

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Essays in Economics

Permalink

<https://escholarship.org/uc/item/5xb5q481>

Author

Romem, Israel Hadas

Publication Date

2013

Peer reviewed|Thesis/dissertation

Essays in Economics

By

Israel Hadas Romem

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Economics
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Enrico Moretti, Chair

Professor David Card

Professor Robert Helsley

Professor Shachar Kariv

Spring 2013

Abstract

Essays in Economics

by

Israel Hadas Romem

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Enrico Moretti, Chair

Chapter 1: Using a novel combination of administrative and proprietary data from 2007 to 2011 on King County, WA (metro Seattle), I estimate the effect of owner-occupants' home equity on their probability of sale and, indirectly, mobility. I exploit plausibly exogenous variation that follows only from changes in local housing price indices, and I account for confounding economic conditions that vary by time and location. The estimates indicate that sales decline dramatically over the combined loan-to-value ratio range from approximately 70% to 100%, well before homeowners reach negative equity levels.

Chapter 2: A known shortcoming of Burdett and Mortensen's celebrated result of pure equilibrium wage dispersion is that it hinges upon the assumption of random matching. This chapter presents a modified version of their model, in which pure equilibrium wage dispersion arises under a form of balanced matching. Rather than modeling workers and firms, the model addresses workers and jobs. Matching is assumed to be random with respect to jobs, which amounts to balanced matching with respect to firms, when firm size is measured in terms of jobs. The bizarre implication of random matching with firms, whereby splitting a firm in two increases its recruitment rate, is eliminated. In addition to increasing employment, speeding up recruitment and reducing worker turnover, higher wages in the modified model also reduce the rate of job vacancy.

Chapter 3: I model the relationship between a central government and its tax-collecting proxies. The Old Kingdom was the first great age of Ancient Egypt, witnessing the construction of such wonders as the pyramids of Giza and the Sphinx. Its downfall was the culmination of a long process, whereby the balance of power gradually shifted from the royal court to an emerging provincial elite. I maintain that the process was born out of state policy: Incapable of perceiving the long-run dynamics governing the equilibrium division of resources between itself and its local proxies, the royal court set the economy on a divergent path that empowered the provincial elite at the state's expense. The underlying strategic interaction between the royal court and its proxies was present throughout the Old Kingdom period. Careful attention is paid to the question of which decisions should be modeled by optimizing behavior and which should not. The paper fills a gap in the literature by addressing why, in the short run, the royal court could not decrease the flow of state resources to the provincial elite.

To my beloved parents and siblings who I miss so much,
to my wonderful little Eitan,
and to my darling Amy.

Contents

1	The Role of Equity in the Housing Market: Empirical Evidence from 2007-2011	1
1.1	Introduction	1
1.2	The home buyer's liquidity constraint	5
1.3	Challenges to identification and solutions	9
1.3.1	Earlier approaches	9
1.3.2	Challenges to identification	9
1.3.3	Solutions	12
1.3.4	Limitations	15
1.4	Data	16
1.4.1	Data sources and construction	16
1.4.2	Sample selection	18
1.4.3	Housing prices in King County	20
1.5	Empirical estimates	20
1.5.1	Unconditional observation	20
1.5.2	Naive and reduced form estimates	26
1.5.3	Instrumental variable estimates	31
1.5.4	Sensitivity of sales to housing price innovations	35
1.5.5	Loss Aversion	36
1.6	Conclusion	38
2	Re-Interpreting Burdett and Mortensen: Equilibrium Wage Dispersion with Balanced Match- ing	43
2.1	Introduction	43
2.2	Model	46
2.2.1	Setup:	46
2.2.2	Preliminaries:	48
2.2.3	Steady-State Job Occupancy:	49
2.2.4	Equilibrium Wage Dispersion:	51
2.3	Conclusion	53

3	The Unraveling of Prolonged Stability: The Fall of the Old Kingdom in Ancient Egypt	54
3.1	Introduction	54
3.2	Stylized Facts	57
3.3	Model	61
3.4	Analysis	66
3.5	What Is Optimized And What Is Not:	70
3.6	Conclusion	72
	References	77
A	Appendix to Chapter 1	78
A.1	Data	78
	A.1.1 Measuring equity	78
	A.1.2 The Zillow housing price index	79
	A.1.3 Inferring owner-occupancy status	79
A.2	Dynamic selection	80
A.3	The “better homes” hypothesis	81
B	Appendix to Chapter 3	87
B.1	A risk-neutral proxy as a simplifying assumption	87
B.2	A myopic proxy as a simplifying assumption	87

Acknowledgments

I thank my advisors, David Card, Bob Helsley and Enrico Moretti for their advice, as well as Barry Eichengreen who provided caring mentorship over the years and Yves Zenou who introduced me to search theory was particularly helpful with respect to Chapter 2. Mark Borgschulte, Carola Conces, Brad DeLong, Jan DeVries, Fernando Ferreira, Willa Friedman, Francois Gerard, Joe Gyourko, Shachar Kariv, Patrick Kline, Victor Lavy, Atif Mian, Roy Mill, Carl Nadler, Valentina Paredes-Haz, David Romer, Jesse Rothstein, Itay Saporta-Eksten, Michel Serafinelli, Edson Severnini, Ity Shurtz, Yannay Spitzer and seminar participants at UC Berkeley, Fannie Mae, the Federal Housing Finance Agency, Freddie Mac and at the 2012 EconCon conference at Princeton University provided many valuable comments and insights. I am grateful to Stan Humphries and Randy Newton for their help with data from Zillow.com and from the Washington State Voter Registry, and I am particularly grateful to Katie Dobbyn for her help with CoreLogic data. Writing chapter 1 would not have been possible without these data. The CoreLogic Academic Research Council (CLARC) generously provided data for chapter 1, although the views expressed throughout the dissertation are my own, as are any remaining errors.

