

## APPENDIX A: METHOD FOR DETERMINING HOME SALES AND PRICES

All of our home sales records were provided by the Oakland County government. They obtained the records from county and local assessor offices. The file provided to us included both residential and non-residential sales. Residential sales included both sales of undeveloped residential land and sales of land with residential structures. This appendix explains steps we took to make the data usable for our analysis.

### I. Preparing the Data for Analysis

#### A. *Breaking Out Residential from Non-Residential Sales*

1. Created SPSS files for the sales record file and the residential file (i.e. file from assessors that contained all parcels known to be residential land use).
2. Matched records in the sales file to those in the residential file to determine which sales are for residential properties (excludes commercial, including apartment buildings).
3. Saved as a sales file for residential properties only.
4. Determined if there were records in the sales file that were actually residential, but not coded as such because they did not have a corresponding record in the residential file (e.g. land sales for parcels numbers that were later sub-divided (original parcel number replaced in the residential file with subdivided parcel numbers) or sales of land that were not coded as residential (and therefore not in the residential file), because no significant improvement to a residential structure existed on the lot (this is how records become added to the residential file). Slightly less than 10,000 of the sales records were not coded as residential through steps 1-3.
  - a. pulled in the land use code from parcel file for all sales not coded as residential to identify ones that are definitely non-residential
    - i. 2047 were definitely not SF residential
    - ii. 2692 had no use code (likely to be residential, but not definite)
    - iii. 2484 had residential use codes
  - b. Checked a 1% sample to make sure they weren't in either the commercial file or the residential master file. None were found in either file.
  - c. Searched for a sample of those w/o use code in the land use file (separate file provided by the county), but none were found.
  - d. Looked at cvt for obvious patterns. 60 of 61 CVT's (Sfld Twp excluded) were represented.
  - e. Per conversation with a representative from the county
    - i. Those w/residential codes s/b considered residential sales even if they aren't in the residential file. They w/n/b in the residential file, because the assessor did

- not use residential improvements to calculate assessed value (see hard copy of e-mails).
- ii. Ones w/o use code may or may not be residential. Would have to check on a parcel-by-parcel basis.
    - 1) Mapped a few of those w/o use code against the land use layer (using parcel number w/o the last digit).
    - 2) No clear pattern emerged.
      - a) In most cases, there were residential parcels in the vicinity of where the old parcel w/b. Many of which were added to the parcel file at a date later than the date the old parcel was sold.
      - b) Most appear to be parcels sold and sub-divided as residential. But some were clearly still commercial/business. Others were commercial/apartment.
    - 3) Could have assumed that a percentage of these are residential land sales subdivided after sale and removed (the original PIN) from the parcel file. But, because they did not have a cvt or acreage amount they would not add much to the sales price analysis. When combined with the uncertainty as to whether or not they were residential sales this made including these data in the residential analysis problematic. So, we excluded them from the residential sales file analysis.
  - iii. Before coding as sales, we ran frequencies for year of sale, year built, and year remodeled (for those land sales in the residential file) to identify data errors (6 for year built, 5 for year remodeled). Errors were corrected using the following simple assumptions.
    - 1) 1002 = 2002
    - 2) 199 = 1999
    - 3) 203 = 2003
    - 4) 204 = 2004
    - 5) 2997 = 1997
    - 6) 91 = 1991
    - 7) 968 = 1968
    - 8) 9990 = 1990
    - 9) 9999 = 1999

### ***B. Determining Sales for PINS (parcels) w/more Than One Structure***

*When one PIN has multiple structures (and those structures have different key characteristics) it is difficult to determine which structure is associated with the sales record. We used the following method to determine which attributes to associate with such records.*

