

# UC Berkeley

## Working Papers

### Title

Do Housing Transactions Provide Misleading Evidence about the Course of Housing Values?

### Permalink

<https://escholarship.org/uc/item/3pk6m1zd>

### Authors

Englund, Peter  
Quigley, John M.  
Redfearn, Christian L.

### Publication Date

1999-02-01



Institute of  
Business and  
Economic Research

Fisher Center for  
Real Estate and  
Urban Economics

---

# **PROGRAM ON HOUSING AND URBAN POLICY**

WORKING PAPER SERIES

**WORKING PAPER NO. W99-004**

## **DO HOUSING TRANSACTIONS PROVIDE MISLEADING EVIDENCE ABOUT THE COURSE OF HOUSING VALUES?**

By

Peter Englund  
John M. Quigley  
Christian L. Redfearn

These papers are preliminary  
in nature: their purpose is to  
stimulate discussion and  
comment. Therefore, they are not  
to be cited or quoted in any  
publication without the express  
permission of the author.

**Do Housing Transactions Provide Misleading Evidence**

**About the Course of Housing Values?**

by

Peter Englund  
Stockholm School of Economics  
(finpe@hhs.se)

John M. Quigley  
University of California, Berkeley  
(quigley@econ.berkeley.edu)

Christian L. Redfearn  
University of California, Berkeley  
(redfearn@econ.berkeley.edu)

February 1999

A previous version of this paper was presented at the workshop sponsored by the Center for Real Estate Education and the John K. Galbreath Chair at Ohio State University, June 1998. This research is supported by Berkeley's Program on Housing and Urban Policy and the Swedish Council of Building Research.

## Abstract

Estimates of the prices of housing and the value of the stock are derived from observations on housing transactions. These transactions may well be a non-random sample of the underlying population of dwellings. For example, it is widely thought that smaller "starter homes" sell more frequently than more expensive properties and that the frequency of transactions on high-valued properties varies over the business cycle.

This paper considers the importance of these selectivity issues in making imputations about housing price trends. We estimate a model of housing price determination and of the non-random selection of observed transactions. We analyze the factors affecting the probabilities that transactions on different houses will be observed, and we estimate the effect of these factors upon housing prices. The analysis considers a variety of plausible selection models. For each of the alternatives, the estimated effect of selectivity upon housing price calculations is quite substantial.

The analysis is based on a unique body of data containing observations of all house sales in Sweden during the period 1981-1993.

## I. Introduction

Estimates of the value of stocks of durable goods are derived from observations on sales. Often the sales represent a small fraction of the stock, and imputations of value may be crude. In the property market, appraisers use sales of houses or other real property to estimate the values of other properties. Sales information is also used to compute price indexes for the housing stock by relying upon a variety of statistical techniques. These aggregate price measures, however, are derived from a very small amount of information. In the U.S. single family housing market, for example, only about seven percent of the standing stock is sold in any year. In most other countries the fraction is even smaller. In the Swedish housing market, the source the data analyzed below, only about three percent of the stock turns over in a given year.

There are several mechanisms that could generate a sample of house sales out of a population of houses during any time interval. First, the observable characteristics of houses or of time periods may affect the trading propensity of dwellings. Life cycle savings behavior may suggest that young households will purchase smaller, less expensive dwellings and will "trade up" several times as circumstances permit. In this case, with a growing population a sample of sales would include a

disproportionate share of these "starter homes."

Second, the unobservable characteristics of houses sold frequently could differ from those sold infrequently. For example, if some defects in dwellings were difficult for potential purchasers to uncover, then as long as the number of transactions on a house were public information, dwellings sold more frequently would sell for less than those sold infrequently (regardless of their underlying quality). This is a standard "lemons" effect arising from the asymmetry of information between buyer and seller (see Akerlof, 1970).

Third, house sales could be a random sample from the stock of houses. People die; they are transferred; they move to other regions. For a variety of idiosyncratic reasons, dwellings appear on the market in any given time interval.

Little empirical evidence exists on potential selectivity. Case, et al. (1997) analyzed the housing characteristics and price appreciation patterns for houses in four U.S. counties. They compared houses which sold more frequently with those sold less frequently, finding significant differences in types of dwellings and patterns of price change. Gatzlaff and Haurin (1997) analyzed house sales in Dade County, Florida. Clapp and Giacotto (1992) analyzed house sales in Connecticut, and Jud and Seaks (1994) analyzed house sales in Greensboro, North Carolina.<sup>1</sup>

---

<sup>1</sup> All of these studies deal with the selection problem within a

These studies provide weak evidence that house sales are not a random sample of the stock of houses. Presumably, failure to account for non-random selection of houses biases statistical analyses based on samples of observed sales.

This paper extends these analyses in two ways. First, it provides a more complete analysis of the nature of non-randomness in samples of housing transactions than has been previously reported. We present and test several models of the selection process. In the most general model, we postulate that the probability that a dwelling is sold at two points in time varies systematically with its physical characteristics and with the *specific time periods* themselves. We also test special cases of this model, including the hypothesis that the number of sales of any dwelling in a given time interval depends only upon the characteristics of the dwelling.

Second, the paper provides a far more complete quantitative analysis of the effects of these forms of selectivity on housing price calculations. We accomplish this by analyzing *all* single family housing transactions in Sweden during a 13-year period; the analysis is based on almost half a million transactions including more than 100,000 repeat sales of owner-occupied

---

repeat-sales framework, i.e., they compare a repeat-sales price index with an index computed from single sales. Di Pasquale and Somerville (1995) analyze the selectivity of single sales compared with unsold dwellings.

dwellings. We estimate the nature and incidence of selectivity in samples of house transactions for each of the eight administrative regions in the country. We use this information to analyze the effects of sample selectivity on measures of housing prices in each of these regions.

We find in all cases that samples of sold dwellings are decidedly non-random samples of the housing markets from which they are selected. In general, the probability that any house sells depends upon the physical characteristics of the dwellings and the time period under consideration. We also find, with one important exception, that this selectivity has substantial effects upon estimates of housing prices. In seven of the eight regions in Sweden, selectivity-corrected price indexes show smaller price increases over the 13-year period investigated. The differences are reasonably large and are consistent across various selection models, suggesting that, over this period, the price appreciation of houses observed to be sold was 5 to 11 percent larger than the unrealized capital gains on elements in the larger stock of unsold dwellings.

