R. and Charles I. Plosser, "Trends and Random Walks in Macroeconomic Time Series," Journal of Monetary Economics 10 (1982) 129-162.

Poterba, James M., "Tax Subsidies to Owner-Occupied Housing: An Asset Market Approach," Quarterly Journal of Economics 99 (1984) 729-752.

Smith, Lawrence B., "Household Headship Rates, Household Formation, and Housing Demand in Canada," Land Economics 60 (1984) 180-188.

Technical Appendix

As applied in this article, the Johansen procedure is based on the error-correction representation of the bivariate model given by equations (5) - (6). This representation can be derived by multiplying (5) by $(1 - \rho_1 L)(1 - \phi_1 L)$, multiplying (6) by $(1 - \rho_2 L)(1 - \phi_2 L)$, and writing the resulting expressions in vector autoregressive form.

$$X_t = \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + \mu + \epsilon_t \tag{A1}$$

where

$$X_t = \begin{bmatrix} s_t \\ p_t \end{bmatrix}$$

$$\Pi_1 = (1 - \alpha\beta)^{-1} \begin{bmatrix} (\rho_1 + \phi_1) - \alpha\beta(\rho_2 + \phi_2) & -\alpha(\rho_1 + \phi_1) + \alpha(\rho_2 + \phi_2) \\ \beta(\rho_1 + \phi_1) - \beta(\rho_2 + \phi_2) & -\alpha\beta(\rho_1 + \phi_1) + (\rho_2 + \phi_2) \end{bmatrix}$$

$$\Pi_2 = (1 - \alpha\beta)^{-1} \begin{bmatrix} -\rho_1\phi_1 + \alpha\beta\rho_2\phi_2 & \alpha\rho_1\phi_1 - \alpha\rho_2\phi_2 \\ -\beta\rho_1\phi_1 + \beta\rho_2\phi_2 & \alpha\beta\rho_1\phi_1 - \rho_2\phi_2 \end{bmatrix}$$

$$\mu = (1 - \alpha\beta)^{-1} \begin{bmatrix} \mu_1 + \alpha\mu_2 \\ \beta\mu_1 + \mu_2 \end{bmatrix}$$

$$\epsilon_t = (1 - \alpha\beta)^{-1} \begin{bmatrix} \epsilon_{1t} + \alpha\epsilon_{2t} \\ \beta\epsilon_{1t} + \epsilon_{2t} \end{bmatrix}$$

The error-correction representation is derived by subtracting IX_{t-1} from both sides of (A1) and then adding and subtracting $(\Pi_1 - I)X_{t-2}$ from the left hand side of the resulting expression. After these manipulations, we have

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} - \Pi X_{t-2} + \mu + \epsilon_t \tag{A2}$$

where

$$\Gamma_1 = \Pi_1 - I$$

$$= (1 - \alpha\beta)^{-1} \begin{bmatrix} [1-(\rho_1+\phi_1)] + \alpha\beta[1-(\rho_2+\phi_2)] & \alpha[1-(\rho_1+\phi_1)] - \alpha[1-(\rho_2+\phi_2)] \\ -\beta[1-(\rho_1+\phi_1)] + \beta[1-(\rho_2+\phi_2)] & \alpha\beta[1-(\rho_1+\phi_1)] - [1-(\rho_2+\phi_2)] \end{bmatrix}$$

$$\Pi = \Pi_1 + \Pi_2 - I$$

$$= (1 - \alpha\beta)^{-1} \begin{bmatrix} -(1-\rho_1)(1-\phi_1) + \alpha\beta(1-\rho_2)(1-\phi_2) & \alpha(1-\rho_1)(1-\phi_1) - \alpha(1-\rho_2)(1-\phi_2)] \\ -\beta(1-\rho_1)(1-\phi_1) + \beta(1-\rho_2)(1-\phi_2) & \alpha\beta(1-\rho_1)(1-\phi_1) - (1-\rho_2)(1-\phi_2)] \end{bmatrix}$$

Equation (A2) is a vector autoregressive model in first differences, except for the presence of the lagged level of X_t which captures the error-corrections component. There are three possibilities for the coefficients on the error-correction component (Π) are to be considered:

(1) $\rho_1 < 1$ and $\rho_2 < 1$, i.e. neither s_t nor p_t contains a unit root so that rank(Π) = 2. In this case, both s_t and p_t are stationary in levels.

