A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Lending

Fernando Ferreira and Joseph Gyourko

The Wharton School

University of Pennsylvania

Large Literature on the Crisis

- Mostly focus on subprime sector
  - ¾ of 30+ papers published on the subject since 2008
  - Ex. Mian and Sufi (2009)
- Mostly loan level data sets
  - Loan Performance—securitized subprime (private label market)
  - Sometimes merged with credit bureau data (Equifax)
  - Ex. Mayer, Pence, and Sherlund (2009)
- Mostly with limited time periods due to subprime focus
  - Typically end no later than 2008
  - Securitized subprime market becomes widespread by 2005
- Findings of very large default rates in subprime sector
  - Explanations: loan traits, securitization incentives, etc.

Traditional Mortgage Default Literature

- Mostly focus on two factors:
  - Negative equity from falling house prices
  - Borrower illiquidity from negative income shock
  - Ex. Foote, et. al. (2010): Double trigger hypothesis

  - More correlated with economic cycle; may be independent of subprime status

Our Paper

- Economic analysis of the foreclosure crisis that integrates both strands of literature

- Can common factors explain subprime/prime differences in propensity to foreclose?
  - Housing traits, household traits, loan traits, local economic shocks, and negative equity-timing of last transaction

- Provide (new?) stylized facts about foreclosure crisis

Our Paper

- How?
  - Create large panel of ownership sequences
    - 800 million quarterly observations on these ownership sequences
  - Examine entire market over full cycle
    - Subprime, FHA/VA, Prime, Cash
    - 96 MSAs from 1993-2012
  - Estimate panel regressions with micro data
    - Deal with previously unobserved heterogeneity as best as possible

Roadmap

1)Intro

2)Creating the panel

3)Stylized facts

4)Panel estimates

5)Conclusion

Creating the Panel

- DataQuick micro data
  - Sales, initial mortgages, refis, and seconds
  - Entire market for 96 MSAs with good data since at least 1998, most data starts in 1993-1994
  - Date of purchase and transaction price, address and census tract code, names of purchaser and seller (including investors/speculators), loan amounts and lender names (3 loans), and house characteristics
  - All non-arms-length transactions included; Foreclosures clearly identified
  - Merged with HMDA to include race, gender, and self-reported income