We find essentially no evidence that the unobserved characteristics of dwellings affect housing prices after controlling for those observable characteristics which influence the frequency of sale. Apparently, the transactions costs of buying and selling are large enough, relative to the cost of



repairing defects, to prevent disappointed purchasers from disposing of lemons.

Section II presents a simple model of housing sales and selectivity. Section III outlines the estimation procedure. Section IV describes the data utilized. Section V presents empirical estimates of the model and reports their implications for the estimation of aggregate housing prices. Section VI is a brief conclusion; an appendix provides more detail on the sample selectivity issue.

## **II. The Repeat Sales Index and Sample Selectivity**

An accurate measure of aggregate housing prices must account for heterogeneity in the stock. We control for quality by utilizing a method which controls for heterogeneity by comparing the observed sales price of the same unit at two points in time (see Bailey, Muth, and Nourse [1963]). With quality held constant, changes in price are attributed solely to the effect of time. However, limiting the sample to dwellings that sell two or more times greatly reduces the fraction of the stock represented in the data. For reasons noted above, this may leave the resulting estimates of the aggregate price index particularly susceptible to sample selection bias.

To analyze this, let  $i$  and  $t$  index dwellings and time

periods, respectively. Define  $P_{it}$  as the logarithm of house value (i.e., selling price),  $X_{it}$  as the set of relevant characteristics of the physical structure, including location,  $D_{it}$  as a set of dummy variables with a value of one for the time period of sale (and zero otherwise), and  $\varepsilon_{it}$  as a well-behaved error term. Then we may express the price as

$$(1) \quad P_{it} = X_{it}\beta + D_{it}\delta + \varepsilon_{it},$$

where  $\beta$  and  $\delta$  represent vectors of hedonic coefficients. The price difference between two sales of the same unit at time  $t$  and  $\tau$  is

$$(2) \quad P_{it} - P_{i\tau} = (X_{it} - X_{i\tau})\beta + (D_{it} - D_{i\tau})\delta + \varepsilon_{it} - \varepsilon_{i\tau}.$$

If the set of physical characteristics remains unchanged over time, i.e.,  $X_{it} = X_{i\tau}$ , then equation (2) simplifies to

$$(3) \quad P_{it} - P_{i\tau} = D_{is}\delta + v_{it\tau},$$

where

$$(4) \quad D_{is} = D_{it} - D_{i\tau} = \begin{cases} 1 & \text{if } s = t \\ -1 & \text{if } s = \tau \\ 0 & \text{otherwise} \end{cases}$$

and,

$$(5) \quad V_{it_t} = \varepsilon_{it} - \varepsilon_{i_t}.$$

Estimates of the effect of time are obtained by regressing the difference in log price on  $D_{is}$ . Because the characteristics of the dwelling unit are identical at the time of the two sales, quality is held constant. Requiring only the transaction prices and dates for two sales from the same unit, the model is a parsimonious means of obtaining estimates of the course of aggregate housing prices.

Following Gatzlaff and Haurin (1997), a sale is observed as the result of two price-generating processes. Let  $P_{it}^O$  be the log offer price made in period  $t$  by a potential buyer of unit  $i$ , and  $P_{it}^R$  be the reservation price held by the owner and potential seller of unit  $i$ . These prices can be described by

$$(6) \quad P_{it}^O = P_{it} + \varepsilon_{it}^O ;$$

and

$$(7) \quad P_{it}^R = P_{it} + \varepsilon_{it}^R .$$

Offer prices reflect buyers' preferences, reservation prices, and perceptions of market conditions. Reservation prices reflect sellers' costs of waiting as well. We assume the errors in equations (6) and (7) are well behaved.

A sale occurs when the price offered by a potential buyer,  $P_{it}^O$ , is at least as large as the reservation price held by the potential seller,  $P_{it}^R$ . Because data are generated only for sold dwellings, the expected transaction price of an observed sale, the expectation of equation (1), is

$$(8) \quad E(P_{it}) = X_i\beta + D_{it}\delta + E(\varepsilon_{it} \mid P_{it}^O \geq P_{it}^R),$$

where  $\beta$  and  $\delta$  are the hedonic coefficient vectors. Estimates of  $\beta$  and  $\delta$  are subject to bias if sample selection is not random. In the repeat sales model, equation (3), an observation is generated only if two sales of the same unit occur, that is, only if

$$(9) \quad P_{it}^O \geq P_{it}^R \text{ and } P_{it}^O \geq P_{it}^R.$$

The expected difference in log price for the sample of observed sales is

$$(10) \ E(P_{it}-P_{i\tau}) = D_{is}\delta + E(v_{it\tau} \mid P_{it}^O \geq P_{it}^R, P_{i\tau}^O \geq P_{i\tau}^R).$$

As in the single-period model, the estimated coefficients in equation (10) are subject to bias if the conditional expectation of the error term is nonzero.

As shown by Heckman (1979), consistent estimates of the coefficients in equation (8) or (10) may be obtained by modeling the process that selects dwellings into the set of observations on sales. Heckman shows that the inclusion of the inverse Mills' ratio, derived from the selection process, in the subsequent regression yields unbiased estimates of the parameters, despite non-random sample selection. The selectivity-corrected repeat sales model associated with equation (10) is

$$(11) \ P_{it}-P_{i\tau} = D_{is}\delta + \psi\lambda_{it\tau} + \omega_{it\tau},$$

where  $\lambda_{it\tau}$  is the inverse Mills' ratio associated with an observation of paired sales at times  $t$  and  $\tau$ , and  $\omega_{it\tau}$  is a well-behaved error term. Thus, unbiased estimates of aggregate price movements in the stock of housing may be based on the non-random sample of dwellings sold two or more times during a time interval.

### III. The Estimation Procedure

As indicated above, a house sale is observed in period  $t$  if and only if the price offered by a potential buyer exceeds the reservation price of the current owner. Let  $S_{it\tau}$  equal one if the  $i$ th dwelling is sold in period  $t$  and also in period  $\tau$ . In general, the probability that  $S_{it\tau}$  equals one depends both on the specific time periods involved - e.g. because mobility varies - and on house characteristics - e.g. because smaller houses, "starter homes" are easier to sell. This may be expressed as

$$(12) \quad \text{prob}(S_{it\tau} = 1) = \text{prob}(f(Z_i, t, \tau) + \eta_{it\tau} > 0)$$

where  $Z_i$  is some set of physical characteristics, and the composite error term  $\eta_{it\tau}$  includes any idiosyncratic characteristics of the sellers and prospective buyers of dwelling  $i$  at  $t$  and  $\tau$ .