1. Used the SPSS merge files command (with the master residential file and the tax parcel file) to determine which sales in the residential sales file were linked to parcels where there are either
  - a. duplicate parcels in the residential file (identified as duplicate on master residential file) or
  - b. polygons with multiple PINS (condos) (pinflg in tax parcel file = 1)
2. Created an SPSS file that contained only the duplicate pins.
3. Examined info for each structure within a duplicate pin record on the residential master file to determine whether or not patterns of other identifying variables were evident that would help us determine how to handle duplicate pins. Key finding—key variables differ enough to prevent development of a universal solution.
  - a. par\_assess—values are the same for all cases w/same pin
  - b. par\_taxabl—values are the same for all cases w/same pin
  - c. par\_hmstdp—values are the same for all cases w/same pin
  - d. res\_stylea—many are same, many differ. Ones with many cases differ the most. Some have no value for this field.
  - e. res\_stylel—all appear to have same value as res\_stylea
  - f. res\_yearbu—most differ, some are same, many have zero value even when a res\_style exists, values tend to be same when res\_style is same
  - g. res\_yearre—few have a date, value tends to be same when res\_yearbu is same
  - h. res\_roomsb—most differ, some are same, some have zero value, value tends to be same when res\_yearbu is same
  - i. res\_roomsf—most differ, some are same, some have zero value, value tends to be same when res\_yearbu is same
  - j. res\_rooms2—most differ, some are same, some have zero value, value tends to be same when res\_yearbu is same
  - k. res\_plumb3—most differ, some are same, some have zero value
  - l. res\_plumb2—most differ, some are same, some have zero value
4. For a sample of the duplicate pin properties we viewed pictures of the properties using the URL for the parcel from the parcel attributes file and Access Oakland web site.
  - a. Many showed only one building. Some showed a primary residential building with another building that might or might not be residential. A few showed two or more residential buildings.
  - b. Determined that these sales (even when there are multiple buildings) pertain to all structures on the property. For ease of analysis, we used the attributes of the largest residential building on the property, but flagged it as a duplicate PIN for further analysis, if necessary.

### **C. Distinguishing Land Sales from Sales of Houses w/Land**

*Sales records did not identify whether or not the property sold contained a house. We used the following method to determine which sales were for properties that contained homes vs. straight land sales.*

1. Assumed sale was for land (or land with a home to be demolished and rebuilt) when
  - a. There was no year built or year remodeled identified (or the value for each was zero) AND there was no square footage listed for the parcel.
    - i. Several had no value or zero value for both yr built and yr remodeled but had square footage listed. The years of sales varied, as did the use code, though many were condos.
    - ii. Those with no value or zero values for both yr fields, but a square footage listed, are coded as a 2 and excluded from the analysis, but mentioned as such. (2 = land sale uncertain)
  - b. There was a year built, AND the sale year preceded the year built
  - c. There was no year built, but there was a year remodeled AND the sales date preceded the year remodeled
  - d. The sales record was in the sales file and had a residential use code in the parcel file, but was not in the residential file. These residential sales are not in the residential file, because they were assessed w/o consideration for residential improvements on the property, because they were insignificant or did not exist as of 12/31.
  - e. *Several sales records were in the sales file, but not in the parcel file or residential file. These were most likely land parcels later subdivided. Most were probably residential, but can't tell with certainty w/o additional data. Also, other than sales price and date, most analytical variables are unavailable for these records. Thus, though these were likely residential land sales, they were not included in the analysis—only mentioned in the general discussion.*
2. Ran crosstabs of land sales with cvt, adddt, and acreage to check for any obvious problems with the procedure used above.
  - a. Seemed generally OK. Many of the land sales are in the outlying areas.
  - b. However, a large number were also in Birmingham (this is accurate if we assume they were sales-teardowns-rebuilds).
  - c. The overall volume (as well as volume in some communities) is somewhat lower than expected (probably because of those that we think are land sales but not included in the analysis—sales not in residential or parcel file—see above).
  - d. The volume of land sales tend to be larger in the more recent years, but the last couple were pretty small.
  - e. Bottom line: Pretty confident that those identified as land sales were land sales, but it also seems like many that were uncertain or determined not to be land sales might be land sales as well.

3. Look at multiple sales for a given pin to see if the land sales end up being housing unit sales in later years. If so, could look at price before subdivision and price after subdivision for those that were subdivided and sold. Could not do this, because I can't determine the new parcel numbers with certainty, even after looking at the addt.

#### **D. Handling Multiple Sales for Same PIN in One Year**

*When there are multiple sales for the same PIN in one year all sales may not be typical market transactions. Some are simply paper sales undertaken to transfer ownership from one party to another (e.g. developer to buyer, husband to wife, divorce decrees, etc.). It is best to back these out where possible to determine true market prices.*

- 1) Ran frequencies to assess the extent of the potential problem.
  - a) 39,543 parcels had more than one sale during the period.
  - b) 32,346 of these had only two sales. 7197 had three or more.
  - c) 3447 parcels had multiple sales in the same year
  - d) 1053 parcels had multiple sales in the same year for the same dollar amount
- 2) Created an "Exclude\_Sale" variable to use to exclude the 1053 parcels (1066 sales records) that were duplicates in the same year with same dollar amount. The first occurrence of the sale is included in the analysis.