(2) $\rho_1 < 1$ and $\rho_2 = 1$, i.e. both s_t and p_t contain a unit root but they are cointegrated with the cointegrating vector $[1\ -\alpha]$, Π becomes

$$\Pi = (1 - \alpha\beta)^{-1} \begin{bmatrix} (1-\rho_1)(1-\phi_1) & \alpha(1-\rho_2)(1-\phi_2) \\ -\beta(1-\rho_1)(1-\phi_1) & \alpha\beta(1-\rho_2)(1-\phi_2) \end{bmatrix}$$

Thus, when s_t and p_t are cointegrated, rank(Π) = 1 because the second row of Π is β times the first.

(3) $\rho_1 = 1$ and $\rho_2 = 1$, there are two unit roots in the system, and s_t and p_t are not cointegrated. In this case, $\Pi = 0$ and hence rank(Π) = 0.

Notice that in a bivariate system, the number of unit roots plus the rank(Π) will always equal 2, and the rank(Π) equals the number of cointegrating vectors.

The Johansen test for cointegration assumes ϵ_t are Gaussian and tests for the rank of Π. It is a least squares procedure and can be briefly described as follows: Regress ΔX_t on a constant and ΔX_{t-1} yielding the estimated residual series denoted R_{0t}.[12] Then, regress X_{t-2} on a constant and ΔX_{t-1} yielding the estimated residual series denoted R_{2t}. Denote the cross-product matrix of these residuals as S and partition this matrix as

$$S = \left[\begin{array}{c|c} S_{00} & S_{02} \\ \hline S_{20} & S_{22} \end{array}\right]$$

[12]By including the constant in these regression, we are using the unrestricted mean versions of these tests, which imposes no restrictions on the drifts in each of the series.

The tests for the rank(Π) involve the estimates of the ordered eigenvalues, $\lambda_1 > \lambda_2$, from the characteristic equation:

$$\left|\lambda S_{22} - S_{20}S_{00}^{-1}S_{02}\right|$$

The first hypothesis to be tested is

H_0: rank(Π) ≤ 1 versus H_a: rank(Π) > 1 (i.e. rank(Π) $= 2$).

The likelihood ratio statistic for this hypothesis, called the trace statistic, is given by

$$-2 \ln Q = -T \ln(1 - \lambda_2)$$

The distribution of this likelihood ratio test statistic is not the usual $\chi^2(1)$, but Johansen and Juselius (1990) provide simulated critical values. Given that we do not reject this hypothesis, the next null hypothesis to be tested is

H_0: rank(Π) ≤ 0 (i.e. rank(Π) $= 0$) versus H_a: rank(Π) > 0

The trace statistic for this hypothesis is

$$-2 \ln Q = -T\,[\ln(1 - \lambda_1) + \ln(1 - \lambda_2)]$$

If we do not reject this null, we infer that rank(Π) $= 0$ so that a cointegrating vector between the two series is not present in the data. That is, the two variable system contains two unit roots implying that their

stochastic trends are not in common. If we reject this null hypothesis, we conclude that the rank(Π) = 1 so that there is a cointegrating vector present, and hence they share a common stochastic trend.

An additional statistic, called the maximum eigenvalue statistic, should provide evidence which confirms any inference reached by using the trace statistic. For example, given that rank(Π) $\leq$ 1, the maximum eigenvalue statistic for the test that rank(Π) = 0 is

$$(-2 \ln Q \mid \text{rank}(\Pi) \leq 1) = -T \ln(1 - \lambda_1)$$

If this statistic is 'small' we do not reject that the two series are not cointegrated. As with the trace statistic, simulated critical values for the maximum eigenvalue statistics are provided by Johansen and Juselius (1990).