Equation (12) may be estimated as a probit and the inverse Mills' ratio  $\lambda_{it\tau}$  computed directly for inclusion in equation (11).<sup>2</sup> In this formulation, the probability of sale of a dwelling

---

<sup>2</sup> The inverse Mills' ratio is defined as  $\lambda_{it\tau} = \phi(\text{prob}[S_{it\tau}]) / \Phi(\text{prob}[S_{it\tau}])$  where  $\phi$  is the standard normal density function and  $\Phi$  is the cumulative normal density function.

in two specific periods is a function of the specific time periods involved and some set of physical characteristics,  $Z$ .

Simpler special cases may be more plausible. For example, suppose the probability of a particular house being sold at  $t$  is independent<sup>3</sup> of its probability of sale at  $\tau$ , i.e.

$$(13) \text{ prob}(S_{it}=1) = \text{prob}(S^*_{it}=1) \times \text{prob}(S^*_{i\tau}=1)$$

where

$$(14) \text{ prob}(S^*_{it}=1) = \text{prob}(g[Z_i, t] + \eta_{it} > 0).$$

Alternatively, and still more restrictively, suppose the probability of sale is a function only of the characteristics of the dwelling itself, i.e.,

$$(15) \text{ prob}(S_{it}=1) = \text{prob}(S^{**}_{it}=1)$$

where

$$(16) \text{ prob}(S^{**}_{it}) = \text{prob}(h[Z_i] + \eta_i > 0).$$

This special case may reflect the belief that "starter homes," are equally likely to sell in any time period and are more likely to sell than larger and more expensive properties (see Case et al., 1997, for a discussion). Note that, in these selectivity models, the probability of sale is a function of characteristics observable to buyers and sellers. See footnote 8 below for evidence from models where we postulate that unobservables also affect the probability of sale and the selling price.

#### **IV. The Data**

The data used in this analysis consist of all sales of owner-occupied housing in Sweden during the period from January 1, 1981 through August 28, 1993. Contract data reporting the transaction price for each sale have been merged with tax assessment records containing detailed information about the characteristics of each house. The merged data set contains 462,749 observations on sales from 393,908 separate dwellings in eight administrative regions. Figure 1 indicates the regional character of the data. The largest conurbations are located in region I (Stockholm), region V (Gothenburg), and region IV (Malmö). Time is recorded in 26 half-year intervals. The data set is exceptional in its detailed description of each dwelling at the

---

<sup>3</sup> See Gatzlaff and Haurin (1997) for a discussion.



date of sale and its identification of repeat sales. These data are described in more detail in Englund et al. (1998).

The selection process is estimated from observations on the attributes of each dwelling and a set of dummy variables indicating two half-years of potential sale. The dependent variable in the most general selection model ( $S_{it}$ , from equation (12)) has a value of one if the dwelling was sold in *both* half-years indicated by the dummy variables. Each dwelling is observed in 325 ( $=26*25/2$ ) pairs of half-year periods. Dwellings are observed to sell up to eight times during the period. Apart from the characteristics of the dwelling, one additional independent variable is included in the analysis: gross migration, the total number of in and out migrants, measured separately by region and half-year interval.

Table 1 indicates the extent to which the reliance on unchanged repeat sales limits the size of the available sample. The large majority of dwellings were sold only once during the 13-year sample period. The tail of the distribution of sales is long but thin -- note that 334,007 dwellings sold once between 1981 and 1993, but only 52,097 dwellings with unchanged characteristics were exchanged twice. Only 7,804 dwellings with unchanged characteristics sold three or more times. Note that we restrict the sample to transactions without changes in physical

characteristics between sales. See Englund et al. (forthcoming) for an indication of the importance of this restriction (which typically cannot be made in repeat-sales studies).

It is quite clear that estimation of house price indices using repeat sales (i.e., equation 3) utilizes data covering but a small fraction of sold homes, and an even smaller fraction of the entire stock. The sample of dwellings sold during the entire sample period represents only about 25 percent of the stock of single-family houses in Sweden. This sample shrinks when restricted to unchanged repeat sales, accounting for only five percent of the housing stock.

Table 2 reports averages of selected housing attributes as a function of the frequency of sale of dwellings in each of the eight regions. The sample is divided into single and repeat sales, and then further restricted to dwellings that sold three or more times. The pattern is clear: newer, smaller, lower quality, and lower priced houses sell more frequently. Repeat-sale dwellings are also more likely to be close to the center of the local labor market, and are less likely to be detached units. The table provides support for the notion that lower priced dwellings sell more often, but it also suggests that the population of repeat sales may not be representative of the larger stock of dwellings.

## V. Sample Selectivity and House Prices

The most general form of the probit selection model, eq. (12), is, assuming linearity,

$$(17) \text{ prob}(S_{it_t} = 1) = \text{prob}(\alpha Z_i + \gamma T_i + \theta_1 M_t + \theta_2 M_t + \eta_{it_t} > 0),$$

where  $Z_i$  is a vector including 11 characteristics of dwelling  $i$ :  $M_j$  represents gross migration at time  $j$ , and  $T_i$  is a vector of 26 variables measuring time periods, with a value of one for each of the two periods in which a sale is recorded, and zero otherwise. The symbols  $\alpha$ ,  $\gamma$ , and  $\theta$  represent estimated coefficients, and  $\eta_{it_t}$  is a composite error term, assumed normally distributed. Each dwelling is observed 325 times, with the periods in which sales occur noted in the vector  $T$ .<sup>4</sup>

The more restrictive models of selectivity, eqs. (14) and

(16), are

---

<sup>4</sup> The probit models in equation (17) are estimated using choice-based samples drawn from the 325 alternative ways in which two sales can be consummated in 26 half-year periods. The samples include observations on all double sales and a random sample of approximately five percent of the alternative time-period combinations in which repeat sales did not occur. These choice-based samples are weighted according to the technique suggested by Manski and McFadden (1981). The sample sizes for the probit results range from 486,000 in Region 7 to 2,219,000 in Region 5.

$$(18) \quad \text{prob}(S^*_{it}=1) = \text{prob}(\alpha Z_i + \gamma \Gamma_i + \theta M_t + \eta_{it} > 0),$$

and

$$(19) \quad \text{prob}(S^{**}_{it}=1) = \text{prob}(\alpha Z_i + \eta_{it} > 0).$$

In equation (18),  $\Gamma_i$  is a vector of 26 time variables, with a value of one in the time period in which a sale is recorded.<sup>5</sup>

Table 3 summarizes the results of the estimated selectivity models using the time-invariant probit model, equation (19). By and large, the probit results confirm the patterns noted in Table 2. Smaller dwellings with fewer amenities are more likely to trade. In general, the results are not sensitive to the choice of model, even though the estimated  $\alpha$  coefficients are generally less significant when time dummies are included. Further, the gross migration variables are only marginally significant when the model includes dummy variables for time, as in equations (17) and (18). The results indicate that the probability of sale, *ceteris paribus*, dropped sharply after 1991.

