Using Monthly Data to Predict Quarterly Output

Robert Ingenito
and Bharat Trehan

Fiscal and Policy Analyst, California Legislature, and Research Officer, Federal Reserve Bank of San Francisco. Robert Ingenito was a Senior Research Associate at the FRBSF at the time this was written. The authors are grateful to John Judd, Joe Mattey, Brian Motley, and Glenn Rudebusch for their helpful comments.

Some time ago, the Commerce Department changed the way it calculates real Gross Domestic Product. In response to that change, this paper presents an update of a simple model that is used to predict the growth rate of current quarter real output based on available monthly data. After searching over a set containing more than 30 different variables, we find that a model that utilizes monthly data on consumption and nonfarm payroll employment to predict contemporaneous real GDP does best.

Although monetary policy actions are usually undertaken with a view to affecting the economy sometime in the future, policymakers are also interested in the current state of the economy. One reason is that estimates of the current state of the economy constitute the starting point for predictions of the future state of the economy. In addition, these estimates can also be used as an input for policy rules whose prescriptions are based on the current state of the economy.[1] Towards this end, a small model to predict current quarter real GDP growth was developed at the Federal Reserve Bank of San Francisco about ten years ago. This model has done reasonably well over this period. For instance, in Trehan (1992) it was shown that real time forecasts from this model outperformed the Blue Chip average forecast (though the sample period available for comparison was relatively short). In fact, the model has been incorporated into the forecasting process of one member of the panel of Blue Chip forecasters.[2]

In late 1995, the Commerce Department changed the methodology they use to calculate GDP, moving from the use of fixed weights to chain weights.[3] In this paper we discuss how the model—known as the monthly indicators model—has been modified in response to this change.

I. The Original Specification

The specification search for the original model was guided by the following considerations: We wanted a method to predict real GDP that did not involve any judgmental adjustments to the forecast; we also wanted the forecasts to be available relatively early in the quarter. The result of our search was a model that predicted current quarter real GDP based on knowledge of nonfarm payroll employment, industrial production, and real retail sales. These series have the virtue of being available monthly; an additional advantage is that all the data we need for a given month become available by the middle of the following month.

1. The rule recommended by Taylor (1993) is a well-known example.

2. See Laurence H. Meyer and Associates (1994).

3. See Motley (1992) for a discussion of some of the issues involved in the change.

Is there any reason to change the specification in response to the change in how real GDP is measured? At first glance, the answer appears to be no. When the original specification is estimated over the 1968–1995 period[4] using the new chain-weighted GDP we obtain the following equation:

$$\underset{}{RGDP_t} = \underset{(4.4)}{1.24} + \underset{(5.6)}{1.01}\,EMP_t + \underset{(3.5)}{.17}\,IP_t + \underset{(5.5)}{.19}\,SAL_t$$

$$- \underset{(-2.7)}{.16}\,RGDP_{t-1} - \underset{(-2.6)}{.15}\,RGDP_{t-2} - \underset{(-3.3)}{.18}\,RGDP_{t-3}$$

where the adjusted $R^2 = .74$, SEE = 1.98, and the *t*-statistics are shown in parentheses. *RGDP* is chain-weighted real GDP measured in 1992 dollars, *EMP* is nonfarm payroll employment, *IP* is industrial production, *SAL* is real retail sales; all variables are entered in growth rates. While these estimates are not too different from prior estimates where GDP was measured in constant 1987 dollars,[5] an examination of the forecasting performance of this equation over the last 10 years shows that it does not do particularly well over this period. Estimating the equation over the last ten years of the sample (actually over the period from 1985.Q1 to 1995.Q3) shows why:

$$RGDP_t = \underset{(2.3)}{1.08} + \underset{(3.1)}{1.04}\,EMP_t - \underset{(-0.1)}{.01}\,IP_t + \underset{(2.8)}{.15}\,SAL_t$$

$$- \underset{(-0.5)}{.07}\,RGDP_{t-1} + \underset{(0.1)}{.01}\,RGDP_{t-2} - \underset{(-1.5)}{.25}\,RGDP_{t-3}$$

where the adjusted $R^2 = .52$, SEE = 1.47. The *IP* variable is no longer significant over this period; the coefficients on the GDP lags are somewhat different as well, suggesting that the dynamics of the process may have changed. While the smaller sample can be expected to lead to larger standard errors, the change in the *IP* coefficient is harder to attribute to the small sample. To establish that this change was the result of the new GDP data, we estimated this equation over the same sample period (1985–1995) using GDP measured in 1987 dollars. We then obtained:

$$GDP87_t = \underset{(2.7)}{1.06} + \underset{(2.6)}{.70}\,EMP_t + \underset{(2.5)}{.22}\,IP_t + \underset{(2.9)}{.13}\,SAL_t$$

$$- \underset{(-1.6)}{.22}\,GDP87_{t-1} + \underset{(0.8)}{.09}\,GDP87_{t-2}$$

$$- \underset{(-0.7)}{.09}\,GDP87_{t-3}$$

where the adjusted $R^2 = 0.63$, SEE = 1.24. As can be seen, *IP* helps predict *GDP*87 over this sample (as it does when the equation is estimated over the entire 1968–1995 sample period).

These results suggest that we would be better off respecifying the monthly indicators model in response to the change in the GDP data. Our goals are the same as before: We would like a small model to forecast real GDP that does not involve judgmental adjustments. It would also be useful to obtain forecasts relatively early in the quarter.

II. SELECTION STRATEGY

One way to select the variables that will be used to forecast GDP is to rely on measures of in-sample performance. For instance, one could select the set of variables that maximizes R^2 in an equation that predicts real GDP or select those variables that have *t*-statistics above a certain value. However, specifications obtained in this way generally do not lead to good forecasts, since attempts to explain in-sample variation often lead to over-fitting. In other words, while the movements of a particular series within any given sample often can be explained by adding additional variables to the regression, relationships "discovered" in this way can sometimes be spurious and fail to hold up outside the sample under study. To minimize the possibility of such an outcome, we will use a strategy based on the results from two different search procedures. First, we use a procedure that selects a set of variables based on within-sample performance. Specifically, we use what Maddala (1977) calls the "Stepwise Regression Procedure" to determine an initial set of variables to be included in the model. Second, we select variables by looking at how well they help forecast real GDP out of sample.[6] In our final specification we place more weight on the second criterion, especially with regard to the number of variables included in the model.

Another set of issues involves the date at which we would like to make a forecast. The underlying issue is a familiar one: A forecast that is available relatively early in the quarter is likely to be less accurate than a forecast that is available later; yet it is possible to wait too long in an effort to get the most accurate forecast. As a practical matter, the date at which we would like to make a forecast will determine the set of variables that will be considered potential candidates for our model. All the variables included in our original specification were available by the 15th of

4. The start date is dictated by the availability of the retail sales data.

5. For instance, based on data through 1991, the coefficients in Trehan (1992) are: $GDP87_t = 1.1 + 0.96\,EMP_t + 0.20\,IP_t + 0.16\,SAL_t - 0.20\,GDP87_{t-1} - 0.10\,GDP87_{t-2} - 0.26\,GDP87_{t-3}$.

6. Here, too, one has to be careful to do more than simply choose the specification with the smallest prediction error; we provide the details of our procedure below.

the following month; for example, December data for all three variables in the original model are usually available by January 15. We have relaxed this constraint somewhat this time and will consider variables that are available by the end of the following month. This has the advantage of introducing variables such as personal income and consumption in the set of candidates that we consider.[7] However, it still excludes potentially important variables, such as inventory accumulation, which become available only about six to eight weeks after the end of the month in question.

III. Implementing the Strategy

We used the stepwise procedure to determine which of the 34 variables shown in Table 1 could be included in an equation that "explains" contemporaneous real GDP growth.[8] This list contains a representative of just about any kind of variable for which data become available within 30 days of the end of the relevant month. (For instance, while we did not consider every interest rate series available, we did make sure that we had both long and short maturities, as well as rates on private and government instruments, etc.) Our sample covers the 1967–1995 period, where the starting date is determined by the availability of the retail sales data. The procedure adds variables to the regression one at a time, choosing the one that has the highest partial correlation with output.[9] Only those variables whose *t*-statistic had a marginal significance level below 0.05 were included; further, if the introduction of a new variable caused a variable that was already in the regression to become insignificant at the 5% level, then the insignificant variable was dropped.

This procedure led to including the following variables in the equation: nonfarm payroll employment, average weekly hours, the number of passenger automobiles sold (the dollar value of which is included in retail sales), personal

7. Note that for variables such as consumption the third month of data for any quarter will generally become available after the preliminary estimate of GDP is released. However, as our results below will demonstrate, the third month of data do not have a large effect on the accuracy of the current quarter forecast. More specifically, the model attains its lowest root mean square error before the preliminary GDP data are released.

8. All variables listed in Table 1 are available on the Citibase data tape.

9. An alternative strategy is to include all the variables we have in the regression and keep dropping variables that are insignificant; this approach is reminiscent of the "general to specific" approach recommended by Hendry. (See Hendry and Mizon 1978, for example.) However, following this procedure leads to including an extremely large set of variables in the model. We chose to follow a more conservative strategy here, for reasons we discuss below.

TABLE 1

LIST OF VARIABLES CONSIDERED FOR INCLUSION IN MONTHLY INDICATORS MODEL

1. Federal Funds Rate
2. 3-Month Treasury Bill Rate
3. 6-Month Commercial Paper Rate
4. 1-Year Treasury Bond Rate
5. 10-Year Treasury Bond Rate
6. Moody's AAA Corporate Bond Rate
7. M2
8. Standard & Poors 500 Composite Stock Price Index
9 Loans and Leases at Commercial Banks
10. Index of Consumer Confidence (University of Michigan)
11. Index of Consumer Confidence (Conference Board)
12. New Privately Owned Housing Units Started, Total
13. The Consumer Price Index
14. Commodity Research Bureau Spot Market Index—All Commodities
15. Retail Sales deflated by the Consumer Price Index
16. National Association of Purchasing Managers' Index
17. New Orders for Durable Goods
18. Total New Passenger Cars Sold
19. Index of Industrial Production (Factories, Mines & Utilities)
20. Capacity Utilization, Manufacturing Sector
21. Real Personal Income
22. Real Consumption
23. Index of Leading Economic Indicators
24. Civilian Unemployment Rate
25. Total Employment (Household Survey)
26. Total Workers on Non-agricultural Payrolls (Establishment Survey)
27. Workers on Manufacturing Payrolls
28. Total Non-farm Payrolls Less Manufacturing Payrolls
29. Average Weekly Hours of Production Workers on Total, Private, Non-farm Payrolls
30. Index of Aggregate Weekly Hours, Production Workers on Non-farm Payrolls
31. Average Weekly Initial Claims for Unemployment Insurance
32. Diffusion Index: Percent of Firms Adding to Non-farm Payrolls (1-Month Span)
33. Gross Average Hourly Earnings, Constant Dollars
34. Gross Average Weekly Earnings

income, and consumption. Three lags of real GDP and a constant also were included in the equation.

Next, we used forecast performance over the 1985.Q1–1995.Q3 period to choose among alternative specifications. The procedure we used was as follows: For each forecast, the GDP equation is estimated up to the prior quarter, and the values of the indicator variables are used to predict the current quarter's output. For example, for the first forecast the equation is estimated through 1984.Q4 and used to forecast real GDP for 1985.Q1 using the contemporaneous values of the indicator variables. Next, the estimates are updated through 1985.Q1 and the equation is used to forecast 1985.Q2. The best specification is defined to be the one that leads to forecasts with the lowest root mean square error (RMSE) over the 1985.Q1–1995.Q3 period.[10] We carried out the search in two steps. We first searched for the set of variables that was the best at predicting real output, conditional on including a given number of variables in the set. We then varied the number of variables in the set, going from two to four.

In Table 2 we show the forecast error statistics for the best set of variables for the three different set sizes. For the two-variable case, the combination of total nonfarm payroll employment and real personal consumption leads to the smallest RMSE. The error falls slightly (from 1.40 to 1.31) when we move to the three-variable specification. The best specification here includes the two variables in the first set plus weekly hours. It turns out, however, that the third variable in the set does not matter very much; any of about a dozen variables when added to the first two lead to about the same size RMSE. Perhaps more to the point, only a dozen of the three-variable specifications actually perform better than the best two-variable case. Given that roughly 6,000 combinations were considered, it seems reasonable to attribute the slightly superior performance of 12 of these to chance. Hence, our conclusion is that the three-variable model does no better than the two-variable version.

We reach the same conclusion regarding the four-variable specification. The best specification there leads to a RMSE of 1.28 over this period; the variables included are real consumption, manufacturing payroll employment, nonmanufacturing payroll employment,[11] and a measure of commodity prices. Given that we looked at more than 46,000 combinations to find the lowest error, the small improvement we obtain does not appear to warrant rejecting the two-variable specification in favor of the four-variable one.

10. This means that the specification we choose could have a nonzero average forecast error.

11. These are the two components of nonfarm payroll employment, which is the variable that is selected in the first two specifications.

TABLE 2

PREDICTING CONTEMPORANEOUS OUTPUT: FORECAST PERFORMANCE OF ALTERNATIVE SPECIFICATIONS SAMPLE PERIOD: 1985.Q1–1995.Q3

BEST SPECIFICATION	MEAN ERROR	MEAN ABSOLUTE ERROR	ROOT MEAN SQUARE ERROR
2 variables	–0.01	1.17	1.40
3 variables	–0.43	1.13	1.31
4 variables	–0.15	1.07	1.28

Of course, the same logic also can be used to question whether the two-variable specification is really any better than a specification that uses a single variable to forecast output. It turns out that among all the specifications that use only one indicator variable to predict output, the one that contains nonfarm payroll employment alone has the smallest RMSE: 1.67 percent. Thus, adding consumption to the equation that contains payroll employment leads to a reduction of about 0.3 percentage points in the RMSE.

IV. THE FINAL SPECIFICATION

Our preferred model is one that contains only two variables: nonfarm payroll employment and real consumption. The estimated equation is:

$$\begin{aligned} RGDP_t = {} & \underset{(0.1)}{0.05} + \underset{(10.7)}{1.41}\,EMP_t + \underset{(7.3)}{.51}\,CONS_t \\ & - \underset{(-3.1)}{.19}\,RGDP_{t-1} - \underset{(-3.4)}{.19}\,RGDP_{t-2} \\ & - \underset{(-4.2)}{.23}\,RGDP_{t-3}\,. \end{aligned}$$

RGDP is real GDP measured in chain-weighted 1992 dollars, *EMP* denotes nonfarm payroll employment, and *CONS* denotes real personal consumption; all variables are in growth rates. The equation is estimated over the 1968.Q2–1995.Q3 period. The adjusted R^2 is 0.71, and Godfrey's (1978) test reveals no evidence of either first or fourth order autocorrelation in the residuals.

The two indicator variables in our model are included in the set of variables selected by the stepwise procedure; they are also usually in the set of variables selected on the basis of our minimum RMSE criterion. A natural question here is: Are two variables enough to forecast contemporaneous output, or should we include additional variables? For instance, as discussed above, the stepwise regression procedure leads to the inclusion of a number of other variables in the set of variables used to forecast real output.

However, it seems to us that significant *t*-statistics alone are not sufficient to include a given variable in the model. This is especially the case because we have not selected the set of variables that we have searched over on any *a priori* basis, but have simply searched over (the relatively large set of) all available variables. Thus, there is a good chance that we will find variables that have large *t*-statistics but that are not really useful in predicting real output. In view of this, it seems desirable to opt for a relatively conservative specification.

V. Predicting the Indicator Variables

So far we have focused on how to predict output when we have all the monthly data we require available to us. However, most of the time, the model will be used to predict GDP when we have only partial data for the quarter. For instance, forecasts of Q4 real GDP made in late November or early December will be based on only one month of data on consumption and payroll employment and will require that we forecast how these variables will evolve over the following two months. In other words, in order to produce a model that predicts real GDP we need to produce an auxiliary model that generates forecasts of the indicator variables themselves.[12]

We begin by presenting the forecast errors that result from univariate models of both nonfarm payroll employment (which we denote by *EMP*) and real personal consumption (denoted by *CONS*). We regressed each variable on a constant and lags of itself, and then generated forecasts for the one-month-ahead to three-months-ahead horizon over the period from January 1985 to September 1995 (129 monthly forecasts). Once again, the forecast from each model was generated after estimating the model through the month prior to the (first) month being forecast; e.g., the forecast for 1988:01 was made after estimating the model from 1967:01–1987:12, while the forecast for 1992:7 was done after estimation through 1992:6. We tested several alternative specifications (by allowing the number of lags to vary) and concluded that using six lags results in the smallest forecast errors for both *EMP* and *CONS*. The forecast error statistics are presented in Table 3; *EMP* errors are in thousands of jobs per month, and *CONS* errors are in billions of dollars.

The first class of alternative specifications we examined combined the two variables into a vector autoregression (VAR). A VAR system models each variable (*EMP* and *CONS*) as a function of a constant and lags of both variables, where the univariate model employs its own lags only. Intuition suggests, for example, that previous changes in nonfarm payrolls contain information that might improve forecasts of consumption. However, forecasts from this specification were consistently worse than its univariate counterparts. For instance, the RMSEs for both variables when using the VAR were higher at all three forecasting horizons. Changing the lag lengths of the two specifications did not change this result.

As a robustness check, we augmented the bivariate VAR system by including real retail sales, various short and long-term interest rates, industrial production and the National Association of Purchasing Managers index. Not a single VAR specification including these variables (or any sub-group) improved upon the forecasting performance of the autoregressive specifications described above.

Our finding that the VARs do not forecast very well has been known for a while. Litterman (1986) suggested that the way to overcome this problem was to impose "Bayesian priors" on the VAR. The priors recommended by Litterman push each equation in the VAR towards a random walk. Specifically, in each equation, the coefficient on the first lag of the dependent variable is pushed towards one while all other lags are pushed to zero. How tightly these priors are imposed depends upon the forecasting performance of the model.[13] We tested several different Bayesian VARs (BVARs), including the bivariate case and several three- and four-variable systems. Imposing a Bayesian prior on the bivariate system produced slightly more accurate (lower RMSE) forecasts of both indicator variables than the univariate regimes, and it was also the most promising of all the BVAR specifications we tried.

Our final step was to see if the forecasts of the indicator variables could be improved by including contemporaneous values of those monthly series that are released before the indicator variables themselves. This is not a significant issue for the employment data, since that is one of the first releases that becomes available to us. However, consump-

12. We are following a two-stage strategy here: First, we search over the set of variables that leads to the best forecasts given that we have all the data we need for the quarter. Second, we try to find the best model to predict the indicator variables themselves. An alternative strategy is to integrate the two stages. This would allow us to compare the forecasting performance of alternative models at different points in the quarter (i.e., when we have partial data for the quarter we are trying to forecast). This latter approach was followed when the model was first estimated; the results obtained were not sufficiently different to justify the effort of the extensive search that would be required. Note that Figure 2 below does provide one comparison of this kind; it also provides a hint of why the more extensive search may not be very useful, since it shows that the model that does best based on three months of data also does best at every other point in the quarter.

13. For a more detailed discussion of how such a prior is imposed, see Todd (1984) or Litterman (1986).

tion data come out rather late, and it is natural to ask if consumption forecasts can be improved by taking account of other data already available to us. After some searching, we found that retail sales data—which are released roughly ten days to two weeks prior to the consumption data—are extremely useful in predicting contemporaneous consumption. The specification we finally settled on contains employment, consumption, and retail sales. The retail sales equation contains contemporaneous employment data, while the consumption equation contains contemporaneous employment as well as retail sales data; the inclusion of contemporaneous values reflects the order in which the data are released. Six lags of each variable are also included in each equation. Once again we have placed Bayesian priors on this system; the only exceptions are the contemporaneous terms, which have been left unrestricted.

The results from this exercise are contained in Table 4. In the first panel we show the forecast errors from an exercise where we assume we have no contemporaneous information when making our forecasts. By contrast, in the second panel we assume that we know the values of employment and retail sales during the first month that we are forecasting. Incorporating this information into the model cuts the RMSE of the *CONS* forecast for the first month by about a half; subsequent months are not affected as much, however. Note also that while the errors for employment in the second and third month seem to have declined noticeably, that is because they really represent one- and two-month-ahead forecasts.

Based on these results, our preferred specification for forecasting the indicator variables is the three-variable system

TABLE 3

Forecasts of *EMP* and *CONS* from Univariate Autoregressive Models (85:01–95:09)

	EMP			*CONS*		
Months Ahead	Mean Error	Mean Absolute Error	Root Mean Square Error	Mean Error	Mean Absolute Error	Root Mean Square Error
1	8.1	90.0	116.4	0.6	12.2	17.0
2	18.7	137.7	178.2	1.1	14.6	19.9
3	31.0	194.9	252.6	1.7	17.2	22.5

TABLE 4

Forecasts of *EMP* and *CONS* from a 3-variable System

	EMP			*CONS*		
Months Ahead	Mean Error	Mean Absolute Error	Root Mean Square Error	Mean Error	Mean Absolute Error	Root Mean Square Error
			Assuming no contemporaneous information			
1	–1.9	89.8	115.4	–1.2	12.1	17.0
2	–3.9	134.2	175.3	–2.0	14.2	19.4
3	–7.7	187.9	246.7	–3.2	15.7	21.7
			Assuming one month of information on *EMP* and retail sales			
1	NA	NA	NA	–3.6	7.6	9.4
2	–2.5	90.2	115.7	–3.9	13.9	18.5
3	–4.5	134.9	176.0	–4.9	15.3	21.5

where we include contemporaneous values of employment and retail sales in the equation for predicting consumption.

VI. MID-QUARTER OUTPUT FORECASTS

We are now in a position to analyze how the performance of the monthly indicator model would change as more and more information became available over the course of the quarter. It is easiest to understand how this works by means of a concrete example. Assume that we are in the second week of November and wish to generate an estimate of Q4 GDP. At this point we are likely to have employment data through October, but no Q4 data for either consumption or sales. Thus, we will use the monthly equations for predicting employment, consumption, and sales to fill out the remainder of the quarter. The quarterly averages of these (actual and forecasted) values can then be incorporated into the real output equation to estimate Q4 output. We can then repeat this exercise for every quarter of our forecast sample (1985–1995 again) and obtain a set of forecasts based on the same amount of information each quarter. Error statistics based on this set of forecasts are plotted as the point E1 in Figure 1. We show the mean absolute error and the root mean square error; the mean error stays close to zero throughout and therefore is not shown.