Chapter 1

The Role of Equity in the Housing Market: Empirical Evidence from 2007-2011

1.1 Introduction

As of mid 2012, fallen home prices left roughly 30% of mortgaged homeowners in the US with negative equity in their homes, while owner-occupant mobility was 30% below its level in 2005.¹ In theory, these observations are linked by the fact that falling home prices erode equity, preventing some homeowners from moving if selling their homes no longer enables them to pay off a mortgage or make a down payment on a new home. Yet, the empirical relationship between equity and mobility and its causal nature remain elusive: Whereas Ferreira et al. (2010, 2012) and earlier literature preceding the recent bust find that negative equity reduces mobility, recent studies such as Schulhofer-Wohl (2011) and Coulson and Grieco (2013) report the opposite.

Both theory and policy hang in the balance. In a seminal contribution to theory, Stein (1995) argues that the liquidity constraint imposed by diminished equity explains a fundamental feature of housing markets, whereby prices and trading volumes are positively correlated.² This feature, and Stein's explanation for it, form a basic tenet of our understanding of housing markets and their cycles. More recently, the possibility that diminished equity is adversely affecting the labor market by dampening mobility has drawn a great deal of attention and scrutiny, and the notion that resolving the housing crisis is inextricably linked with the relaxation of equity-driven constraints to mobility looms large.^{3,4}

¹Zillow.com estimates 30.9% of US homeowners with mortgages had negative equity in the second quarter of 2012. Aggregate data from the CPS and the ACS indicate the annual rate of owner-occupant mobility fell by 33.1% and 29.4%, respectively, from 2005 to 2011.

²Positive price-volume correlation in housing markets is documented for the US in Stein (1995), Berkovec and Goodman (1996) and Ortalo-Magne and Rady (2006), the UK in Andrew and Meen (2003), Ortalo-Magne and Rady (2004) and Benito (2006), Sweden in Hort (2000) and Hong Kong in Leung et al. (2002).

³The role of labor mobility in adjusting to economic shocks has long been recognized (Blanchard and Katz, 1992). Recent studies proposing this theme and questioning it include Sterk (2010), Aaronson and Davis (2011), Donovan and Schnure (2011), Estevao and Tsounta (2011), Karahan and Rhee (2011), Kothari et al. (2012), Modestino and Dennett (2012), Nenov (2012), Sahin et al. (2012) and Valletta (2013).

⁴If higher equity returns repeat buyers to the housing market, and if the historic correlation between

Resolving the debate concerning the empirical relationship between equity and mobility is essentially a matter of rigorous identification. Using a novel combination of recent administrative and proprietary data on homes in King County, Washington (metro Seattle), I estimate the effect of owner-occupant equity on the probability of sale and, indirectly, on mobility. I use changes in local housing price indices to isolate the component of equity that is most plausibly exogenous. Equity is jointly determined by owners' financial decisions, which are fraught with endogeneity concerns, and by changes in home value. Setting the down payment amount, for example, is a financial decision that influences equity and is likely to be endogenous. Larger down payments offer savings and a bargaining advantage, but they require liquidity and so they are likely to reflect a more prosperous home buyer and lower mobility.⁵ Changes in home value, on the other hand, consist of aggregate housing price changes and idiosyncratic property-level changes. As shown in Case et al. (1997) and confirmed in this study, properties that sell more frequently tend to experience idiosyncratic changes in home value that are more accentuated, reflecting that some homes are (unobservably) "better" than others and therefore exhibit less "churn" and more stable demand over time.⁶ This suggests that properties which lost more value in the recent bust and whose owners therefore have less equity are also likelier to sell at any moment, implying that idiosyncratic changes in home value, too, are endogenous. Aggregate changes in housing prices, on the other hand, are driven to a large extent by factors far removed from the individual homeowner and often even from the local housing market. To harness these changes I construct an instrument for equity based purely on housing price indices that I refer to as *predicted* equity, but it too presents identification challenges.

A key identification challenge stems from economic conditions that vary by time and location, and which influence both home prices and sales. Consider for example a recession or an economically distressed area in which job security is low. Insecurity can deter commitment to long term housing debt, reducing both the price and the quantity of homes sold. But even though fewer home sales coincide with lower prices and diminished equity, it would be wrong to attribute reduced sales to diminished equity. Time- and location-varying economic conditions can shift home prices and sales in a confounding manner through other channels as well, such as household formation and separation rates (Farnham et al. (2011)), the leverage cycle (Geanakoplos (2010)) and prevailing expectations with respect to future housing price trajectories.⁷

I address this challenge by conditioning estimates on the time and location of observation,

price and volume also reflects causality in the opposite direction, *from the volume of traders to prices* (e.g. because of more frequent multiple bidder situations), then there is scope for policy to launch home prices on an upward spiral.

⁵Additional borrowing against a home once it has appreciated is another financial decision that influences equity and is likely to be endogenous. Such borrowing is only worthwhile provided the owner intends to stay in the home beyond a certain time horizon, suggesting that when such borrowing takes place, lower mobility could be a cause of low equity rather than a consequence.

⁶This finding has known implications in terms of sample selection bias for repeat sales housing price indices.