---

<sup>5</sup> The probit models in equation (18) are estimated using samples which include all sales during the period, but without distinguishing multiple sales of any property. The dependent variable for these analyses is the sale or nonsale of each

Table 4 summarizes the implications of these models of selectivity for the estimates of housing prices. The table reports the coefficient of the inverse Mills' ratio in the equation estimating the selectivity-corrected price index (i.e., the coefficient  $\psi$  in equation 11). It also summarizes the difference between prices computed from the uncorrected estimator, equation (3), and the selectivity-corrected estimator, equation (11). These results are reported for each of the three selectivity models, equations (17), (18), and (19).

The coefficients of the inverse Mills' ratio based upon these selection models are large and highly significant in the estimation of the price index -- at least for all regions outside Stockholm. This indicates that sample selectivity "matters" in the computation of the appropriate housing price index.<sup>6</sup> The inverse Mills' ratio is significantly positive and important for all three formulations of the selection model.<sup>7</sup>

---

dwelling in each time period.

<sup>6</sup> The selection specification in equation (18) seems more plausible to us, but not to some others who have read preliminary versions of this paper.

<sup>7</sup> Strictly interpreted, the standard selection-correction method which underlies the results reported in Table 4 (and Appendix Table A1) requires that the errors in equations (10) and (12) be jointly normally distributed. A test of this restriction was made using a nonparametric technique suggested by Newey, Powell, and Walker (1990). (See also Ahn and Powell, 1993). The structure of the selection correction terms,  $\lambda_{it_r}$ ,  $\lambda_{it}$ ,  $\lambda_i$  in the different models is approximated through a series of basis functions, whose arguments are the single-valued index function  $Z\delta$ . Numerous combinations of approximations were included in the second step

Panels B and C in the table summarize the extent of the differences between the biased estimates of housing prices and the selectivity-corrected estimates. In this comparison, we normalize the indexes at 100 at the beginning of the period, 1981:I for each region, and compare the subsequent estimates. We report this comparison for each of the three selection models. The average discrepancy between the uncorrected and corrected price indexes is negligible in Stockholm (region I) but quite large in all other regions -- ranging from two to eleven percentage points depending on model and region. Figure 2 is based upon equation (18); it presents the biased and unbiased estimates of housing prices for Stockholm, Gothenburg and Malmö, Sweden's three largest metropolitan regions, during the period 1987-93. As the figures illustrate, the selectivity correction adds little in Stockholm, but the selectivity-corrected measures of housing prices are quite a bit lower in the other two metropolitan areas. The differences peak towards the end of the period, when the selectivity-corrected indexes are 2-5 percent below the uncorrected indexes<sup>8</sup>.

---

regressions reported in Table 4 (and in Appendix Table A1 as well) without any significant change in the estimated coefficients in equation (11) or the resulting house price indexes.

<sup>8</sup> Note that in all three selectivity models, the probability of sale in any period is a function of the observable characteristics of dwellings and time. We analyzed the extent to which unobservables affected house prices by adding a variable indicating the number of times each house was sold to equations

The value of owner-occupied homes comprises about two thirds of household net wealth in Sweden. This suggests that correcting for sample selectivity lowers the estimate of 1993 household wealth by 2-3 per cent, relative to its value in 1981. The overvaluation accumulates gradually over time but is arrested in 1991 when the housing price cycle in Sweden reached its peak. There is a slight tendency in the opposite direction after 1991, suggesting that the direction of the bias might be related to the housing price cycle itself.<sup>9</sup> Generally the differences in time patterns between the two index series are not dramatic. The differences in rates of change do not exceed 1.4 percent in any half-year.

The appendix presents the results from a special case of the time-invariant model of selectivity, equation (19), corresponding to a Poisson process generating house sales from the population. In this case, the average deviation between the uncorrected and the selectivity-corrected price index ranges

---

(17), (18) and (19). After controlling for the observable characteristics of dwellings which affect the frequency of sale, the additional variable reflects any unmeasured characteristics of dwellings which affect sale frequencies. For seven of the eight regions, the coefficient estimate was negative but in only one case was the estimate significantly less than zero, providing only quite weak evidence that lemons behavior is important in this market. Of course, if transaction costs are 5-10 percent of sales prices, it would require the concealment of very expensive defects to induce high turnover in the housing market.

<sup>9</sup> The time span covered by the data is too short to allow us to distinguish this from the alternative interpretation that the biased indexes tend to overestimate consistently the rate of

between 5 and 10 percentage points in the regions outside Stockholm. The maximum deviation approaches 24 percentage points.

## **VI. Conclusion**

In this paper, we have examined the nature of the selection process that distinguishes dwellings which are sold frequently from the entire stock of sold dwellings. Specifically, we consider the influence of time and a dwelling's physical attributes on its probability of sale at two points in time. We have also explored the impact of these relationships on the measurement of aggregate housing prices.

We find, using a sample of essentially all arm's-length sales in Sweden during a 13-year period, that the selection process governing dwelling unit sales is distinctly non-random, confirming earlier suggestive work. We also find that the appropriate correction for the selection process implies that housing price appreciation is otherwise overstated in a price change.conventional repeat-sales price index.

The ramifications for national housing wealth may be substantial. The results indicate average deviations in the estimated indexes attributable to sample selection ranging

---

price change.



between two and eleven percent towards the end of the period, a substantial difference given the size of the housing stock. The implications are clear: the use of transactions data requires careful consideration of the process that generates the observations, and the non-random nature of the selection process has a significant impact on measured aggregate housing prices.

