Wall Street and the Housing Bubble

Appendix A Data

Ing-Haw Cheng^{\dagger}, Sahil Raina^{\ddagger}, and Wei Xiong^{\$}

September 2013

[†]Tuck School of Business, Dartmouth College, Hanover, NH 03755, email: ing-haw.cheng@tuck.dartmouth.edu, http://www.dartmouth.edu/~icheng.

[‡] Ross School of Business, University of Michigan, Ann Arbor, MI 48109-1234, email: sraina@umich.edu, http://webuser.bus.umich.edu/sraina.

[§] Department of Economics and Bendheim Center for Finance, Princeton University, Princeton, NJ 08540, e-mail: wxiong@princeton.edu, http://www.princeton.edu/~wxiong.

This appendix contains more details about sampling methods and data construction. Section A1 provides details about our securitization agent sample, while Sections A2 and A3 provide details for the equity analyst sample and lawyer sample. Section A4 contains details about the computation of transaction intensities.

A1. Securitization agents

The analyses in this paper rely on contrasting characteristics and behaviors of participants at the 2006 American Securitization Forum (ASF 2006) with those of two control groups: equity analysts and lawyers. As a first step, we found a comprehensive list of the participants in ASF 2006 on the ASF website and used it to build a sample of securitization agents based off of a randomly selected list of conference participants. The participant list includes each participant's first name, last name, company for whom they work, and whether they work on the buy (investor) or sell (issuer) side. As of the present date, this list of ASF 2006 participants is no longer available on the ASF website, but we retained copies of the list of ASF 2006 meeting participants. In that meeting, there were 1,045 investors and 714 issuers, for a grand total of 1,759 participants in total. Of these 1,759, we randomly sample enough participants so that our dataset contains 400 securitization agents with matched records in Lexis/Nexis Public Records. We make sure to oversample ASF 2006 participants from prominent institutions associated with the financial crisis. For our purposes, these are AIG, Bank of America, Bear Stearns, Citigroup, Countrywide, JP Morgan Chase, Lehman Brothers, Merrill Lynch, Morgan Stanley, Washington Mutual, Wachovia, Barclays, Deutsche Bank, HSBC, UBS, Credit Suisse, and Mellon Bank.

Next, we collected data on the sampled ASF 2006 meeting participants in Lexis/Nexis Public Records. Lexis/Nexis Public Records (from herein L/N) is a service that aggregates information from various county-, state-, and national-level public records into a cohesive, searchable database. The data provided in L/N is vast, and, in general, contains many types of public records data available in the United States. The data in L/N are organized such that, upon identifying a particular person, organization, or location, the L/N user is able to find all the public records data associated with that person, organization, or location. For instance, as shown in Figure A2, if the L/N user were to successfully identify a person, the user would know the month and year of the person's birth, the first five digits of their social security number, and any locations associated with that person. For some people, additional information is available, including gender, current phone number, criminal records,

divorce filings, bankruptcy filings, past and present employers, voter registrations, and professional certifications.

In order to collect Lexis/Nexis data on the securitization sample, we first uniquely identify each person in Lexis/Nexis. As the ASF 2006 participant list provided only the person's name and employer, in order to uniquely identify each person, we first find background details about each person based on information available on the Internet. For most participants, this means searching for the person and firm on Google and using the search results to ascertain the person's location and approximate age. Our search for such biographical information was simplified tremendously by the LinkedIn profiles maintained by several of the people in our sample, often containing the year of college graduation, which we use to form initial bracketed estimates of age as a filter within L/N. Using data found via the Internet, we input the name, location, and initial bracket of age estimates of each of the sample ASF participants into the Lexis/Nexis web form for Comprehensive Person Search (see Figure A1) and uniquely identify 400 securitization agents. This process required sampling 613 ASF 2006 participants ex post in order to uniquely identify 400 securitization agents on L/N. The 205 participants not in our securitization sample are either not found at all in L/N (29), not uniquely identifiable on L/N given the information from our Internet searches (50), not involved in real estate mortgage loan securitization (94), not mid-level managers (13), or are found to be living outside the US (27).

Once we found unique L/N records for 400 securitization agents, we used L/N to collect information on their real estate-related personal transactions. We used the following public records data from Lexis/Nexis: deed transfer records, property tax assessment records, utility connection records, and mortgage records. All of these records are available to Lexis/Nexis users via a Property Report webpage (see Figure A3). For each person, we first locate all properties owned anytime between 2000 and 2010. This is done by finding all properties owned between 2000 and 2010 by the person, by someone with whom the person has owned another property (e.g., a spouse), or by a trust of which the person is a beneficiary.¹ 58 of the 400 analysts do not own any properties at all. For each of the properties owned by the remaining 342 analysts, we collect the following information: location, purchase loan type, refinance loan amount, purchase loan interest rate, refinance loan period, period, purchase loan type, sale date

¹ Living trusts generally make abundantly clear its beneficiaries in the name of the trust. For instance, if John Smith is the only beneficiary for a living trust which owns a particular property, the deed record for the property purchase will have "Smith John Living Trust" listed as the purchaser. Very often, they also put the beneficiaries name in a separate field.