⁷Note that any factor channeling contemporary local economic conditions into housing demand contributes to the positive price-volume correlation in housing markets.

using quarter by zip code area fixed effects. Doing so corresponds to a thought experiment in which I compare the outcomes of similar homes observed at the same time and in the same place - and which are therefore subject to the same economic conditions - but which have experienced different changes in local housing prices only because they were bought at different times.⁸ Consider for example the homes at 311 and 324 NW 48th street in Seattle's popular Fremont neighborhood. They are within 150 feet of each other, have the same number of bedrooms and bathrooms, nearly identical square footage, and they were built in 1940 and 1949, respectively.⁹ When I compare them at later dates, say in late 2011, their probability of sale is subject to the same economic conditions - namely, those prevailing in Seattle's Fremont neighborhood in late 2011. Commonly held housing price expectations for the neighborhood at that time, for example, or the income, job security or family situation of typical Fremont homebuyers and sellers around that time all affect the sale probabilities of these two homes similarly. The crucial difference between the homes is in the timing of their last purchase. 311 NW 48th street was purchased in late 2009, but the home on 324 NW 48th street was purchased in mid 2006, when prices in the neighborhood were substantially higher and were still rising. By late 2011 the zip code area housing price index was 7% below its late 2009 level and 15% below its level in mid 2006, implying that the owner of 324 NW 48th street was likely to have lost twice as much equity as his neighbor due to price changes.¹⁰

However this comparison raises the concern that homes, owners and loans may differ systematically in ways that affect mobility depending on the time of purchase. The typical home buyer profile, for example, may vary over time reflecting changing lending standards (e.g. "subprime" lending), and homeowners buying when housing prices are expected to fall may be selected upon having a low propensity to sell in the short term. I account for this possibility by controlling for a flexible function of the time of purchase. The corresponding thought experiment then becomes comparing the sales of similar homes observed at the same time and in the same place, which differ in their experienced housing price changes only because they were bought at different times, while accounting for the average sale rate of homes bought in any given time period as they are realized in *all* periods of observation.

My data combine county assessor records with proprietary mortgage data from CoreLogic for the years 2007 to 2011. The merged dataset contains sale and equity histories, and rich home and loan characteristics, which I augment with owner characteristics from voter registration records and from Home Mortgage Disclosure Act (HMDA) files. The final dataset spans roughly 107,000 owner-occupied properties with appropriate attributes.¹¹ I

⁸The similarity of homes refers to conditioning on observables, which include a flexible function of homeowners' tenure duration and numerous owner, home and loan characteristics.

⁹All details given here are publicly available from the King County assessor's office. For privacy reasons I do not specify owner or loan characteristics in this illustrative example.

¹⁰Note that experienced changes in the local housing price *index* are orthogonal by construction to any difference between owners, homes and loans observed in the same quarter and zip code area, except through the time of purchase.

¹¹I limit properties to those whose owners have tenure duration of 1 to 10 years. The total number of quarterly observations is just under 1.5 million.

construct predicted equity - the instrument for equity based purely on housing price indices - using zip code area housing price indices from Zillow.com.

I find that the probability of sale falls substantially at lower equity levels. Reduced form estimates using predicted equity indicate that the probability of arm’s length sale falls over the range of combined loan-to-value (CLTV) ratios from 70% to 100%, and estimates using predicted equity as an instrument for actual equity place the magnitude of this effect at roughly 2.1 percentage points.^{12,13} Given that the average unconditional rate of arm’s length sales for mortgaged homes observed with CLTV ratios up to 70% is only 3.07%, this estimate implies a dramatic reduction in mobility. In contrast, naive OLS estimates using actual equity indicate an implausibly non-monotonic effect, and in particular an interval of high equity in which greater equity appears to reduce mobility, raising the concern that these estimates capture systematic unobservable differences between homeowners.

Whereas the public debate tends to dwell on the share of homeowners with negative equity, the estimates reveal that mobility is reduced well before homeowners reach negative equity levels (over 100% CLTV). Thus, the share of homeowners with negative equity substantially understates the share whose mobility is impaired by insufficient equity.¹⁴ This finding also indicates the range of equity - 70% to 100% CLTV - in which policies manipulating homeowner equity are most likely to effectively shift sales and mobility.

In directly related earlier work, Henley (1998) observes 3,500 British households from 1992 to 1994 and finds that higher equity mildly lowers mobility, but that negative equity reduces it sharply. Chan (2001) observes 5,800 adjustable rate mortgages in New York’s tri-state area from 1989 to 1994 and also finds that shifting from 70% to 100% CLTV reduces mobility sharply, by roughly 80%. Observing 2,400 US households from 1979 to 1996, Engelhardt (2003) argues that loss aversion - not insufficient equity - is responsible for reducing mobility when home prices fall. More recently Ferreira et al. (2010) observe 20,000 US households over the years 1985 to 2007 and estimate that negative equity reduces mobility by roughly one third.¹⁵ I make three contributions with respect to this literature, of which the first two involve identification. First, earlier work shares the identifying assumption that conditional on controls, equity levels are uncorrelated with omitted variables that influence mobility. Given that actual equity is likely to be endogenous, whereas predicted equity captures only

¹²Unless stated otherwise, “sales” refer arm’s length sales throughout this paper. An arm’s length sale is one in which the involved parties are independent of each other and enter the agreement freely. Sales between relatives or between a firm and its subsidiary, for example, are not conducted at arm’s length. In the absence of a sharper definition classifying sales as being at arm’s length is ultimately a matter of discretion. I adopt a conservative approach, as detailed in appendix section ??.

