HOUSING MARKET DYNAMICS: SOME FINANCE-RELATED APPLICATIONS
Patrick Bayer
April 4, 2011
SUMMARY OF DATASETS – Part I

Dataquick

• Coverage
  o 100+ million properties in United States
  o Some metro areas from 1988-2010 [e.g., LA, SF, BOS, SEA, PHOE]
  o Most from 1997-2010
  o Some cities/states missing [e.g., Texas]

• Data
  o Transactions
    ▪ Sales Price
    ▪ Names of Buyer and Seller
  o Assessor
    ▪ Current, Detailed House Characteristics
  o Mortgage Information – New Purchases and Refinances
    ▪ Lender, loan amount for each loan
SUMMARY OF DATASETS - Part II

MLS
- Coverage
  - Weekly snapshot of houses listed for sale
  - Several metro areas from 2006-2010 [e.g., LA, SF, BOS, SEA]
- Data
  - Asking Price (each week listed)
  - Address and House Characteristics
- Can merge with Dataquick by address
SUMMARY OF DATASETS – Part III

HMDA
• Coverage
  o Every mortgage application in US 1990-2009
• Data
  o Lender, Loan Amount, Zip code, Race and Income of Applicant
• Can merge with Dataquick by lender, loan amount, zip code

McDash
• Coverage
  o Data drawn from largest loan service centers
  o National sample of mortgages from 2005-2010
• Data
  o Detailed information at time of origination
    ▪ Structure of mortgage
    ▪ Some characteristics of borrower (e.g., FICO, income)
  o Monthly performance data – including regular checks of FICO
• Can merge with Dataquick only for geographic areas
CONSTRUCTING PRICE INDICES: REPEAT SALES ESTIMATION

- Natural heterogeneity in the housing market – every property is a unique asset.

- Because individual houses sell infrequently, the construction of any price index or return series must of necessity draw on a wider set of houses.

- Some basic notation:
  - $t$ – indexes time periods from $1, \ldots, T$.
  - $h$ – indexes complete set of $H$ houses in the metropolitan area.
  - $j$ – indexes sample of $J$ houses that sell multiple times throughout the sample period
  - $j(s)$ - a repeat sale of house $j$ in period $s$.
  - $p_{jt}$ - price of an individual transaction of house $j$ at time $t$,
  - $\pi_t$ - to indicate the estimated metropolitan price index at time $t$, and
Estimating Standard Price Indices

- The existing literature uses the sample of all repeated sales $j$ to estimate price indices, which can be turned into a return series.
- Can estimate in levels or differences

Regression in levels with house fixed effects
- Regression of log prices (in levels) on house fixed effects and time dummies:

$$\ln(p_{j(s)}) = \xi_j + \omega_s + \varepsilon_{j(s)}$$

- Because house fixed effects, $\xi_s$, are included in this regression, only houses that sell multiple times help in identifying the price index.
- With this approach, there is one observation for every transaction $j(s)$ of a house in the repeat sales sample but a whole vector of ancillary parameters to be estimated, the house fixed effects.
- Following estimation, a price index $\pi$ is formed by exponentiating the time effects, $\omega$, estimated above: $\pi_t = \exp(\omega_t)$. The estimated return between any two periods $t$ and $t+k$ is simply: $r_{t+k,t} = (\pi_{t+k} - \pi_t)/\pi_t$. 
Regression in first differences

- Imagine that each repeat sales sold exactly twice at time periods $s'$ and $s''$

\begin{equation}
\ln(p_{j(s'')}) - \ln(p_{j(s')}) = \omega_{s''} - \omega_{s'} + \varepsilon_{j(s'')} - \varepsilon_{j(s')}
\end{equation}

- Estimates of the time fixed effects $\omega_t$ using OLS in equations (1) and (2) would be identical.

- Now only one observation for every house $j$ in the repeat sales sample $\leftrightarrow$ no need to estimate house fixed effects in order to estimate the time effects.

- If some houses sell more than twice, one can also take differences between successive repeat sales of house $j$ to form observations. Properly weighted, this approach can be used to return exactly the same estimates as in equation (1).

- Following estimation, a price index and returns series can be formed in exactly the same way as in the first approach.
Case-Shiller Indices

- The construction of the S&P Case-Shiller indices is based on a variant of the estimation of equation (2).

- The estimated Case-Shiller indices are proper price indices only if one of the following two conditions hold:

  (i) the subset of houses that sell at each point in time is representative of the housing stock as a whole or

  (ii) the return process is homogeneous for all of the houses in the metropolitan area.

- Both assumptions are clearly rejected in the data.
Constructing a House-Specific Price Index

- Construct a distinct price index for each house $h$ in the full sample $H$ - including those that never sell in the sample period.

- Information on the full stock of houses is drawn from assessor files.

- $j$ continues index the set of houses that sell multiple times.

- Let $X_h$ characterize a set of observable attributes of the house including, for example, precise location, square footage, year built, and lot size.

- For each house $j$ in the sample of houses that sell multiple times in the study period, we construct a weight based on how similar it is to house $h$: $w_{hj} = w(X_h, X_j)$.

- These can be chosen in such a way to place strong weights on properties in close proximity in both geographic and characteristic space (i.e., in real-estate parlance, comparable sales).
Locally Weighted Repeat Sales Estimation

- To construct the price index for each house, we estimate a *locally-weighted* version of the repeat-sales regression shown in (1) continuing to use all houses that sell multiple times but weighting observations by $w_{hj}$.

\[
\ln(p_{js}) = \xi_j + \omega_{hs} + \varepsilon_{js}
\]

- The resulting time dummies $\omega_{ht}$ are now a function of the exact weights used and form the basis for the house-specific price index for house $h$.

- Notice that the only thing that changes in constructing this price index for different houses in a metropolitan area is the weight on each house $j$ in the repeat sales sample.

- In practice, these weights will be essentially zero for a large fraction of houses; those distant in either geographic or characteristic space.
**Constructing More Aggregate Price Series**

- With house-specific price indices and return series in hand for each house in the full sample $H$, we can now properly construct returns for any of a number of more aggregate portfolios.

- In essence, we now have exactly the kind of data that researchers working with standard financial data have.

- Aggregating over all the houses in the metropolitan area (including those that do not sell) gives rise to a metropolitan price index that properly accounts for the heterogeneity in the subset of houses that sells at each point in time.

- Price indices can also be constructed for particular types of properties (based on observable) and for particular locations within the metropolitan area.
Value-Based Portfolios

- For all houses that sell at least once during the sample, it is straightforward to turn the estimated price indices into an estimate of the value of the house at each point in time.

- Straightforward to bin houses on the basis of their estimated value and estimate the return in the subsequent period.

- One can rebalance value-based portfolios (e.g., the top quartile of the houses in the metropolitan area) at each point in time and construct price (and return) series for these portfolios in a way that is exactly analogous to stock portfolios such as the S&P 500.
Recent Job Market and Prospectus Papers

"Uncertainty, Learning and the Value of Information in the Residential Real Estate Market"
Elliot Anenberg, 2011, Federal Reserve Board, Washington D.C.

"New Evidence on the Microdynamics of Neighborhood Transition"
Marcus Casey, 2011, University of Illinois-Chicago, Economics Dept

"Risk in Housing Markets: An Equilibrium Approach"
Aurel Hizmo, 2011, New York University, Stern Finance

“Feedback Between the Mortgage and Housing Markets: A Dynamic Equilibrium Approach"
Edward Kung, 2012

“Dynamic Pricing with Network Externalities"
Timothy Schwuchow, 2013