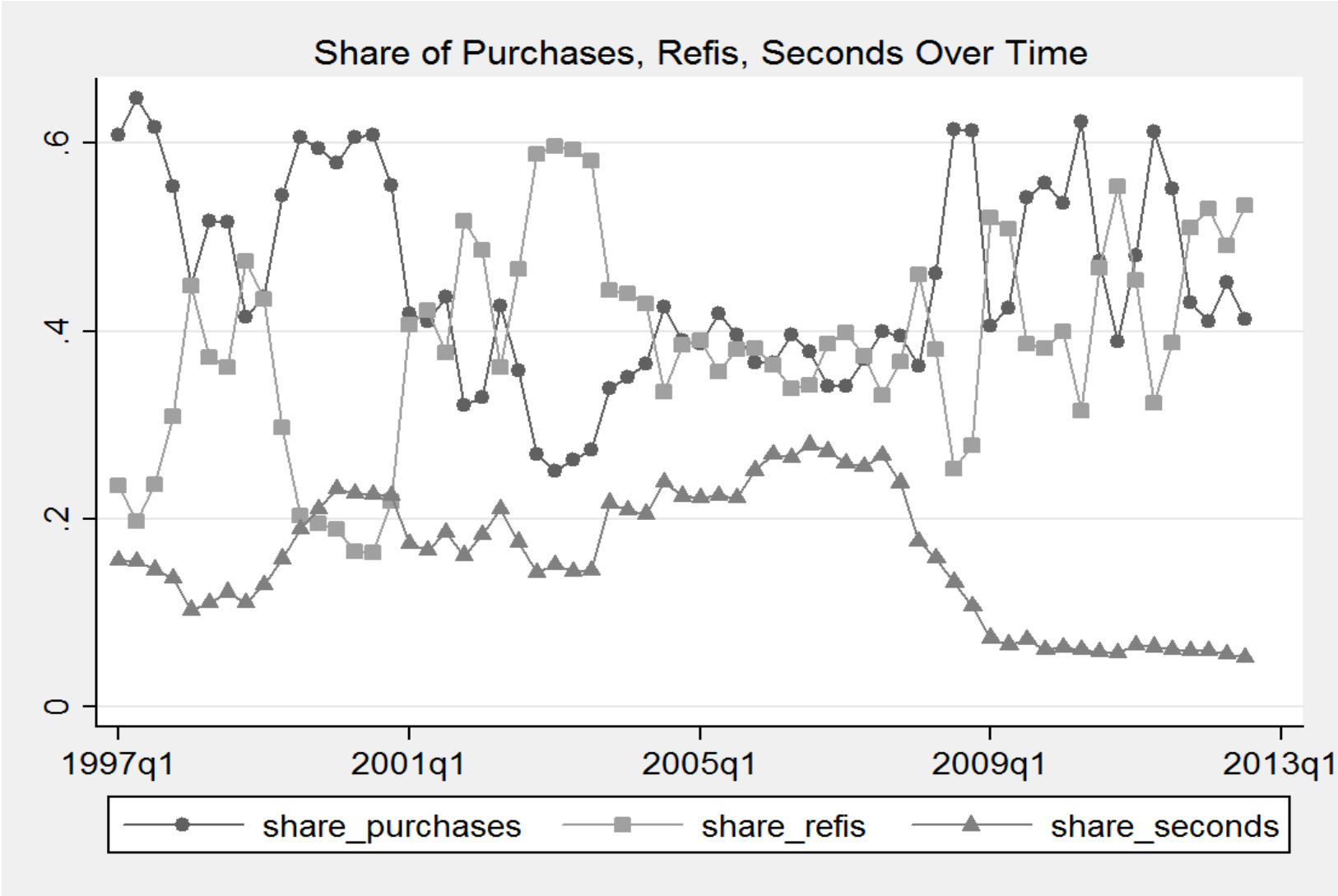
Data Representativeness

	All U.S. (1)	DataQuick (2)	Final (3)
Number of MSAs	362	269	99
Population of MSAs	642,486	809,386	1,322,485
	(1,485,668)	(1,691,640)	(2,520,843)
% East	0.21	0.22	0.24
	(.34)	(.35)	(.37)
% Midwest	0.22	0.20	0.17
	(.44)	(.41)	(.31)
% South	0.33	0.32	0.19
	(.49)	(.49)	(.44)
% West	0.24	0.26	0.40
	(.41)	(.44)	(.51)
% White	0.73	0.72	0.69
	(.14)	(.14)	(.15)
% Black	0.13	0.13	0.12
	(.11)	(.11)	(.09)
% College Degree	0.24	0.24	0.26
	(.07)	(.07)	(.07)
Median Family Incom	$53,574	$54,017	$56,252
	(9,497)	(9,564)	(10,382)
Median House Value	$149,545	$153,381	$186,629
	(60,794)	(62,683)	(75,842)

Creating the panel (cont'd.)

- Types of Transactions
  - Sales/Purchases
    - Arms-length trades between HHs (~80% of all sales/purchases)
    - Sales of new homes from builders (~11%)
    - Sales out of foreclosure (~9%)