¹³Adhering to the norm of using lending industry terminology, I measure equity in terms of CLTV ratios. An $x\%$ CLTV ratio is identical to $(100 - x)\%$ equity, so 70% CLTV means 30% equity and CLTV ratios above 100% imply negative equity. The word “combined” in CLTV refers to the summing of *all* loans secured against a home, and is meaningful given that from 2005 to 2010 over 25% of housing units in the US - and 40% of properties in my sample - served as collateral for two or more loans (ACS estimate).

¹⁴The same Zillow.com report citing that 30.9% of mortgaged US homeowners had negative equity in their homes in the second quarter of 2012 also indicates that approximately half of mortgaged US homeowners have CLTV ratios above 80%.

¹⁵Ferreira et al. (2011) update this estimate with data from 2009.

a plausibly exogenous component of it, my version of the identifying assumption is likelier to hold and yield estimates whose causal interpretation is valid. Second, by conducting the analysis within quarter by zip code area cells I control for time- and location-varying economic conditions that pose a key challenge to identification. I am able to do so by constructing data that are an order of magnitude larger than those used in earlier studies. The third contribution of this paper is in bringing the literature up to date with estimates from the current housing crisis.

This paper also relates to the influential study of the Boston condo market in the early 1990's by Genesove and Mayer (2001) and to a more recent study of the San Francisco Bay Area housing market by Anenberg (2011), who observe and attribute higher asking prices and lower sale probabilities primarily to owners' loss aversion. A crucial distinction between our papers is that Genesove and Mayer (2001) and Anenberg (2011) observe homeowners conditional on having listed their homes for sale, whereas I observe them unconditionally (in this respect), and examine an outcome that incorporates the decision to list a home for sale. This distinction implies that there is no contradiction in our findings. Rather, the contrast between our findings suggests it is possible that even though loss aversion plays a more decisive role once the sale process is underway, equity plays an important role in owners' earlier decision to list their homes for sale.¹⁶ Like Genesove and Mayer (2001), I contribute to the literature studying the causal relationships underlying the positive price-volume correlation in housing markets, and in particular causality running from prices to trade volumes. Other papers disentangling the web of causal relationships underlying this correlation include Leung et al. (2002), Clayton et al. (2010) and Anenberg (2012).

The paper is organized as follows: section 1.2 briefly formalizes the liquidity constraint underlying the role of equity, section 1.3 describes the empirical strategy and section 1.4 describes the data. Section 1.5 reports empirical estimates and section 1.6 concludes.

1.2 The home buyer's liquidity constraint

How might falling home prices and subsequent diminished equity affect homeowner mobility? An example adapted from Stein (1995) illustrates the idea.

Example: consider a family whose home was initially worth \$100,000, has an outstanding mortgage of \$85,000 and no other assets, and suppose the family wants to move. If housing prices have fallen so that the home is worth less than \$85,000, the family's equity becomes negative and it cannot even afford to pay off the mortgage. Under the circumstances, the family cannot move at all in the traditional way of selling one home and buying another.¹⁷

¹⁶The decision to list a home for sale cannot be analyzed separately with either of our data sets, because the data in Genesove and Mayer (2001) and Anenberg (2011) consist of sale listings, so are conditioned on the decision to sell, and the sale outcome in my data reflects the decision to sell only in conjunction with ultimate success of the sale process.

¹⁷Foreclosure, leaving behind a vacancy or engaging in two-way rental may remain viable options depending on the family's income flow, though these options preclude owner-occupancy.

Notice, however, that contrary to the common focus on “underwater” homeowners the family’s mobility might be affected even without having negative equity. At a value of \$90,000 the family can pay off the mortgage and keep \$5,000 of the proceeds. Supposing a minimum down payment requirement of 10% this amount is sufficient for a \$50,000 home, but not for one that is equivalent to the old home (that would require a \$9,000 down payment). Whether the family chooses to make such a move depends on the trade off between the want - or need - to move and the reduction in housing service consumption implied by moving into a less valuable home.

Of course, there is also a potential upside to the family’s leverage. Provided it has the necessary future income, an increase in value to \$110,000 allows the family to pay off the mortgage and keep \$25,000 of the sale proceeds - enough for a \$250,000 home. Upgrades such as this are referred to as trade-up buying.

Formalizing the liquidity constraint: Suppose a *first-time home buyer* with savings S has no other assets at his disposal. If he is unable to take out a mortgage, he can afford a home H , where H is the measure of housing services obtained from the home and P is their unit price, provided that his savings satisfy $S \geq PH$. A mortgage allows the buyer to borrow funds with which to buy a home in exchange for a guarantee to repay the funds in the future and to make a down payment of at least γPH immediately, with $\gamma \in [0, 1]$.¹⁸ With the option of a mortgage in hand and provided sufficient future income, the buyer can afford any home H such that his savings exceed the minimum required down payment,

$$S \geq \gamma PH. \tag{1.1}$$

Re-arranging, this condition implies that the housing services provided by his new home, which I refer to as its “size”, is limited by the leveraged buying power of his savings:¹⁹

$$H \leq \frac{S}{\gamma P}. \tag{1.2}$$

This is the first-time buyer’s liquidity constraint. Clearly, the buyer can afford less housing if its price increases.