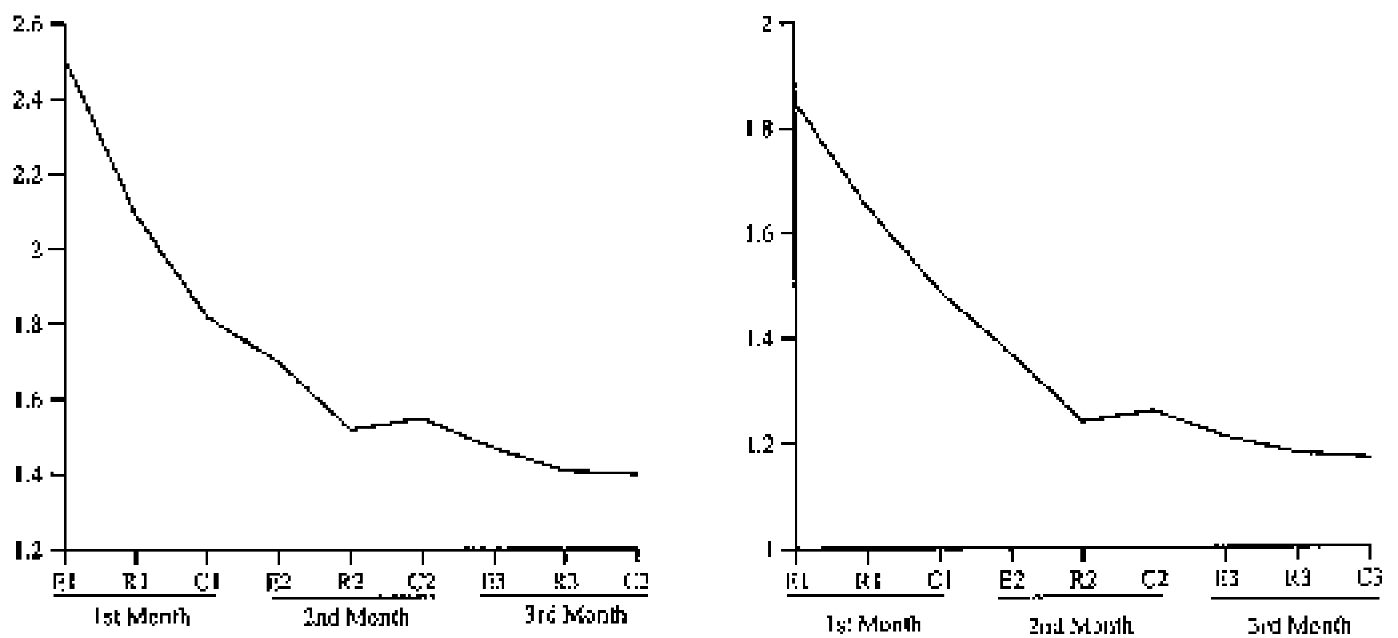
Successive points on the figure show how the performance of the model changes as more data become available. Thus, the point labeled R1 shows the forecasting performance of the model once the first month of retail sales data become available (that is, based on one month of employment and retail sales data), while the point labeled C2 shows the performance of the model once consumption data for the second month become available. The RMSE is 2.5% when employment data for the first month are received; it falls below 1.8% when we receive consumption data for the first month and is 1.5% based on complete data for the second month. The RMSE hits its minimum when we obtain retail sales data for the third month of the quarter.

Another issue has to do with the timeliness of the forecast. Use of consumption data in this version of the monthly indicators model means that a forecast based on complete data for the month will be available relatively late; for instance, if we had used retail sales we would have had a comparable forecast available about two weeks earlier. It is natural to wonder whether the new specification means that we will be worse off during the period between the re-

FIGURE 1

ERRORS IN PREDICTING REAL GDP AT VARIOUS DATES IN THE QUARTER

ROOT MEAN SQUARE ERROR | MEAN ABSOLUTE ERROR



Note: E1 is the date at which employment data for the first month become available; R2 is the date at which retail sales for the second month is released; and C3 is the date at which consumption data for the third month are published. Errors are measured in annualized growth rates.

ceipt of the retail sales data and the consumption data. Figure 2 provides an answer to this question. It compares the forecast errors from this specification to a specification where we use employment and retail sales to forecast output.[14] The figure shows that the errors from the specification that uses consumption to forecast output are never greater than those from the specification that uses retail sales. (Of course, both specifications also use employment.) The only time the RMSEs are close, for instance, is upon receipt of the first month of data on retail sales. Thus, this exercise does not suggest that the use of consumption instead of retail sales in the equation to predict real GDP leads to a less accurate forecast during the period in which we have retail sales data but do not have consumption data.

VII. SOME RELATED ISSUES

It is worth discussing two other issues before concluding this paper. The first one has to do with the use of initial versus revised data. All the results we have presented here have been based upon data as it existed at the time this project was first started (in early 1996).[15] It is not likely that we would obtain the same results using data that would actually be available to us in real time. Unfortunately, since the required data are not available to us, it is not possible to determine how the model would perform under these circumstances. (Recall also that the chain-weighted GDP data are new.) However, it is possible to get some sense of how the error statistics might change with data revisions by looking at the historical performance of the *original* model (with GDP measured in 1987 dollars). Over this small sample of 16 forecasts, we obtain a RMSE of 1.1% when values of the monthly variables as they exist today are used to forecast GDP87 data as they exist today. The real time forecast error is the same, that is, when the forecasts that the model actually made over this period are

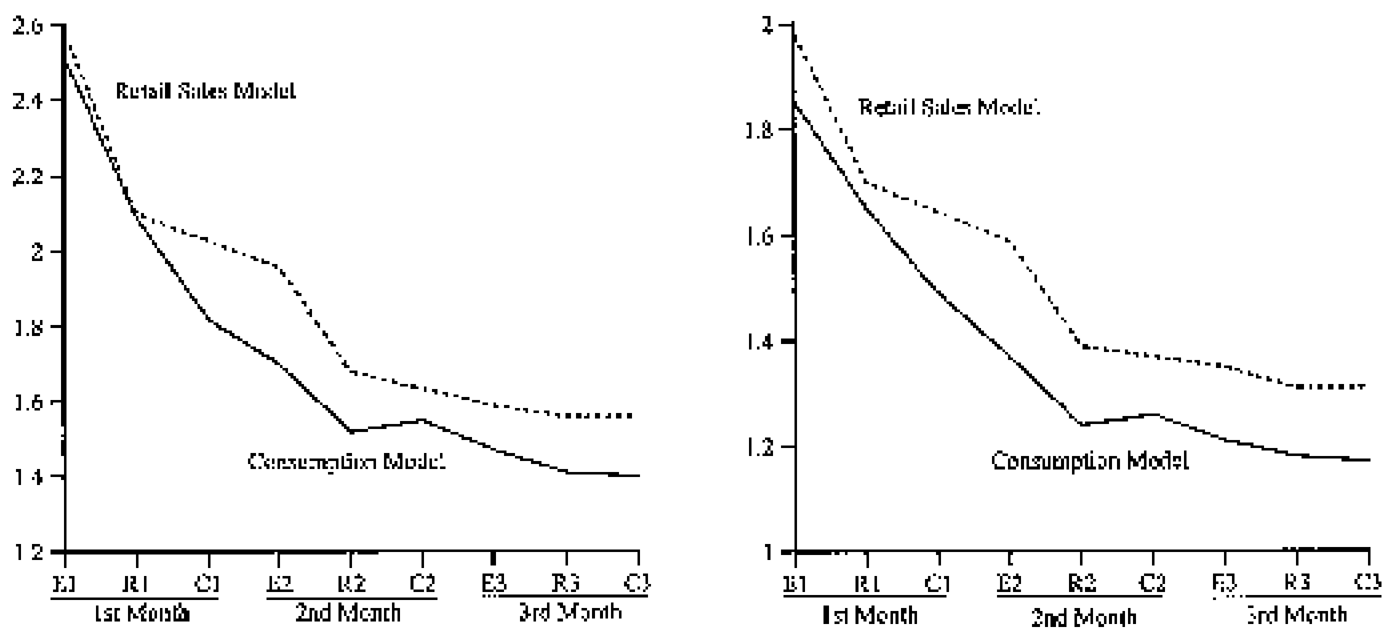
14. Note that this requires monthly forecasts of retail sales; these forecasts are obtained from the same 3-variable system that is used to forecast employment and consumption.

15. The monthly consumption data are current as of June 1996.

FIGURE 2

COMPARISON OF FORECAST ERRORS

ROOT MEAN SQUARE ERROR | MEAN ABSOLUTE ERROR



Note: See Figure 1.

compared to the initial estimates of GDP87 the RMSE is 1.1% as well. However, when the model's historical forecasts are compared to currently available GDP87 data, the RMSE is 1.6%.

A final issue has to do with the stability of the estimated equation. It is well known that estimated macroeconomic relations shift over time, and it is quite possible that the coefficients of our estimated equation will change as well. This suggests that it might be better to forecast using a specification based on time-varying coefficients. We tried a number of such specifications, including several that assume that the coefficients follow a random walk and others that assume that the coefficients can move around but tend to return to some fixed value. We were unable to come up with a specification that outperformed the fixed coefficient version.[16] In fact, the only time we were able to get RMSEs smaller than those from the base version (which involves Kalman filtering) was when we set the coefficients equal to their final period value at the beginning of the forecast period and held them there throughout.

VIII. SUMMARY AND CONCLUSIONS

In this paper we have presented a revised version of a small model that is used to forecast current quarter GDP. We have shown that a specification based on two indicator variables does about as well at forecasting GDP as specifications that contain three or four variables. In addition, we have searched over a larger set of indicator variables this time, allowing for variables that are available up to one month after the month to which the data pertain. As a result, we found that monthly consumption data provide key information about contemporaneous output. There is a potential trade-off here: While forecasts based on the consumption data are more accurate, we have to wait longer to get the relevant consumption data. So there is a period of time when a model based on consumption could make forecasts that are worse than a model that does not contain consumption (because the latter model will have more current information over this period). It seems that we do not have to pay such a price, because retail sales data help forecast consumption.

16. This finding is consistent with what Stock and Watson (1996) find. They analyze bivariate regressions based on a data set of 76 monthly series (5,700 relationships) and conclude that "...in over half the pairs, random walk TVP models or rolling regressions perform better than fixed coefficient or recursive least squares, although the gains typically are small." In other words, time-varying parameter models do no better than fixed coefficient models about half the time.

REFERENCES

Godfrey, L.G. 1978. "Testing against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables." *Econometrica* 4, pp. 1293–1302.

Hendry, D.F., and G. E. Mizon. 1978. "Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study by the Bank of England." *Economic Journal* 88, pp. 549–563.

Laurence H. Meyer & Associates. 1994. *The U.S. Economic Outlook.* September 7.

Litterman, Robert B. 1986. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience." *Journal of Business and Economic Statistics* (January) pp. 25–38.

Maddala, G.S. 1977. *Econometrics.* New York: McGraw Hill.

Motley, Brian. 1992. "Index Numbers and the Measurement of Real GDP." Federal Reserve Bank of San Francisco *Economic Review* 1, pp. 3–14.

Stock, James H., and Mark W. Watson. 1996. "Evidence on Structural Instability in Macro Time Series Relations." *Journal of Business and Economic Statistics* (January) pp. 11–30.

Taylor, John B. 1993. "Discretion Versus Policy Rules in Practice." *Carnegie–Rochester Conference Series on Public Policy* 39, pp. 195–214.

Todd, Richard M. 1984. "Improving Economic Forecasting with Bayesian Vector Autoregression." Federal Reserve Bank of Minneapolis *Quarterly Review* (Fall) pp. 18–29.

Trehan, Bharat. 1992. "Predicting Contemporaneous Output." Federal Reserve Bank of San Francisco *Economic Review* 2, pp. 3–12.

Explaining Differences in Farm Lending among Banks

Mark E. Levonian

Research Officer, Economic Research Department, Federal Reserve Bank of San Francisco. I am grateful to Jennifer Soller for research assistance and to the editorial committee (especially Joe Mattey) for helpful suggestions that changed the focus of the paper.

Do small, rural banks lend to farmers because they are small, or because they are rural? This paper combines a new measure of the extent of agricultural activity in banking markets with an appropriate statistical framework to examine causes of interbank variation in agricultural production loans. The results show that a bank's size and head office location both matter to some extent, but that the size of a bank's branches in agricultural areas is the single most important factor determining agricultural loan levels. Other variables, such as ownership structure and charter type, have no significant effects. While far from definitive, the results suggest that industry consolidation and mergers may have little effect on agricultural credit, as long as they do not lead to the outright closure of branches in rural areas.

Banks differ substantially in their agricultural lending. Most banks do none. Banks that *are* agricultural lenders vary in their degree of emphasis, with most doing little but some devoting 50 percent or more of their assets to farm loans. One possible explanation for such variation in the composition of bank loan portfolios is that location matters, especially in farm lending. As an industry, agriculture is notably tied to particular, typically rural, locations. Banks located in such areas might specialize in farm loans, while banks in urban areas might not.[1]

This explanation, while simply stated, is not so simply tested. When is a bank "located" in a farm area? When it has its head office in a farm area, or when it has branches in farm areas? If the latter, *how many* branches must it have in farm areas? And what exactly is meant by a "farm" area? Even given answers to these questions, other complications arise. For example, it is part of banking folklore that small banks are more likely to lend to farmers, all else equal, and rural banks tend to be smaller. Could differences in farm lending by location actually be a size effect? Or could apparent size effects simply be due to differences in bank location?

Answering these questions requires, as a first step, a measure of how "agricultural" a bank's market area is. For banking, a sensible measure of the degree to which a market is "agricultural" is the quantity of farm loans demanded within that area. This paper begins by constructing a proxy for agricultural loan demand within the area served by a bank, based on the geographic distribution and relative size of its branches. As a second step, this paper develops a model of bank agricultural loan decisions consistent with the observation that most banks do no farm lending, and in which

1. The idea that bank loan portfolios vary by location to reflect the industrial composition of the local market area assumes that proximity matters in lending, that banks have a greater tendency to lend to nearby potential borrowers than to those farther away. Such a tendency might reflect the importance of information for credit analysis and monitoring on the bank's side or convenience-related effects on the borrower's side. However, while these considerations seem reasonable, their empirical importance in bank lending is an open question. A finding that loan portfolios reflect the nature of the local market would provide evidence that these things have an observable effect.

location, size, and other bank characteristics play a role in determining two facets of lending behavior: whether or not a bank becomes involved in agricultural lending at all, and the quantity of farm loans if it does get involved. Appropriate statistical techniques are applied to account for the apparent "censoring" of the data (that is, the fact that most banks hold no farm loans).

The model is applied to banks in four important western agricultural states: California, Idaho, Oregon, and Washington. The results allow an assessment of factors that might determine differences in farm lending by commercial banks, including differences in the nature of markets and differences in bank characteristics, such as size and ownership structure. In addition, the results have implications for other issues. For example, concerns have been raised over the effects of structural changes in the banking industry, such as merger waves or interstate banking. Suppose that large banks, or those owned by out-of-state holding companies, tend to do less agricultural lending than otherwise similar banks not owned by such companies. In that case, an industry trend toward bigger banks, or toward acquisition of independent banks in agricultural states by out-of-state banking firms, might tend to reduce the amount of credit flowing to agriculture. The loan mix of acquired banks would change to match the acquiring companies' profiles. However, branch *locations* usually change little following mergers or other structural changes.[2] If the results show that location and market composition are what matter for farm lending, then structural change in banking at the industry level might have little effect on lending, because it would not change the composition of markets. Any institution acquiring a particular branch in an agricultural area likely would continue to lend to farmers.

The next two sections provide background, summarizing relevant aspects of existing research and presenting information on agricultural lending in the sample states. Four following sections describe the measure of local market demand for agricultural loans, the model of loan decisions, the econometric framework, and the data used in the study. Section VII discusses the results, Section VIII assesses the implications for the relative importance of bank and market characteristics, and a final section concludes.

2. Some branches may be closed outright following a merger or acquisition. However, most often the acquirer consolidates unwanted branches with other branches in the same market; in that case, the combined presence within the market is unchanged. Occasionally, unwanted branches are sold to other banking firms rather than closed or consolidated, which similarly maintains the same lending capacity in the market. Frequently, the same employees remain at a branch as it changes hands.

I. Previous Research

Three previous papers dealt with various aspects of the issues raised here. Gilbert and Belongia (GB 1988) examined the effects of bank size and holding company affiliation on agricultural lending. They attempted to eliminate the effects of location through sample design, using only banks in counties that were not part of any Metropolitan Statistical Area (MSA), that had high ratios of agricultural loans to total loans, and that were in one of nine states with restricted branching in 1985. GB found that agricultural loans comprised a significantly smaller share of assets for banks owned by bank holding companies than for other banks, and the holding company effect was greater the larger the parent company.

Laderman, Schmidt, and Zimmerman (LSZ 1991) looked at the effects of location on agricultural lending by banks. LSZ found that banks headquartered in MSAs had significantly lower ratios of agricultural loans to total loans. The sample consisted of banks surveyed each quarter from 1981 through 1986 by the Federal Reserve as part of the Survey of Terms of Bank Lending to Agriculture; this group of banks, varying in number from 168 to 188 depending on the date, has been deemed to be representative of farm lenders. The only bank-specific variables in the LSZ model were total assets and the ratio of deposits to loans; no measures of ownership structure were included, so it is not clear to what extent the results were driven by structural differences rather than location. LSZ found that size had a negative but insignificant effect on farm lending.

A paper by Whalen (1995) covered small agricultural loans as part of a more general analysis of small-business lending by banks. Whalen's sample consisted of 1,377 banks in the states of Illinois, Kentucky, and Montana (all of which had restricted branching as of his June 1993 sample date). Whalen looked specifically at the effects of bank size, holding company ownership, and out-of-state ownership; he found some evidence that small banks not owned by bank holding companies have higher ratios of agricultural loans to total assets than do other banks. However, Whalen acknowledged that the difference might reflect location rather than structure, since he found no significant size- or affiliation-related differences in mean agricultural loan ratios among banks in non-MSA areas.

II. Agricultural Lending by Banks

Banks provide two broad types of agricultural credit: loans secured by agricultural real estate and other agricultural loans. Agricultural producers generally use loans secured by real estate to acquire physical capital, including land, equipment, and livestock. Nonbank lenders, especially insurance companies, are active competitors for this type of lending. Prices and quantities of loans secured by farm real estate depend heavily on land values and only indirectly on agricultural prices and output.

This paper considers the second category, loans not secured by real estate; banks are the dominant supplier of such loans. These loans are referred to as agricultural production loans, generally financing variable production costs such as seed, fertilizer, and labor. Demand for production loans is driven primarily by agricultural output. The loans tend to be shorter term and have a strong seasonal element, with a clear trough in the first quarter of each year.

The four states covered in this study—California, Idaho, Oregon, and Washington—account for about 90 percent of agricultural production lending by banks in the Twelfth Federal Reserve District. They comprised about 16 percent of agricultural output for the United States in 1992 as measured by the market value of sales (see Table 1) and similar percentages of total U.S. farm debt and bank agricultural production loans outstanding. California is the largest of the four in terms of market value of agricultural sales.

The importance of banks as agricultural lenders varies somewhat across the states, with banks supplying about 54 percent of production credit in Oregon, but nearly 70 percent in the state of Washington; except for Oregon, all four are above the national average. Oregon has a higher proportion of farm debt secured by real estate. Viewing production loans as an input to the agricultural production process, California and Oregon have the highest output per dollar of bank loans, with Idaho below the national average. Taken as a group, banks in these four states have a higher than average share of production lending, production loans are a slightly smaller share of total farm debt, and the value of output is higher relative to total production loans, but the differences from the rest of the country are not remarkable.

One aspect of agricultural production in these states that may limit the generality of the results is that a relatively high proportion of production is concentrated in larger farms. For example, production units with annual sales exceeding $500,000 accounted for 80 percent of the total market value of sales in California in the 1992 Census of Agriculture, compared to 47 percent for the United States as a whole. The difference stems in part from an emphasis on production of higher value crops, but also reflects an above average number of large agricultural enterprises. Large farms in Idaho, Oregon, and Washington had somewhat smaller shares of state output—61, 54, and 60 percent, respectively—but all are above the national average. On the other hand, these states are not so far from the norm that they are completely unrepresentative; for example, Florida is comparable to California in the dominance of large farms, and traditional farming states like Kansas and Colorado have higher percentages than the Pacific

TABLE 1

AGRICULTURE AND BANK LENDING

	CALIFORNIA	IDAHO	OREGON	WASHINGTON	FOUR STATES	U.S.
Market value of agricultural sales	17,052	2,964	2,293	3,821	26,130	162,608
Market value of sales as % of US total	10.5	1.8	1.4	2.4	16.1	100.0
Bank share of ag production loans, in %	55.1	62.9	53.7	69.4	58.4	54.5
Ag production loans as % of total farm debt	41.9	48.6	33.0	49.7	42.8	45.8
Ratio of sales to ag production loans	3.1	2.4	3.0	2.7	2.9	2.6

Notes: Market value of sales in millions of dollars, from 1992 Census of Agriculture.
Total ag production loans and total farm debt from 1992 USDA Farm Balance Sheets by State.
Bank ag production loans from December 1992 Call Reports.

Northwest states. Nevertheless, readers should be cautious in using results from this sample to draw inferences for the rest of the county.[3]

III. How Agricultural Is the Market?

Gauging the importance of location requires a measure of the degree to which banks' markets are agricultural. Previous papers (such as GB, LSZ, and Whalen) used the location of banks' head offices, typically comparing banks headquartered in MSAs with those headquartered outside of MSAs. The reasoning is that non-MSA areas are more rural, and hence probably more agricultural.

This distinction based on MSA/non-MSA headquarters is sensible if the location of the head office adequately portrays the location of the bank's business, and if MSA areas are in fact less agricultural than non-MSA areas. However, both conditions frequently are violated. In the western states, and increasingly in recent years in the rest of the country, banks can and do branch statewide; the result is that the head-office location of the bank is not a good indication of the location of its branches. The characterization of MSAs as less agricultural than non-MSA areas also is not necessarily accurate in the western states. The top agricultural counties as measured by total agricultural production in California, Oregon, and Washington are all MSA counties; in California, nine of the top ten counties in agricultural production are within MSAs. (MSA definitions are based on boundaries of single counties or groups of contiguous counties.)