  - Financings Subsequent to Purchase (and before sale)
    - Refinances (34 million cases) - Rule-based definition: 50% of outstanding loan amount or imputed property value
    - Junior debt (i.e., seconds; 14 million cases) - If not a refinancing, it's a second

Creating the Panel (con’d.)



Creating the Panel (cont'd.)

- Unique ownership sequences
  - This is the complete span of time a given owner owns a specific residence
  - 55.7 million ownership sequences; 32.2 million housing units
    - 2010 Census indicates 31.4 million owner-occupied units in our 96 MSAs
- Final sample
  - 42.4 million ownership sequences; 20.9 million housing units
  - Sequences dropped if we cannot impute current LTV
    - Happens if we do not observe a valid price or if house bought prior to 1993

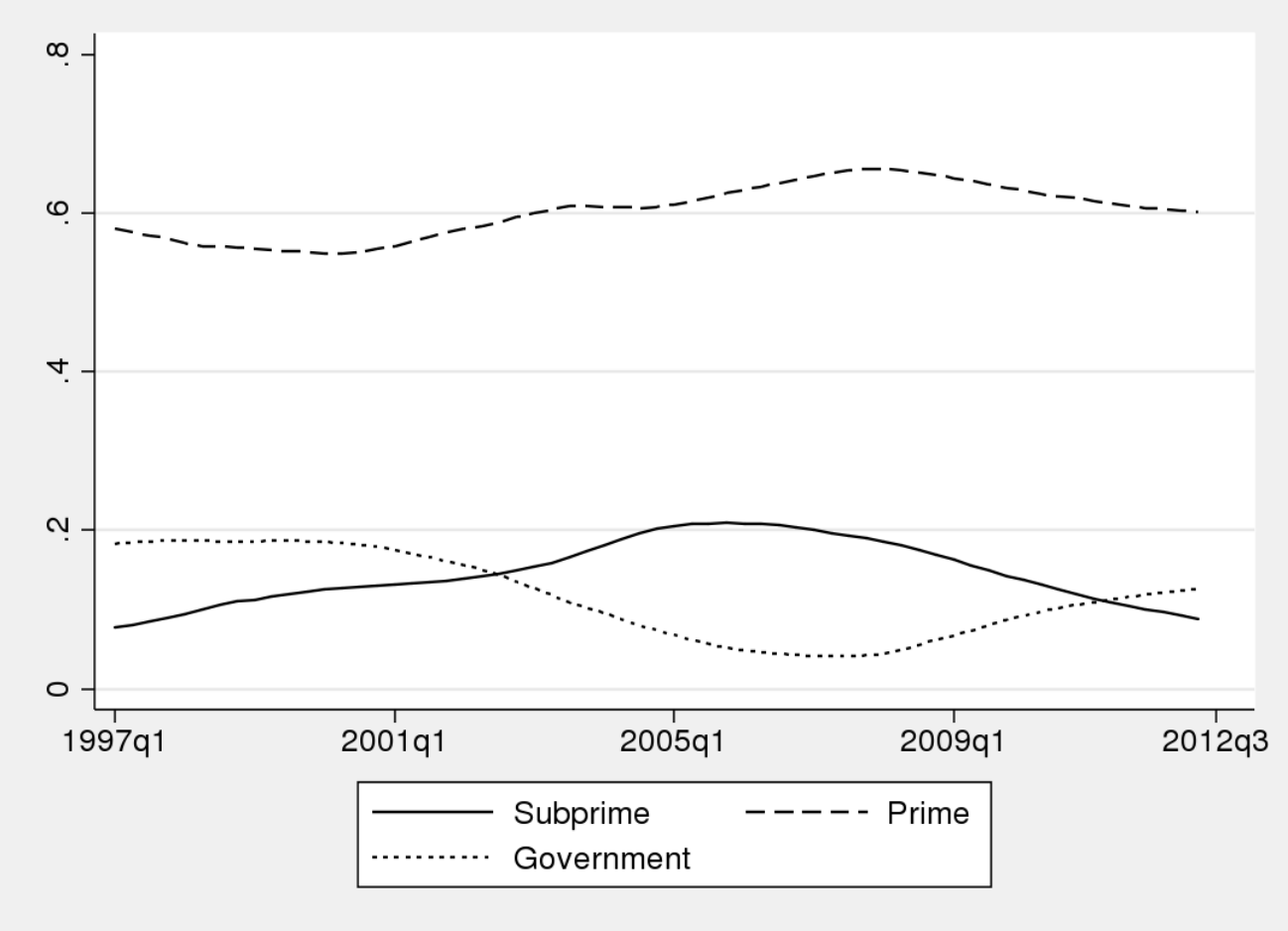
Creating the Panel (cont'd.)