Now suppose the buyer is a *repeat buyer* who already owns a home, H_0 , and has outstanding mortgage debt $M \geq 0$ on his old home. Contemporary mortgages almost always include a due-on-sale clause, obligating the borrower to repay the outstanding balance of the mortgage in the event of sale.²⁰ Thus, provided sufficient future income, the buyer can

¹⁸Of course, mortgages also involve interest, amortization and the use of the home as collateral, not to mention reams of paperwork.

¹⁹The term *leveraged* buying power refers to the fact that every dollar of savings permits buying $1/\gamma$ dollars of housing with the help of a mortgage.

²⁰This clause became prevalent in the late 1970’s in response to rising interest rates. These prompted lenders to phase out the previously ubiquitous *assumable* mortgage, which allowed buyers to retain previous owners’ lower rates.

afford any home H such that

$$PH_0 + S \geq M + \gamma PH, \quad (1.3)$$

meaning that he can only afford a home if his savings and sale proceeds are enough to pay off his old mortgage and make the minimum down payment on that home. Expressing this constraint as

$$H \leq \gamma^{-1} \left(H_0 + \frac{S - M}{P} \right). \quad (1.4)$$

reveals that the “size” of the new home is limited by the leveraged buying power of the sum of the sale proceeds and the buyer’s net financial assets. This is the repeat-buyer’s liquidity constraint. Whether rising home prices relax or tighten the constraint now depends on the sign of the buyers’ net financial assets, $S - M$. Crucially, if his mortgage debt outweighs his savings, i.e. $S - M < 0$, then *rising home prices relax the liquidity constraint*.^{21,22} This situation is typical of most repeat buyers, who tend to be within years - not decades - of purchasing their old home.

Note that if the mortgage debt exceeds the sum of sale proceeds and savings, i.e. $M > PH_0 + S$, then the buyer cannot afford to pay off the old mortgage, so he cannot afford any positive amount of housing and is unable to move in the traditional manner. Ignoring the role of savings, this situation is referred to as being “underwater” and amounts to having negative equity.

Suggestive stylized facts: An increase in house prices reduces the amount of housing that first time home buyers can afford, but it increases the amount that repeat buyers whose mortgage debt exceeds their savings can afford.²³ If enough repeat buyers have mortgage debt in excess of their savings, we should expect to see the share of first-time buyers decrease and the share of repeat buyers increase when housing prices rise, and the opposite when they fall. Figure 1.1 shows that the share of repeat buyers in the US increased until 2006 and then decreased, in tandem with housing prices.

Figure 1.2 goes further and shows that *within* the set of repeat buyers, the share reporting savings and the share reporting home equity as sources of funding increased and decreased, respectively, as housing prices fell from 2006 to 2010, suggesting that buyers whose wealth was in the form of financial assets rather than home equity were increasingly the ones capable of moving.²⁴

²¹An implicit assumption is that all homes’ prices change in proportion to their “size”, H . Exceptions to the above rule may occur if this assumption does not hold. In this case, (1.4) becomes $\frac{H_1}{P_1} \leq \frac{H_0 + S - M}{P_0}$, where P_0 and P_1 are the prices of the old and new homes, H_0 and H_1 , respectively. A straightforward example would be a differential increase in P_0 and P_1 such that the liquidity constraint tightens.

²²Note that non-mortgage debt - omitted from the above formulation above for simplicity - is really no different than mortgage debt in this context, and can be included in M .

²³Repeat buyers whose savings exceed their mortgage debt are affected similarly to first time buyers.

²⁴Savings are here defined broadly to include stocks and other financial assets. Equity refers primarily to the proceeds of selling a home, but also, e.g., home equity loans on additional homes owned.



Figure 1.1: The share of repeat home buyers. Whereas first-time home buyers can afford more housing when home prices are lower, repeat buyers whose mortgage exceeds their savings can afford more (less) housing when home prices are higher (lower). The housing price shown is the real average sale price of all US home sales (in 2012 dollars).

Source: National Association of Realtors and US Census.

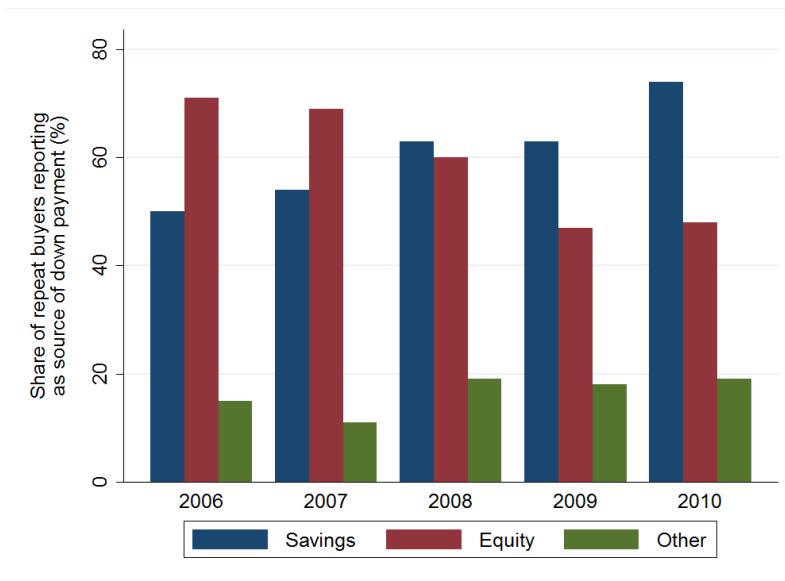


Figure 1.2: Sources of down payment funding reported by repeat buyers. As housing prices declined from their peak, a decreasing (increasing) share of funding came from housing equity (savings).

Source: National Association of Realtors.