## Bibliography

- Ahn, Hyungtaik, and Hames L. Powell, "Semiparametric Estimation Of Censored Selection Models with a Nonparametric Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 1993: 3-29.
- Akerlof, George, "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84, 1970: 481-500.
- Bailey, Martin J., Richard F. Muth, and Hugh O. Norse. "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*, 58, December 1963: 933-42.
- Case, Bradford, Henry O. Pollakowski and Susan M. Wachter, "Frequency of Transaction and House Price Modeling," *Journal of Real Estate Finance and Economics*, 14, 1997: 173-87.
- Clapp, John M. and Carmelo Giacotto, "Estimating Price Trends for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods," *Journal of Real Estate Finance and Economics*, 5, 1992: 357-374.
- DiPasquale, Denise and C. Tsurriel Somerville, "Do House Prices based on Transactions Represent the Entire Housing Stock? Evidence from the National and Metropolitan AHS, 1974-1991," *Journal of Housing Economics*, 4, 1995: 100-110.
- Englund, Peter, John M. Quigley and Christian L. Redfearn, "Improved Price Indexes for Real Estate: Measuring the Course of Housing Prices," *Journal of Urban Economics*, 44, 1998: 171-196.
- Englund, Peter, John M. Quigley and Christian L. Redfearn, "The Choice of Methodology for Computing Housing Price Indexes: Comparisons of Temporal Aggregation and Sample Definition," *Journal of Real Estate Finance and Economics*, forthcoming.
- Gatzlaff, Dean H., and Donald R. Haurin. "Sample Selection and Repeat Sales Index Estimates," *Journal of Real Estate Finance and Economics*, 14(1-2), January-March 1997: 33-50.
- Gatzlaff, Dean H., and Donald R. Haurin. "Sample Selection and

- Repeat Sales Index Estimates," *Journal of Urban Economics*, 43(2), March 1998: 33-50.
- Hausman, Jerry A. and David A. Wise. "Social Experimentation, Truncated Distributions, and Efficient Estimation," *Econometrica*, 45(4), May 1977: 919-38.
- Heckman, James. "Sample Selectivity Bias as a Specification Error," *Econometrica*, 47(1), January 1979: 153-61.
- Jud, Donald G. and Terry G. Seaks. "Sample Selection Bias in Estimating Housing Sales Prices," *Journal of Real Estate Research*, 9(3), Summer 1994: 289-98.
- Manski, Charles F. and Daniel McFadden, *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: MIT Press, 1981.
- Newey, Whitney K., James L. Powell, and James R. Walker, "Semiparametric Estimation of Selection Models: Some Empirical Results," *American Economic Review*, 80 (3), May 1990: 324-328.

## Appendix A

The simple time-independent form of selection equation (19) suggests a more powerful method of estimating this model using the total number of sales of each dwelling during the 13-year analysis period. Specifically, if  $Y_i$  is the number of sales of dwelling  $i$  during the period and if this count follows a Poisson process, then the truncated Poisson distribution describes the probability that  $Y_i$  equals the count of sales observed during the period:

$$(A1) \quad \text{prob}(Y_i = y | y \geq 1) = \frac{\Lambda_i^y}{y! (e^{\Lambda_i} - 1)}$$

where

$$(A2) \quad \log(\Lambda_i) = k(X_i) + \eta_i^\#$$

The Poisson arrival parameter,  $\Lambda_i$ , is estimated for the sample period for each dwelling  $i$ . The arrival rate of sales for a single period is then  $\Lambda_i/T$ , where  $T$  is the number of periods. The probability of sale in periods  $t$  and  $\tau$  is

$$(A3) \text{ prob}(S_{it_t}) = \text{prob}(S_{it}) \times \text{prob}(S_{i_t}) = (1 - \text{prob}[Y_i=0])^2 \quad .$$

Equation (A1) is estimated by maximum likelihood methods. Equation (A3) can be computed directly from equations (A1) and (A2). When the selectivity correction is based upon equation (A3), the Mills' ratio is again highly significant in seven of the eight regions. The average deviation between the uncorrected and the selectivity-corrected price index ranges between 5 and 10 percentage points.

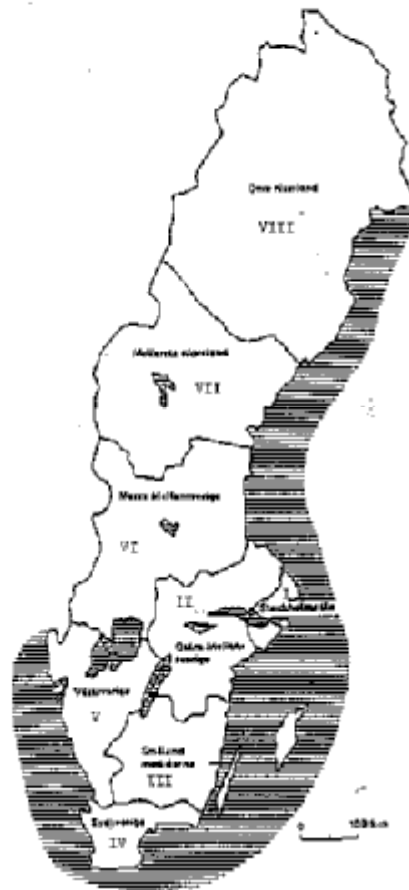
Appendix Table A1 reports the implications of the sample selectivity model based upon the Poisson model. The coefficients are similar to those reported in Table 3. The t-ratios of the selectivity parameter are somewhat higher than those reported in the text. (Again, there is no evidence of sample selectivity in Stockholm.) The average deviation estimated by this selectivity model is somewhat larger and the maximum deviation is substantially larger.

Appendix Table A1  
 Implications of Poisson Model of Sample Selectivity on House Price Estimates

	Region							
	I	II	III	IV	V	VI	VII	VIII
A. Estimated Coefficient of the Inverse Mills Ratio in Price Index Equation: (t-ratio in parentheses)	0.004 (0.84)	0.026 (7.50)	0.046 (8.53)	0.049 (10.57)	0.030 (8.08)	0.048 (9.18)	0.046 (5.23)	0.042 (5.86)
B. Average deviation between biased and selectivity-corrected price index: (in percentage points)	0.93	5.38	8.40	10.21	6.66	9.30	8.52	8.56
C. Maximum deviation between biased and selectivity-corrected price index: (in percentage points)	1.99	11.64	19.71	23.84	13.96	21.46	19.91	18.90

FIGURES AND TABLES

Figure 1  
Boundaries of Major Economic Regions in Sweden



Source: Central Bureau of Statistics



Figure 2  
Effects of Selectivity upon House Price Estimates  
(Equation 18)