(if sold), sale price (if sold), property type, building square footage, bedrooms, bathrooms, and property acreage. While not all of this information is available for all properties, L/N provides a relatively complete picture of the real and financing transactions associated with each residential property for all of our samples. For most of the properties, deed transfer records such as in Figure A4 provide the purchase date, purchase price, sale date, and sale price information for our sample members' ownership of a property. For some of the properties, tax assessment records such as in Figure A5 fill in missing purchase and sale information. The tax assessment records also provide information regarding the type of the property (single-family houses, condominiums, vacant land, commercial properties, multi-family dwellings, office buildings, planned developments, and total acreage. The mortgage records (such as in Figure A6) provide information on all financings associated with each house, including the loan amount, loan interest rate, loan period, and loan type (fixed-rate or adjustable-rate). We collect financing information for both purchase loans and refinancings.

For properties where we are missing purchase or sale date information, we proceed as follows. For purchase dates, we confirm that there is no information in either tax assessment or mortgage records that may provide purchase date information. If there is not, then we assume that the property was purchased by our person prior to the start of our data window (i.e., prior to January 1, 2000). If a property is missing a sale date, we use tax assessment records to confirm that no one has owned the property after our person. If no one has, then we assume that our person still owns the property.

After collecting all the necessary data from L/N for our securitization agents, we collect additional information on them from LinkedIn. As explained above, many of the participants from the ASF 2006 meeting maintain LinkedIn profiles. As such, we are able to develop an employment history for each analyst from 2000 to 2010. The employment history is composed of two parts. First, we create an annual indicator of whether the person had no changes in employment that year, gained new employment that year, lost employment that year, or gained and lost employment that year. We code a job loss in a year as occurring if employment was lost during that year in any way.

We use the ArcGIS geocoding software to round out our information on the securitization agents' properties. In particular, we use ArcGIS to get the census tract of each property, the distances between

any two properties owned by an agent, and a graphical representation of the geographic distribution of the properties owned by all securitization agents.

A2. Equity analysts

The first control group for our analyses consists of equity analysts. Equity analysts for our sample are chosen at random out of the universe of financial analysts covering in 2006 any companies included in the S&P 500 in 2006 that are not part of the homebuilding sector. To do this, we download from I/B/E/S the names and firms of equity analysts covering in 2006 companies included in the S&P 500 in 2006 and not in sectors with SIC codes 152, 153, and 154. The complete universe of these analysts numbers 2,978. From that, we randomly sample enough such analysts to create a sample of 400 equity analysts with data from L/N; ex post, we needed to draw 469 such equity analyst names. The 69 analysts that we drew but are not in our final sample are either living outside the United States (25), deceased (1), not identifiable uniquely in L/N (27), or not found in L/N (16). Having identified our 400 equity analysts in L/N, we proceed with the same data collection exercise that is detailed above for securitization agents, pulling information from L/N, GIS, and LinkedIn.

A3. Lawyers

The second control group for our analyses consists of lawyers. In order to construct our sample of lawyers, we match each member of the securitization sample on age and location with lawyers drawn from the Martindale-Hubbell Law Directory (herein referred to as M-H). The Martindale-Hubbell Law Directory is a directory of lawyers that has been in publication since the mid-19th century and provides biographical and professional data on every lawyer in the United States. In particular, each entry of M-H provides information such as the lawyer's name, employer, position, address of the employer, date of birth, legal fields of specialization, and the law school from which the lawyer graduated.

Our matching process is as follows. After collecting property data for our 400 securitization agents from L/N, our first step is to generate a list of potential matching lawyers for each securitization agent, matched on age and the work location of the lawyers.² To match on location, we look for lawyers with work locations in all counties associated with all properties owned by the securitization agent in 2000.³

 $^{^{2}}$ Home locations are not available in L-N. We implicitly assume that a lawyer's work location is a proxy for their home location.

³ For agents who do not own property in 2000, we use the property owned most recently after the year 2000.

For properties in CBSAs, we also include all counties associated with a property's metropolitan division (if the CBSA is divided into such divisions), or all counties associated with the CBSA if the CBSA is not divided into metropolitan divisions. We include these extra counties to generate lists with a reasonable number of potential matches; including only the specific counties of a securitization agent's properties sometimes generates lists with very few potential matches which may be empty once intersected with our age matching criterion. To match on age, we look for lawyers with an age at most five years older or younger than the securitization agent, where age is computed as of 2011. For each of the 342 securitization agents with property information, we feed M-H a custom search query that generates a list of lawyers who simultaneously match the securitization agent on both of these location and age dimensions.