A better measure of the nature of any bank's market area can be based on the actual geographic distribution of its branches and the amount of agricultural activity in the branch locations. From a bank's perspective, a market is more agricultural if more of the loans in that market are used to finance agricultural production. Assume that agricultural loan demand in a county c at any point in time is proportional to farm output as measured by the total value of sales reported by farms in that county:

$$LD_c = \gamma \cdot Q_c\,, \tag{1}$$

where LD is loan demand, Q is farm output, and γ is a proportionality factor that may vary over time depending on bank interest rates, the price of substitute forms of credit, and other factors, but is constant at any point in time over the counties in which the bank operates. (In the empirical work to follow, the factor γ is allowed to vary by state, to reflect state differences in agricultural production functions and credit market conditions. It is held constant within any given state to allow estimation.) Assume that the share of this loan demand faced by bank i is equal to the bank's share of the deposit market in a county:

$$LD_{ci} = \gamma \cdot Q_c \cdot \left(\frac{D_{ci}}{D_c} \right), \tag{2}$$

where D represents deposits. Summing across counties for bank i yields a measure of the agricultural loan demand facing the bank, based on the extent of agricultural production in the counties in which the bank actually operates:

$$LD_i = \gamma \cdot \sum_{c=1}^{N} Q_c \cdot \left(\frac{D_{ci}}{D_c} \right) = \gamma \cdot MARKET_i\,, \tag{3}$$

where $MARKET_i$ is a weighted average of agricultural production in all of the counties in which the bank has branches.

IV. Lending Decisions

Banks can and do invest in many different kinds of assets, including various types of loans. However, most do not invest in every type of asset available to them; they go through management decision processes that result in positive amounts of some assets and zero of others. In the case of agricultural lending, some banks invest in farm loans and others do not, despite the fact that the market areas of almost all banks include at least some agricultural production.

One way to explain such a pattern is to posit a decision process in which bank management takes prices (or interest rates) as given by the market, and then sets threshold levels for investments in various types of assets, including farm loans. Thresholds might arise because different types of loans require different approaches to marketing, credit evaluation, and monitoring; as a result, a particular type of lending can require specific investment in systems and staff, leading to quasi-fixed costs that must be incurred regardless of the quantity of lending done. Unless the quantity of any given category of lending is high enough, the costs cannot be covered and that type of loan is unprofitable for the bank. After setting a threshold, a bank then calculates a profit-maximizing quantity of each asset type; if this quantity exceeds the threshold, the bank invests (making loans, in the case of lending), and otherwise it does not.

More formally, a bank sets (on some basis not explicitly modeled here, but possibly depending on characteristics of the bank) a threshold T for agricultural lending. Independently of the threshold, the bank determines a profit-maximizing quantity of farm loans L^*, based partly on the

3. Zimmerman (1989) discusses differences between the West and the rest of the country in agricultural lending.

demand for such loans. The bank then compares L^* to T; if L^* is at least as large as T, the bank holds farm loans in the amount L^*; otherwise, the bank holds no farm loans, thereby avoiding the costs of gearing up to manage such a specialized type of asset. Both T and L^* may depend partly on characteristics of the bank and partly on factors that are common to all banks in a particular market or region. However, while L^* depends on the demand for agricultural loans in the market, T does not; in essence, the bank sets a threshold, then looks around its markets to see if the quantity of lending it can actually do would meet or exceed that threshold.

V. ESTIMATION FRAMEWORK

For empirical work, both L^* and T are modeled as linear random functions of observable variables. Factors that affect all banks or markets within a state equally (such as interest rates) are captured through binary dummy variables for each state. To allow for idiosyncratic variation at the firm level, a disturbance term is added to each equation:

$$L_i^* = \beta_0 + \beta_1 X_i + \beta_2 STATE_i + \beta_3 MARKET_i + u_{Li}\ , \tag{4}$$

$$T_i = \alpha_0 + \alpha_1 X_i + \alpha_2 STATE_i + u_{Ti}\ , \tag{5}$$

$$\begin{aligned} L_i &= L_i^* \quad \text{if} \quad L_i^* \geq T_i \\ L_i &= 0 \quad\ \ \text{if} \quad L_i^* < T_i \quad , \end{aligned} \tag{6}$$

where X is a vector of bank-specific characteristics that might influence loan decisions, u_T and u_L are bank-specific disturbances assumed to be jointly normal with means of zero, standard deviations σ_T and σ_L respectively, and covariance σ_{LT}, and $STATE$ is a vector of state dummies (with one state, California, omitted and picked up in the intercept instead). The coefficient β_3 captures two effects, the impact of loan demand LD_i on L^*, and the influence of agricultural output on loan demand, as measured by γ above; γ implicitly is incorporated into β_3.

The threshold T cannot be observed; thus, this is a model in which the data are censored, and the censoring variable is endogenous, stochastic, and unobserved. As with any censored regression model, estimation using only the banks with nonzero values of L would give biased estimates, because the errors would not have zero mean.[4] However, following Heckman (1976), it is possible to estimate a well-behaved probit model, from which the conditional mean of the residuals can be computed and used as an adjustment in an ordinary least squares regression to explain variations in loan quantity.

Specifically, let I be an indicator variable that takes the value $I_i = 1$ if $L_i^* \geq T_i$, and zero otherwise:

$$\begin{aligned} I_i = 1 \text{ if } & (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1)X_i \\ & + (\beta_2 - \alpha_2)STATE_i + \beta_3 MARKET_i \\ & + (u_{Li} - u_{Ti}) \geq 0\ , \\ I_i = 0 & \text{ otherwise.} \end{aligned} \tag{7}$$

The disturbance term $u_L - u_T$ is normal with zero mean and variance $\sigma^2 = \sigma_L^2 + \sigma_T^2 - 2\sigma_{LT}$. Probit estimation of this "selection" equation yields consistent estimates of β_3 and of the differences in all of the other coefficients. Most importantly, it can be used to compute estimates of the inverse Mills ratio, which is related to the probability that each observation is censored; this ratio can be used as a regressor in an ordinary least squares "quantity" regression based on the observations with positive farm lending:

$$\begin{aligned} L_i = \beta_0 + \beta_1 X_i + \beta_2 STATE_i \\ + \beta_3 MARKET_i + \beta_4 IMR_i + \varepsilon_{Li}\ , \end{aligned} \tag{8}$$

where IMR is the inverse Mills ratio computed from the probit results. The coefficient estimates measure the impact of each variable on the optimal quantity of agricultural loans in the bank's portfolio, conditional on the bank engaging in such lending. The adjustment for censoring introduces an element of heteroskedasticity which must be corrected, but the corrections are relatively straightforward (see Maddala 1983). With consistent estimates of the β coefficients from the quantity regression, estimates of the α coefficients can be recovered from the probit coefficients in the selection equation (7), thereby providing information about determinants of agricultural loan thresholds.

The Heckman censored regression framework used here is similar to a Tobit regression. The major differences are that with a Tobit, the factors determining the lending threshold must be the same as those determining the optimal level of lending, and the coefficients on the variables in the selection and quantity equations must be constrained to be identical (that is, the coefficients in the threshold equation must be constrained to zero). The two-step Heckman procedure is preferable because it relaxes those unnecessary constraints.[5]

4. LSZ explicitly assumed this problem away, while GB do not appear to have dealt with the issue at all.

5. Gunther and Siems (1995) applied a related approach to an analysis of banks' exposure to derivative financial instruments; I am grateful to them for pointing me in this direction.

VI. DATA

The sample of banks includes all 527 commercial banks with branches in any of the four states of the sample as of June 1994.[6] Of these, 229 reported having farm loans on their books. Data on agricultural production loans at the banks come from the Reports of Condition (Call Reports) filed by banks, Schedule RC-C Line 3, "Loans to finance agricultural production." Use of the June reporting date avoids the seasonal trough in farm production lending.

Several variables are used to describe bank characteristics that may be related to either the loan threshold or the profit-maximizing loan quantity or both:

BHC = 1 if the bank is owned by a bank holding company, 0 otherwise

FOREIGN = 1 if the bank is owned by a foreign entity, 0 otherwise

OSBHC = 1 if the bank is owned by a holding company from a state other than the state in which the bank is headquartered (but not a foreign entity), 0 otherwise

MSAHQ = 1 if the head office of the bank is in a Metropolitan Statistical Area, 0 otherwise

NATIONAL = 1 if the bank has a national charter, 0 if state-chartered (reflecting possible differences in supervision)

MC = 1 if the bank has branches in multiple counties, 0 otherwise

BRANCHES = Number of branches of the bank, including the head office

SIZE = Natural log of total assets (in thousands of dollars) of the bank

The first five items come from Federal Reserve bank structure data. *MC* and *BRANCHES* are constructed from data in the FDIC Summary of Deposits. The asset figures come from the Call Reports. Regardless of the source, all data are reported as of June 30, 1994.

Figures for the market value of agricultural sales in each county are used as a proxy for agricultural production from the 1992 Census of Agriculture. The other elements needed to construct the *MARKET* variable for each bank are the deposits at each branch of the bank and the locations of the branches. Such branch-level data are available from the FDIC Summary of Deposits for June 1994.

6. Five banks were excluded from the sample because they reported having no loans or no deposits or both.

The composition of the sample of banks is summarized in Table 2; the first column shows the number of banks in each group, and the second column shows the percentage of those banks that report holding agricultural production loans. Only a few banks in the sample are either foreign-owned or owned by an out-of-state BHC. A large number have branches in only one county. Most of the banks are in California, and most are headquartered in MSAs. A notably larger proportion of non-MSA banks engage in agricultural lending compared to MSA-headquartered banks. A smaller percentage of banks located in California hold farm loans in their portfolios, and banks owned by foreign entities also are less likely to be agricultural lenders.

VII. ESTIMATION RESULTS

Initial estimates using the two-step model revealed substantial size-related heteroskedasticity, a common problem in banking research: Larger banks in the sample may have loan levels that are many times greater than the total assets of the smaller banks, and hence tend to have much larger regression residuals. Experimentation with various

TABLE 2

COMPOSITION OF SAMPLE

	Number of Banks	Percentage with Agricultural Loans
Total	527	43
Owned by BHC	209	49
Not owned by BHC	318	40
Owned by foreign entity	20	20
Not owned by foreign entity	507	44
Owned by out-of-state BHC	44	52
Not owned by out-of-state BHC	483	43
Headquartered in MSA	444	36
Not headquartered in MSA	83	86
National charter	165	38
State charter	362	46
Branches in more than one county	195	56
Branches in only one county	332	36
Headquartered in: California	375	31
Idaho	20	85
Oregon	45	78
Washington	87	69

size variables and specifications revealed that the residuals were most strongly related to the natural log of total assets. Dividing by the log of assets to rescale all variables is a simple correction for this source of heteroskedasticity and is applied throughout the rest of this paper. Thus, the estimation framework in Section V should be understood as applying to the rescaled data.[7]

Estimated coefficients for each variable are presented in Table 3; standard errors are in parentheses immediately below each coefficient, with asterisks denoting various levels of significance for a test of the hypothesis that the coefficient is zero. The first column shows estimates for the first-stage probit selection equation, using all 527 observations. (In the notation of equation (7), the reported coefficients are actually $(\beta - \alpha)/\sigma$, as is standard in probit estimation.) The results can be interpreted as identifying factors that affect the probability that a bank will engage in agricultural lending.

The variable that measures the degree to which banks' markets are agricultural, *MARKET,* has a positive and strongly significant effect. The larger a bank's presence in highly agricultural areas according to this measure, the more likely that the bank does at least some farm lending. This result directly supports the hypothesis that a bank's decision to engage in a particular type of lending reflects the composition of its local markets. Equally important, with this variable included in the model, the effects of various bank-specific factors on the probability of engaging in farm lending can be evaluated separately from the confounding correlation between those characteristics and location.[8]

The significant negative coefficient on *SIZE* shows that larger banks are less likely to engage in agricultural lending than are smaller banks. This is a true size effect, since other factors such as location and ownership structure that may be related to the size of the bank have been separately controlled.

Of the structural variables, head office location has a significant effect: Banks headquartered in MSA counties are significantly less likely to lend to farmers than those headquartered in non-MSA areas. This is *not* because

7. The precise form of this correction turns out to have little practical effect on the results; use of other scaling variables changes none of the conclusions regarding the effects of any explanatory variables. Other studies have used a similar (but usually implicit) scaling, generally based on total assets.

8. Of course, this assumes that the location of branches does not depend on these other characteristics. While unlikely to be strictly true, such an assumption is a reasonable working approximation in the absence of a fully developed theory of bank branch location.

TABLE 3

ESTIMATION RESULTS

	PROBIT	*L**	*T*
MARKET	0.0004*** (0.0001)	0.2751*** (0.0208)	—
SIZE	–2.49** (1.02)	1454 (1775)	3083
BHC	–1.14 (1.70)	–1849 (3524)	–1101
FOREIGN	–17.64 (16.17)	77836*** (18194)	89380
OSBHC	–5.59 (5.59)	14889* (8036)	18548
MSAHQ	–10.47*** (2.41)	–1939 (4405)	4912
NATIONAL	2.01 (1.69)	2156 (3655)	838
MC	2.76 (1.95)	3938 (3594)	2132
BRANCHES	0.71** (0.33)	–251** (101)	–715
INTERCEPT	24.01** (11.04)	–27174 (20665)	–42882
IDAHO	11.16** (5.14)	11036 (7192)	3731
OREGON	10.62*** (2.90)	7741 (5234)	792
WASHINGTON	9.31*** (2.09)	11664*** (4432)	5572
Number of banks	527	229	—
Log likelihood	–229.82	—	—
R^2	—	0.901	—

Notes:
*** Significantly different from zero at 1% level
** Significantly different from zero at 5% level
* Significantly different from zero at 10% level

banks with rural head offices tend to be smaller or in more heavily agricultural markets, since the *SIZE* and *MARKET* variables have captured the influence of size and market composition. The coefficient on number of branches is significantly positive: Having more branches raises the probability that a bank will do at least some farm lending. Thus, if two banks of identical size and ownership structure have the same market share in the same array of counties, the one with more branches is more likely to be holding farm loans in its portfolio. This may reflect enhanced monitoring capability for the lender or possible convenience-related effects on loan demand.

Notably, ownership structure—whether by an in-state or out-of-state BHC or a foreign entity—has no statistically significant effect on whether a bank is a farm lender; independent banks are neither significantly more nor significantly less likely to have agricultural production loans in their portfolios. As for the state dummies, banks in Idaho, Oregon, and Washington are all more likely to hold agricultural loans than are California banks, even with other factors held constant.

The second column shows the estimated coefficients for the loan quantity equation; recall that L^* is the profit-maximizing, or desired, quantity of farm loans the bank would hold if there were no threshold. These coefficients come directly from the second stage least squares regression using the 229 banks with $L > 0$, incorporating the *IMR* variable as an estimate of the degree of censoring of each observation. (The coefficient on *IMR* was 831.2, with a standard error of 431.5.) Coefficients from the probit selection equation can be multiplied by the standard deviation of the residuals (654.3) and combined with coefficients from the quantity regression to derive implied coefficients for a loan threshold equation; these implied values are presented in the last column of the table.

The significantly positive coefficient on *MARKET* shows that banks with a greater presence in agricultural counties do more lending to farmers, all else equal. The desired L^* also increases with asset size—larger banks aim for higher loan quantities—but the coefficient is not significantly different from zero. The coefficient on *SIZE* in the threshold equation implies that T rises with size faster than L^* does; larger banks run larger agricultural loan portfolios if they have farm loans, but require still higher levels of activity if they are to engage in farm lending in the first place. The net effect is that larger banks are significantly less likely to engage in lending at all, as the probit coefficient indicates.

Foreign ownership has a significant and positive effect on the quantity of agricultural loans in the portfolio, implying that foreign-owned banks engaging in farm lending do more of it, for any given combination of market composition, size, and other bank characteristics. However, this turns out to be an anomaly due to a single large foreign-owned bank, Sanwa Bank of California. Sanwa reported \$360 million in agricultural production loans, accounting for 64 percent of the total farm loans of foreign-owned banks in the sample. If Sanwa is deleted from the sample, the coefficient on *FOREIGN* drops to 2347 and becomes insignificant, with little change in the other coefficients.

Having a head office in an MSA lowers the desired quantity of agricultural loans, as the negative coefficient on *MSAHQ* shows. An MSA head office also raises the lending threshold substantially. These two effects reinforce each other, thereby significantly reducing the probability that an MSA-headquartered bank will engage in farm lending. Once again, this is *not* because MSA-headquartered banks necessarily operate in markets with less agricultural loan demand. These effects on the threshold and desired loan quantity are related specifically to the location of the head office as opposed to the branches, perhaps indicating that the physical location of key decisionmakers has an important influence on the type of lending a bank will do.

The effect of the number of branches is significantly negative, a surprising conclusion in view of the probit result that additional branches significantly raise the probability of being a farm lender. The explanation lies with the coefficient on *BRANCHES* in the threshold equation; additional branches lower T by more than they lower L^*. Banks with more branches are willing to engage in agricultural lending in much smaller amounts than otherwise similar banks with fewer branches. Each additional branch reduces a bank's desired quantity of agricultural production loans by \$251,000 on average.

Out-of-state ownership is positive and mildly significant, indicating that out-of-state banks have higher desired loan levels; however, their thresholds are also higher, so they are less likely to actually lend to farmers (the probit point estimate is negative, although insignificant). Neither BHC ownership nor charter type has a significant effect on agricultural loan quantity decisions. Banks in the state of Washington hold significantly more farm loans than do banks in other states, a finding that is consistent with Washington banks' large share of the total agricultural production loan market in that state (see Table 1). At the risk of being repetitious, this is *not* because markets in Washington are more agricultural than in other states, since that characteristic is separately controlled in the estimation.

Gilbert and Belongia (1988) found that the size of the parent BHC had a significant impact on agricultural lending for banks that were owned by holding companies. To test for such an effect in these data, the two-step model was reestimated with *BHC* replaced by three separate binary variables, each taking the value of one for banks owned by holding companies if the parent's consolidated total assets were under \$150 million, between \$150 million and \$1 billion, or over \$1 billion, respectively. The results (Table 4) show no significant effect of parent size on either the quantity of agricultural loans or the probability of engaging in farm lending. The impact of other variables is largely unaffected by the change in specification, except that the influence of out-of-state BHC affiliation on loan quantity diminishes to insignificance.

VIII. RELATIVE IMPORTANCE OF LOCATION AND BANK CHARACTERISTICS

The preceding section focused on the statistical significance of coefficients. However, statistical significance does not directly address the quantitative impact of interbank differences in these variables on farm lending. As noted above, effects flow through two channels: the decision to engage (or not engage) in farm lending, and the decision regarding quantity of agricultural loans for banks that choose to hold such loans. Since the directions of these two effects may be opposing (as, for example, in the cases of *BRANCHES* or *MSAHQ*), it is important to have a summary measure of the impact of each variable, combining the two channels.

One possible summary measure is the effect of each variable on the conditional expectation of farm lending at a representative bank. The conditional expectation is equal to the *unconditional* expected value of farm loans multiplied by the probability of being above the threshold level. If Z is the matrix of all variables included in the model, β is the vector of *OLS* coefficients, and γ is the vector of probit coefficients, then expected farm lending is:

$$E(L) = E(L^*(Z\beta)) \cdot N(Z\gamma)\,, \tag{9}$$

where $E(L^*)$ is the expected profit-maximizing quantity of farm loans (the unconditional expectation) and $N(\cdot)$ is the cumulative normal density function.

The effects of each variable z_i included in Z can be evaluated through the elasticity:

$$\frac{\partial E(L)}{\partial z_i} \cdot \frac{z_i}{E(L)}\,, \tag{10}$$

which can be interpreted as the approximate percentage change in expected farm loans for a 1 percent change in z_i. The elasticity must be calculated for a representative bank; in this case, consider an "average" bank, for which all variables are equal to their sample means.

Following McDonald and Moffitt (1980), $\partial E(L)/\partial z_i$ can be decomposed into a "quantity effect" due to the impact on the quantity of farm loans for banks above the threshold, and a "selection effect," the change in the expected

TABLE 4

ESTIMATION RESULTS, WITH BHC PARENT SIZE DUMMIES

	PROBIT	L^*	T
MARKET	0.0004*** (0.0001)	0.2760*** (0.0208)	—
SIZE	–2.36** (1.06)	1116 (1857)	2672
BHC < \$150 mil	–0.05 (0.18)	–4 (375)	28
BHC \$150 mil–\$1 bil	–0.25 (0.24)	–319 (456)	–151
BHC > \$1 bil	–0.15 (0.32)	551 (613)	649
FOREIGN	–18.76 (17.09)	84911*** (19007)	97262
OSBHC	–5.28 (5.80)	8806 (9728)	12286
MSAHQ	–10.38*** (2.41)	–1920 (4401)	4911
NATIONAL	1.93 (1.71)	1690 (3705)	419
MC	2.74 (1.98)	3813 (3604)	2012
BRANCHES	0.76** (0.35)	–267** (102)	–767
INTERCEPT	22.52** (11.32)	–23209 (21455)	–38030
IDAHO	11.06** (5.15)	9684 (7228)	2401
OREGON	10.54*** (2.91)	7487 (5255)	548
WASHINGTON	9.20*** (2.11)	11051** (4456)	4994
Number of banks	527	229	—
Log likelihood	–229.44	—	—
R^2	—	0.901	—

Notes:
*** Significantly different from zero at 1% level
** Significantly different from zero at 5% level
* Significantly different from zero at 10% level

value of farm loans due to the change in the probability of being above the threshold and therefore engaging in farm lending. McDonald and Moffitt develop this decomposition for a Tobit, but extension to the current case is fairly straightforward.

The results are in Table 5. The figures can be interpreted as percentage changes in expected agricultural loans for the average bank. With the exception of *MSAHQ*, the quantity effects are larger in absolute value than the selection effects, implying that the major impact of differences in each variable come through their effect on the size of the loan portfolio held by banks that are in the farm loan business, not through the impact on the probability of being farm lenders. Based on either the selection elasticity or the quantity elasticity or the two combined (the total elasticity), market composition has a relatively large impact on cross-sectional differences in expected agricultural loans. A 1 percent increase or decrease in agricultural loan demand leads to a corresponding 1.7 percent increase or decrease in the expected value of farm production loans, with about 0.3 percentage points of that arising from the increase in the probability that desired lending will exceed the bank's threshold.

TABLE 5

IMPACT OF VARIABLES ON EXPECTED VALUE OF BANK AGRICULTURAL LOANS

VARIABLE	SELECTION ELASTICITY	QUANTITY ELASTICITY	TOTAL ELASTICITY
MARKET	0.294	1.427	1.721
SIZE	–0.458	3.583	3.125
BHC	–0.007	–0.102	–0.109
FOREIGN	–0.011	0.484	0.474
OSBHC	–0.008	0.220	0.212
MSAHQ	–0.141	0.023	–0.118
NATIONAL	0.010	0.087	0.097
MC	0.016	0.199	0.215
BRANCHES	0.141	–0.779	–0.638
IDAHO	0.007	0.053	0.059
OREGON	0.014	0.076	0.090
WASHINGTON	0.025	0.258	0.283

Note: "Selection Elasticity" reflects the change in expected loans due to changes in the probability of engaging in farm lending. "Quantity Elasticity" reflects the change in expected loans due to changes in the expected value of the unconditional profit-maximizing loan quantity. "Total Elasticity" is the sum of the two. All are expressed as elasticities, the approximate percentage change in expected loans for a percentage change in the explanatory variable.