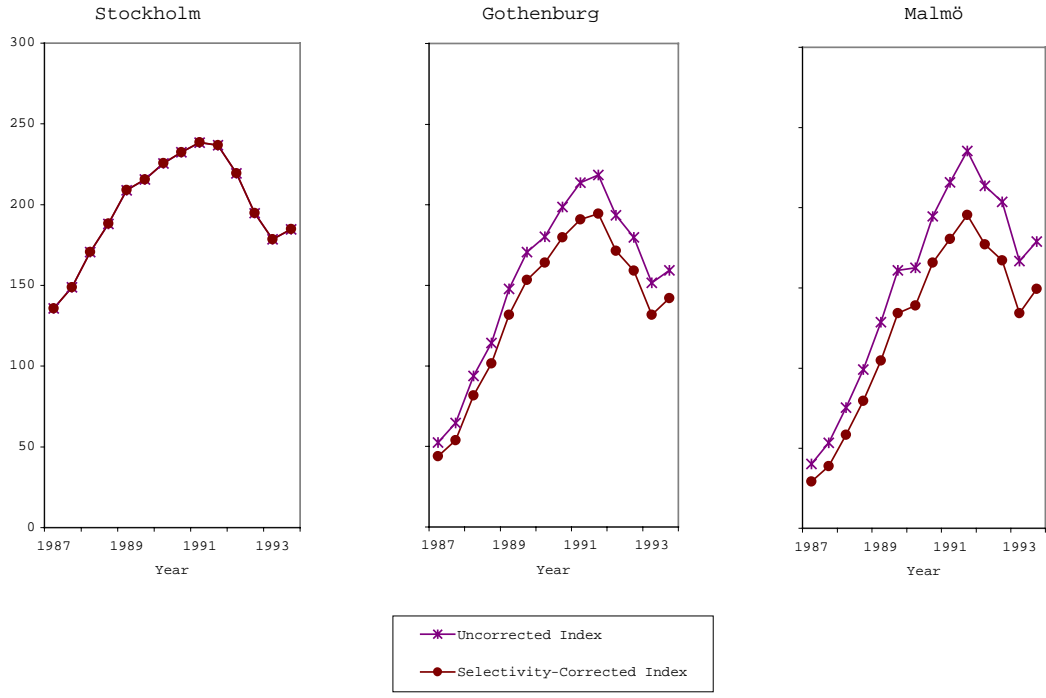


Table 1  
Frequency of Sales:  
Once-Sold Dwellings & Dwellings Unchanged Between Sales

Number of sales	Region								Total number of dwellings	Total number of transactions
	I	II	III	IV	V	VI	VII	VIII		
Sold Once:										
1	47,100	59,170	34,013	54,806	67,014	38,440	14,455	19,009	334,007	334,007
Unchanged between sales:										
2	6,766	9,576	5,197	9,449	10,034	5,670	2,285	3,120	52,097	104,194
3	811	1,301	697	1,238	1,317	760	304	438	6,866	20,598
4	112	154	73	120	170	73	39	52	793	3,172
5	14	34	5	15	18	2	7	11	106	530
6	2	12	1	3	1	1	2	5	27	162
7	1	6	0	0	1	0	1	1	10	70
8	0	1	0	0	0	0	1	0	2	16
Unchanged units sold two or more times										
	7,706	11,084	5,973	10,825	11,541	6,506	2,639	3,627	59,901	128,742
sales										
	63602	83133	46821	77991	91816	52368	20155	26863		
houses										
	54,806	70,254	39,986	65,631	78,555	44,946	17,094	22,636		

Table 2  
Average Characteristics of Dwellings as a Function of Sales Frequency  
(Standard Deviations in parentheses)

Region	I			II			III			IV		
	1	2+	3+	1	2+	3+	1	2+	3+	1	2+	3+
Number of times sold												
Price (thousands of SEK)	780.83 (20.7)	758.38 (19.6)	715.18 (18.5)	482.36 (24.8)	463.49 (24.3)	452.43 (23.4)	377.41 (26.4)	378.02 (26.1)	371.21 (25.8)	443.98 (28.3)	430.56 (28.4)	410.76 (28.4)
Year built (19xx)	59.25 (20.7)	61.08 (19.6)	62.79 (18.5)	55.36 (24.8)	56.28 (24.3)	57.57 (23.4)	50.87 (26.4)	52.17 (26.1)	52.80 (25.8)	47.66 (28.3)	46.14 (28.4)	44.08 (28.4)
Interior size (square meters)	123.39 (36.8)	119.59 (34.3)	115.87 (32.3)	119.91 (36.6)	118.78 (36.0)	117.15 (35.8)	122.61 (39.8)	122.07 (39.5)	120.76 (39.3)	120.58 (40.0)	118.31 (39.7)	115.39 (39.3)
Parcel size (square meters)	884.25 (851.8)	728.13 (732.9)	643.30 (700.9)	1176.67 (1176.3)	1076.30 (1070.4)	986.26 (1037.7)	1337.94 (1165.9)	1212.97 (1087.9)	1154.44 (1096.8)	1108.89 (1107.5)	1047.29 (1038.0)	1004.13 (973.7)
Two car garage (fraction)	0.048 (0.21)	0.045 (0.21)	0.042 (0.20)	0.082 (0.27)	0.072 (0.26)	0.061 (0.24)	0.071 (0.26)	0.064 (0.25)	0.046 (0.21)	0.044 (0.21)	0.043 (0.20)	0.041 (0.20)
Tiled bath (fraction)	0.121 (0.33)	0.115 (0.32)	0.101 (0.30)	0.091 (0.29)	0.090 (0.29)	0.081 (0.27)	0.094 (0.29)	0.097 (0.30)	0.091 (0.29)	0.139 (0.35)	0.148 (0.36)	0.146 (0.35)
Sauna (fraction)	0.220 (0.41)	0.212 (0.41)	0.186 (0.39)	0.214 (0.41)	0.219 (0.41)	0.205 (0.40)	0.172 (0.38)	0.186 (0.39)	0.185 (0.39)	0.124 (0.33)	0.119 (0.32)	0.120 (0.33)
Detached House (fraction)	0.693 (0.46)	0.612 (0.49)	0.541 (0.50)	0.818 (0.39)	0.788 (0.41)	0.734 (0.44)	0.886 (0.32)	0.860 (0.35)	0.831 (0.37)	0.872 (0.33)	0.854 (0.35)	0.834 (0.37)
Stone/Brick ext. (fraction)	0.242 (0.43)	0.222 (0.42)	0.186 (0.39)	0.354 (0.48)	0.351 (0.48)	0.337 (0.47)	0.350 (0.48)	0.340 (0.47)	0.333 (0.47)	0.557 (0.50)	0.535 (0.50)	0.517 (0.50)
Laundry Room (fraction)	0.841 (0.37)	0.844 (0.36)	0.846 (0.36)	0.813 (0.39)	0.824 (0.38)	0.823 (0.38)	0.784 (0.41)	0.805 (0.40)	0.806 (0.40)	0.783 (0.41)	0.784 (0.41)	0.772 (0.42)
Fireplace (fraction)	0.389 (0.49)	0.332 (0.47)	0.274 (0.45)	0.351 (0.48)	0.327 (0.47)	0.293 (0.46)	0.475 (0.50)	0.454 (0.50)	0.412 (0.49)	0.262 (0.44)	0.255 (0.44)	0.256 (0.44)
Winter walls/wndws. (fraction)	0.168 (0.37)	0.154 (0.36)	0.144 (0.35)	0.196 (0.40)	0.182 (0.39)	0.167 (0.37)	0.172 (0.38)	0.170 (0.38)	0.151 (0.36)	0.182 (0.39)	0.173 (0.38)	0.151 (0.36)
Electric radiator (fraction)	0.385 (0.49)	0.426 (0.49)	0.463 (0.50)	0.306 (0.46)	0.349 (0.48)	0.376 (0.48)	0.296 (0.46)	0.341 (0.47)	0.372 (0.48)	0.313 (0.46)	0.338 (0.47)	0.352 (0.48)
Good kitchen (fraction)	0.210 (0.41)	0.176 (0.38)	0.173 (0.38)	0.226 (0.42)	0.203 (0.40)	0.186 (0.39)	0.268 (0.44)	0.236 (0.43)	0.238 (0.43)	0.282 (0.45)	0.275 (0.45)	0.284 (0.45)
Good/excellent roof (fraction)	0.692 (0.46)	0.638 (0.48)	0.599 (0.49)	0.788 (0.41)	0.759 (0.43)	0.709 (0.45)	0.828 (0.38)	0.799 (0.40)	0.778 (0.42)	0.684 (0.47)	0.653 (0.48)	0.614 (0.49)
Distance to center (kilometers)	4.806 (6.04)	4.636 (6.18)	4.236 (5.89)	5.762 (6.87)	5.558 (6.80)	5.348 (6.60)	8.295 (9.58)	7.282 (9.15)	6.672 (8.43)	5.261 (5.26)	5.405 (5.36)	5.353 (5.33)