- Five types of financing
  - Subprime loans (15%)
    - Lender lists (we do not have credit scores)
      - Annual HUD lists since 1997
      - Inside Mortgage Finance since 1990
    - Loan not insured by FHA or VA
  - Government loans (10%)
    - FHA/VA-insured loans
      - Separate variable in DataQuick identifying these loans
      - Sometimes directly identified in lender codes

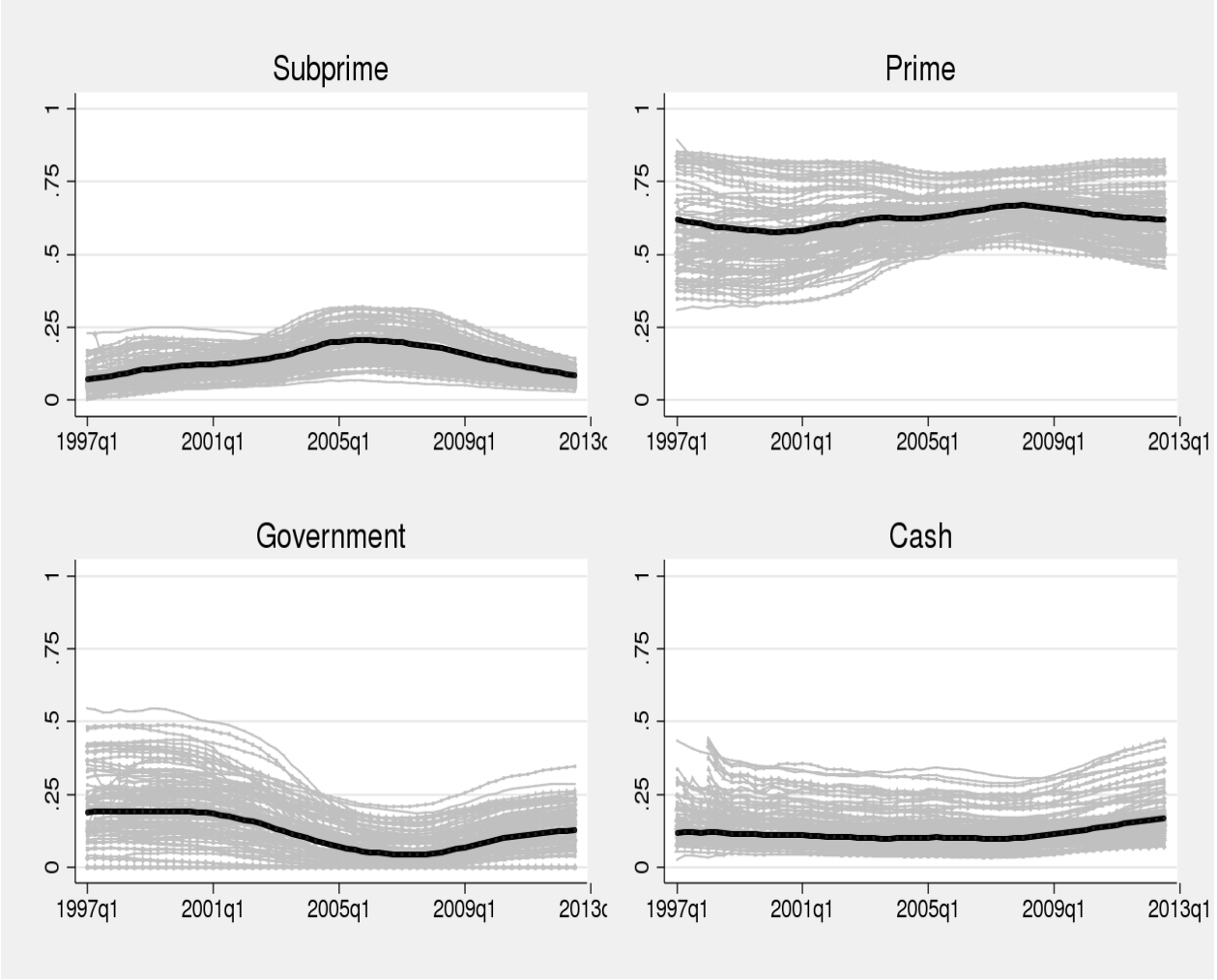
Creating the Panel (cont'd.)