Bank size has a large negative selection effect, but the total elasticity is positive and relatively large, due to the quantity effect. However, these size results are hard to interpret for two reasons. First, the quantity elasticity depends heavily on the size coefficient in the *OLS* equation, which (from Table 3) has a large standard error. A shortcoming of the elasticity-based analysis as developed by McDonald and Moffitt is that it does not reflect the standard errors of the estimated parameters and therefore does not explicitly recognize that some coefficient estimates are noisier than others.[9] Second, the *SIZE* variable in the regression is the log of assets; a 1 percent change in *SIZE* corresponds to a much larger percentage change in actual bank assets.[10] The elasticity of $E(L)$ with respect to the average bank's *total assets* rather than log of assets is only 0.195, based on the estimates in Table 3. For these reasons, the large calculated elasticity with respect to *SIZE* should be viewed with some skepticism.

The negative effect of the number of branches on loan quantity overwhelms the positive effect on selection, while the opposite is true for *MSAHQ*. Despite the strong statistical significance of *MSAHQ*, the net quantitative impact on farm lending is relatively small. The net impact of out-of-state ownership is large and puzzling, and may be driven by a small number of large multistate organizations in the sample. Foreign ownership has a substantial measured effect, but as noted above this reflects the influence of a single bank, Sanwa.

IX. CONCLUSION

This paper has considered determinants of cross-sectional differences in agricultural production lending by banks in four western states, distinguishing between the influence of

9. In principle, each quantity and selection elasticity could be treated as a statistical estimate, and more precisely estimated elasticities could be given additional weight. An extension along these lines is left for possible future work.

10. Put differently, a 1 percent change in the log of assets is very large relative to its cross-sectional sample variation. A 1 percent change corresponds to a change of 0.085 standard deviations in SIZE, whereas a 1 percent change in MARKET corresponds to only 0.0016 standard deviations for that variable.

structural characteristics of banks and attributes of the markets in which the banks operate. A new measure of the importance of agriculture in each bank's market was developed, based on county-level agricultural production data and the distribution of each bank's branches across those counties. To account for apparent censoring in the farm loan data, the empirical analysis was based on a model of bank decisionmaking in which a bank lends only if the quantity of loans the bank can make exceeds a bank-specific threshold.

How "agricultural" a bank's local markets are is the single most important variable influencing agricultural lending. The proxy for agricultural loan demand, which is based on agricultural output in a bank's market areas, is highly significant in a statistical sense; moreover, of the variables that are statistically significant, it has the greatest quantitative impact on expected farm loans for a typical bank. This "market composition" variable is most influential in determining the quantity of farm loans held by banks that decide to engage in lending. Thus, the results strongly suggest that banks, even in these statewide-branching states, tend to lend to borrowers located near the banks' branches.

A number of additional factors influence the choice of whether or not to engage in farm lending. Notably, the results support the "folklore" that large banks are less likely to hold farm loans, even when they have branches in agricultural areas. Moreover, banks with head offices in MSAs are less likely to engage in farm lending, even after controlling for differences in size and in the agricultural composition of their markets, suggesting that the physical location of key decisionmakers plays an important role. Both large and MSA-headquartered banks are significantly less likely to hold farm loans because they appear to set higher threshold levels for engaging in agricultural lending.

Results related to ownership structure are important, in view of concerns raised by banking industry consolidation; although the cross-sectional results presented here do not directly address the effects of mergers, they are suggestive. The analysis shows that whether or not a bank is owned by a holding company, the size of that holding company, and whether or not it is headquartered in-state or out-of-state have no significant effect on either the probability of engaging in agricultural lending or the quantity of loans held. The absence of such effects, and the overwhelming importance of market characteristics as opposed to bank structure in determining agricultural loan patterns, makes it unlikely that changes in ownership structure resulting from mergers and acquisitions will have a substantial effect on farm credit.

REFERENCES

Gilbert, R. Alton, and Michael T. Belongia. 1988. "The Effects of Affiliation with Large Bank Holding Companies on Commercial Bank Lending to Agriculture." *American Journal of Agricultural Economics* Vol. 70, no. 1, pp. 69–78.

Gunther, Jeffrey W., and Thomas F. Siems. 1995. "The Likelihood and Extent of Bank Participation in Derivatives Activities." Unpublished manuscript, Federal Reserve Bank of Dallas.

Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economics and Social Measurement* Vol. 5, no. 5, pp. 475–492.

Laderman, Elizabeth S., Ronald H. Schmidt, and Gary C. Zimmerman. 1991. "Location, Branching, and Bank Portfolio Diversification: The Case of Agricultural Lending." Federal Reserve Bank of San Francisco *Economic Review* Winter, pp. 24–38.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press.

McDonald, John F., and Robert A. Moffitt. 1980. "The Uses of Tobit Analysis." *The Review of Economics and Statistics* Vol. 62, pp. 318–321.

U.S. Department of Agriculture. *Farm Balance Sheets by State.* Economic Research Service Documents 44013–44062.

U.S. Department of Commerce. *1992 Census of Agriculture, Geographic Area Series.*

Whalen, Gary. 1995. "Out-of-State Holding Company Affiliation and Small Business Lending." Comptroller of the Currency Economic & Policy Analysis Working Paper 95-4.

Zimmerman, Gary C. 1989 "Agricultural Lending in the West." Federal Reserve Bank of San Francisco *Weekly Letter* (December 22).

Intervention, Sterilization, and Monetary Control in Korea and Taiwan

Ramon Moreno

Senior Economist, Federal Reserve Bank of San Francisco.
I would like to thank Mark Spiegel for helpful comments. Research assistance by Barbara Rizzi is gratefully acknowledged.

This paper uses a four-variable vector autoregression model to explore how monetary authorities responded to shocks in Korea and Taiwan over the period 1981.1–1994.12. The analysis reveals that sterilization is an important element of the response to shocks to foreign assets in both economies. In particular, monetary authorities do not appear to be prepared to accept fluctuations in the exchange rate and the money supply that may result from changes in foreign assets, but more readily accept fluctuations in these variables that result from domestic credit shocks. There are also differences in the responses of Korea and Taiwan that suggest that the former may be more insulated from extrernal shocks.

In recent years, the monetary effects of sharp increases in central bank foreign assets—associated with large current account surpluses and capital inflow surges—have attracted much attention, particularly in Asian economies. As discussed by Glick and Moreno (1995), these changes in central bank foreign asset holdings, resulting from efforts to stabilize the exchange rate, have adversely affected monetary control.

In spite of the interest in this subject, there has been relatively little empirical analysis of the characteristics of shocks to foreign assets and the implications for monetary control (an exception is the comparison of Germany and Japan by Glick and Hutchison 1994). Such an analysis can be used to shed light on a number of interesting questions. In particular, it is of interest to inquire how monetary authorities respond to shocks in countries that seek to stabilize the exchange rate. This includes assessing the relative importance of foreign and domestic influence in explaining fluctuations in foreign assets, the degree of sterilization and its effectiveness in limiting the monetary impact of shocks, and how foreign assets respond to changes in domestic credit.

This paper seeks to shed light on these questions by estimating vector autoregression models of Korea and Taiwan. The rest of the paper is organized as follows. Section I motivates the empirical analysis by highlighting the implications of certain balance of payments and central bank accounting identities. Section II discusses model estimation and identification. Section III discusses the results of the model. Section IV provides conclusions and indicates possible areas for future research.

I. Intervention, Sterilization, and Monetary Reserves

To motivate the empirical analysis that follows, it is useful to recall two identities. First, the balance of payments identity implies that the sum of the current (trade) and the capital account balance equals the change in the (net) foreign assets of the central bank.

(1) $$CA + CAP = \Delta FA$$

where Δ is the first difference operator. To illustrate how balance of payments disequilibria can come about, consider an economy in which at the prevailing exchange rate (holding everything else constant) the current account is balanced ($CA = 0$). However, the expected returns on this economy's domestic assets are larger than the expected return on other countries' domestic assets, so there is a tendency for capital inflows in a foreign currency (U.S. dollars). This puts pressure on the exchange rate to appreciate, as holders of U.S. dollars seek to convert to acquire domestic assets. If the local central bank wants to prevent currency appreciation, it will intervene in the foreign exchange market by purchasing U.S. dollars, thus increasing its holding of foreign assets. The outcome is positive capital inflows and an increase in foreign assets ($CAP > 0$, $\Delta FA > 0$). If the central bank does not intervene, the exchange rate will appreciate freely to the point where it is unprofitable for capital to flow in. In this case, balance of payments equilibrium is achieved with $CAP = 0$ and $\Delta FA = 0$. It is apparent from the preceding that changes in foreign assets of the central bank reflect the balance of payments conditions (current account and capital account) of a country and its exchange rate policies.

Second, the simplest version of the central bank balance sheet implies that a change in reserve money (H), or the (monetary) liabilities of the central bank, is identically equal to the change in its assets, which in turn equals the sum of changes in domestic credit (DC) and in foreign assets (FA) of the central bank.

$$\Delta H \equiv \Delta DC + \Delta FA. \tag{2}$$

Equation (2) implies that a change in foreign assets tends to change the supply of money. In many countries, central banks attempt to prevent changes in foreign assets from affecting reserve money by implementing offsetting changes in domestic credit, a policy known as *sterilization.* For example, Asian countries in the 1980s and 1990s responded to large increases in foreign asset holdings by implementing sharp reductions in domestic credit (see Glick and Moreno 1995). Sterilization can also work the other way. In 1994, the Mexican central bank offset the monetary contraction caused by declining foreign asset holdings by increasing domestic credit.[1] In the discussion that follows, foreign assets and domestic credit will at times be described as "monetary variables" because of their close relationship to reserve money.

The preceding discussion illustrates how efforts to stabilize the exchange rate when the balance of payments is not in equilibrium can adversely affect monetary control. It is worth noting, however, that the extent to which balance of payment disequilibria will arise and be reflected in changes in foreign assets depends in part on the degree of capital mobility. In some cases, capital controls may limit balance of payments imbalances and their monetary effects. For example, returning to our previous example, an exchange rate peg that raises the expected return on a country's domestic assets above the foreign rate may not result in capital inflows and increases in central bank foreign assets if capital controls prevent foreigners from investing in domestic assets.[2]

Korea and Taiwan: Policies and Experiences

As discussed above, the extent to which an economy is exposed to balance of payments imbalances, and concomitant pressures on the money supply, will depend on a country's exchange rate policies and the characteristics of capital controls.

Exchange rate policies may range from a rigid peg to a single currency to more flexible basket pegs or managed floats, the latter being closer to the policies adopted in Korea and Taiwan. Korea maintained a multi-currency basket peg until March 1990, when it switched to a market average exchange rate system. Taiwan maintained a managed float for most of the 1980s, also using a market average system.[3] It switched to a pure float for transactions exceeding a minimum amount in April 1989.

1. There is an extensive literature on whether sterilized intervention in foreign exchange markets is effective in stabilizing the exchange rate (this requires that domestic and foreign assets not be perfect substitutes). However, this question is not the focus of the present study.

2. In other cases, however, capital or foreign exchange controls may accentuate balance of payment imbalances. For example, suppose capital controls require exporters to surrender their foreign exchange earnings and prevent domestic residents from investing abroad. If there is a current account surplus at the current exchange rate targeted by the government, the government will in effect purchase the foreign assets accumulated by the economy's trade sector. Foreign exchange controls in this case result in a tendency for the money supply to increase.

3. In Korea, since the introduction of a market average rate (MAR) system on March 2, 1990, the won–U.S. dollar rate has been determined on the basis of the weighted average of interbank rates for won–U.S. dollar spot transactions of the previous day. During each business day, the Korean won–U.S. dollar exchange rate in the interbank market is allowed to fluctuate within fixed margins of plus or minus 1% against the MAR of the previous day. The exchange rates of the won against currencies other than the U.S. dollar are determined in relation to the exchange rate of the U.S. dollar against these currencies in the international market. The buying and selling rates offered to customers are set freely by foreign exchange banks. In Taiwan, the spot central rate of the U.S. dollar against the NT dollar was set daily on the basis of the weighted average of interbank transaction rates on the previous business day. Daily adjustment of the spot rate was limited to 2.25% of the central rate on the previous business day.

Although exchange rate policies allowed some flexibility in currency movements, both economies appear to have been exposed to shocks that contributed to balance of payments disequilibria and complicated efforts at monetary control. For example, the U.S. dollar depreciation against major industrial country currencies in the second half of the 1980s was associated with increases in the current account balances of both economies, in spite of significant appreciation in both the Korean won and the NT dollar against the U.S. dollar. In the case of Taiwan, the expansionary monetary effects of current account surpluses were exacerbated by speculative short-term capital inflows. As is well known, declining U.S. interest rates in the early 1990s were associated with a surge in capital flows to emerging market economies.

The vulnerability of these economies to shocks was to some extent influenced by policies affecting capital mobility in the two economies, which are described in some detail in the Appendix to Moreno (1993). It seems accurate to describe Korea's capital controls as being generally more restrictive than Taiwan's. However, while foreign exchange controls have been very gradually and steadily liberalized in Korea, in Taiwan, the path to liberalization has included some significant policy reversals.

At the beginning of the 1980s, Korea and Taiwan both had restrictions affecting capital flows, including controls on foreign exchange availability for current account transactions, controls on capital flows, restrictions on foreign exchange market transactions (such as forward or futures transactions, swaps or options) and restrictions on foreign access to the domestic financial sector. However, there are at least two important differences. First, in Taiwan, restrictions on current account transactions were eliminated in 1987. In Korea, licenses are still required to obtain foreign currency for current account transactions, and there are limits on the foreign currency holdings of firms (proportional to the size of their international trade activities).

Second, Korea traditionally limited capital inflows as well as outflows (particularly via the banking sector), whereas until 1987, Taiwan restricted only outflows. For example, government approval was needed in Korea in the 1980s for any external borrowing by firms exceeding specified limits: US$200,000 before October 1982, and US $1,000,000 thereafter. There were no comparable limits in Taiwan until 1987, when a surge in speculative capital inflows prompted the government to freeze external bank borrowing. This ceiling on Taiwan banks' external liabilities was lifted gradually in the years that followed.

Differences in capital controls may have been reflected in the components of the balance of payments that led to changes in foreign assets and in the size of the imbalances. For example, as discussed by Glick and Moreno (1995), both Korea and Taiwan experienced increases in foreign assets in the second half of the 1980s as a result of current account surpluses. However, in Taiwan the increase in foreign assets was much larger than in Korea, partly because of significant short-term capital inflows in Taiwan that were interrupted only by the imposition of controls on capital inflows.

II. MODEL ESTIMATION AND IDENTIFICATION

In order to capture key elements of Korea's and Taiwan's balance of payments and monetary sectors, a vector autoregression (VAR) model for each economy was estimated using monthly data. As the primary focus of this study is to assess the implications of balance of payments imbalances for monetary control, the VAR model includes two macroeconomic variables that are likely to affect balance of payments conditions: the end-of-month nominal exchange rate in domestic currency units per U.S. dollar (an increase is a depreciation of the local currency, labeled *XR*, IMF International Financial Statistics (IFS) line ae), and the domestic *CPI* (IFS line 64). It also includes two central bank balance sheet variables that capture the actions of policymakers: foreign assets (labeled *FA*, IFS line 11), and a measure designed to capture the variation in domestic credit (labeled *DC*). The domestic credit measure was estimated by taking the difference between the log of central bank reserve money (IFS line 14) and the log of gross foreign assets (this corresponds to a ratio of the two series in levels).[4] All the other variables were entered in logs.

One potential disadvantage of the measure of foreign assets used is that it does not explicitly take into account changes in the valuation of foreign assets (which is measured in domestic currency) that result from changes in the

4. The measure of domestic credit is approximate for two reasons. First it is the difference of the log reserve money and log foreign assets, rather than the difference in the levels of these series implied by the central bank accounting identity. The measure used in this paper still captures the variation in domestic credit, but allows taking logs even when estimated domestic credit is negative. In the case of Korea, the correlation of the measure used in this paper with the log of the accounting measure over the periods when the accounting measure was positive is 86% (period 1981.1–1988.6) and 82% (1989.12–1994.12). In the case of Taiwan, the accounting measure is negative over most of the sample. However, if the accounting measure is scaled up by a constant to make all values positive and logs are then taken, the correlation with the measure used in this paper is 77%. Second, the measure includes the foreign liabilities of the central bank. However, foreign liabilities are small and vary less frequently in comparison to foreign assets in Asian economies, and do not always appear to be consistently reported; therefore I have chosen not to take them into account explicitly.

value of the exchange rate. This effect, discussed by Takagi (1991), could cloud the interpretation of some of the findings of the model. However, as reported later, the correlation between the residuals of the exchange rate and foreign asset equations is relatively small, which suggests that, on a monthly basis, the effect of changes in the exchange rate on the value of foreign assets is probably not very large. In addition, a similar model was also estimated using total reserves less gold, denominated in U.S. dollars, which are unaffected by changes in the domestic exchange rate vis-à-vis the U.S. dollar. The results do not appear to be too sensitive to this change in variable.[5]

The data span the period 1981.1–1994.12, when large swings in balance of payments conditions in these two economies, as well as in other economies (notably the United States and Japan), were observed. The data set begins in 1981, rather than 1980, to avoid Korea's transition from a fixed exchange rate to a multiple currency basket peg in 1980. As unit root test results are consistent with the data over the sample being trend stationary, the model was estimated using OLS with the series entered in levels, and a linear trend term added to each equation.

Interpreting the VAR Model

Glick and Hutchison (1994) show how a model similar to that estimated in this paper can be derived as a reduced-form representation of an open economy portfolio model with sluggish portfolio adjustment, intervention, and sterilization. In their model, changes in foreign assets and domestic credit reflect changes in private asset demand (which may be attributable to factors such as changes in foreign interest rates) and in domestic credit. They also show that the contemporaneous correlations between foreign assets and domestic credit and the adjustment responses depend on private sector asset demand parameters, the asset speed of adjustment and central bank intervention parameters, as well as on the underlying disturbance. For example, consider a situation where the government seeks to dampen currency fluctuations and then sterilizes the monetary effects of intervention. Then if foreigners decide to acquire more domestic bonds, this will be associated with an initial increase in central bank holdings of foreign assets (due to intervention) and a fall in domestic credit (due to sterilization). The dynamics of convergence to the long-run equilibrium will depend on the various factors cited above.

Identification

The VAR model was identified by orthogonalizing the variance-covariance matrix of the residuals of the four equations using the Choleski decomposition. In performing the decomposition, the causal ordering *XR*, *CPI*, *FA*, *DC* was used. That is, in the current period, the exchange rate is assumed to affect the remaining variables but is unaffected by them; the CPI is affected contemporaneously by the exchange rate and affects foreign assets and domestic credit. Of particular importance is the assumption that shocks to *FA* contemporaneously affect *DC* rather than the other way around.

The sensitivity of the results to the ordering can be assessed by examining the contemporaneous correlations of the residuals. As can be seen in Table 1, correlations with *XR* and *CPI* are relatively small, suggesting that the results are not sensitive to the ordering. However, there is a strong negative correlation between foreign assets and domestic credit (–74 percent for Korea and –49 percent for Taiwan), so the results reported below are sensitive to the assumption that places foreign assets before domestic credit in the causal ordering.

The macroeconomic variables (*XR* and *CPI*) are placed first partly to reflect the focus of the paper on describing the responses of the central bank to shocks. However, this ordering may also be justified by considering the likely sources of contemporaneous covariation in the series. It is unlikely that much of the within-month variation of a highly volatile series like the exchange rate is the result of monthly changes in the CPI, which justifies ordering *XR*

TABLE 1

CONTEMPORANEOUS CORRELATIONS OF RESIDUALS IN PERCENT

	CPI	FOREIGN ASSETS	DOMESTIC CREDIT
KOREA			
Exchange Rate	–6.4	–17.2	13
CPI		–5.1	4.3
Foreign Assets			–74.1
TAIWAN			
Exchange Rate	–7	6.8	–22.9
CPI		6	–4.5
Foreign Assets			–48.9

5. The model in this case consisted of the following variables in logs: exchange rate, the CPI, total reserves less gold (in U.S. dollars), and reserve money.

before *CPI*. Placing *CPI* before *FA* and *DC* also is plausible, as the CPI generally responds to monetary variables with a lag, while policymakers may respond to CPI innovations in the same month. It is less obvious that *XR* should be ordered prior to *FA* and *DC*. However, due to the relatively low correlations of *XR* with these variables cited above, the results are not likely to be sensitive to this assumption.

Placing *FA* before *DC* is consistent with assuming that exogenous shocks to foreign assets lead to offsetting contemporaneous changes in domestic credit, reflecting sterilized intervention by central bankers (see the discussion above and the more formal exposition by Glick and Hutchison, 1994). This causal ordering can be justified by the observation that episodes of balance of payments imbalances in Korea and Taiwan appear to have been triggered by certain discernible international events. For example, in Taiwan, both the large dollar depreciation in the mid-1980s and the period of declining U.S. interest rates in the early 1990s appear to have been associated with unusually high foreign asset levels. In the reduced-form specification of the model such events would be captured in variations in foreign assets that would be contemporaneously associated with sterilization. In addition, in the case of Korea, it can be argued that restrictions on capital flows may make it less likely that policymakers would need to offset changes in domestic credit with changes in foreign assets contemporaneously.

However, the reader should bear in mind that an alternative interpretation of the correlation is that changes in domestic credit lead to changes in the exchange rate that the government seeks to avoid through intervention. Given the high correlation between foreign assets and domestic credit, adopting such an interpretation to identify the model (by reversing the causal ordering) affects the results reported below.

III. MODEL RESULTS

The model was used to address two broad questions. First, how do policymakers respond to shocks to the economy, as indicated by the behavior of foreign assets and domestic credit? Second, what are the main sources of variation in these policy variables? Of particular interest is whether domestic credit fluctuations reflect disturbances in foreign assets, which would indicate that sterilization is an important element in domestic credit policy. Also of interest is whether foreign asset fluctuations reflect changes in domestic credit, as this would suggest that policymakers intervene in foreign exchange markets to offset the effects of domestic monetary policy.

To address the first question, the dynamic responses to selected one-standard-deviation shocks are illustrated in Figures 1 to 3. To address the second question, statistical tests were first performed to determine which variables help predict foreign assets and domestic credit in the respective regression equations. The results of the tests (null hypothesis that the block of lagged coefficients is zero) are reported in Table 2. In addition, the contribution of each variable to the variance of the forecast error of foreign assets and domestic credit also was examined. The results of these decompositions are reported in Tables 3 and 4.