Table 2 - continued  
Average Characteristics of Dwellings as a Function of Sales Frequency  
(Standard Deviations in parentheses)

Region	V			VI			VII			VIII		
	1	2+	3+	1	2+	3+	1	2+	3+	1	2+	3+
Number of times sold												
Price (thousands of SEK)	506.68 (23.7)	478.75 (22.7)	450.10 (22.4)	383.73 (24.8)	384.57 (24.6)	371.59 (24.4)	376.13 (25.2)	383.35 (24.4)	375.04 (23.6)	405.69 (24.8)	402.70 (23.7)	387.52 (23.3)
Year built (19xx)	55.25 (23.7)	57.27 (22.7)	58.00 (22.4)	52.22 (24.8)	53.69 (24.6)	53.66 (24.4)	51.24 (25.2)	54.01 (24.4)	54.87 (23.6)	55.80 (24.8)	57.89 (23.7)	58.18 (23.3)
Interior size (square meters)	119.03 (38.3)	116.93 (36.8)	114.24 (38.3)	116.17 (37.9)	116.64 (37.5)	114.26 (37.0)	116.77 (38.6)	116.10 (36.1)	113.11 (33.1)	117.74 (36.0)	117.83 (35.1)	115.59 (31.6)
Parcel size (square meters)	1140.73 (1145.4)	1011.04 (1041.4)	952.31 (1016.1)	1405.92 (1240.6)	1271.49 (1144.5)	1212.83 (1126.2)	1469.36 (1309.0)	1214.25 (1088.9)	1120.85 (1068.3)	1280.36 (1177.3)	1082.97 (959.9)	974.13 (765.7)
Two car garage (fraction)	0.064 (0.24)	0.052 (0.22)	0.045 (0.21)	0.076 (0.26)	0.068 (0.25)	0.056 (0.23)	0.061 (0.24)	0.048 (0.21)	0.042 (0.20)	0.106 (0.31)	0.094 (0.29)	0.079 (0.27)
Tiled bath (fraction)	0.113 (0.32)	0.104 (0.31)	0.090 (0.29)	0.082 (0.27)	0.080 (0.27)	0.072 (0.26)	0.074 (0.26)	0.073 (0.26)	0.057 (0.23)	0.069 (0.25)	0.067 (0.25)	0.064 (0.25)
Sauna (fraction)	0.180 (0.38)	0.171 (0.38)	0.163 (0.37)	0.186 (0.39)	0.191 (0.39)	0.182 (0.39)	0.207 (0.41)	0.203 (0.40)	0.172 (0.38)	0.377 (0.49)	0.405 (0.49)	0.397 (0.49)
Detached House (fraction)	0.808 (0.39)	0.743 (0.44)	0.698 (0.46)	0.897 (0.30)	0.853 (0.35)	0.829 (0.38)	0.874 (0.33)	0.816 (0.39)	0.784 (0.41)	0.885 (0.32)	0.838 (0.37)	0.791 (0.41)
Stone/Brick ext. (fraction)	0.296 (0.46)	0.275 (0.45)	0.253 (0.44)	0.248 (0.43)	0.234 (0.42)	0.210 (0.41)	0.165 (0.37)	0.163 (0.37)	0.156 (0.36)	0.190 (0.39)	0.189 (0.39)	0.171 (0.38)
Laundry Room (fraction)	0.804 (0.40)	0.821 (0.38)	0.816 (0.39)	0.738 (0.44)	0.763 (0.43)	0.746 (0.44)	0.743 (0.44)	0.773 (0.42)	0.792 (0.41)	0.818 (0.39)	0.832 (0.37)	0.823 (0.38)
Fireplace (fraction)	0.354 (0.48)	0.313 (0.46)	0.276 (0.45)	0.386 (0.49)	0.366 (0.48)	0.355 (0.48)	0.357 (0.48)	0.326 (0.47)	0.297 (0.46)	0.295 (0.46)	0.275 (0.45)	0.263 (0.44)
Winter walls/wndws. (fraction)	0.201 (0.40)	0.185 (0.39)	0.172 (0.38)	0.189 (0.39)	0.192 (0.39)	0.186 (0.39)	0.210 (0.41)	0.212 (0.41)	0.198 (0.40)	0.286 (0.45)	0.291 (0.45)	0.291 (0.45)
Electric radiator (fraction)	0.339 (0.47)	0.394 (0.49)	0.437 (0.50)	0.294 (0.46)	0.322 (0.47)	0.344 (0.48)	0.353 (0.48)	0.396 (0.49)	0.399 (0.49)	0.350 (0.48)	0.409 (0.49)	0.441 (0.50)
Good kitchen (fraction)	0.259 (0.44)	0.225 (0.42)	0.213 (0.41)	0.307 (0.46)	0.271 (0.45)	0.266 (0.44)	0.285 (0.45)	0.237 (0.43)	0.220 (0.42)	0.236 (0.43)	0.189 (0.39)	0.185 (0.39)
Good/excellent roof (fraction)	0.798 (0.40)	0.746 (0.44)	0.702 (0.46)	0.786 (0.41)	0.764 (0.43)	0.745 (0.44)	0.609 (0.49)	0.595 (0.49)	0.583 (0.49)	0.501 (0.50)	0.431 (0.50)	0.375 (0.48)
Distance to center (kilometers)	5.947 (5.85)	5.719 (5.76)	5.625 (5.64)	5.859 (9.14)	5.330 (8.07)	5.098 (7.62)	11.629 (14.47)	10.399 (13.48)	9.918 (12.83)	7.243 (14.10)	5.834 (12.20)	5.355 (11.18)