Next, within each of the lists of matched lawyers, we exclude all lawyers who possess one or more real estate-related specializations (i.e., we exclude any lawyers who have one or more specializations with the words "Real Estate" in the name). Next, we remove any lawyer entries that are duplicated across lists so that we do not have any lawyers that are used as a match for more than one securitization agent.

This leaves us with 342 cleansed lists of lawyers matched to each of our securitization agents with property data. From each of these lists, we randomly choose one lawyer each, providing us with 342 lawyers in our lawyers control group. Since we do not possess location data on the 58 securitization agents without property data, we randomly choose 58 of the 342 lists of matched lawyers from which to sample an additional lawyer, which brings our lawyer sample up to the desired size of 400. This approach essentially samples matching lawyers for the 58 securitization agents who never own property from the empirical distribution of locations of securitization agents who do own property.

As M-H provides detailed information on each of the lawyers we sample, it is relatively easier to uniquely identify the lawyers in L/N. As a result, in order to attain the 400 lawyers required for our sample, we only need to draw 406 lawyers from the lawyer lists. For the 6 instances in which the initially sampled lawyer is not uniquely identified in L/N, we randomly choose another lawyer from the same list, so that the new lawyer also matches the securitization agent. The 6 instances where we could not find the initial lawyer arise from situations where we could not identify the lawyer uniquely in L/N (3). Having identified 400 lawyers in L/N, we proceed with the

data collection exercises for L/N and GIS data detailed above. We do not collect LinkedIn data for the lawyers sample.

A4. Transaction intensities

In the simplest conceptual setup where a person may only engage in one transaction per year, a basic estimate of the intensity of transaction type k occurring in year t is the number of people who conduct transaction k in year t divided by the number of people who could have conducted that transaction in that year. In this setup, the number of people eligible for each type of transactions at the beginning of the year is given in Table A1.

However, one person may engage in more than one type of transaction per year. For example, a non-homeowner at the start of year t may buy a first and second home during the year. In this case, the person was a non-homeowner at the beginning of year t and bought a second home in year t. On the one hand, this may suggest that everyone in each sample is eligible to make every type of transaction each year. However, measuring the number of people eligible each year as the whole sample implicitly assumes that each person i in the sample has an equal probability of conducting transaction k irrespective of her homeowner at the beginning of the year has a much lower probability of buying a second home during the year than a homeowner, since the non-homeowner must buy two houses. Taking the whole sample as the number of eligible people ignores valuable conditioning information about whether she is a homeowner and will mix together two distinct sets of outcomes.

A full treatment of this problem requires creating multiple new transaction types – for example, buying a second home when beginning the year as a non-homeowner, buying a first home during the year when beginning the year as a homeowner, and so on. Since these types of multiple-transaction outcomes are infrequently observed, we instead modify our framework by counting the number of "adjusted homeowners," defined as the number of homeowners at the beginning of year t plus the number of non-homeowners who bought a first home during year t. The number of people eligible to buy a second home or swap a home during year t is this adjusted homeowners group. Although this still mixes the two channels, it mitigates the issue by only including the non-homeowners who in fact buy a first home during the year.

Similarly, we create an "adjusted non-homeowners" group, which adds together people who are not homeowners at the beginning of year *t* with the number of people who divest their last property during the first six months of the year, and use this as the number of eligible people for buying a first home. Note that the number of adjusted homeowners plus the number of adjusted non-homeowners may exceed the total number of people in each sample.

We split the number of adjusted homeowners into those with two houses or more at any point during year t (or on the last day of year t-1) and label them adjusted multiple homeowners. Adjusted multiple homeowners are eligible to divest a second home. Because one may sell off houses in rapid succession, we take all adjusted homeowners as eligible to divest their last home. We summarize the adjustments for homeowners, non-homeowners, and multiple homeowners in Table A2 and the people eligible for each type of transaction after accounting for the possibilities of multiple transactions per year in Table A3.