An alternative cointegration testing strategy is the two-step procedure recommended by Engle and Granger (1987). In this procedure, the cointegrating vector is estimated by ordinary least squares in the first step, and unit-root tests are performed on the residuals in the second step after correcting for autocorrelation in the residuals. The Johansen procedure is similar to the two-step procedure, but the Johansen procedure performs these two steps simultaneously. Because the correction for autocorrelation is performed simultaneously with the cointegration test, the Johansen procedure will preform better in small samples if the maintained assumptions are true, e.g. assumptions regarding the order of the autoregressive component and the Gaussian errors.

If we find evidence for cointegration, we can also use the Johansen methodology to test restrictions on the cointegrating vector. The specific

restriction of interest here is that the housing stock is proportional to the population, i.e. the cointegrating vector is [1 -1]. This test can be performed by replacing the vector of regressions of X_{t-2} with the single regression of the difference between the housing stock and population, $s_t - p_t$, in the procedure outlined above. In this case, there will be a single eigenvalue, λ_1^r. If this eigenvalue is not significantly smaller than the corresponding unrestricted eigenvalue, then we do not reject the restriction on the cointegrating vector. The likelihood ratio statistic for this test is

$$-2 \ln Q = T \ln [(1 - \lambda_1^r)/(1 - \lambda_1)]$$

which is distributed as the usual $\chi^2(1)$.

TABLE 1

REEXAMINATION OF MANKIW-WEIL RESULTS
(Annual data 1947-1985)

Housing Stock Variable (logs)	Demographic Variable (logs)	Number of Cointegrating Vectors	Johansen Test Statistics: Trace Statistic	Johansen Test Statistics: Max eigenvalue Statistic	Coefficient on Demographic Variable
VSTOCK	HD	1 or less	1.13	1.13	1.64
		0	15.14	14.01+	
VOWNREN	HD	1 or less	1.12	1.12	1.69
		0	15.40	14.27+	
VOWN	HD	1 or less	0.51	0.51	1.34
		0	15.57	15.05*	

* significant at the .05 level indicating that we can reject the hypothesis that there are the indicated number of cointegrating vectors or less, or that the coefficient on the demographic variable is not different from one. Critical values are from Johansen and Juselius (1990) Table A.2.
+ significant at the .10 level.

TABLE 2
COINTEGRATION OF VALUE OF RESIDENTIAL CAPITAL STOCK AND AGE GROUPS
(Annual data 1948-1985)

Housing Stock Variable (logs)	Demographic Variable (logs)	Number of Cointegrating Vectors	Johansen Test Statistics: Trace Statistic	Johansen Test Statistics: Max eigenvalue Statistic	Coefficient on Demographic Variable
VSTOCK	POP20	1 or less	2.82	2.82	1.26
		0	14.14	11.33	
VOWNREN	POP20	1 or less	2.78	2.78	1.25
		0	14.89	12.11	
VOWN	POP20	1 or less	1.20	1.20	1.08
		0	17.45+	16.25*	
VSTOCK	POP25	1 or less	0.57	0.57	1.06
		0	14.22	13.65*	
VOWNREN	POP25	1 or less	0.55	0.55	1.02
		0	15.14	14.59+	
VOWN	POP25	1 or less	0.05	0.05	0.59
		0	17.55+	17.56*	

* significant at the .05 level indicating that we can reject the hypothesis that there are the indicated number of cointegrating vectors or less, or that the coefficient on the demographic variable is not different from one. Critical values are from Johansen and Juselius (1990) Table A.2.
+ significant at the .10 level.

TABLE 3
COINTEGRATION OF HOUSING UNITS AND AGE GROUPS
(quarterly data 1965:I-1989:II)

Housing Stock Variable (logs)	Demographic Variable (logs)	Number of Cointegrating Vectors	Johansen Test Statistics: Trace Statistic	Johansen Test Statistics: Max eigenvalue Statistic	Coefficient on Demographic Variable
HSTOCK	POP20	1 or less	3.25	3.25	1.74
		0	15.26	12.01	
HOWNREN	POP20	1 or less	3.81	3.81	1.34*
		0	18.07*	14.27+	
HOWN	POP20	1 or less	10.43*	10.43*	1.30*
		0	30.53*	20.09*	
HSTOCK	POP25	1 or less	0.65	0.65	1.12*
		0	10.60	9.95	
HOWNREN	POP25	1 or less	2.90	2.90	1.09*
		0	10.15+	13.25+	
HOWN	POP25	1 or less	5.35	5.35	1.04
		0	22.02*	16.67*	

* significant at the .05 level indicating that we can reject the hypothesis that there are the indicated number of cointegrating vectors or less, or that the coefficient on the demographic variable is not different from one. Critical values are from Johansen and Juselius (1990) Table A.2

+ significant at the .10 level.