- Types of financing (cont'd.)
  - Cash (11%)
    - Bought your home with no debt
  - Small lenders and typos (2%)
    - Lenders with less than 100 loans issued during complete time period
    - Lenders with personal names
  - Prime (61%)
    - If you took out debt and you are not Subprime, Gov't, or Small then you are Prime

Shares of Ownership Types Over Time



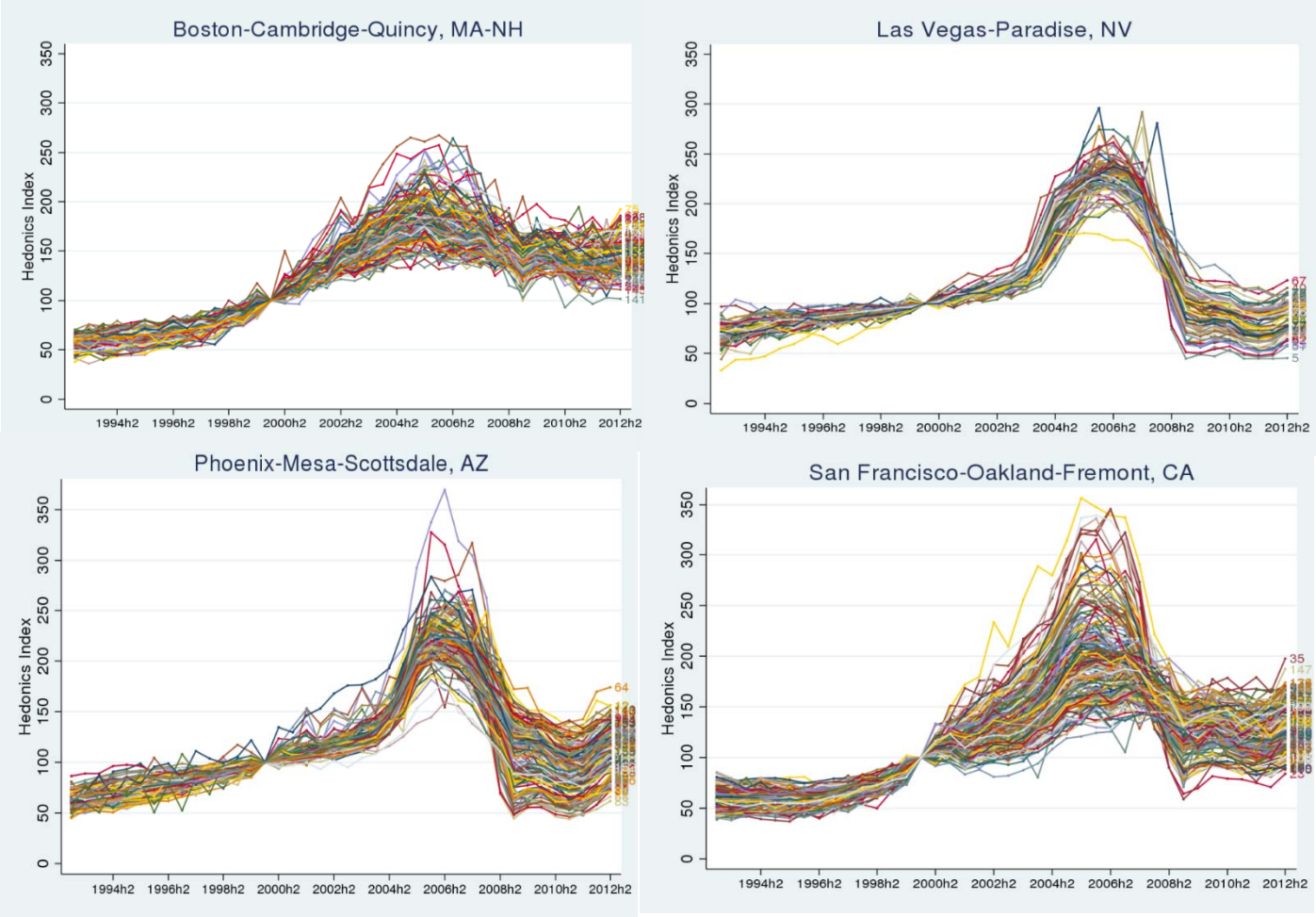
Shares of Ownership Types Over Time: Aggregate (bold line) and MSA-Level (gray lines)



Creating the Panel (cont'd.)