🔴 LexisNexis® Custom Solutii ×										
← → C #	https://w3.nex	is.com/hottopic	s/Indp/						প্ল	2
Individual Com	pany Public Recor	rds Dossier	Advanced Sources	History &	Alerts Sign Out	What's New		LexisNexis® for Development Pro	fessional	's
Public Records	Comprehensive	e Person Report	•			Make this my Start Page	View Tutorial			-
Person Report Business	People •	Businesses	▼ Location ▼	All Sea	rches 🔻	Co	overage Tips			
Report Location Report	Enter informatio	Enter information in at least one field below.								
		Strict	Search (Find only exact	names ar	nd addresses)					
	Last Name				Street Address					
	First Name	E Find s	imilar-sounding last nan	nes	City					
		🔽 Find n	icknames of first name		State	All Available States	•			
	Middle Name,	/Initial			Zip Code					
	Link ID				Previous City					
	SSN			1	Previous State	All Available States	•			
	Date of Birth	1			Other Previous State	All Available States	•			
	Age Range	•	То		Talaakaaa	30 Miles				
	Alias Last Na	ime			relephone	e.g., 456-7890 or (123)	456-7890			
	Relative's Fir	rst Name]						
	Other Relativ Name	ve's First								
					Save Search	Search Clear	Form			
	Your DPPA Permissible Use: <u>No Permissible Use</u> Your GLBA Permissible Use: <u>No Permissible Use</u>									
	Important: The Public Records and commercially available data sources used in Smartlinx® Reports have errors. The data collected by government agencies or private data vendors may sometimes be entered poorly, processed incorrectly and could contain factual defects. This system should not be relied upon as definitively accurate. Before relying on any data this system supplies, it should be independently verified. For Secretary of State documents, the information contained in the Smartlinx® Reports is for information purposes only and is not an official record. Certified copies may be obtained from that individual state's Department of State.									
	Copyrights 2012 LexisNexis Terms & Conditions Contact Us Copyrights 2012 LexisNexis, a division of Reed Elsevier Inc. All rights reserved.					•				

Figure A1: Lexis/Nexis Development Professional Person Search Interface



Figure A2: Person Report

· → C f https://w3.nexis.com/hot	topics/Indp/	<u>ک</u>
ndividual Company Public Records Dossi	er 🗸 Advanced 🗸 Sources 🗸 History & Alerts 🗸 Sign Out 🗸 What's New	LexisNexis® for Development Professiona
ublic Records View: <u>Return to Results Vi</u> erson Report usiness eport <u>Search: Comprehensive Per</u> first-name	sualize Report son Report Search Results > Source Documents ast-name state(ALL) radius(30) (<u>Edit Search New Search</u>)	Create Al Create Al Select for Delive
ocation Report Top of Report Location Information (1) Address Variations (2) APN(s) (3) Property Information (2) Neighbors (5) Associated Entities (10) Sources (100) Sources (100)	NO. 1 Legal Description: Land Usage: CONDOMINIUM Tax Year: Data Source: A Property Sale Information Sale Date: Sale Price: Assessed Land Value: Assessed Total Value: No. 2 Legal Description: Tax Year: Data Source: B Property Sale Information Sale Date: Sale Price: Assessed Information Assessed Land Value: Assessed Information Sale Date: Sale Price: Assessed Information Assessed Land Value: Assessed Information Assessed Land Value: Assessed Information Contemporter Cont	
	Associated Entities - 10 records found No. Current Owner(s)/Resident(s) Current Owner(s):	Neighbor Phone Actions Get Report Get Report Get Report Get Report Get Report Get Report
	2. Current Resident(s): 3. Previous Owner(s): 4.	Get Report Get Report Get Report

Figure A3: Property Report



Figure A4: Deed Record



Figure A5: Tax Assessment Record

Figure A6: Mortgage Record



Transaction Type	Eligible People That Year
Buy a first home during the year	Non-homeowners at beginning of year t
Buy a second home during the year	Homeowners at beginning of year t
Swap a home (up, down or missing) during the year	Homeowners at beginning of year t
Divest any home during the year	Homeowners at beginning of year t
Divest a second home during the year	Homeowners with multiple homes at beginning
	of year t

Table A1: Eligible People for Different Transactions

Table A2:	Adjustments	to Different	Groups
-----------	-------------	--------------	--------

Group	Definition
Adjusted homeowners at beginning of year t	Homeowners at beginning of year t plus non-
	homeowners who buy a first home during year t
Adjusted non-homeowners at beginning of year t	Non-homeowners at beginning of year <i>t</i> plus those
	who divest their last property in the first six months
	of year t
Adjusted multiple homeowners at beginning of	Adjusted homeowners at beginning of year t who
year t	have more than two houses at any point during the
	year t

Table A3: Eligible People for Each Type of Transactions after Adjustments

Transaction Type	Eligible People That Year
Buy a first home during the year	Adjusted non-homeowners at beginning of year t
Buy a second home during the year	Adjusted homeowners at beginning of year t
Swap a home (up, down or missing) during the year	Adjusted homeowners at beginning of year t
Divest any home during the year, including the last	Adjusted homeowners at beginning of year t
Divest a second home during the year	Adjusted homeowners with multiple homes at
	beginning of year t