Dynamic Responses to Shocks

In order to assess policymakers' responses to shocks in Korea and Taiwan as well as the implications for the money supply, we examine the responses to shocks of foreign assets and domestic credit, and the net effect on reserve money.[6] As this study is largely concerned with conditions in foreign exchange markets, only the dynamic responses over a 60-month horizon to shocks to the exchange rate, foreign assets, and domestic credit are illustrated (in Figures 1, 2, and 3, respectively). The underlying series are in logs, so the dynamic responses can be interpreted as log deviations from the baseline path. The shaded areas define a confidence band that excludes the upper and lower 1 percent fractiles.

Responses to an exchange rate shock. The point estimates in Figure 1 indicate that a shock to the Korean exchange rate (depreciation) is associated with a decline in foreign assets and an offsetting increase in domestic credit. In contrast, in Taiwan, an exchange rate depreciation appears to be associated with an *increase* in net foreign assets and an offsetting reduction in domestic credit. The decline in Korean foreign assets indicates that Korean monetary authorities intervene to offset changes in the exchange rate, that is, they "lean against the wind," while in Taiwan intervention apparently "leans with the wind" in the months that follow the shock to the exchange rate.[7] In both economies, the offsetting movement in domestic credit indicates that intervention is largely sterilized. These policy actions are reversed gradually, after a period of about four

6. Since domestic credit is defined as the difference between the log reserve money and log foreign assets, the response of log reserve money can be computed by adding the respective responses of foreign assets and domestic credit

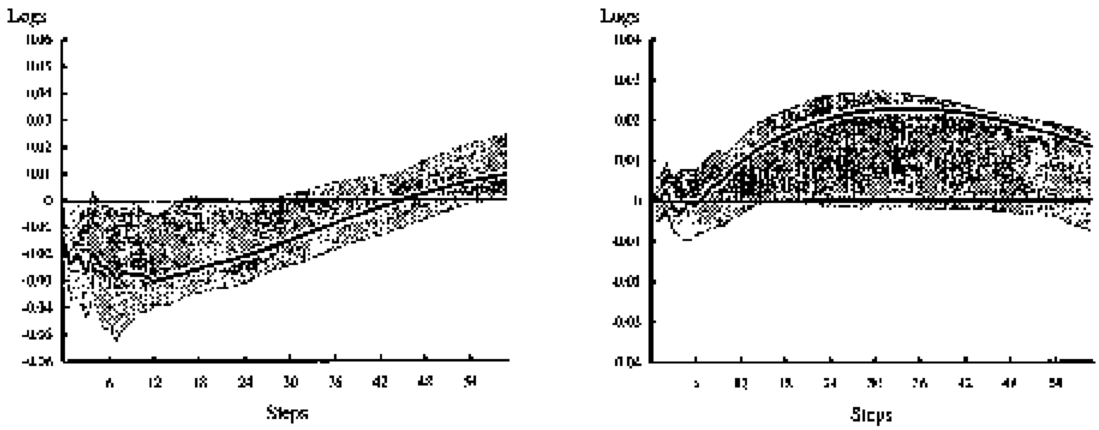
7. In interpreting this result, the reader may bear in mind that responses to shocks to the exchange rate do not fully describe intervention policy in response to disturbances in foreign currency markets. As discussed below, there are also responses to shocks to foreign assets.

FIGURE 1

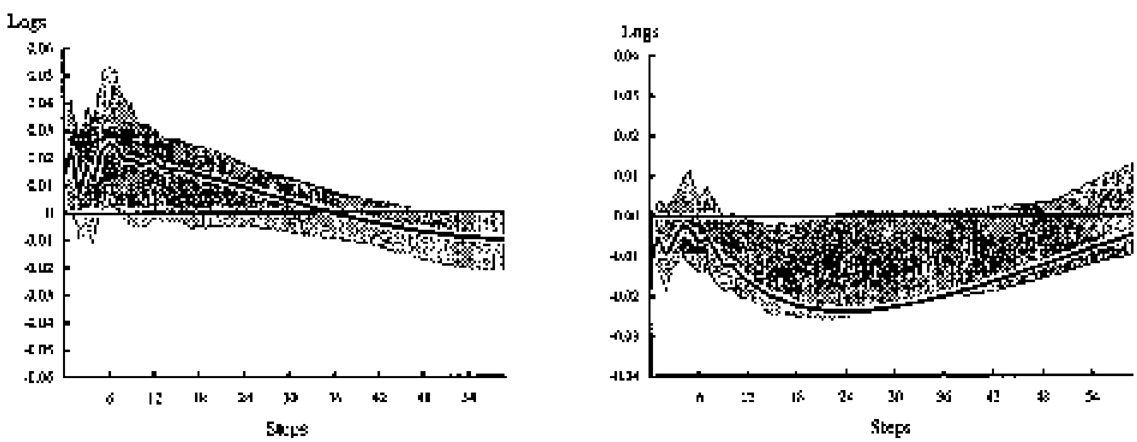
RESPONSES TO A SHOCK IN EXCHANGE RATE

KOREA	TAIWAN

FOREIGN ASSETS



DOMESTIC CREDIT



MONETARY RESERVES

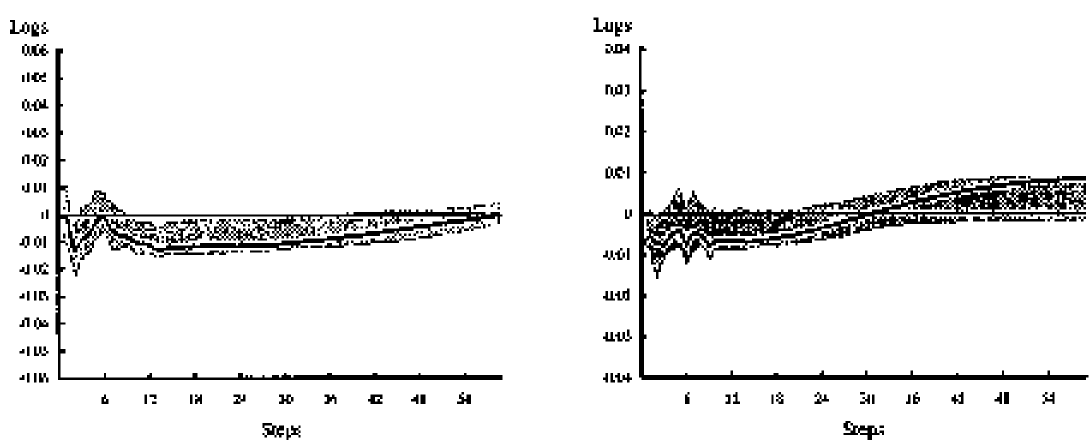
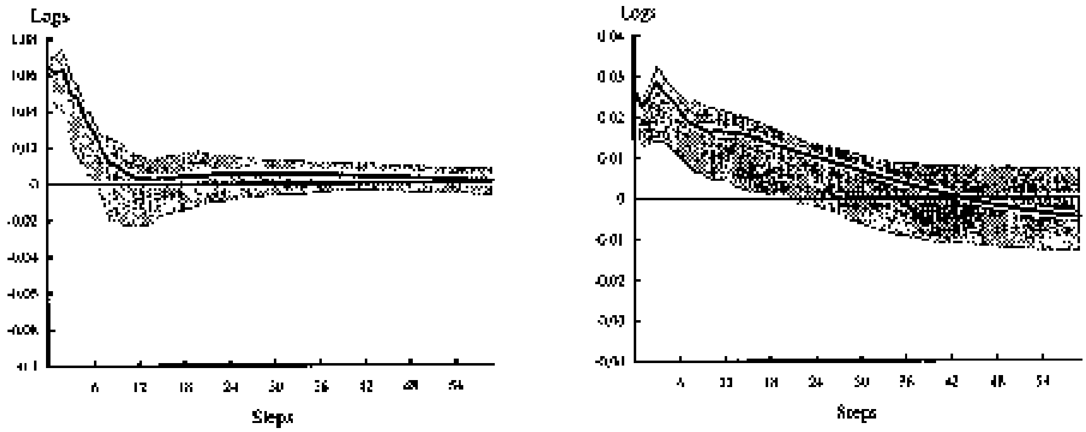


FIGURE 2

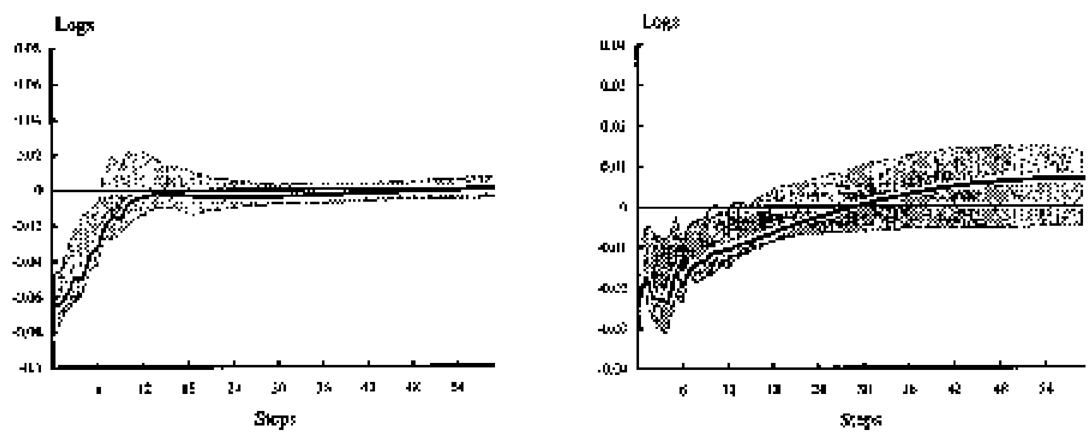
RESPONSES TO A SHOCK IN FOREIGN ASSETS

KOREA	TAIWAN

FOREIGN ASSETS



DOMESTIC CREDIT



MONETARY RESERVES

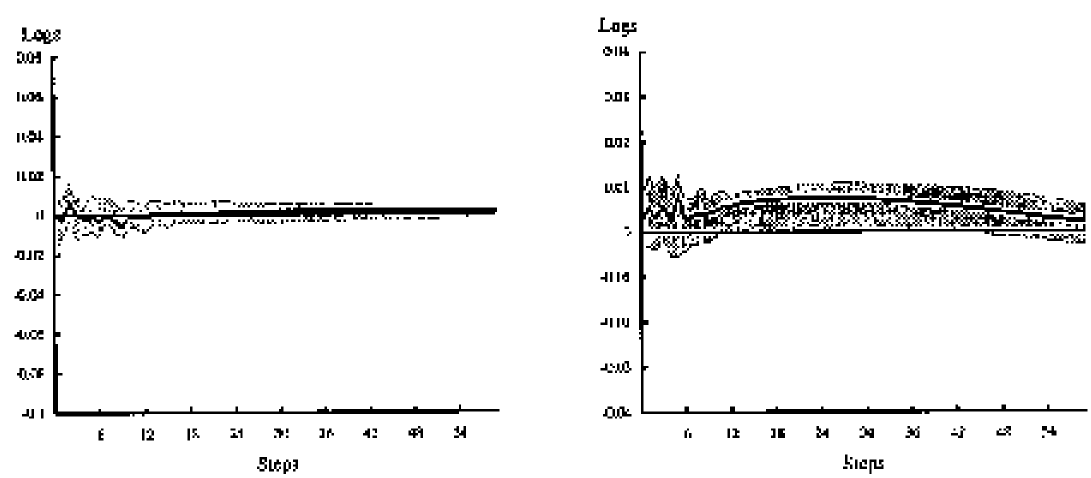
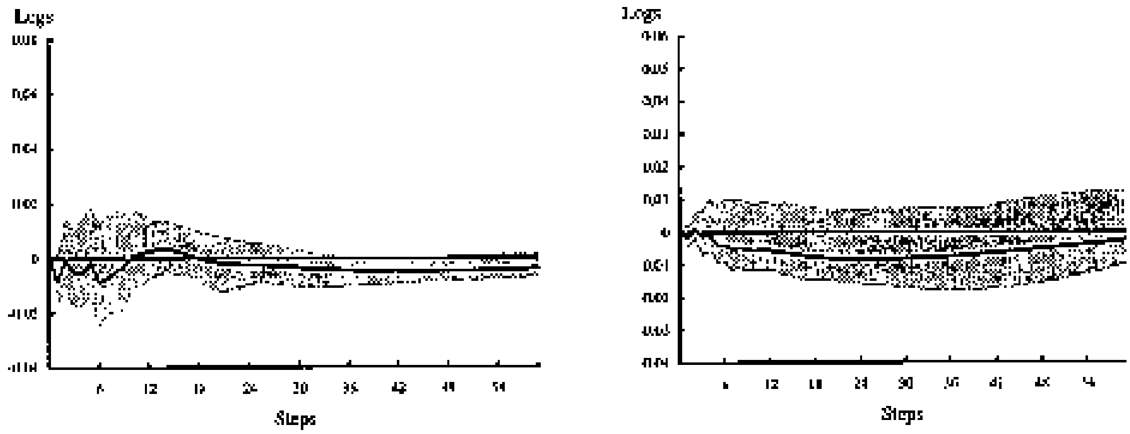


FIGURE 3

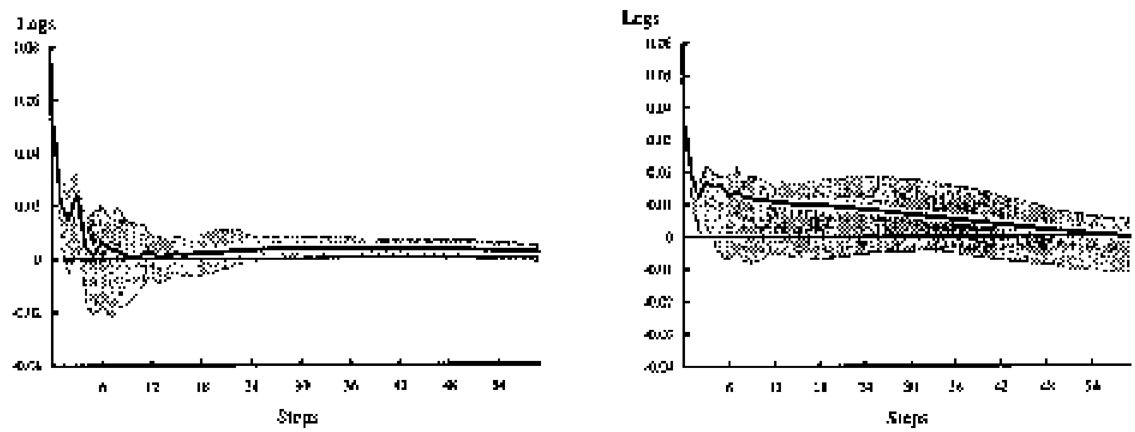
RESPONSES TO A SHOCK IN DOMESTIC CREDIT

KOREA	TAIWAN

FOREIGN ASSETS



DOMESTIC CREDIT



MONETARY RESERVES

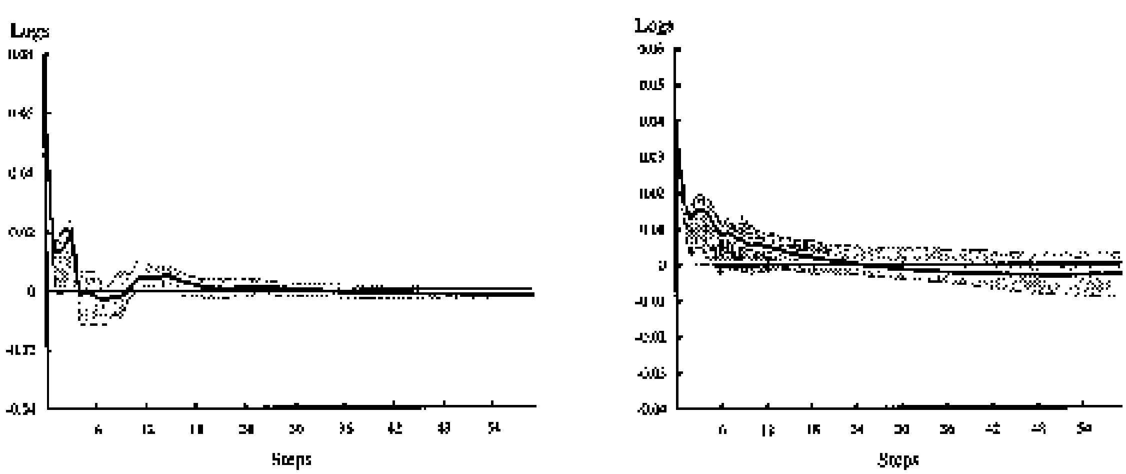


TABLE 2

TESTS OF PREDICTIVE ABILITY

	FA	*DC*
KOREA		
XR	1.3***	5.9*
CPI	23.2	0.0***
FA	0.0***	0.7***
DC	35.4	0.0***
TAIWAN		
XR	29.7	17.2
CPI	0.0***	0.8***
FA	0.0***	0.3***
DC	41.6	0.0***

Notes:
*** Significant at 1%
* Significant at 10%

TABLE 3

VARIANCE DECOMPOSITIONS FOR FOREIGN ASSETS

	Step	*XR*	*CPI*	*FA*	*DC*
KOREA					
	1	4	0	96	0
	24	40	9	50	1
	60	42	12	44	2
TAIWAN					
	1	1	1	98	0
	24	19	25	52	5
	60	48	22	25	6

TABLE 4

VARIANCE DECOMPOSITIONS FOR DOMESTIC CREDIT

	Step	*XR*	*CPI*	*FA*	*DC*
KOREA					
	1	2	0	53	45
	24	20	5	60	15
	60	22	9	54	15
TAIWAN					
	1	5	0	26	69
	24	34	17	24	25
	60	53	14	16	17

years. As shown in Figure 3, the net effect on reserve money is on balance negative in both economies.

While the apparent contrasting responses suggested by the point estimates are interesting, the confidence bands indicate that actual differences in policy responses may be less stark than those indicated by the point estimates. The hypothesis that the response of Korean foreign assets is zero can be rejected over some interval of the dynamic response, but this same hypothesis is generally not rejected in the case of the other responses in Figure 1.[8]

Responses to a shock to foreign assets. The point estimates in Figure 2 indicate that an unexpected increase in foreign assets in Korea is associated with an offsetting decline in domestic credit, evidence of a strong sterilization response. The responses to the shock are reversed relatively quickly, within about twelve months. On balance, the reserve money response fluctuates around zero. In contrast to the responses to exchange rate (and CPI) shocks, the responses of foreign assets and domestic credit to a foreign asset shock are estimated with sufficient precision so that, based on the confidence bands, the hypothesis that they are zero is easily rejected over some interval. However, the hypothesis that the response of reserve money is zero cannot be rejected, indicating that sterilization of shocks to foreign assets is complete in Korea.

In the case of Taiwan, an increase in foreign assets is also associated with an offsetting decline in domestic credit. However, the reversal in the gross responses is far more gradual than in Korea, taking 30 to 40 months. One possible explanation for the persistence of a shock to foreign assets is that restrictions on capital flows are more limited in Taiwan. For example, as noted earlier, in 1986–1987 Taiwan experienced large capital inflow surges that persisted for months, as speculators sought to capture gains from expected continued NT dollar appreciation. Such persistent surges have not been observed in Korea. In further contrast to Korea, in Taiwan the point estimates indicate that sterilization does not fully offset the shock to foreign assets. A foreign asset shock is followed by an increase in reserve money that dies out very gradually. While this result should be interpreted with caution due to wide

8. Differences between the two economies are also apparent in the responses to a shock to the CPI (not shown). In Korea, such a shock is associated with a temporary increase in foreign assets and a contraction in domestic credit, while in Taiwan it is associated with a temporary (but persistent) decline in foreign assets and an increase in domestic credit. On balance, the monetary reserves response to a CPI shock in Korea is positive, while it appears to be negative in Taiwan. The confidence bands once again indicate that any inferences should be made with caution, as the null that the dynamic response is zero in many cases cannot be rejected.

confidence bands, it suggests that Korea may be able to achieve greater stability in the exchange rate with less net change in the money supply, possibly because of greater capital controls.[9] Indeed, in Figure 4, a one standard deviation shock to foreign assets in Korea results in an exchange rate change that is smaller than (according to the point estimate) or at least as large as (according to the confidence band) in Taiwan.

Responses to a shock to domestic credit. A shock to domestic credit in Korea is associated with an erratic response in foreign assets that is not significantly different from zero. The result is an increase in reserve money that is quickly eliminated. A similar pattern of responses is apparent in Taiwan. Thus, in sharp contrast to the tendency to dampen the monetary impact of shocks to foreign assets, shocks to domestic credit by and large are not offset by intervention in either Korea or Taiwan. This result may be viewed in the context of traditional models of balance of payments crises and abandonment of exchange rate pegs originally developed by Krugman (1979). In these models, domestic credit creation leads to a depletion in foreign assets that eventually leads to the abandonment of an exchange rate peg. Such an effect appears to be absent in both Korea and Taiwan, suggesting that monetary authorities in these two economies are prepared to accept fluctuations in the exchange rate and the money supply that may result from changes in domestic credit, even if they are not prepared to accept fluctuations in these variables that result from changes in shocks originating in the external sector. The effect also may indicate that shocks to domestic credit do not significantly affect the balance of payments for other reasons, such as imperfect capital mobility.

Predictive Ability and the Relative Importance of Shocks

While the preceding discussion gave an idea of the qualitative responses of foreign assets and domestic credit to economic shocks, it did not explicitly identify the main factors that drive these two variables. To shed light on this question, we first identify which variables help predict foreign asset and domestic credit behavior by testing exclusion restrictions. We also assess the contributions of different variables to the variance of the forecast errors of foreign assets and domestic credit.

The tests of exclusion restrictions are presented in Table 2. Both foreign assets and domestic credit are predicted by their own lags and by either the lagged exchange rate or the lagged CPI. One interesting result that emerges from the table is that in both Korea and Taiwan, lagged domestic credit does not help predict foreign asset behavior, while lagged foreign assets do help predict domestic credit.