- Use simple hedonics to create constant quality price series
  - Able to create neighborhood-level price series
    - Groups of 4-6 census tracts
- MSA-level series very highly correlated with repeat sales indexes
- Use neighborhood-level series to impute LTVs
  - Presume all debt is 30yr, FRM product
    - Almost certainly leads to understatement of LTVs, especially for subprime borrowers

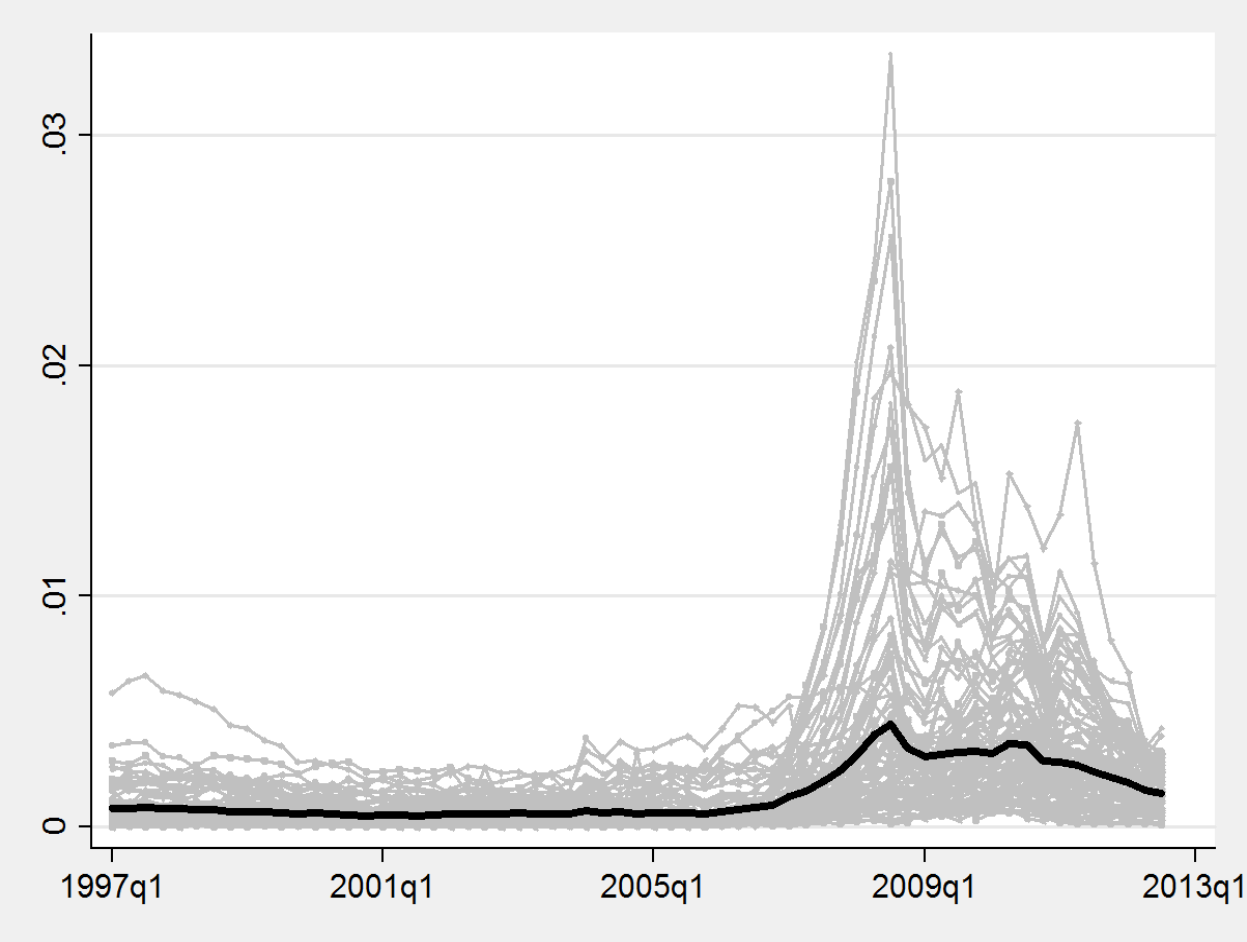
Neighborhood-Level Constant Quality Prices: Boston, Las Vegas, Phoenix and San Francisco