Figure 1

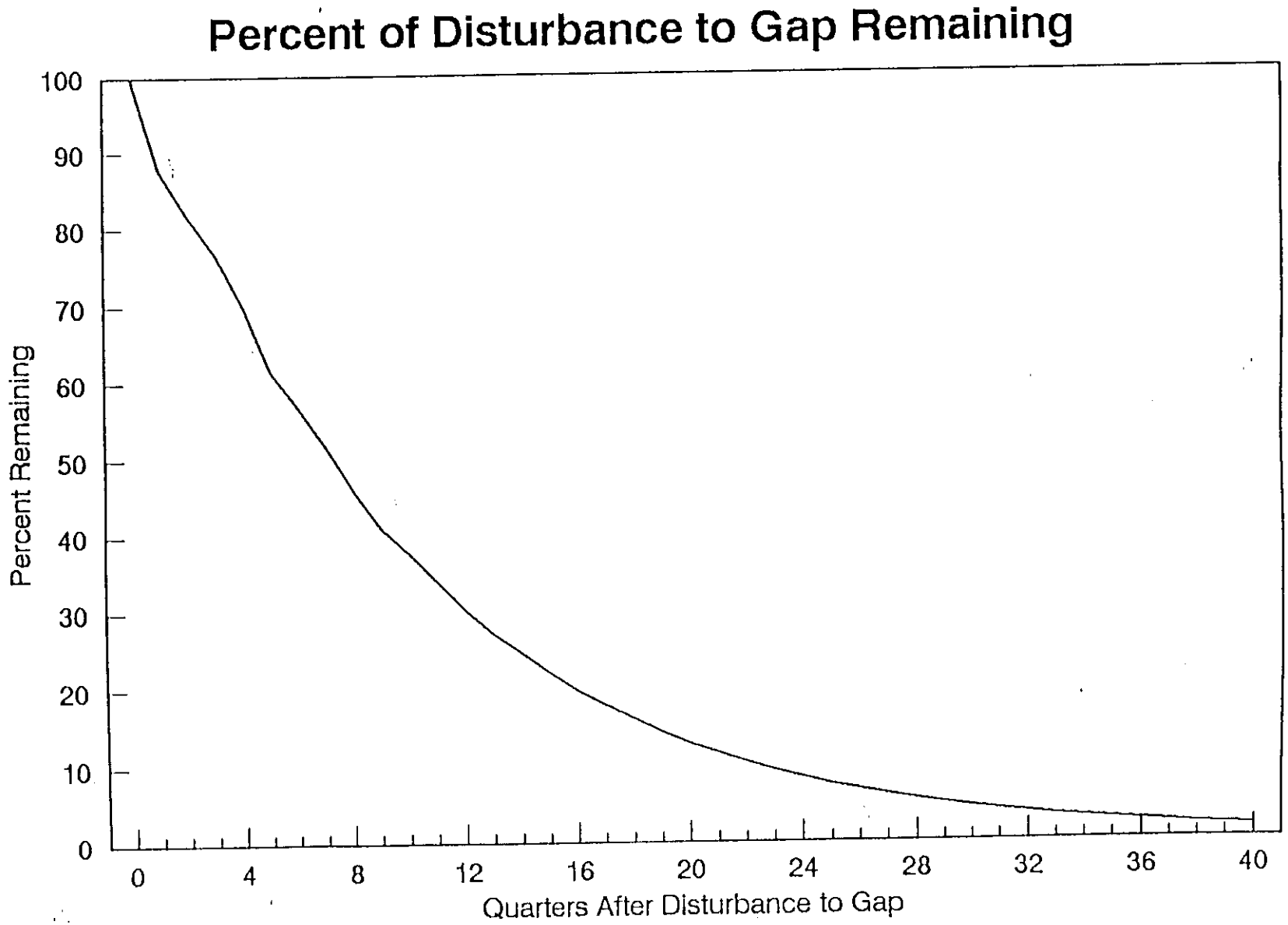

Percent of Disturbance to Gap Remaining
Percent Remaining
100
90
80
70
60
50
40
30
20
10
0
0
4
8
12
16
20
24
28
32
36
40
Quarters After Disturbance to Gap