#### **E. Determining Key Variables to Use in Sales Analysis<sup>1</sup>**

1. Matched sales records to descriptive variables from residential file using the data, merge files command. For duplicate PINS matched to the first instance of the duplicate PIN. Later, I replaced the data for duplicate pins with the info for the largest residential property listed for the duplicate PINS.
2. Ran frequencies to determine useful grouping variables for the sales analysis.
  - a. res\_stylea and b. Ran frequencies to see if I could create broad categories (e.g. 1-story, 1 to 2-story, 2-story, and 3-story). The large number of parcels classified as "other" (almost 17,000) and "single-family" (almost 17,500) along with the acknowledged inconsistencies in determination of styles at the local level render these variables useless for this analysis.
  - b. res\_yearbu: 22 (less than .1 percent) are missing (at least some because there is no building). This is a useful variable. Some dates must be corrected.
  - c. res\_yearre: This is a useful variable. Some dates must be corrected.

---

<sup>1</sup> From this point forward, references to the number of records that fell into specific categories are slightly off. They reflect accurate counts before we added residential sales that were not in the original residential file provided by the county (see step A.4. above). The counts became inaccurate once we added these newly identified residential sales into our analysis.

- d. res\_roomsb: Based on conclusions for other res\_roomsf variables (see below), this is not a useful variable.
  - e. res\_roomsf: Many records have zero for this (more than just those with no year built). This is not a useful variable.
  - f. res\_rooms2: Based on conclusions for other res\_roomsf variables (see below), this is not a useful variable.
  - g. res\_plumb3 (i.e. full bath) and res\_plumb2 (i.e. half-bath): There is nothing to indicate that these variables are not useful. Use them if they add value to the analysis.
  - h. Res\_totflr: Yes, this is the square footage for the residence.
    - i. Created categories after running frequencies and percentiles.
      - 1. < 1000
      - 2. 1000-1249
      - 3. 1250-1499
      - 4. 1500-1749
      - 5. 1750-1999
      - 6. 2000-2499
      - 7. 2500-2999
      - 8. 3000 or more
    - i. par\_propcl: Useful primarily for identifying condos, lakefront property, and farm property.
    - j. cvt: Useful—determine from PIN.
3. Looked at data available in parcel attributes and tax parcel files.
- a. Most are in the residential file or not useful. Compared values in the master files with those in residential to see if one was better than the other.
    - i. cvt—pin2 in sales file (later changed to cvt\_code) has more identified than cvtcd in parcel attribute file, because parcel file does not contain some of the pins in sales file (probably those that were subdivided)
    - ii. property use/classification code—par\_propcl in sales file has more identified than usecd in parcel attributes (22 missing vs. 294 missing). Again, the missing ones in parcel attributes are those where pins don't appear in parcel file (subdivided parcels?). However, the usecd categories are fewer than the par\_propcl and no erroneous codes appear. Par\_propcl has more identified as condo or lake property than parcel attributes file. Keep both for now.
  - b. sctaxcd and schdesc: Yes. This is the school district. Could use it to run analysis by district and/or to run analysis for sales values by different school districts within the same jurisdiction (e.g. Southfield/Beverly Hills/Birmingham). Could also run an analysis for the few areas where the school district changed to see if there was a noticeable increase/decrease in values.