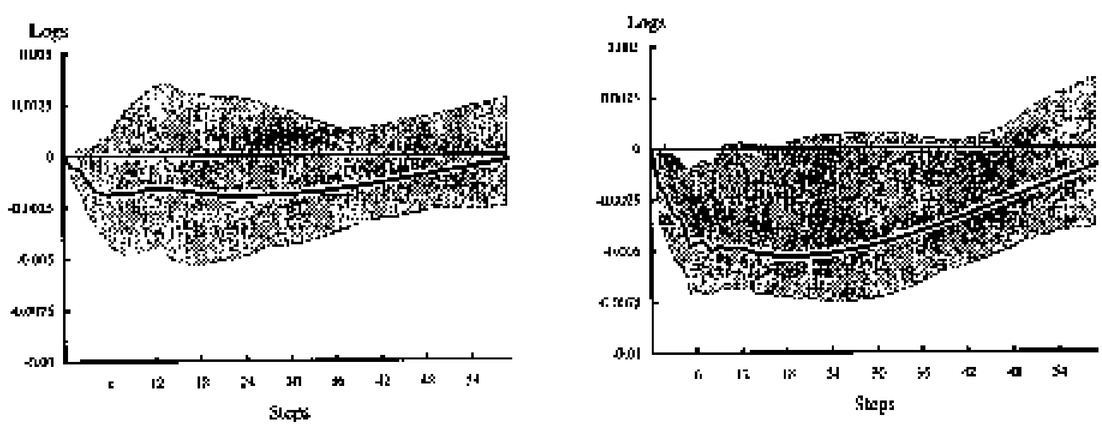
We can exploit the identifying restrictions of the model to estimate the contribution of innovations in each of the variables to the variance of the forecast error in foreign assets and domestic credit. The results are reported in Table 3.

Table 3 reveals that, in the very short run, the variance in foreign assets in both Korea and Taiwan cannot be explained by other variables. After 24 months the exchange rate accounts for about half of the variance in foreign assets in Korea, with innovations in foreign assets accounting for most of the rest. In the case of Taiwan, shocks to foreign assets account for most of the variance up to 24

FIGURE 4

Exchange Rate Responses to a Shock in Foreign Assets

Korea | Taiwan



months; at 60 months shocks to the exchange rate and the CPI also play a role.

Table 3 also reveals that at a 24-month horizon, foreign assets account for about half of the variance in domestic credit in Korea and nearly a fourth of the variance in Taiwan. Innovations in the exchange rate also play a role, particularly in Taiwan, where they account for about half of the variance in domestic credit at a 60-month horizon.

Thus, the variance in foreign assets appears not to reflect innovations in domestic credit, reinforcing the impression conveyed by the dynamic responses, namely, that the extent to which monetary authorities intervene in foreign exchange markets to offset changes in domestic credit is small. In contrast, the variance in domestic credit appears to reflect innovations in foreign assets as well as in the exchange rate. The influence of foreign assets in domestic credit may reflect sterilization policies, while the influence of the exchange rate may indicate that policymakers rely not only on intervention, but also on domestic credit creation, to stabilize the exchange rate.

IV. CONCLUSIONS

This paper has developed an empirical model to analyze the monetary implications of intervention and sterilization policies in Korea and Taiwan. At least two interesting results emerge from the empirical analysis.

First, sterilization is an important element of the response to shocks to foreign assets in both economies. Shocks to foreign assets were largely offset by shocks to domestic credit, and were therefore generally associated with little net change in reserve money, particularly in the case of Taiwan. In line with this, a significant proportion of the variation in domestic credit reflects innovations in foreign assets.

It is interesting that the converse was not true. In general, shocks to domestic credit were not associated with fully offsetting movements in foreign assets, indicating that in contrast to the traditional description of unsustainable exchange rate pegs (Krugman 1979) domestic credit creation does not lead to asset depletion in this empirical model of Korea and Taiwan. Neither is domestic credit contraction associated with unsustainable asset accumulation in this model (Grilli 1986). In particular, it appears that monetary authorities in these two economies are prepared to accept fluctuations in the exchange rate and the money supply that may result from changes in domestic credit, while they are not so prepared to accept fluctuations in these variables that result from changes in foreign assets.

Second, there are some differences in the responses of Korea and Taiwan that indicate that Korea may be more insulated from foreign asset shocks. This is consistent with institutional practices that suggest that Korea may have had more restrictive capital controls over the sample period. Shocks to foreign assets are quickly reversed in Korea, while they appear to be much more persistent in Taiwan. This may reflect more persistent foreign exchange market speculation in Taiwan made possible by less restrictive capital controls. In addition, Korea has tended to sterilize shocks to foreign assets more fully than has Taiwan, achieving a smaller exchange rate change with a far smaller change in the money supply.

The preceding conclusions are sensitive to the identifying assumptions used in this model, specifically the assumption that foreign assets are contemporaneously exogenous to domestic credit. However, this assumption does not appear to be unreasonable, since episodes with large swings in the balance of payments and in foreign assets appear to have been associated with certain identifiable international events, such as the dollar depreciation of 1985–1987.

A number of additional questions warrant further research. It would be of interest to investigate further the apparent asymmetry in policymakers' responses. Policymakers appear to be concerned with offsetting the monetary and exchange rate effects of balance of payments shocks, but less concerned with offsetting domestic credit shocks. It would also be interesting to examine to what extent foreign asset behavior and domestic monetary conditions are influenced by a more disaggregated set of external shocks, such as the value of the U.S. trade-weighted dollar against the currencies of major industrial countries or U.S. interest rates. This can be done by expanding the VAR model explicitly to take account of these variables.

REFERENCES

Glick, Reuven, and Ramon Moreno. 1995. "Capital Flows and Monetary Policy in East Asia." In *Monetary and Exchange Rate Management with International Capital Mobility,* (ed.) Hong Kong Monetary Authority. Hong Kong.

__________, and Michael Hutchison. 1994. "Foreign Reserve and Money Dynamics with Asset Portfolio Adjustment: International Evidence." Federal Reserve Bank of San Francisco Center for Pacific Basin Monetary and Economic Studies, Working Paper No. PB94-09.

Grilli, Vittorio. 1986. "Buying and Selling Attacks on Fixed Exchange Rate Systems." *Journal of International Economics* 20, pp. 143–156.

Krugman, Paul. 1979. "A Model of Balance of Payments Crises." *Journal of Money, Credit and Banking* 11, pp. 311–325.

Moreno, Ramon. 1993. "Exchange Rate Policy and Insulation from External Shocks: The Experience of Taiwan and Korea, 1970–90." Federal Reserve Bank of San Francisco Center for Pacific Basin Monetary and Economic Studies Working Paper PB 93-05.

Takagi, Shinji. 1991. "Foreign Exchange Market Intervention and Domestic Monetary Control in Japan, 1973–1989." *Japan and the World Economy* 3, pp.147–180.

Hedonic-Based Price Indexes for Housing: Theory, Estimation, and Index Construction

Nancy E. Wallace

Associate Professor of Real Estate, University of California at Berkeley and Visiting Scholar. I gratefully acknowledge the comments and suggestions of Brian Motley, Joe Mattey, and seminar participants at the Federal Reserve Bank of San Francisco. Any errors and omissions are my own.

Housing price indexes should not confound the effect of changes in quality with the effects of changing house prices. A recent nonparametric regression technique, loess, *allows flexible estimation of the hedonic price function and centers the estimation at fixed points, such as the beginning or ending period housing characteristics. Indexes using these estimates are consistent with the requirements of Laspeyres and Paasche price indexes. The technique is used to obtain indexes for fifteen municipalities in Alameda County from 1970:Q1 through 1995:Q1. The nonparametric hedonic-based indexes provide better controls for the effect of quality evolution on price movements than alternative methods.*

Residential real estate accounts for about 70% of the wealth portfolio of the average U.S. household. Residential housing assets also provide the collateral support for the residential mortgage market with an outstanding stock of about $3.2 trillion in 1996. In addition to the size of the housing market, the market is also notable because it is prone to boom and bust cycles. The unpredictability of these cycles introduces considerable volatility into the wealth positions of the average household and to the mark-to-market value of residential mortgages held in portfolio by financial institutions, pension funds, insurance companies, and individual investors.

In the last four or five years there has been a concerted effort to develop valuation methods that give market participants more accurate information about residential real estate price levels and returns over time. One reason for this interest is growing investor demand for measures of value and return that are comparable to the wide variety of indexes available for the bond and stock markets. A second reason is the increased sophistication of real estate investors and the more widespread use of modern tools of financial analysis, such as portfolio allocation models, option pricing models, and advances in structuring real estate investment vehicles through securitization. A final reason is the search for cost efficiencies in mortgage lending and real estate portfolio management. Cost efficiency has led to the increased use of automated appraisal and underwriting technologies and reliance on capital-at-risk models which require accurate measures of risk and return by asset class. For these reasons, many practitioners would like housing price indexes that are transaction-based and that can be produced with high levels of reporting frequency and accuracy.

Most currently available housing price indexes are transaction-based; reporting frequency and accuracy, however, remain unresolved issues. All the available strategies must contend with the fact that transactions are infrequent and that information on the terms of sale and the characteristics of the properties are costly to obtain. Choosing among existing methods to obtain housing price indexes must be done on the basis of the desired application. The choices here would include whether the index is intended to proxy the price per unit of the housing stock, whether it is intended to estimate the changing price level (or returns) of a "representative" house over time, or whether what is sought is an estimate of the value of a particular house or a portfolio of houses over time.

The purpose of this paper is to consider hedonic-based indexes of housing prices. The indexes are evaluated using a comprehensive transaction-based data set for residential sales from first quarter 1970 through first quarter 1995 for fifteen municipalities in Alameda County (171,131 transactions). The intent of this review is not to demonstrate the superiority of the hedonic-based method, but rather to highlight the empirical importance of the theoretical assumptions that underlie it. For some applications, the hedonic-based indexes would not be expected to differ greatly from strategies such as repeat sales indexes. This would be the case for applications in which there are large numbers of repeat transactions in a housing market and the market is characterized by low levels of production or remodeling. The repeat sales and hedonic methods would be expected to be equivalent if it is reasonable to assume that both the levels and prices of the underlying housing attributes, such as bathrooms and bedrooms, have remained the same over time. For other applications, however, the differences between the methods are important both theoretically and substantively. This would be the case in markets for which it is not reasonable to assume that attribute prices and levels are constant over time

The advantages of hedonic-based methods must also be evaluated relative to their cost of application. These costs vary greatly by state. States such as California have a number of high quality vendors of residential transaction data, while other states do not have these services commercially available. Thus, the appropriate choice of price index methodology also depends upon data availability.

The paper is organized into five sections. In Sections I and II, I will survey the theoretical framework for hedonic price indexes and housing price index number construction. The purpose of this overview is to highlight the assumptions required to obtain econometrically estimable price indexes and the economic theory that supports these assumptions. This conceptual framework is important because it establishes guidelines for the estimation methods and allows for meaningful interpretation of empirical results. In Section III, I will discuss two non-parametric formulations for price index composition using hedonic price functions. I apply these strategies using transaction data from Alameda County and evaluate the results. Section IV provides a graphical evaluation of the price indexes constructed from hedonic-based methods and those using repeat sales. Section V concludes.

I. HEDONIC PRICE FUNCTIONS FOR HOUSING

In economics, housing is usually treated as a heterogeneous good, defined by a set of characteristics such as square footage, bathrooms, public service amenities, and location, among many others. The number of such characteristics is indexed by j and the number of houses produced by n. The price of housing is defined by a hedonic price function, which is a mathematical relationship between the prices of the composite housing assets and the quantities of characteristics embodied in them. Thus,

$$P = h(x), \tag{1}$$

where P is an n-element vector of house prices, x is a $j \times n$ matrix of house-specific characteristics.

In the housing market, the economic decisionmaking behavior of market participants (behavior related to what is being demanded or supplied) really pertains to housing characteristics. A housing transaction is a tied sale of a set of characteristics.

To formalize the assumption that characteristics are the true arguments of the consumption- and/or production-optimization strategies of economic agents, assume for simplicity that there is only one heterogeneous good, housing, and the utility function for a household can be written as:

$$Q = Q(q(x), c), \tag{2}$$

where Q is utility, $q(\cdot)$ is a function over the housing characteristics, and c is all other homogeneous consumption goods. The production of housing assets can be represented as the joint output of a bundle of housing characteristics. Assuming the usual capital, labor, and materials (KLM) production function this can be written as:

$$t(x, K, L, M) = 0, \tag{3}$$

where $t(\cdot)$ is a transformation relationship in production.

It is well-established that the hedonic price function, $h(\cdot)$, does not represent a "reduced form" for supply and demand functions derived from the utility or production functions (Rosen 1974, Epple 1987). Instead the hedonic, $h(\cdot)$, should be thought of as the binding constraint in the optimization problems of producers and purchasers of housing.[1] Rosen (1974) shows that as long as there is increasing marginal cost of characteristics for producer/sellers and a

1. Rosen (1974) identifies special cases in which the hedonic price surface can be identified. These cases include; (1) when there is only a single type of buyer the $q(\cdot)$'s are identical so that the $h(\cdot)$ is uniquely identified by the functional form of $q(\cdot)$ and (2) when there is only a single type of seller the $t(\cdot)$'s are identical so that the $h(\cdot)$ is uniquely identified by the functional form of the $t(\cdot)$. In the former case the hedonic

constraint on unbundling the attribute package, the hedonic function is likely to be nonlinear. The nonlinearity of the hedonic constraint implies that relative characteristics prices are not fixed and instead are uniquely determined for each buyer by the buyer's location on the hedonic surface.

To illustrate the problem, consider Figure 1. It shows two nonlinear hedonic price contours for houses with two characteristics ($P_1 = h(x_1,x_2)$ and $P_2 = h(x_1,x_2)$) at a given time period. The $P_1(P_2)$ contour describes all possible types of houses that sell for price $P_1(P_2)$ and are composites of the two characteristics, x_1 and x_2 , such as square footage and number of rooms. The slope of the $P_1(P_2)$ contour defines the marginal purchase costs for the respective characteristics.

Buyers *l* and *k* in this market select the house type with characteristics that are closest to optimal. The point *A* represents the tangency of q_l and t_E with the hedonic price surface P_1 for consumer *l* and producer *E*, and the point *B* represents the tangency of q_k and t_F with the hedonic price surface for consumer *k* and producer *F*. The total expenditure on characteristics, the price of quality, is the slope of the hedonic surface above an expansion path such as *AA′* shown in Figure 1. Figure 1 also shows that housing types with different characteristics, though available at the same price, are chosen by different consumers. As shown, buyer *l* purchases house type *A* with characteristic level x_{1A} and x_{2A}. Rosen (1974) shows in markets with many buyers and sellers, the hedonic contours will trace out an envelope of tangencies between the bid and offer prices of the buyers and sellers. The realism of the nonlinear hedonic constraint requires that housing characteristics must be bought and sold in tie-in sales. We would expect tie-in sales for housing because housing characteristics cannot be unbundled from the geographic location of the house.

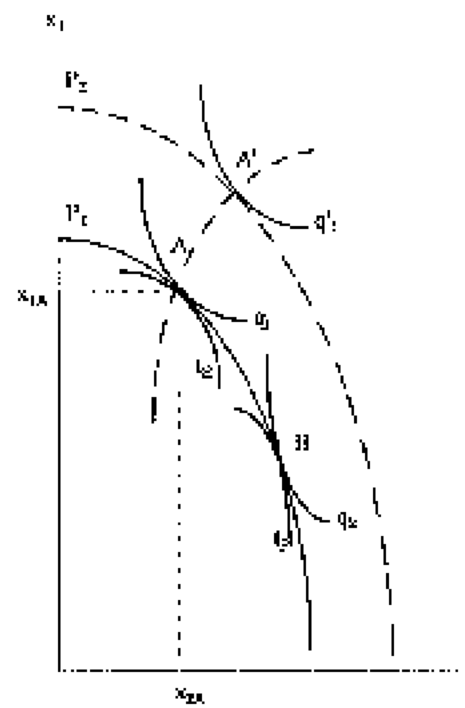
The discussion above and Figure 1 suggest that functional forms used to estimate hedonic prices should allow for the possibility of nonlinearity in the relationship between the price of the house and the prices and quantities of the underlying attributes. They also suggest that the divergence of tastes and technologies is an essential part of the theory of hedonic price functions and that "representative consumer" models may not describe market outcomes well.

The derivation of hedonic price functions outlined above views the price of houses as determined in a flow market—where housing supply comes from producers of housing and price equilibrates the demand for new houses to the supply of new housing. An alternative view focuses on the stock of existing housing. In this case prices, again defined for attributes, guide both bids and offers for locational choices with respect to packages of housing characteristics (Alonso 1964, Muth 1969). The hedonic price function is determined by market clearing conditions in which the tie-in sales of attributes at each location equal the amount demanded by buyers. In equilibrium buyers and sellers are perfectly matched, and again the hedonic price surface is likely to be nonlinear.

The primary implication of the theoretical literature is that hedonic price functions are likely to be nonlinear because locational uniqueness leads to tie-in sales. Thus, observed housing prices reflect both the implicit prices of characteristics in housing packages and the quantities of characteristics embedded in the housing units sold. The theoretical structure of the market-clearing mechanisms for housing does not suggest that it can be assumed either that at a given market period the relative implicit prices for attributes are the same or that across market periods the implicit characteristics prices for the same packages of housing services remain constant. This inherent difficulty in interpreting observed housing price levels presents a particular problem for solving the index number problem for

FIGURE 1

Hedonic Frontiers



frontier would be concave to the origin following classical utility theory, and in the latter case the frontier would be convex to the origin because it is a production transformation curve. Neither of these two cases is particularly helpful in the housing market since neither condition would be expected to be true.

housing—how to measure average price level changes across time periods.

II. EMPIRICAL HEDONIC PRICE INDEXES FOR HOUSING

A comprehensive review of the economic theory of index numbers and their use in housing markets is beyond the scope of this paper.[2] In brief, the index number problem for housing has much in common with the problem of index number construction for other goods and services. In a given base period, a "representative" consumer takes base period prices as given and buys a utility maximizing combination of goods and services, including housing services. In later periods, the consumer faces new sets of prices and selects alternative bundles of goods and services. The index number problem is to determine how much the cost of living has changed between periods if the consumer retains the original standard of living. The theory of hedonic price indexes for housing follows this literature, with the only modification being that economic agents select across composite characteristics.

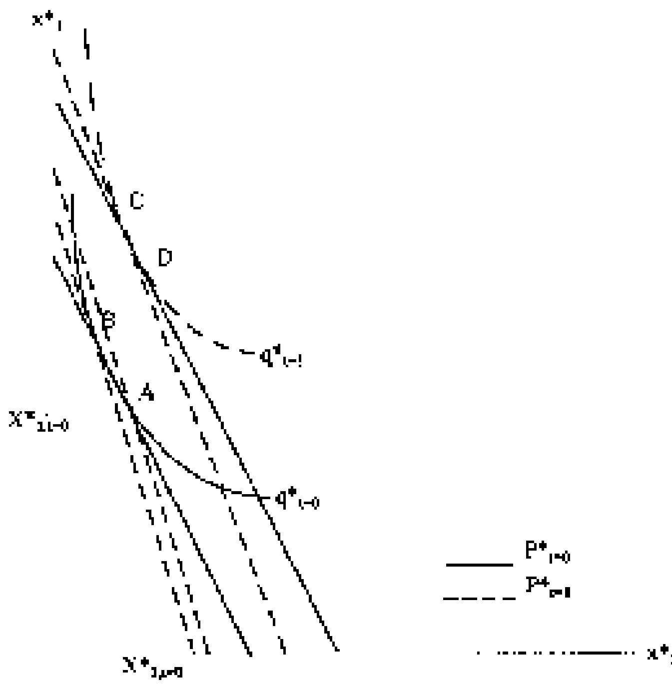
Figure 2 shows the price index problem for the more standard homogeneous goods case. Figure 2 shows optimal consumption in the two-commodity case, x_1^* and x_2^* over two periods, $t = 0$ and $t = 1$. As shown, there are several ways to measure relative price level changes. The first way, which is called Laspeyres price indexes,[3] holds the base period commodity bundle ($x^*_{1,\,t=0}$, $x^*_{2,\,t=0}$) fixed at point *A* and measures how much the base period bundle would cost at the subsequent period prices, $P^*_{t=1}$. The problem with the measure is that it does not account for the fact that at the new prices, $P^*_{t=1}$ the consumer would be expected to substitute to a new combination of goods, point *B*, while holding the level of well-being, or standard-of-living, $q^*_{t=0}$, constant. Because the Laspeyres price index is weighted on the initial bundle, point *A*, it does not account for the substitution effect and thus has an upward bias as a measure of the cost to the consumer of keeping the initial standard of living once prices have changed.

The alternative measure, the Paasche index, is similar except that it uses the subsequent period consumption bundle as its reference point, point *C*, for the subsequent period standard of living, $q^*_{t=1}$, and measures how much the subsequent period's consumption bundle would cost at the previous period's prices, $P^*_{t=0}$.[4] Here again, because the Paasche index weights on the $t = 1$ period's optimal consumption bundle, point *C*, it does not account for the substitution effect, point *D*, and thus has a downward bias as a measure of the cost to the consumer of keeping the $t = 1$ period standard of living, $q^*_{t=1}$. This bias arises because the bundle represented by point *C* was not the one actually chosen by the consumer in the base period, so computing its costs at the new prices overstates the cost of living in that period.

If one knew the consumer's preferences, either $q^*_{t=1}$ or $q^*_{t=0}$, one could measure the substitutions that would be made in order to maintain a constant level of well-being subsequent to a shift in relative prices for the two commodities. In fact, it would be possible to measure exactly

FIGURE 2

LASPEYRES AND PAASCHE PRICE INDEXES



2. For excellent discussions about the theory of index numbers and cost-of-living indexes, see Motley (1992), Pollak (1991), and Diewert (1983).

3. The general *n*-commodity Laspeyres price index measures the increase in prices from base period 0 to period *t* holding the initial consumption level constant:

$$Index_{Laspeyres} = \frac{\sum_{n=1}^{N} p^*_{nt}\, x^*_{n0}}{\sum_{n=1}^{N} p^*_{n0}\, x^*_{n0}} .$$

4. The general *n*-commodity Paasche price index measures the increase in prices from base period 0 to period *t* holding the t^{th} period consumption level constant:

$$Index_{Paasche} = \frac{\sum_{n=1}^{N} p^*_{nt}\, x^*_{nt}}{\sum_{n=1}^{N} p^*_{n0}\, x^*_{nt}} .$$

the difference in the minimum costs of obtaining any fixed level of satisfaction at any given set of prices. Such an exact cost-of-living index would be a measure of the true cost of maintaining a fixed level of satisfaction. The Laspeyres and Paasche indexes thus would be only approximations to the hypothetically exact cost-of-living indexes because they hold the observable consumption bundles fixed rather than the unobservable constant levels of satisfaction.