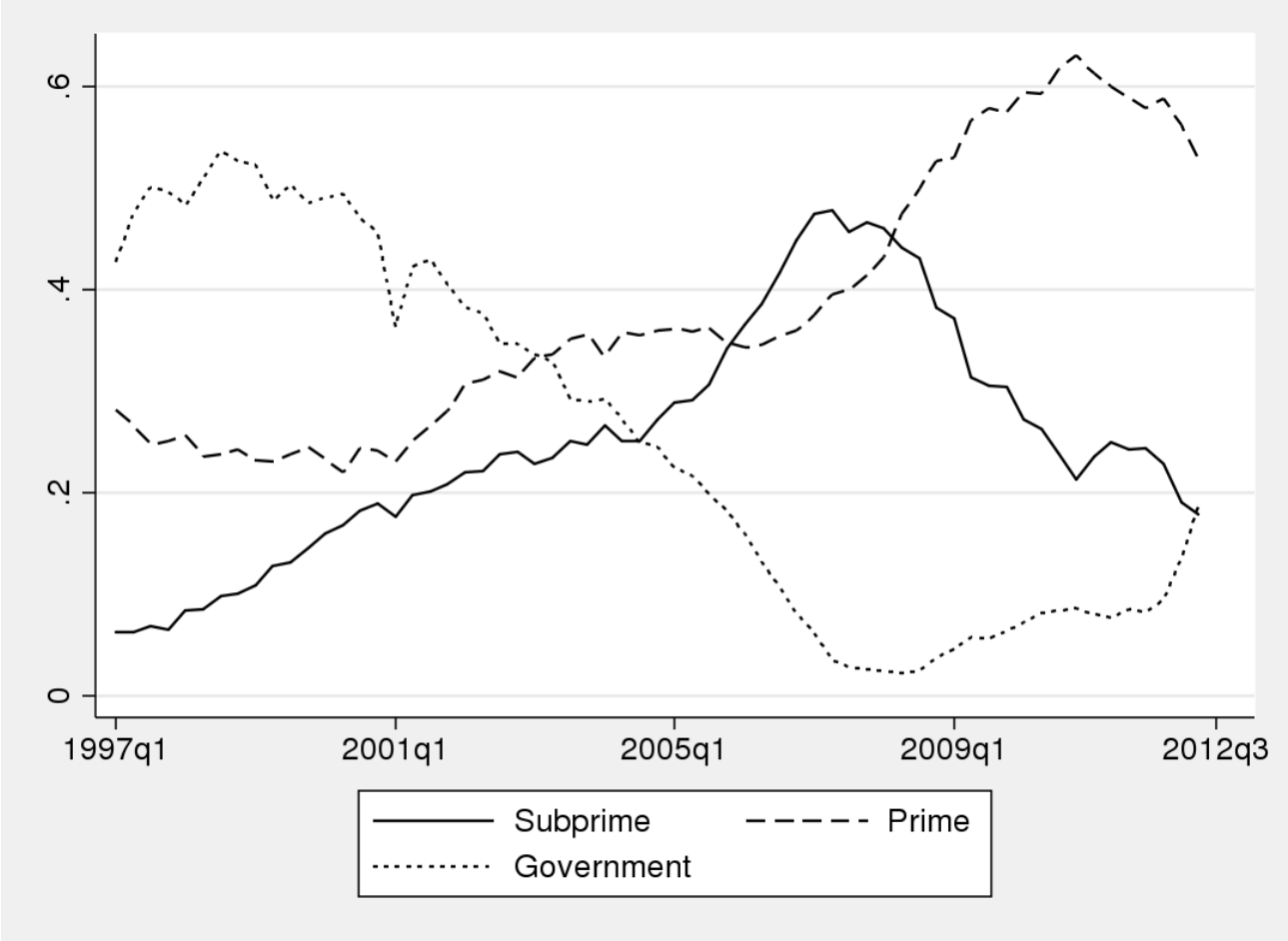
Creating the Panel (cont'd.)

- Measures of distress
  - Foreclosures clearly identified in DataQuick with special distress code (2.1 million cases, or 0.26% of all observations)
    - Can confirm this by looking at name of 'buyer'; Typically some type of financial institution (bank, RMBS pool number, special servicer, etc.)
    - Local tax authority or other local public entity for Cash; Non-payment of taxes appears to underlie these losses to foreclosure
  - Short-sales (~1 million cases, or 0.12% of all observations)
    - Inputted by DataQuick via proprietary information and model
    - We also used our own decision rule: transactions that occur at 90% current LTV or less
    - Both measures are highly correlated; prefer our measure due to better coverage

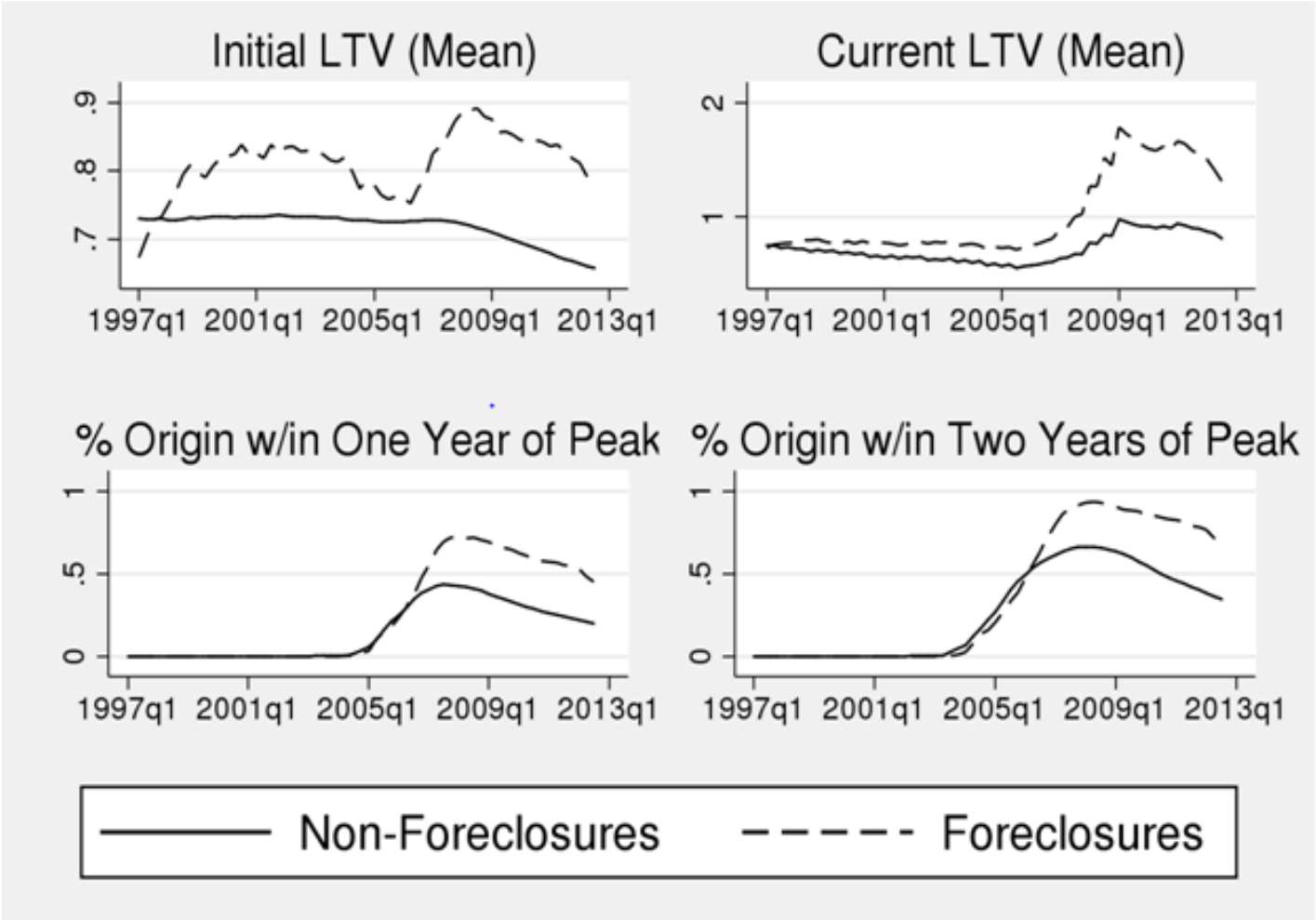
Quarterly Foreclosure Rates Over Time: Aggregate (bold line) and by MSA (gray lines) (per owner-occupied unit)