- i. Used the code for analysis, but pulled in the description for easy reference later.
  - ii. Some PINS in sales file are not in the parcel file and won't have a school district reference. Determined these manually.
    - 1. Mapped school districts over CVTs and parcels.
    - 2. Searched for parcels near the missing pin by dropping the last digit of the pin.
    - 3. Looked at the district these pins fell in and assigned that district to the missing pin.
  - iii. Re-coded district codes to remove sub-categories for DDAs, TIFAs, etc. within districts. Made all sub-categories the same as the broad category (e.g. Farmington DDA becomes Farmington Public Schools).
- c. acreage: Yes. This is a useful value for analysis. Probably will need to use as a sub-set along with square feet, rather than a totally separate variable.
- i. 25,987 sales have no value for acreage
    - 1. 280 are pins in sales/residential files, but not in parcel file
    - 2. 25,351 (97.6%) are for condos
    - 3. Not worth trying to resolve remaining ones. Simply exclude from the analysis by acreage and mention in narrative.
  - ii. Created groups of acreage based on common logic and analysis of frequencies and percentiles of acreage values.
    - 1. Less than one-fifth
    - 2. One-fifth to less than one-third
    - 3. One-third to less than one-half
    - 4. One-half to less than one
    - 5. One to less than one and one-half
    - 6. One and one-half to less than three
    - 7. Three to less than five
    - 8. Five to less than ten
    - 9. Ten or more
- d. adddt: This is the date when new parcels are added to land records. Most w/b (through subdivision), but some c/b old ones combined(?). S/b useful for looking at value of land after subdivision and or land vs. house sales, but can't figure out a good way to use it.
- i. Compared to sales records not in parcel file to see if any pattern of sales dates and adddt emerged for these parcels and those with similar numbers. Could not find a pattern.

- ii. Sorted by adddt and compared to other info, such as sales dates, but couldn't find anything useful.
- iii. Mapped to see if spatial patterns emerged. None obvious in relation to sales, but might use this for general descriptives on subdivisions, etc.

## **F. Handling Errors and Outliers**

*Checked values of key analytical variables (those listed above, plus price) that seem incorrect or extreme so that they can be corrected and/or removed from the analysis.*

1. Ran frequencies for all key variables in the residential sales file (after matching it to the key variables from other files using the merge files command). Where questions arose (e.g. high sales price in unexpected city or for small square footage) looked at pictures or other variables in parcel file to determine legitimacy.
  - a. Sales Price
    - i. exclude anything under \$1000 (Exclude\_Sale=1)
      1. includes dollar sales (recorded as \$1 for public purposes, but actual sales price was higher)
      2. includes parcels where the price may be wrong or under-reported or where the parcel was added as part of a sale of an adjacent property
    - ii. almost all the properties costing \$1,000,000 or more were 3000 square feet or more and located in CVTs where we would expect them to be.
      1. one was \$40,000,000—exclude from analysis
      2. two were in Southfield
        - a. one was a wooded parcel >10 acres
        - b. one was a business property on Northwestern (but in residential file)—deleted from file
    - iii. sales over \$500,000 looked OK too
  - b. all properties over 5000 square feet seemed legitimate (e.g. one for 22,184 sf in Oakland Twp for Chairman of Board for Flagstar, but it was a land sale originally)
  - c. year built and year remodeled fine
  - d. All other variables seemed OK.

## **II. Estimating Monthly Cost of Mortgage**

*Needed to estimate cost in two ways—income required to afford homes sold and number of home sales affordable to specific income levels. Below, we outline the procedure used to create these estimates.*

## A. Overriding Cost Elements

1. Elements of up-front cost
  - a. Downpayment: Use three estimates
    - i. 5% (max for most of those for whom we are concerned about)
    - ii. 10% (mid-range figure)
    - iii. 20% (high end)
  - b. Points
    - i. Will vary, can obtain average from Freddie Mac site
    - ii. Assumed zero for all (ease of analysis and consistent with affordability focus—i.e. lower income levels that have less money to pay up-front). Can increase interest rate by 1/8 of a percent for each average point (based on Freddie Mac data).
2. Elements of monthly cost
  - a. Principal
    - i. Monthly payment for total sale price less the downpayment amount
    - ii. Calculate using automatic payment calculators in Excel
  - b. Interest
    - i. Interest rate divided by 12 and multiplied by the outstanding principal balance
    - ii. Used the average monthly rate for a 30-year mortgage (for the year of sale) obtained from Freddie Mac survey
    - iii. Adjust the Freddie Mac rate to eliminate points (ease of analysis—see 1.a.ii.2. above).
  - c. Private Mortgage Insurance (PMI)
    - i. Zero for mortgages w/20% downpayment
    - ii. Standard ratio of (.5%) (obtained from mortgage broker) for all loans w/10% down
    - iii. Standard ratio of (1%) (obtained from mortgage broker) for all loans w/5% down
    - iv. Technically, should back out for years after 20% equity is achieved, but this is too complicated for this analysis. Plus, we are looking at immediate affordability; so, costs in initial years is most relevant. To figure later costs we would also probably assume some income growth, which is beyond the scope of this study.
  - d. Property Insurance Escrow—standard ratio of (4%)—obtained from mortgage broker informant.
  - e. Property Tax Escrow