1.3 Challenges to identification and solutions

1.3.1 Earlier approaches

To gauge the effect of owner-occupants' equity on mobility, earlier studies estimate the following generic form,

$$\Pr(\text{move}_{it}|\text{duration}_{it}) = f(\text{CLTV}_{it}, \mathbf{X}_{it}, \epsilon_{it}), \quad (1.5)$$

where move_{it} is an indicator for some measure of mobility from property i at time t , duration_{it} refers to the owner's tenure duration at the time, \mathbf{X}_{it} is a vector of covariates and ϵ_{it} is a remainder term. The measure of mobility move_{it} , the model reflected by $f(\cdot)$ and the form in which CLTV enters it take on different forms in each study, as summarized in Table 1.1. The covariates observed in each study differ depending on the data used and are too numerous to list here, but they broadly tend to consist of owners' demographic and socio-economic characteristics, observable financial traits and environmental variables such as contemporary local unemployment rates.²⁵

Table 1.1: Earlier studies of the effect of equity on mobility

Study	Mobility measure	Model	Functional form of CLTV
Ferreira et al. (2010, 2011)	Sale	Probit, LPM	Equity ≥ 0
Engelhardt (2003)	Direct	LPM, PH	Initial CLTV in 80-90%,..., >95%
Chan (2001)	Mortgage prepayment	PH	Cltv in 40-50%,..., >95%
Henley (1998)	Direct	MPH	Piecewise linear in equity ≥ 0

Notes: LPM refers to a linear probability model, PH to a proportional hazard model, and MPH to a mixed proportional hazard model that accounts for unobserved heterogeneity.

The concern in estimating (1.5) is that CLTV_{it} might be correlated with the remainder term, ϵ_{it} . Such correlation may be present for several reasons.

1.3.2 Challenges to identification

The challenge posed by household-level determinants of equity: A property's CLTV ratio is fully determined by four factors: the initial down payment, payments made towards principal (as per the amortization schedule), adjustments of the combined loan amount and changes in the value of the property.²⁶ Each of these factors provides scope for correlation.

A larger down payment requires more liquidity, indicating greater wealth that may correlate with characteristics that independently influence homeowners' likelihood of moving,

²⁵Some studies include year and MSA fixed effects, but not the cross of the two or more local or frequent fixed effects.

²⁶Adjustments of the combined loan amount include, for example, pre-payment, refinancing with cash-in or cash-out, or borrowing through any of the multitude of additional types of loans secured against the home.

like their stage in the life cycle or the location of relevant employment opportunities if, say, their wealth is derived from a highly specialized occupation. The concern remains even conditional on observable demographic and socio-economic characteristics: a home buyer who is more prosperous in the narrow financial sense is likelier to also be prosperous in the broad sense of being satisfied and secure in his or her present life circumstances. This type of broad prosperity is practically unobservable and is likely to correspond with lower mobility and greater willingness to make a larger down payment (and in so doing to reap financial and bidding advantages). Similarly, the size of amortized payments towards principal increases with shorter mortgage terms and larger payments may indicate greater wealth, suggesting a re-iteration of the previous line of reasoning. Also along these lines, delinquent payments may indicate a lack of wealth.

Adjustments of the combined loan amount may also be problematic. Refinancing, prepayment and additional borrowing against a home all involve substantial transaction costs. Thus, even if the circumstances otherwise justify an adjustment - for instance if refinancing is appealing because interest rates have declined - an adjustment is only worthwhile for a homeowner who plans to stay put for a long enough time. This means that observing a homeowner make or not make adjustments provides a signal that their propensity to move, at least in the short run, is respectively lower or higher than it would be otherwise, for reasons that are independent of the subsequent level of equity. If adjustments to mortgages disproportionately involve additional borrowing which updates CLTV ratios upwards, then the effect of higher CLTV ratios on the probability of sale is biased downwards.

Changes in property value, too, may be endogenous, especially inasmuch as individual property values are considered. When an owner-occupant renovates, for example, the owner's propensity to move is likely to fall while the property value rises. Changes in *aggregate* local housing prices such as those measured by local housing price indices are not subject to concerns like this one, but even variation in equity that owes only to changes in aggregate local housing prices presents identification challenges.

The key challenge posed by confounding economic conditions: A key challenge to identification is that local housing prices and home sales are co-determined by factors that essentially reflect the business cycle. Because rising home prices raise equity levels, these factors make higher equity coincide with higher rates of home sale, regardless of any causal effect of equity on sales. In other words the business cycle, and differences in its manifestation across locations, confound estimates of the effect of equity on home sales.

To be more concrete, consider the following examples. Job security tends to vary over time owing to the business cycle, and may be more or less pronounced in different locations depending on the local composition of industry and on the typical demographic and socio-economic characteristics of residents, e.g. a blue-collar neighborhood versus a white-collar one. If job security facilitates commitment to long-term housing debt then it increases housing demand (and supply, inasmuch as home buyers are also sellers), influencing both the price and quantity of homes sold irrespective of equity. Voluntary job turnover, which has been shown by Akerlof et al. (1988) to be pro-cyclical, can have the same confounding

effect if with some probability every job switch involves relocation or a housing upgrade. So can household formation and separation rates, which vary systematically along the business cycle, too.²⁷ The leverage cycle (Geanakoplos, 2010) can also render equity endogenous. When housing prices rise (fall) lenders may become less (more) wary of the risk of default and are likely to provide credit at higher (lower) leverage levels, thereby influencing the price and quantity of home sales in a way that is correlated with equity, but not caused by it.²⁸ Finally, housing price expectations are likely to vary given the past - and especially the recent - trajectory of housing prices, feeding into the demand for housing, and hence into housing prices and sale volumes.²⁹