The economic theory of price indexes for heterogeneous goods such as housing follows the same logic as that of homogeneous goods, represented in Figure 2 (Triplett 1987, 1989). Instead of considering the consumer's optimal consumption combinations of commodities, we would consider their optimal consumption combinations of characteristics of housing. This implies that the axes x_1^* and x_2^* of Figure 2 should be redefined as composite characteristics of the housing asset, and the budget constraints, $P_{t=0}^*$ and $P_{t=1}^*$ should be redrawn as nonlinear functions. Construction of approximate hedonic cost-of-living indexes (or more appropriately sub-indexes) would then proceed analogously to the homogeneous goods framework. However, now both the preferences, $q_{t=0}^*$ and $q_{t=1}^*$, and the true nonlinear hedonic surfaces are unobservable.[5]

Empirical estimates of the hedonic price function can be obtained for alternative price regimes, and these can be evaluated using either fixed characteristics weights from the beginning period, a fixed-weight Laspeyres-type index, or using fixed characteristics weights from the end of the period, a fixed-weight Paasche-type index. Price indexes obtained in this manner can be interpreted as approximations to the exact cost-of-living index for housing. They are approximations in the sense that they contain only information about the hedonic at a fixed set of characteristics between two time periods, whereas the true indexes also require information about preferences or levels of satisfaction. The Laspeyres-type and Paasche-type cost-of-living indexes for housing will therefore suffer from the same substitution bias found in their counterparts for homogeneous goods. The Laspeyres-type housing index would be expected to be biased upward and the Paasche-type index would be expected to be biased downward (Diewert 1983).

Triplett (1987) speculates, though does not prove, that empirical hedonic-index approximations may provide bounds on the true characteristics price index in the same way that the Laspeyres and Paasche indexes do in the homogeneous goods case. Diewert (1978) argues that if the empirical Laspeyres and Paasche indexes lie "close" to each other then the Fisher Ideal index[6] should be "close" to a reference exact price index that lies between the exact Paasche and Laspeyres price indexes.

Although there appear to be a number of similarities between the hedonic approximations for the empirical Laspeyres and Paasche cost-of-living indexes and those obtained for homogeneous goods, Triplett (1987) argues that there are also important differences. First, the form of the hedonic surface (the implicit prices of the characteristics) must be estimated empirically and, other than nonlinearity, there are no theoretical guidelines about appropriate functional forms. Second, the usual statistical procedures produce estimates for a shift in the whole hedonic surface rather than an estimate for shifts in a single selected budget hyperplane (the shift in the prices holding characteristic levels constant) as required in the fixed-weight empirical indexes.

Another problem is the goodness of the approximation. The empirical index numbers, such as the Laspeyres, Paasche, or Fisher's Ideal, use only price and quantity information, not the unobservable preferences. Thus, they are approximations to the theoretically correct, or exact, index numbers because they only approximately hold utility constant over the index comparison periods. With approximations, an error of indeterminable size is introduced into the index every time the fixed utility assumption is violated by changes in relative characteristics prices. As discussed, recent empirical and theoretical work indicates that good approximations to exact indexes can be computed from fixed-weight formulae. Thus, the criterion for the "goodness" of these index approximations in empirical applications is the extent to which the computed index takes account of, and controls for, variation in housing characteristics or quality. Quality variation is measured as the characteristics sets that are embodied in the housing stock from period to period. The fixed-weight approximations must fix these characteristics sets at either the beginning or end of the analysis period.

Several conclusions from cost-of-living index theory have practical implications for the empirical task of constructing housing price indexes. Pollak (1991, p. 168) suggests that it is useful to view the theoretical implications by distinguishing between the "estimation stage" concerning the appropriate specification of the hedonic price function (equation (1) above) and the "composition stage" in

5. The nonlinearity of the hedonic boundary constraint invalidates the usual strategy used for constructing cost-of-living indexes for homogeneous goods, in which it is assumed that the budget constraint (defined in consumption goods space) is a bounding hyperplane whose linearity assures that there is a duality between the utility function and the consumption cost function.

6. The Fisher Ideal price index is defined as the geometric average of the Laspeyres and Paasche price indexes

$$Index_{Fisher} = \sqrt{\left(Index_{Paasche}\right) \cdot \left(Index_{Laspeyres}\right)} \ .$$

which the estimated hedonics are used to obtain price indexes. For the estimation stage, it was shown above that the hedonic function is, in Rosen's terminology, an estimate of the minimum price of any package of characteristics (Rosen, 1974, p. 37) and thus, it is the empirical counterpart to the characteristics cost function. In a market such as housing, with a continuous variety spectrum, the functional form for the hedonic is an empirical question. In general, however, the characteristics price (the partial derivatives of the characteristics cost function) are themselves functions whose value depends on the particular point in the characteristics space where they are evaluated. This suggests that empirical specifications should allow for maximum flexibility of functional form. Theory also has little to say about the elements of the characteristics set used to estimate the hedonic. Theoretically, the chosen set should include all characteristics that can reasonably be assumed to enter household preferences.[7] Finally, it would be desirable to use estimation strategies that provide local approximations to the hedonic price function at fixed characteristics levels in each time period. In this way, it would be possible to control for a fixed consumption bundle and obtain better estimates for either the Laspeyres or the Paasche index approximations.

The second implication of the theory concerns the "composition stage" of the price index. Once an empirical estimate of the hedonic is obtained, what is the appropriate composition of the price index? It was argued that the theoretically exact index could not be uncovered due to the nonlinearity of the hedonic and lack of information about preferences. Thus, suitable approximations are measures of the effects of relative price changes when the beginning point, or end point, of the characteristics bundle is fixed. This strategy ignores the substitution effects from price changes. The practical empirical task is to obtain estimates of the hedonic price surface such that unbiased estimates of the prices of fixed sets of characteristics can be computed.

III. ESTIMATING HEDONIC-BASED PRICE FUNCTIONS

The primary theoretical objectives for the estimation of hedonic housing functions are that the estimation strategy should allow for the nonlinearity of the hedonic contours and that it should provide an accurate accounting of, and control for, variations in characteristics, or quality, over time. The primary criticism that has been raised against the hedonic methodology concerns the appropriate way to meet these theoretical objectives in the usual regression framework. The first complaint is that the "correct" set of characteristics must be selected to achieve an unbiased estimate of the hedonic function. The second complaint is thata priori assumptions concerning the "correct" functional form must be imposed to estimate the hedonic function in a regression framework. A final complaint is that hedonic price function estimates are likely to suffer from sample selection bias because they are obtained from samples of transactions that may not be random samples of the population of house prices.

Alternative Specifications to Control for Characteristics

An important alternative recommended strategy is the repeat sales methodology, which was first introduced by Bailey, Muth, and Nourse (1963) and further developed by Case and Shiller (1987, 1989). This method focuses on price changes rather than price levels, and it restricts estimation to a subsample of houses that have not changed their characteristics set and have sold at least twice. The primary advantage of this strategy is that it avoids the specification of the characteristics set for houses and the functional relationship of characteristics to price. The argument is that first differencing the log of house prices and using only houses that have been sold at least twice and have not changed their characteristics produces a perfect control for the entire set of relevant characteristics.

The primary advantage of the repeat sales methodology also imposes important theoretical restrictions on the admissible class of characteristics cost functions that can be considered. It can also be shown, (Meese and Wallace 1996, Wang and Zorn 1995) that the estimated coefficients in the repeat sales framework are complicated frequency weightings of the simple means of logarithm of the ratio of final transaction price to the initial transaction price over relevant time periods.[8] These weights do not have a "fixed-weight" interpretation in the sense discussed above because

7. It is this point that Shiller (1993) identifies as the greatest weakness of the housing price indexes composed from hedonic price function estimates. He argues that these decisions are necessarily arbitrary because they involve "...not only the decision of which quality variable to include, but there are also decisions to make about allowing nonlinear effects of each and interaction effects..."(p. 129). He also asserts that the lack of available characteristics data leads to problems with sample size and misspecification due to omitted characteristics.

8. For example, in a three-period sample with possible repeat sales between periods 1 and 2, 1 and 3, and 2 and 3, the least squares estimators for the logarithm of the index number for periods 2 and 3, respectively, are:

$$\hat{\phi}_2 = \frac{n_{12}(n_{13}+n_{23})\bar{r}_{12} + n_{13}n_{23}(\bar{r}_{13} - \bar{r}_{23})}{n_{12}(n_{13}+n_{23}) + n_{13}n_{23}}$$

the computed means reflect different subsamples of unobservable characteristics bundles. Thus the estimated price relatives do not provide an estimate of the characteristics cost frontier at a fixed package of characteristics as required in the usual formulation of approximations to exact cost-of-living price indexes. It has this interpretation only if it is assumed that the true hedonic contours shrink toward the origin in a homogeneous fashion.

The repeat sales strategy also assumes that the characteristics levels for houses do not change and those that do can be "correctly" identified. This assumption leaves the measure vulnerable to the same misspecification concerns that the hedonic methodology must contend with. Finally, the repeat sales method requires careful testing of sample selection assumptions because the sample is by definition more restrictive than those used in the hedonic methodology.[9] These trade-offs suggest that further refinements may be required for both methods. These refinements include development of hybrid methods that combine features of both methods (Quigley 1995).

Flexible Nonparametric Estimation Strategies

Flexible nonparametric estimation strategies directly address the problem of imposing a priori specifications on the hedonic functional forms or using grid search methods over a limited class of functional forms. They also allow local approximations to the hedonic surface at fixed points, which is more in keeping with the requirements of price index formation.

Following Meese and Wallace (1991), suppose the natural log of house price in period t, $P(n,t)$, varies with the natural log of its characteristics, $x(n,t)$, according to a hedonic function:

$$(4) \qquad P(n,t) = m(t) + \beta_t' G[x(n,t)] + u(n,t),$$

where $m(t)$ accounts for the changing residual mean in house prices, β_t denotes a $(j \times 1)$ vector of parameters, $x(n,t)$ is a set of j housing characteristics observed for the nth transaction at time t, G is a function of the characteristics, and $u(n,t)$ is an additive error term. The nonstationary mean in housing prices is attributed to the drift, $m(t)$, which is modeled as:

$$(5) \qquad m(t) = \alpha(t) dum(t)$$

where $dum(t)$ is a dummy variable equal to one for each quarterly observation period t and zero otherwise, $\alpha(t)$ is the regression parameter measuring period t residual mean price change between periods once the mean changes in characteristics costs have been accounted for, and $e(t)$ is the time-series error component that is assumed to be white noise. Combining (4) and (5) yields a fully general hedonic function:

$$(6) \qquad P(n,t) = \beta_t' G[x(n,t)] + \alpha(t) dum(t) + (e(t) + u(n,t)).$$

From a theoretical perspective, the preferred method to estimate equation (6) is a strategy that imposes the fewest a priori restrictions on the functional form of $G[\cdot]$.[10] Nonparametric methods allow for the greatest possible flexibility in estimating functional forms and allow empirical estimation of data contours over a wide range of smooth functions. A particularly suitable nonparametric method is regression by *loess* which was first introduced by Cleveland and Devlin (1988) and Cleveland, Devlin, and Grosse (1988). *Loess* also allows for local approximations to the $G[\cdot]$ function at fixed points in the data surface.

Loess is a technique for estimating a regression surface in a moving average manner and can approximate a wide range of smooth functions. Meese and Wallace (1991, 1996) use a version of the regression model in equation (6):

$$(7) \qquad P(n,t) - P(mean,t) = \beta_t' G[X(n,t)] + v(n,t), \quad n = 1,...,N(t),\ t = 1,...,T$$

where $P(mean,t)$ is the quarterly mean of the logarithm of housing prices and $v(n,t)$ is the composite error term. Because nonparametric local fitting strategies require stationary dependent and independent variables, I remove the trend in $P(n,t)$ by subtracting the quarterly mean of the dependent variable each quarter and then standardize the variable by dividing by the quarterly sample standard deviation. I also standardize all the characteristics variables by subtracting the global mean and dividing by the sample standard deviation.

$$\hat{\phi}_3 = \frac{n_{13}(n_{12}+n_{23})\bar{r}_{13} + n_{12}n_{23}(\bar{r}_{12}+\bar{r}_{23})}{n_{12}(n_{13}+n_{23}) + n_{13}n_{23}}$$

where $\bar{r}_{12}$, $\bar{r}_{13}$, and $\bar{r}_{23}$ are the means of the logarithm of the ratio of final transaction prices to the initial transaction prices in the subscripted time interval, and n_{12}, n_{13}, and n_{23} are the sample frequencies for repeat sales in the subscripted time interval.

9. Meese and Wallace (1996) test the repeat sales assumption that the characteristics prices are time-invariant using a second order Taylor series approximation to the hedonic function and a transaction data set from Alameda County. They reject the assumption for all municipalities, suggesting that this is not an innocuous maintained hypothesis.

10. Meese and Wallace (1991) test for parametric flexible functional forms such as the translog and the log-log function as do Halvorsen and Pollakowski (1981). Their findings suggest no consistent preference for one specification across municipalities.

I employ two centering strategies. The first strategy I call the nonfixed-centering *loess* estimator. It uses the vector of mean characteristics for quarter t, $X(m,t)$, to center the local fitting of $G(\cdot)$. *Loess* uses a fraction n^*, $0 < n^* < 1$, of the total number of observations closest to $X(m,t)$, where proximity is measured using the Euclidean distance between all points in the sample and $X(m,t)$. The distance metric is defined by:

(8) $D[X(m,t), X(n,t) = [\Sigma X(m,t) - X(n,t)]^2]^{1/2}$,

where the summation runs over the j-elements of the set of housing characteristics.

The hedonic surface is approximated locally at $X(m,t)$ by a weighted least squares regression for the n^* observations nearest $X(m,t)$. The weights are defined by Cleveland and Devlin (1988) as:

(9) $W = V[D(X(m,t)\ X(n,t)) / D(X(m,t), X(n,*))]$,

where $D(X(m,t), X(n,*))$ is the distance from the mean X in a given quarter to its n^* nearest neighbors. Following Cleveland and Devlin (1988) and Meese and Wallace (1991) I use the "tricube" functional form for $V[\cdot]$.[11] This strategy provides estimates for the curvature of the hedonic price function at the mean characteristics over the 101 quarters in the data set.[12] I set n^* at 0.33 in an effort to balance the trade-off between bias and sampling error.

There are two problems with this strategy. The first is that the quarterly means are used to detrend the price data and these means reflect both price changes and changes in the set of characteristics traded. The second problem is that *loess* is estimated by centering at the quarterly means, whereas the desired price estimates should be centered at a fixed characteristics set from the beginning or end of the period.

The second strategy addresses these problems. The fixed-centering *loess* estimator centers at two fixed characteristics sets: the first set is fixed at the mean, $X(m_L,1)$, of the characteristics for 1970:Q1, a Laspeyres-type estimator, and the second set is fixed at the mean, $X(m_P,101)$, of the characteristics for 1995:Q1, a Paasche-type estimator.[13] Thus, these estimators replace the term $X(m,t)$ with the appropriate fixed characteristics set. The estimation is then carried out for the n^*_Q nearest neighbors for each quarter using the same weighting strategy as in the non-fixed centering strategy.

There remains one problem with the fixed-centering *loess*. For smaller municipalities, the *loess* weightingscheme leads to an insufficient number of observations in the neighborhood of the initial characteristics set in some quarters. It is thus necessary to smooth across quarters, although in the applications here one never has to smooth over more than two quarters. The primary advantage of the fixed estimator is that it is consistent with the requirements of empirical Laspeyres and Paasche-type price indexes.

The transaction data used in this analysis included four characteristics: number of bathrooms, number of bedrooms, square footage of the living area, and the age of the dwelling. I constructed a variable bedrooms/living area to account for possible nonlinearities from adding more bedrooms onto a home of a given square footage. Because homes with a high ratio of bedrooms to living area are likely to be rental property, often for student habitation, I expected that higher ratios would reduce house prices. The other characteristics, except for the age of the dwelling, were expected to have positive effects on housing prices. I did not have strong priors on the effect of age on house price. The age variable may well proxy for other unmeasured features of the dwelling such as architectural design. For example, many older California craftsman homes sell at a premium due to their distinctive design characteristics; on the other hand, age could account for the effects of deterioration or a lack of modern room organization.

The results for the nonfixed and fixed-centering *loess* estimates are reported in Tables 1 and 2.[14] The price elasticities are obtained by taking the derivative of the estimated housing price function with respect to each characteristic and evaluating the derivative at the appropriate characteristics set. The Tables report two types of elasticities for 1970:Q1 and for 1995:Q1. Reading down the columns for each municipality, the Laspeyres-type elasticities are evaluated at the mean characteristics set for the first quarter of 1970 for each municipality. The Paasche-type elasticities

11. The tricube $V[\cdot] = (1 - s^3)^3$, if $s < 1$; it is equal to 0 otherwise. The advantage of the tricube is that it allows smooth contact with 0 and 1 endpoints.

12. The 1970 through first quarter 1988 Alameda County data were obtained from the California Market Data Cooperative and account for about 98% of all arm's-length transactions over the period. The 1988 through 1995 data were obtained from Property Sciences, Inc. and TRW.

13. This assumes that the observed quarter 1 sample is a random sample of the characteristics set for houses traded in 1970 and similarly for the 1995:Q1 sample.

14. The skewness and kurtosis measures for the residual distributions from these estimates have close to symmetric distributions, although they have fatter tails than would be expected under normality. The White test for heteroskedasticity in the residuals indicates that there remains contemporaneous heteroskedasticity for several of the municipalities. These diagnostics suggest that the more efficient estimates of the characteristics prices should be considered. A dynamic model might include allowance for serial correlation and/or ARCH in the time-series component of the composite error, or explicit consideration of the speed of adjustment of prices to changes in market fundamentals or levels of housing characteristics.

TABLE 1

ESTIMATED PRICE ELASTICITIES FOR THE NONFIXED-CENTERING *LOESS*
ALAMEDA COUNTY MUNICIPALITIES: 1970:Q1–1995: Q1

ALAMEDA COUNTY	BATHROOMS		BEDROOMS/ TOTAL LIVING AREA		TOTAL LIVING AREA		AGE OF HOUSE	
	Paasche	Laspeyres	Paasche	Laspeyres	Paasche	Laspeyres	Paasche	Laspeyres
ALAMEDA[3]								
1970	9,563	1,210	3,995	520	126	12	740	112
1995	9,699	1,228	3,995	501	126	12	744	113
ALBANY								
1970	20,477	1,294	36,354	5,576	179	18	2,350	–1,026
1995	36,860	2,330	–7,383	1,132	155	16	307	133
BERKELEY								
1970	7,560	763	–8,836	–1,211	199	19	827	100
1995	11,190	1,129	–14,073	–1,929	203	20	727	89
CASTRO VALLEY[1,2,3]								
1970	7,947	1,404	4,681	9,521	40	8	1,290	311
1995	17,118	3,024	4,057	825	68	13	1,402	339
DUBLIN[1,2,3]								
1970	18,470	2,814	–2,463	–361	88	12	2,414	831
1995	25,574	3,897	–36,041	–5,292	167	23	4,161	1,433
HAYWARD[1]								
1970	7,239	1,233	–2,614	–416	75	11	980	238
1995	13,444	2,290	–5,809	–926	93	14	1,004	243
FREMONT[1,2,3]								
1970	12,356	1,310	–5,734	–749	70	10	1,171	259
1995	22,582	2,395	1,720	225	77	11	1,273	282
LIVERMORE[1,2,3]								
1970	10,389	1,298	3,538	649	42	8	1,717	306
1995	13,961	1,745	1,927	353	53	18	1,787	319
NEWARK[1]								
1970	12,788	1,099	3,444	470	13	1	2,260	441
1995	33,154	2,849	–898	–122	61	6	3,002	585
OAKLAND								
1970	15,933	2,418	–574	–98	119	20	1,248	204
1995	32,432	4,924	3,148	541	121	21	–365	–59
PIEDMONT[1,2,3]								
1970	58,737	3,565	–2,588	–135	105	10	2,896	101
1995	136,857	8,306	–48,785	–2,560	199	19	–5,494	–192
PLEASANTON[1,3]								
1970	36,939	4,439	80	974	121	16	1,487	1,449
1995	49,514	5,947	15,382	1,852	151	20	694	676
SAN LEANDRO								
1970	8,001	1,551	988	304	40	7	765	161
1995	10,784	2,091	751	155	92	16	–655	–138
SAN LORENZO[2]								
1970	7,344	1,327	–6,670	–661	5	1	859	127
1995	9,487	1,714	6,107	1,025	38	8	–3,390	–225
UNION CITY[1]								
1970	9,695	1,326	–6,670	–661	5	1	859	127
1995	16,967	2,321	–3,174	–315	9	2	1,364	202

1. Statistically significant at 5% level, positive trend in bathrooms.
2. Statistically significant at 5% level, positive trend in bedrooms.
3. Statistically significant at 5% level, positive trend in living area.

TABLE 2

ESTIMATED PRICE ELASTICITIES FOR THE FIXED-CENTERING *LOESS* ALAMEDA COUNTY MUNICIPALITIES: 1970:Q1–1995: Q1

Alameda County	Bathrooms		Bedrooms/ Total Living Area		Total Living Area		Ageof House	
	Paasche	Laspeyres	Paasche	Laspeyres	Paasche	Laspeyres	Paasche	Laspeyres
ALAMEDA								
1970	11,384	692	135	–1,694	109	11	1,053	–122
1995	16,394	1903	–9,480	–891	107	10	1,340	203
ALBANY[3]								
1970	9,829	156	108	–172	122	16	–456	154
1995	44,232	3,262	749	344	127	13	–124	–97
BERKELEY								
1970	13,105	712	3,600	–10,991	162	16	1,204	142
1995	25,203	3,058	–295	224	126	12	2,668	240
CASTRO VALLEY[1,2,3]								
1970	10,087	216	–4,486	–515	48	9	2,258	389
1995	11,310	378	3,765	1,309	77	14	864	159
DUBLIN[1,2,3]								
1970	7,104	584	–2,225	–2,132	71	7	–574	–205
1995	28,416	2,923	–10,112	–1,942	50	16	1,242	617
HAYWARD[1]								
1970	4,653	150	8,714	1,466	68	9	1,578	258
1995	11,634	748	726	1,041	71	12	514	23
FREMONT[1,2,3]								
1970	22,156	2,124	–1,762	337	87	12	–886	–252
1995	18,321	2,395	–5,447	824	90	13	–920	–189
LIVERMORE[1,2,3]								
1970	12,337	1,055	–3,660	1,388	57	7	964	52
1995	15,260	1,371	–5,856	–895	72	15	164	43
NEWARK[1]								
1970	12,314	1,628	4,192	–1,104	99	5	873	441
1995	14,682	2,320	898	–286	98	7	1,457	218
OAKLAND								
1970	14,311	2,558	1,481	–95	66	11	918	144
1995	22,898	3,234	–1,297	–445	66	12	739	115
PIEDMONT[1,2,3]								
1970	43,378	1,329	–44,802	–3,866	114	8	1,536	156
1995	100,103	2,172	57,248	1,593	218	19	1,769	–132
PLEASANTON[1,3]								
1970	13,143	849	–1,596	–1,462	112	13	–553	–316
1995	19,649	4,625	–2,264	–564	108	13	623	–2,526
SAN LEANDRO								
1970	10,784	1,146	2,819	4,272	76	12	940	163
1995	18,697	1,281	–1,691	–7	71	13	1,257	221
SAN LORENZO[2]								
1970	9,793	1,382	3,327	1,876	12	7	12	68
1995	12,548	1,880	11,645	1,916	48	11	–1,629	–121
UNION CITY[1]								
1970	5,508	884	–1,118	–239	66	14	1,784	121
1995	8,080	1,824	–11,771	–955	128	32	–2,410	420

1. Statistically significant at 5% level, positive trend in bathrooms.
2. Statistically significant at 5% level, positive trend in bedrooms.
3. Statistically significant at 5% level, positive trend in living area.

are evaluated at the mean characteristics set for the first quarter of 1995 for each municipality.