Mortgage Type Foreclosure Share Over Time



Summary Statistics by LTV and Timing of Loan Origination



Creating the Panel (last “data” slide!)

- Caveats
  - Unbalanced panel since data does not have ownerships that started prior to 1993
    - Estimate all models with data since 1997
  - No information on timing of default
    - But can observe loan that went bad
  - No household level information on employment status
    - Main limitation of data set (and literature)
    - Use household fixed effects to deal with all fixed factors, including propensity to get unemployed (more on this below)

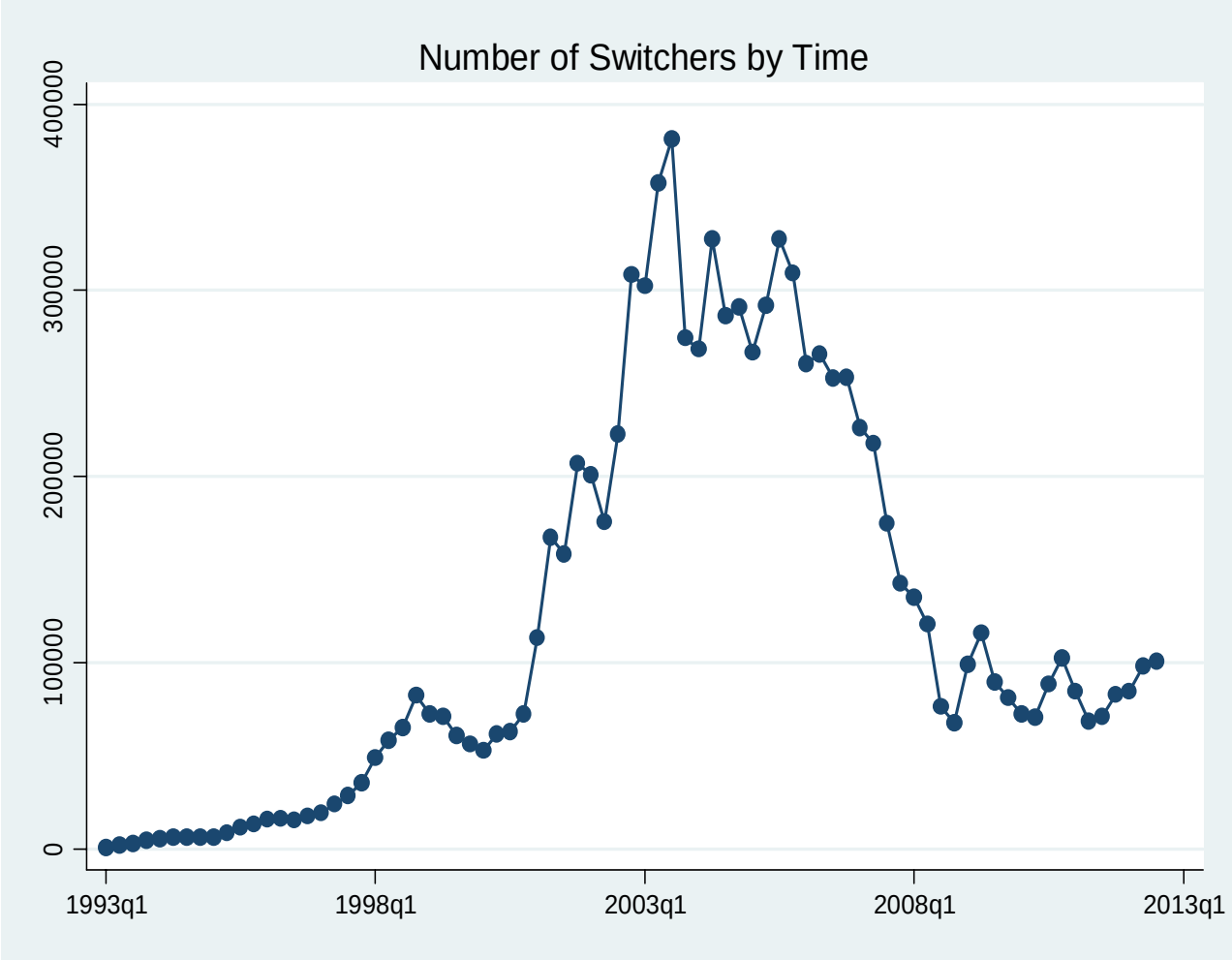
Econometric Model 1

- Panel Estimation #1: Foreclosure is a function of:
  - Type of financing (Subprime, Government, Prime, Small – Cash is the omitted category)
    - Shows unconditional differences in propensity to default
  - Then add groups of traits:
    - Housing: size, # bedrooms, # bathrooms.
    - Household: race, gender, self-reported income, speculator
    - Loan: refi, second, initial LTV, age of the loan
    - Local economic conditions: tract by quarter fixed effects
    - Negative equity and timing of last origination: current LTV and fixed effects for origination cohort
  - Compare evaluation of conditional differences

Econometric Model 2

- Leads to estimation of second specification with household fixed effects
  - Controls for permanent component of omitted factors such as wealth, employability, etc.
    - Black box - difficult to distinguishing among those factors
  - This specification is identified from variation among those who switch financing types (from prime to subprime or from subprime to prime for example) within their ownership sequence
    - Large number of switchers (~30% of all ownerships)
    - Not random sample (all refi, more likely to be minority (25% instead of 21%), less self-reported income (7% less self-reported income), more concentrated in California)

Number of Mortgage Type Switchers Over Time



Results: Average Estimates (Panel Model #1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Subprime	0.0059	0.0059	0.0038	0.0060	0.0068	0.0057	0.0027
	(0.00006)	(0.00006)	(0.00004)	(0.00008)	(0.00007)	(0.00001)	(0.00001)
Prime	0.0019	0.0019	0.0006	0.0024	0.0032	0.0027	0.0004
	(0.00005)	(0.00004)	(0.00003)	(0.00007)	(0.00005)	(0.00000)	(0.00001)
Govt	0.0023	0.0023	-0.0027	0.0024	0.0020	-0.0012	-0.0015
	(0.00004)	(0.00004)	(0.00004)	(0.00007)	(0.00004)	(0.00001)	(0.00001)
Observations	797,634,223	797,633,416	797,634,223	797,634,223	797,634,223	797,633,416	797,633,416
House Traits	No	Yes	No	No	No	Yes	Yes
Loan Traits	No	No	Yes	No	No	Yes	Yes
Household Traits	No	No	No	Yes	No	Yes	Yes
TractxQtr FE	No	No	No	No	Yes	Yes	Yes
Origination Qtr FE	No	No	No	No	No	No	Yes
Current LTV	No	No	No	No	No	No	Yes
Unexplained Foreclosure Rate Gaps							
Subprime-Prime	**0.0040**	**0.0040**	**0.0032**	**0.0036**	**0.0036**	**0.0030**	**0.0023**

Results: Household Fixed Effects (Panel Model #2)

	(1)	(2)
Subprime	0.0011	-0.0006
	(0.00003)	(0.00003)
Prime	0.0000	-0.0013
	(0.00003)	(0.00003)
Govt	-0.0020	-0.0025
	(0.00003)	(0.00003)
Observations	797,634,225	797,634,225
YrxQtr FE's	Yes	Yes
HH FE's	Yes	Yes
Current LTV	No	Yes
Unexplained Foreclosure Rate Gaps		
Subprime-Prime	**0.0011**	**0.0007**

Interpretation

- Previously unobserved heterogeneity important
  - Large origination cohort and negative equity effects
    - Common across subprime and prime borrowers
    - Surge in prime foreclosures appears due to 'bad luck'
  - Smaller but important role for loan traits
    - We are investigating importance of refi dummy
  - Very small role for observed housing and household traits
  - Fixed effects for local economic conditions have little effect as well
    - Need individual employment status
  - Household fixed effects largely eliminates subprime/prime gap
    - Subprime status is irrelevant for the propensity to foreclose of these households

Conclusions and Future Work

- Foreclosure crisis not solely one of subprime
  - More prime borrowers lost their homes—just with a lag
  - Interesting differences between our Subprime and Government groups, too
- Macroprudential regulation focused on loan traits of subprime sector
  - May not mitigate much cyclical risk
- How to prevent homeowners from buying homes with debt financing near the peak of the cycle?
- More work is needed to test borrower illiquidity assumption

Conclusions and Future Work

- Next Up:
  - Linear probability models reported in this version
    - Due to computing constraints arising from size of sample
    - Moving data and programs to AWS – maybe some hope for non-linear models
  - Heterogeneity and robustness
    - By geography and time
    - By subprime lenders
  - More on understanding switchers
    - Selection or random choice?