The challenge posed by the cohort of purchase: Although they are more subtle, exploiting variation in equity derived only from changes in aggregate local housing prices entails additional identification challenges as well. Conditional on the time and place of observation, changes in local housing prices since a home was purchased could reflect systematic differences in unobservables between homeowners who bought at different times, and who have chosen whether or not to sell every period since. An important example concerns *housing price expectations*. When housing prices are expected to fall, potential buyers with intentions of selling in the short-run are less likely to buy. Thus, the pool of buyers when housing price expectations are dire is likely to be characterized, on average, by a lower propensity to sell than the pool of buyers in more optimistic times. To the extent that homeowners' current predicted equity levels reflect their cohort of purchase, such average tendencies of cohorts may confound the effect of equity on home sales.

The challenge posed by dynamic selection: A closely related identification challenge is dynamic selection, which is related to homeowners' repeated sale decision each period from the time of purchase to the time of observation. A stylized example can help clarify matters. Suppose that the pool of homeowners consists of two types, starter-home folk who have a high propensity to move, and dream-home folk who have a low one. Every period, the share of starter-home folk in each cohort dwindles, as starter-home folk sell and move (possibly re-entering the analysis in a different cohort). Inasmuch as this process is homogeneous across cohorts, it can be accounted for by including a polynomial of tenure duration as a control. However, if the decision to sell is influenced by other factors, such as home equity, then the "weeding out" process of starter-home folk occurs at a faster pace in cohorts experiencing higher equity, and is not adequately accounted for simply by controlling for a polynomial of tenure duration.

²⁷Note that in addition to divorce, household separation includes young adults' decision to leave their parents' homes, too. With respect to divorce, Farnham et al. (2011) show that the way in which it is affected by the business cycle is anything but clear, a-priori.

²⁸Central banks' management of interest rates in response to the business cycle can play a similar role, too.

²⁹Case et al. (2012) provide valuable empirical information on housing price expectations.

The challenge posed by loss aversion: A remaining concern stems from the fact that falling home values reduce equity mechanically, so experiencing a net loss in individual property value often coincides with low equity. Genesove and Mayer (2001) show that sellers experiencing a loss are less likely to succeed in selling their homes, suggesting that falling home values reduce sale rates by invoking loss aversion, rather than through diminished equity. Fortunately this concern has testable implications. If it is only loss aversion that reduces home sales when prices fall, then a sample consisting only of homeowners experiencing *gains* should not exhibit any correlation between equity and home sales.

1.3.3 Solutions

To gauge the effect of owner-occupants’ equity on mobility, I adapt the generic form in (1.5) and estimate linear probability models of the following type:

$$sale_{it} = \sum_j \beta_j \mathbf{1}\{CLTV_{it} \in \mathcal{B}_j\} + \mathbf{X}_{it}\delta + \theta_{lt} + \epsilon_{it}, \quad (1.6)$$

where $sale_{it}$ is an indicator for the arm’s length sale of property i in quarter t , the function $\mathbf{1}\{CLTV_{it} \in \mathcal{B}_j\}$ is an indicator that the owner’s CLTV ratio falls in the interval (“bin”) $\mathcal{B}_j \in \mathbb{R}$ belonging to some set $\{\mathcal{B}_j | j = 1, 2, 3, \dots\} \in \mathbb{R}^J$, the vector \mathbf{X}_{it} consists of property-level covariates, θ_{lt} is a saturated set of quarter by zip code area fixed effects where $l \equiv l(i)$ is property i ’s zip code area, and ϵ_{it} is a remainder term. The covariate vector \mathbf{X}_{it} includes the observed owner, home and loan characteristics listed in Table 1.2.

In what follows, I explain how this framework addresses the various identification challenges one by one.

Addressing the challenge posed by household-level determinants of equity: Recall that a property’s CLTV ratio is fully determined by four factors: the initial down payment, payments made towards principal, adjustments of the combined loan amount and changes in the value of the property, which may be specific to an individual home or aggregate in nature. To isolate variation in equity that comes only from aggregate changes in property values, I construct an instrument for actual CLTV using only changes in local housing prices indices, which I refer to as *predicted CLTV*, as follows.

$$CL\tilde{T}V_{it} \equiv \frac{0.8 \cdot hpi_{lc}}{hpi_{lt}}, \quad (1.7)$$

where hpi_{lt} is the value of the local housing price index for zip code area $l \equiv l(i)$ in the quarter of observation, t , and hpi_{lc} is the value of the same local housing price index in the quarter of property i ’s previous purchase, or cohort, $c \equiv c(i, t)$. The intuition behind predicted CLTV is that it mimics the CLTV ratio that would result for property i at time t if its value identically tracked that of the local housing price index, and if it had been purchased at time c with a down payment of 20%. Figure 1.3 is a scatterplot comparing predicted and actual CLTV ratios. Predicted and actual CLTV ratios are positively correlated, but they

Table 1.2: Owner, home and loan characteristics included as controls

Owner	Home	Loan
Tenure duration ³	Year built ³	Initial CLTV ratio at purchase ³
Cohort of purchase ³	Square footage ³	Mortgage term
Age ³	Bedroom count ²	Adjustable rate mortgage (ARM)
Average race and ethnicity of local buyers in cohort	Bathroom count ²	Going interest rate
Local median real income ³	Real purchase price ³	
Local average household size		
Local % college graduates		
Local % high-school graduates		