For the non-fixed centering *loess* reported in Table 1, the parameter estimates underlying the Paasche-type and Laspeyres-type elasticities are the same for each year (e.g., 1970:Q1 has one set of estimates and 1995:Q1 another).[15] Thus, the differences in the magnitudes of the elasticities come from the growth in the attribute sets from 1970:Q1 and 1995:Q1. The parameter estimates for the Paasche-type and Laspeyres-type elasticities are estimated separately for the fixed-centering *loess*. Thus, these elasticities reflect both changes in prices and growth in the characteristics set over the analysis period. The footnotes indicate whether there was a statistically significant trend in the mean levels of characteristics for bathrooms, bedrooms, and living area over the quarters. As shown, ten of the fifteen municipalities experienced statistically significant positive trend in the mean levels of these characteristics over the 101 quarters. Additionally, as expected in some municipalities, increasing the ratio of bedrooms to total living area reduces the value of the house. There is, however, quite a lot of variability in this result across the municipalities. The effect of age also varied across the municipalities; however, for most municipalities, increasing the age of the dwelling led to increases in housing prices.

The differences between the Paasche-type and Laspeyres-type elasticities by characteristics by municipality reflect changes in the magnitudes of the mean level of the characteristics set between 1970:Q1 and 1995:Q1. The nonfixed-centering *loess* estimation reported in Table 1, however, does not account for differences in the coefficient estimates at different mean levels of characteristics on the hedonic surface within a quarter. In Table 2, however, the fixed-centering *loess* estimates provide a local approximation to the hedonic at either the fixed 1970:Q1 characteristics level or the fixed 1995:Q1 level. Thus, the Table 2 elasticities control for the growth in the mean value of characteristics over the quarters, the changes in price levels of mean characteristics across quarters, and the differences in price levels within a quarter for different mean characteristics levels. The Table 1 elasticities control for only the growth in the mean value of characteristics over the quarters and the changes in price levels of mean characteristics across quarters. They do not control for differences in mean price levels within each quarter for different characteristics bundles.

For example, Castro Valley has experienced considerable growth in the mean levels of characteristics in houses sold from 1970:Q1 to 1995:Q1; the nonfixed-centering *loess* Laspeyres-type and Paasche-type price elasticities for bathrooms indicate about a 115% increase in the elasticities over the analysis period. The fixed-centering *loess* elasticities reported in Table 2, in contrast, indicate that the Paasche-type elasticity, holding the characteristics mean fixed at 1995:Q1 levels, experienced only a 12% increase and the Laspeyres experienced only a 75% increase from the 1970:Q1 mean level of characteristics. Similar differences appear in the elasticity of square footage. Fremont and Piedmont also experienced growth in mean characteristics levels over the period. Here again, the price elasticity for bathrooms increased by 82% for Fremont using the Table 1, nonfixed-centering *loess* results, whereas the price elasticity of bathrooms fell by 17% using the Paasche-type fixed *loess* estimates. The Piedmont elasticity of bathrooms increased by 132% using the Table 1 estimates, however, the elasticity growth found for the Table 2 fixed estimates was between 131% and 63%. The results for the square footage elasticities were similar. The elasticity results for the ratio of bedrooms to total rooms is similar in many municipalities, although it is difficult to interpret the negative changes in Livermore. A reasonable conclusion from comparing Tables 1 and 2 is that the differences in the results are most pronounced for the municipalities that experienced the most growth in the mean levels of the characteristics, such as Castro Valley, Dublin, Fremont, Livermore, and Piedmont.

Oakland did not experience statistically significant growth in the mean level of characteristics over the period; however, there is also evidence of the effects of confounding characteristics level growth with price changes. The Oakland price elasticity for bathrooms increased about 104% using the Table 1 estimates, whereas it grew only 60% for the Paasche-type elasticity and 26% for the Laspeyres-type elasticity. Thus, even in a municipality in which the growth of the mean characteristics was not sustained there appears to be confounding of the growth in the mean levels of characteristics with changes in the relative price levels. The fixed *loess* results appear to control better for the confounding effects of growth in the level of the characteristics.

15. To reiterate, the difference between the two estimation strategies is that the non-fixed centering *loess* uses the mean level of characteristics in each quarter and then selects the nearest neighbors from all the data, whereas the fixed centering *loess* estimation obtains two estimates: one centered at the mean of the 1970:Q1 characteristics and the other centered at the mean of the 1995:Q1 characteristics, and the nearest neighbor is determined within a quarter. The price elasticities are then obtained using the coefficients for the mean initial and end-of-period characteristics set. The fixed estimation evaluates the elasticities for the beginning quarter and ending quarter coefficients using either the first quarter mean characteristics (a Laspeyres-type measure) or the last quarter mean characteristics (a Paasche-type measure).

I conclude that the hedonic price surfaces can consistently be estimated with both *loess* strategies, although the fixed strategy is somewhat more consistent with the theoretical structure of empirical Laspeyres and Paasche price indexes. The most important difference between the two strategies is found for the characteristics price for housing attributes that have changed the most over the 25-year period. Finally, the "hedonic" or characteristics effects account for a substantial part of the change in house prices.

IV. CONSTRUCTING HOUSING PRICE INDEXES

As previously discussed, consistent estimates of the hedonic surface can be used to construct estimates of the Laspeyres-type and Paasche-type price indexes. The theoretically desirable Fisher Ideal price index can be computed from the geometric average of these two bounds. As Diewert (1978) has shown, if the Laspeyres-type and the Paasche-type price indexes are very close to one another, the Fisher Ideal can be considered as a close approximation to an exact price index defined in characteristics. The usual sense in which price indexes are considered to be close approximations relates to the degree to which they control for fixed levels of characteristics in the construction of the price index. An advantage of the fixed-centering *loess* estimation is that it allows for local approximations to the hedonic price at fixed mean levels of characteristics. Thus, the fixed-centering *loess* seems to be the preferable estimation strategy given the empirical results summarized in Tables 1 and 2 and the theoretical requirements for close approximation strategies for index number construction.

Figures 3–8 compare the fixed and nonfixed-centering *loess* Fisher Ideal price indexes with repeat sales indexes, quarterly means, and quarterly medians for three municipalities: Oakland, Fremont, and Piedmont. Oakland experienced relatively little growth in the mean level of housing characteristics over the analysis period and Fremont and Piedmont experienced considerable growth in mean housing characteristics. Figure 3 compares the fixed-centering *loess* Fisher Ideal price index with the quarterly means of house prices and a repeat sales price index for Oakland. The quarterly means exceed the fixed Fisher Ideal index and the repeat sales index for nearly all the quarters. Repeat sales accounted for only 19% of all sales over the sample period, and the repeat sales index appears to underestimate the price index consistently. The fixed Fisher Ideal appears to account for the confounding effects of the mean levels of characteristics from the changes in relative prices of the characteristics. Figure 4 is consistent with the results in Tables 1 and 2 in that the fixed Fisher Ideal shows a smaller

FIGURE 3

OAKLAND: FIXED *LOESS* FISHER IDEAL, MEAN, AND REPEAT SALES PRICE INDEXES

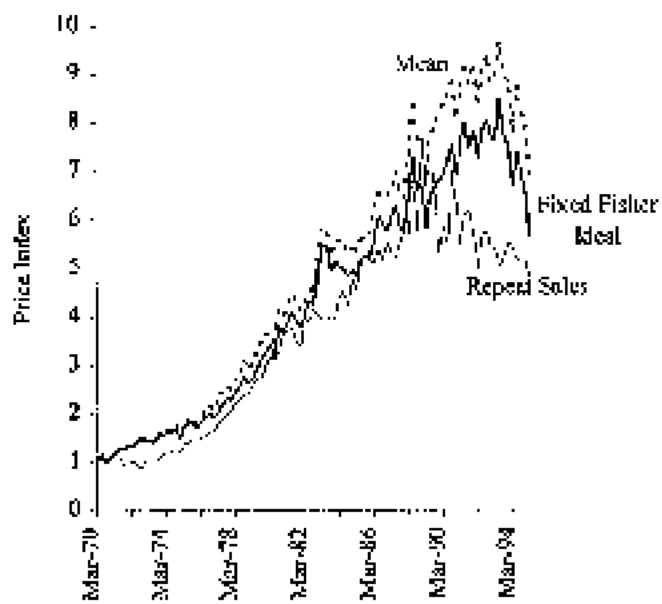


FIGURE 4

OAKLAND: FIXED AND NONFIXED CENTERING *LOESS* FISHER IDEAL PRICE INDEXES

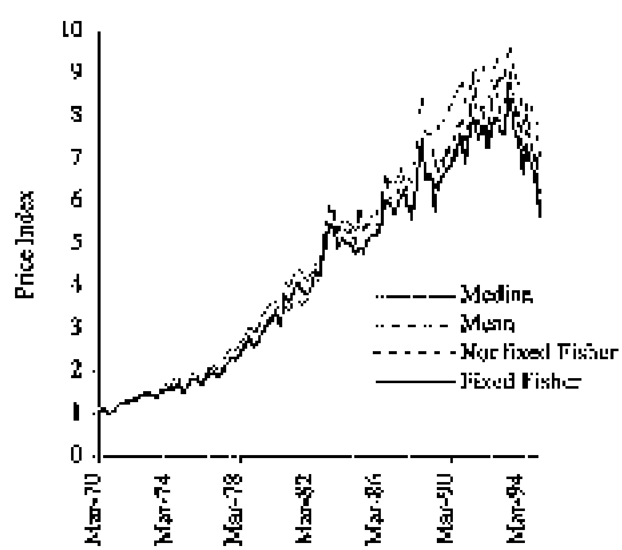


FIGURE 5

FREMONT: FIXED *LOESS* FISHER IDEAL, MEAN, AND REPEAT SALES PRICE INDEXES

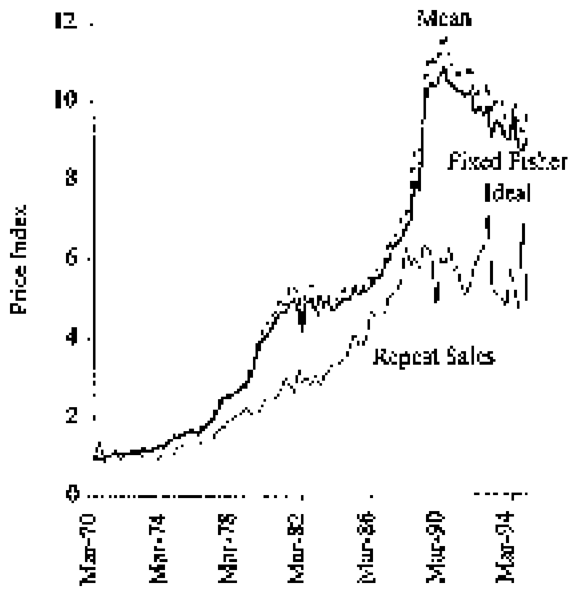


FIGURE 6

FREMONT: FIXED AND NONFIXED CENTERING *LOESS* FISHER IDEAL PRICE INDEXES

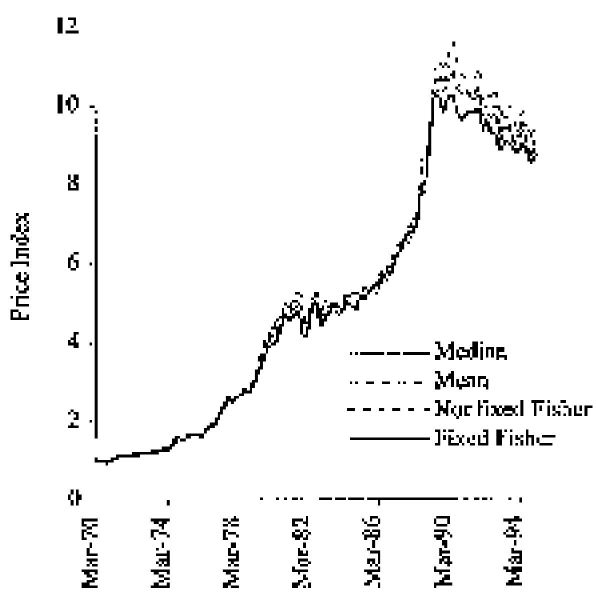


FIGURE 7

PIEDMONT: FIXED *LOESS* FISHER IDEAL, MEAN, AND REPEAT SALES PRICE INDEXES

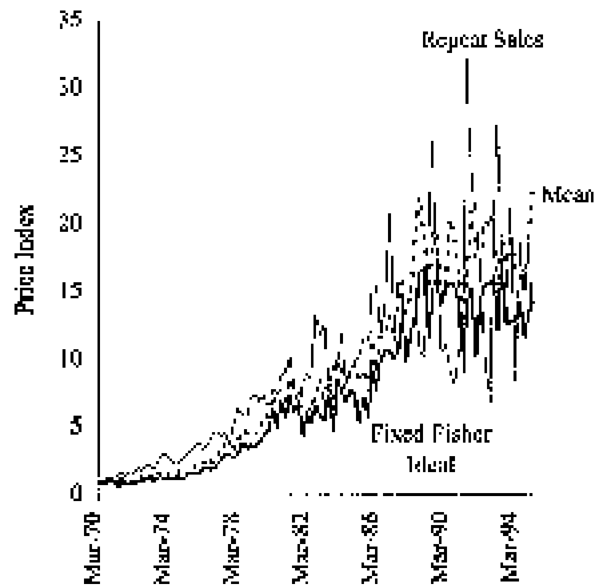
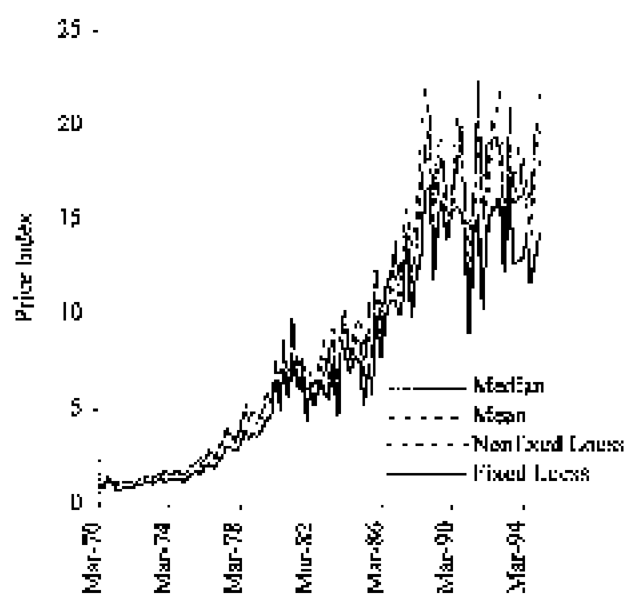


FIGURE 8

PIEDMONT: FIXED AND NONFIXED CENTERING *LOESS* FISHER IDEAL PRICE INDEXES



change in relative prices than either the quarterly means or the nonfixed-centering Fisher Ideal index. The differences between the two Fisher Ideal indexes are important only after the first quarter of 1989, when the fixed-centering Fisher Ideal falls below the nonfixed-centering index.

Figures 5 and 6 provide the same information for Fremont. The results are similar to the Oakland graphs, although the quarterly means more closely track the fixed-centering Fisher Ideal. The repeat sales index again substantially underestimates the relative price changes compared to the quarterly means and the fixed-centering *loess* Fisher Ideal. Repeat sales account for about 18% of the total sales over the period in Fremont. Figure 6 compares the fixed and nonfixed-centering *loess* Fisher Ideals with the quarterly mean and median indexes. The fixed Fisher Ideal is consistently below the nonfixed Fisher Ideal, as expected from the results of Tables 1 and 2.

Figures 7 and 8 provide the index construction results for the city of Piedmont. Piedmont is an exclusive residential community that is entirely surrounded by the city of Oakland, but all its public service systems, including schools, are separate from those of Oakland. Piedmont has experienced growth in the mean levels of characteristics of housing sold during the period as well as very substantial price appreciation of attributes. Figure 7 compares the repeat sales index with the quarterly mean index and the fixed-centering Fisher Ideal. Again, the quarterly mean index appears to overestimate the appreciation of house prices. The repeat sales index is wildly erratic, most probably due to the small sample size for repeat sales in Piedmont, only 630 homes. The fixed Fisher Ideal index appears to control for the confounding effects of the growth in the mean levels of characteristics and is considerably less erratic, due to the larger sample size. Figure 8 compares the fixed and nonfixed-centering Fisher Ideal indexes with the quarterly median and mean indexes. Again the fixed Fisher Ideal lies everywhere below the nonfixed index, which more closely tracks the mean and median indexes.

These graphical results appear to indicate that accounting for the growth in the mean levels of characteristics gives a rather different view of house price increases in Alameda County municipalities. The fixed-centering *loess* Fisher Ideal index is particularly appealing because it allows local approximations of the hedonic surface as prescribed by the theory of cost-of-living indexes and allows for the construction of Fisher Ideal price indexes that are the geometric average of the beginning and ending period characteristics levels. The elasticities derived from the estimation of the hedonic appear to suggest that fixing the point of approximation may be necessary to avoid confounding the growth in the levels of characteristics, which can be viewed as a measure of quality, from the changes in the relative prices of characteristics. The results from comparing the constructed Fisher Ideal indexes also indicate that the fixed approximation may be preferable.

V. Conclusions

This paper reviewed basic principles of price index construction for heterogeneous goods such as housing, where differing levels of characteristics (quality) lead to important differences in prices. The price/quality relationship is described by the housing price hedonic, which is likely to be nonlinear. Nonparametric econometric techniques are particularly suitable for the hedonic price function estimation problem because they allow for many classes of functional forms. I show how one nonparametric technique, *loess*, allows for the added feature of centering the estimation to fixed points, such as the beginning or ending period characteristics sets consistent with the requirements of Laspeyres-type and Paasche-type price indexes.

The *loess* estimates for the hedonic contours were used to construct Fisher Ideal price indexes. These indexes appear to have important differences from repeat sales indexes that rely on mean prices that may not control for quality levels. I also found differences between fixed and nonfixed characteristics estimates, and I attributed these to the additional control for the level of characteristics in the fixed *loess* strategy. These differences suggest that in dynamic markets, such as Alameda County, where new housing construction and high levels of remodeling have led to changes in the mean characteristics levels of the housing stock, it is important to control for the confounding effects of price changes and quality changes both in the estimation of the hedonic and in the price index construction. In less dynamic markets, these differences may not be as important.

References

Alonso, W. 1964. *Location and Land Use.* Cambridge: Harvard University Press.

Bailey, M., R. Muth, and H. Nourse, H. 1963. "A Regression Method for Real Estate Price Index Construction." *Journal of the American Statistical Association* 58, pp. 933–942.

Case, K., and R. Shiller. 1989. "The Efficiency of the Market for Single-family Homes." *American Economic Review* 79, pp. 125–137.

Cleveland, W.S., and S. J. Devlin. 1988. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association* 83, pp. 596–610.

_______, _______, and E. Gross. 1988. "Regression by Local Fitting." *Journal of Econometrics* 37, pp. 87–114.

Diewert, W.E. 1983. "The Theory and Measurement of the Cost-of-living Index and the Measurement of Welfare Change." In *Price Level*

Measurement, eds. W.E. Diewert and C. Montmarquette, pp. 163–233. Ottawa: Statistics Canada.

_______. 1978. "Superlative Index Numbers and Consistency in Aggregation." *Econometrica* 46, pp. 883–900.

Epple, D. 1987. "Hedonic Price and Implicit Markets: Estimating Demand and Supply Functions." *Journal of Political Economy* 95, pp. 58–80.

Griliches, Z. 1971. "Introduction: Hedonic Price Indexes Revisited." In *Price Indexes and Quality Change,* ed. Z. Griliches. Cambridge: Harvard University Press.

Meese, R., and N. Wallace. 1996. "The Construction of Residential Housing Price Indices: A Comparison of Repeat Sales, Hedonic Regression, and Hybrid Approaches." *Journal of Real Estate Finance and Economics.* Forthcoming.

_______, and _______. 1994. "Testing the Present Value Relation for Housing: Should I Leave My House in San Francisco?" *Journal of Urban Economics* 35, pp. 245–266.

_______, and _______. 1991. "Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices." *Journal of the American Real Estate and Urban Economics Association* 19, pp. 308–332.

Motley, B. 1992. "Index Numbers and the Measurement of Real GDP." Federal Reserve Bank of San Francisco *Economic Review* 1, pp. 3–13.

Muth, R. 1969. *Cities and Housing.* Cambridge: Harvard University Press.

Pollak, R.A. 1991. *The Theory of the Cost-of-living Index.* New York: Oxford University Press.

Quigley, J. 1995. "A Simple Hybrid Model for Estimating Real Estate Price Indexes." *Journal of Housing Economics* 4, pp. 1–12.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82, pp. 34–55.

Shiller, R. 1993. *MacroMarkets: Creating Institutions for Managing Society's Largest Economic Risks.* London: Clarendon Press.

Triplett, J.E. 1989. "Price and Technological Change in a Capital Good: A Survey of Research on Computers." In *Technology and Capital Formation,* eds. D.W. Jorgenson and R. Landau, pp. 127–213. Cambridge: MIT Press.

_________. 1987. "Hedonic Functions and Hedonic Indexes." In *The New Palgrave: A Dictionary of Economics,* eds. J. Eatwell, M. Milgate, and P. Newman, pp. 630–634. London: Macmillan.

Wang, F., and P. Zorn. 1995. "Repeat Sales: An Introductory Primer." Unpublished manuscript. Federal Home Loan Mortgage Corporation.