Figure 2

Housing Units for Owner Occupancy

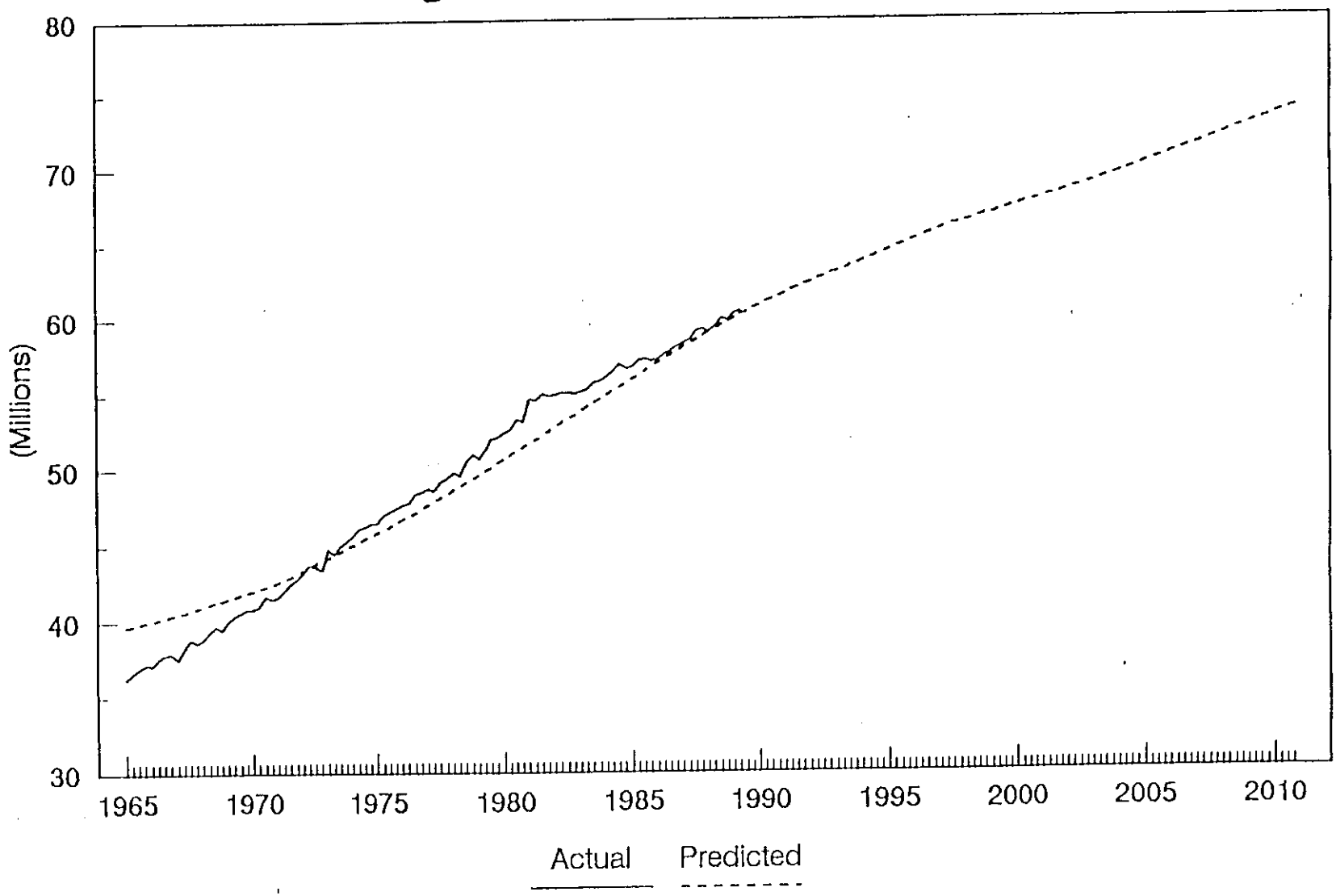