Table 3  
 Estimated Coefficients from Time-Invariant Probit Selection Model, equation (19)  
 ( t-statistics in parentheses )

Region	I	II	III	IV	V	VI	VII	VIII
Intercept	-1.271 (13.43)	-1.412 (16.78)	-1.275 (10.17)	-1.413 (15.93)	-1.246 (17.38)	-1.196 (10.07)	-0.980 (5.84)	-1.204 (8.07)
Interior size <sup>a</sup> (square meters)	-0.027 (4.29)	-0.010 (1.92)	-0.009 (1.32)	-0.014 (2.73)	-0.016 (3.30)	-0.002 (0.39)	-0.004 (0.43)	-0.006 (0.61)
Parcel size <sup>a</sup> (square meters)	-0.067 (2.27)	-0.051 (2.07)	-0.086 (2.37)	-0.043 (1.64)	-0.091 (4.25)	-0.118 (3.49)	-0.167 (3.44)	-0.095 (2.20)
Square of parcel size <sup>a</sup> (square meters)	0.004 (1.66)	0.003 (1.64)	0.005 (1.98)	0.002 (1.29)	0.006 (3.74)	0.007 (3.08)	0.010 (2.80)	0.005 (1.57)
Tiled bathroom (1 = yes)	0.021 (3.87)	0.005 (0.86)	0.005 (0.74)	0.016 (3.50)	0.004 (0.86)	0.003 (0.49)	0.005 (0.44)	-0.001 (0.05)
Sauna (1 = yes)	0.013 (2.94)	0.009 (2.35)	0.009 (1.60)	0.004 (0.77)	0.006 (1.61)	0.007 (1.29)	0.000 (0.00)	0.012 (2.01)
Single detached house (1 = yes)	-0.017 (2.90)	-0.013 (2.38)	-0.007 (0.92)	-0.018 (2.98)	-0.012 (2.24)	-0.017 (2.17)	0.006 (0.50)	-0.008 (0.76)
Laundry room (1 = yes)	-0.005 (0.94)	0.003 (0.60)	0.004 (0.64)	0.000 (0.11)	0.005 (1.20)	0.003 (0.62)	0.005 (0.59)	-0.007 (0.94)
"Winter Quality" walls/windows (1 = yes)	0.015 (3.21)	0.015 (3.79)	0.011 (2.06)	0.013 (3.08)	0.010 (2.70)	0.010 (1.86)	0.016 (2.00)	0.010 (1.69)
Electric furnace (1 = yes)	0.009 (1.57)	0.003 (0.56)	-0.001 (0.07)	0.007 (1.21)	0.003 (0.55)	0.003 (0.49)	-0.001 (0.07)	-0.012 (1.73)
Slate/copper roof (1 = yes)	0.001 (0.14)	-0.003 (0.77)	-0.005 (0.89)	-0.006 (1.75)	-0.010 (2.66)	0.004 (0.88)	0.012 (1.85)	0.001 (0.26)
Distance from City Center (Kilometers)	0.000 (1.30)	0.000 (0.54)	-0.001 (2.15)	0.001 (2.88)	0.000 (0.22)	0.000 (0.37)	0.000 (0.02)	0.000 (1.83)

Note: a - Variable measured in logarithms.

Table 4  
Implications of Alternate Models of Sample Selectivity on House Price Estimates

Selection Model	Region							
	I	II	III	IV	V	VI	VII	VIII
A. Estimated Coefficient of the Inverse Mills Ratio in Price Index Equation: (t-ratio in parentheses)								
Equation (17)	-0.151 (0.57)	1.124 (5.39)	2.587 (6.72)	2.496 (9.12)	1.680 (7.04)	2.084 (7.18)	1.460 (3.01)	1.693 (4.64)
Equation (18)	-0.037 (0.08)	2.258 (6.61)	3.991 (7.20)	4.901 (10.64)	3.092 (8.09)	3.920 (7.73)	3.294 (3.74)	3.782 (5.35)
Equation (19)	0.017 (0.06)	1.577 (7.69)	2.568 (8.01)	2.907 (10.94)	1.895 (8.70)	2.624 (8.44)	2.229 (4.42)	2.206 (5.48)
B. Average deviation between biased and selectivity-corrected price index: (in percentage points; 1981:l=100)								
Equation (17)	-0.33	2.07	4.19	4.60	3.26	3.83	2.54	3.24
Equation (18)	-0.07	3.83	5.71	8.39	5.45	6.21	5.09	6.40
Equation (19)	0.07	5.50	7.92	10.51	7.17	8.62	7.23	7.99
C. Maximum deviation between biased and selectivity-corrected price index: (in percentage points; 1981:l =100)								
Equation (17)	0.68	4.26	10.10	10.25	7.05	7.97	5.36	6.53
Equation (18)	0.16	8.26	12.91	19.93	12.00	13.83	11.78	15.05
Equation (19)	0.14	11.91	18.58	24.52	15.01	19.97	16.80	17.54