*Based on taxable value, property tax rate, and homestead exemption at time of purchase.*

- i. Taxable Value
  - 1) immediately after sale s/b 50% of some fraction of the sales price
  - 2) For this analysis assume 50% of 85% of the actual sales price. Since sales price is adjusted for inflation, but income will be based on actual values for the year of sale, we used actual sales price, not the price adjusted for inflation
- ii. Property tax rate
  - 1) For CVT calculations (96-04) use the average homestead rate by CVT as calculated from the yearly millage rate spreadsheets provided by the Oakland County Equalization Office, via Ryan Runnels and Lori Norris.
  - 2) In some cases these rates differ from those found in the Apportionment of Local Tax Rates reports produced by the Equalization Office and used by the Quality of Life Subcommittee in their reports of 02 and 03 rates.
    - a) Main difference is the special assessments included for some CVTs in the millage rate files.
      - i) Only a few communities are impacted (e.g. 4 in 2003)
      - ii) Most of the differences are minor (e.g. 1 to 2 mills)
      - iii) One exception is Royal Oak Township, which had a special assessment of 30.33 in 2003 (and maybe other years?).
      - iv) Birmingham differed substantially in 2003, and it was counter to the trend for all other years. Upon investigation, we manually adjusted the value to make it consistent with the apportionment report (i.e. only included the B-ham school district).
    - b) Other minor differences exist for a few CVTs. Source of these differences is unknown, but impact is insignificant.
  - 3) For CVT calculations (95 and 04) the millage files are either unavailable or different in format.
    - a) entered the data from the apportionment reports and checked for major differences that could be attributed to the source of data used rather than actual changes in rates.
    - b) No major differences from 95 to 96; so, I used the apportionment data for 95. Similar for 03-04.
  - 4) For countywide calculations, used the average of the CVT averages
- iii. Homestead rate
  - i) We have the data
  - ii) For ease of analysis and since we were considering hypothetical purchasers, we assumed all sales are 100% homestead properties.
- f. Utilities
  - i. Could assume a range of average monthly costs (e.g. 150, 200, 250)

- ii. For ease of analysis, we used \$175, but included some different values for aggregate figures mentioned in the text to show differing impact.

### ***B. Incomes Required to Afford a Home***

1. Using the procedure for estimating cost components outlined above (except using actual tax and interest rates for each year), we calculated the monthly cost of for a \$50,000 home. We repeated this for \$1000 price increments all the way up to \$203,000.
2. We then calculated the annual income that would be necessary to ensure that this cost did not exceed 30% of income.
3. We did this for every CVT and each year from 1995 to 2004 (using interest rates and tax rates for the given year, rather than an average for the entire period).
4. We did this for the county as a whole using average tax rates.
5. We then calculated an income to price ratio for each CVT and year by dividing the home purchase price (i.e. \$50,000 to \$203,000 at \$1000 increments) by the “affordable income” calculated in step 2. We then calculated the average ratio for all price levels. This gave us a generic price-to-income affordability ratio (countywide and by CVT) that we could apply to any home sales price to determine if the price would have been considered affordable at the time of sale.
6. We repeated this procedure assume a 5% downpayment, 10% downpayment, and 20% downpayment amount.
7. We applied this ratio (and/or the average ratio from 1995 to 2004) to the means, medians, and percentiles computed above (2004 dollars and unadjusted)

### ***C. Homes affordable to specific income levels***

1. We used HUD’s income limits for a family of 4 to determine the maximum income that a family of 4 could earn and be considered extremely low-, very low-, low-, or median income in each year from 1995 to 2004.
2. We applied the price-to-income affordability ratios by CVT and year (as calculated above) to these income levels to determine the maximum price that each of these income levels could afford in each CVT and year.
3. In the sales data file, we calculated a new variable that compared the price of the house sold to the affordable price for that year and CVT (as determined in step 2). We coded it a “1”, if it was affordable (i.e. actual price was less than or equal to the affordable price).