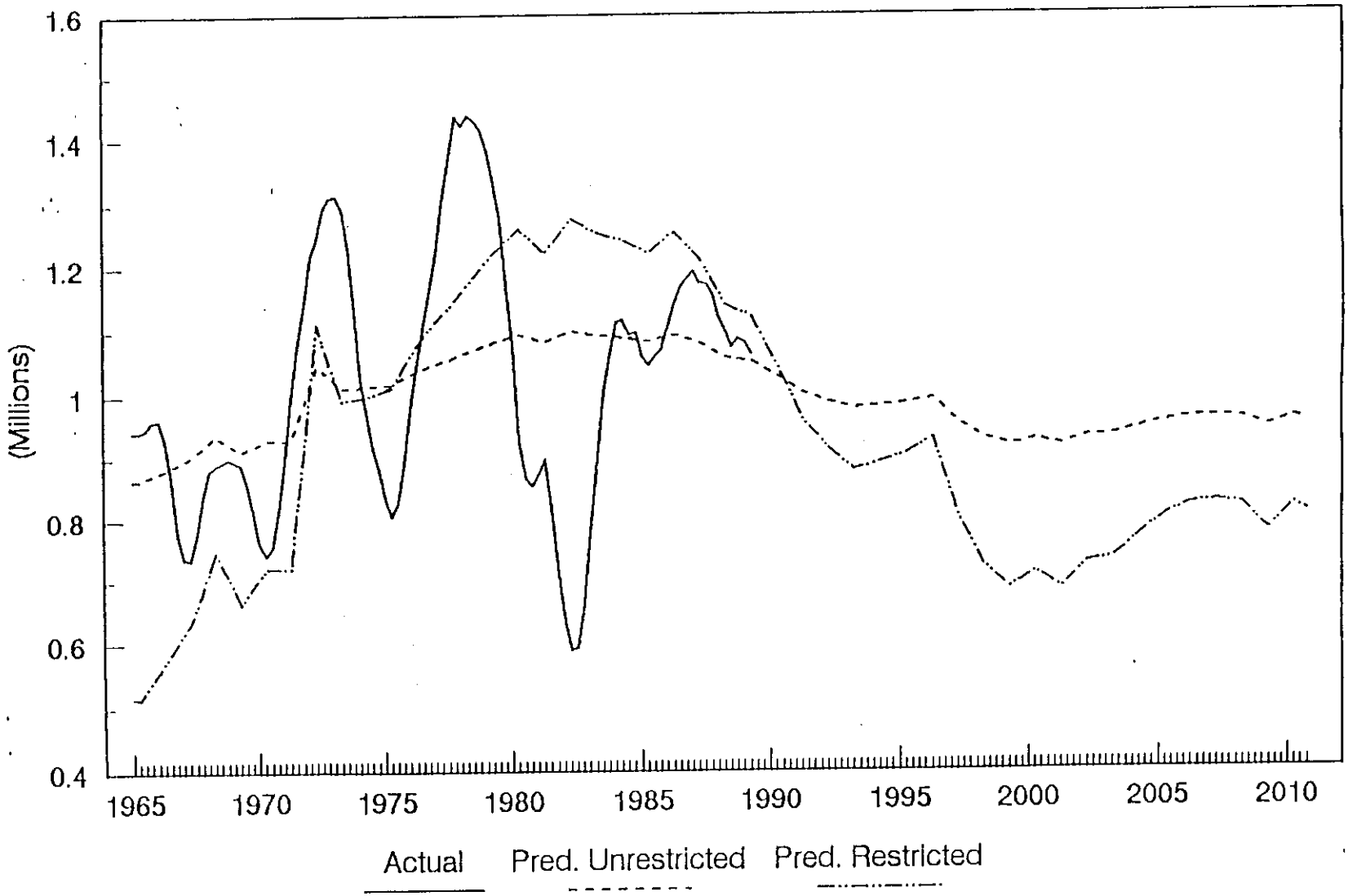


Figure 3

Single Family Housing Starts