Notes: Superscripts $(\cdot)^2$ and $(\cdot)^3$ indicate the inclusion of second or third order polynomials, respectively. Tenure duration is measured in quarters. Cohort refers to the quarter in which the current owner purchased the home. The age variable measures the age of the oldest actively registered voter living in a property. The race and ethnicity variables are the mean shares of successful mortgage applicants in a property’s census tract and its owner’s year of purchase who are asian, black, hispanic and white. Local median real income, average household size and percent of college and high school graduates are observed at the census block-group level from the 2000 Census. Real purchase price is obtained by deflating the actual purchase price with respect to the Zillow.com housing price index for King County, WA, with the aim of obtaining an indication of value that is comparable over time. The mortgage term is captured by indicators for 10, 15, 30 and 40 year mortgages. The going interest rate refers to the average Freddie Mac rate for the appropriate term fixed rate mortgage, 30 days prior to origination (obtained by Corelogic).

are far from identical. Whereas actual CLTV is contaminated by endogenous, household-level determinants of equity such as homeowners’ financial decisions or renovation decisions, predicted CLTV is free of these influences.

Addressing the key challenge posed by confounding economic conditions: To address the key identification challenge I condition estimates on the time and location of observation. The quarter by zip code area fixed effects, θ_{it} in (1.6), effectively transform the analysis into one that is within time by location cells.³⁰ The thought experiment that parallels this strategy is comparing the sale of homes observed at the same time and in the same place, but whose owners nevertheless differ in their equity levels. Note that while

³⁰Conditioning on this set of fixed effects requires sufficient observations in each time by location cell. With insufficient observations, one might condition only additively on time fixed effects and location fixed effects, or even just on a set time fixed effects. While doing so is far better than ignoring the issue, it fails to sweep away all of the endogenous variation. The reason is that location fixed effects account only for time-invariant differences between locations, and time fixed effects account only for location-invariant differences over time (i.e. a common time trend). Realistically, however, variation is neither time-invariant across locations, nor location-invariant over time. Consider for example the housing markets in California’s hard-hit Central Valley and in far less impacted San Francisco: controlling for a common time trend and for time-invariant differences between these two locations can hardly account for the differential impact of the housing crisis on both prices and home sale volumes in the two locations.

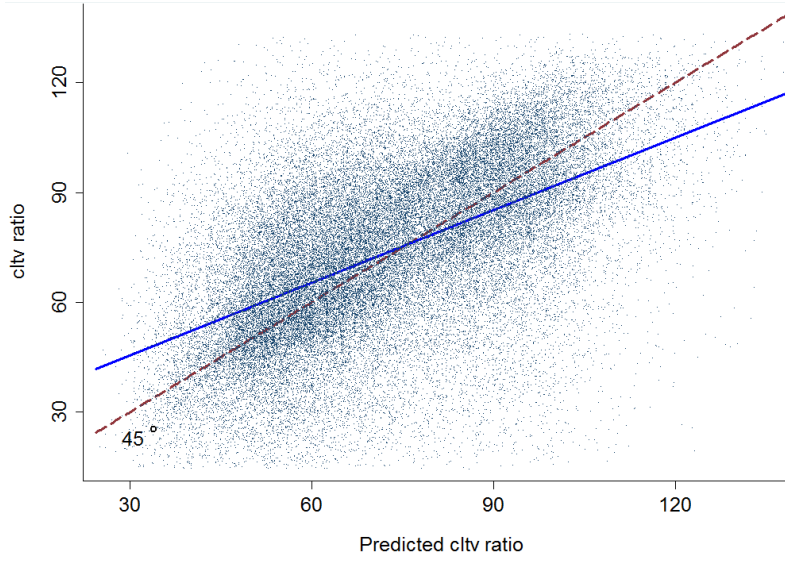


Figure 1.3: Predicted vs. actual CLTV ratios. By construction, predicted CLTV reflects only the component of actual CLTV driven by changes in local housing price indices. Actual CLTV also reflects changes in individual property values, as well as homeowners’ financial decisions such as their initial down payment and changes in their combined loan amount (e.g. due to further borrowing against equity), and amortized payments towards principal. The linear fit shown is $actual\ CLTV = 39.17(0.92) + 0.45(0.01) \times predicted\ CLTV$, and has an R^2 of 0.30 (standard errors clustered by zip code area).

actual CLTV may differ within time by location cells because of any of the household-level determinants of equity, predicted CLTV can only vary within these cells due to homes’ time of purchase.

Addressing the challenge posed by the cohort of purchase: To account for the possibility that homeowners’ current CLTV levels reflect omitted variables correlated with their cohort of purchase, I include in the vector \mathbf{X}_{it} a third-order polynomial of the cohort of purchase. This control accounts for selection into purchase cohorts due to housing price expectations at the time of purchase, as well as any other unobservable owner, home or loan characteristics that vary systematically by cohort of purchase.

Addressing the challenge posed by dynamic selection: The extent to which dynamic selection is a concern is not a-priori clear. I address dynamic selection in an appendix section by gauging its extent and direction.

Addressing the challenge posed by loss aversion: I address the possible role of loss aversion by estimating the effect of equity on home sales in a sample limited only to home-

owners who have experienced a net gain in their home value. While effects observed in the full sample may reflect the influence of loss aversion, any effect observed in this “gain-only” sample cannot be driven by loss aversion.³¹

1.3.4 Limitations

The empirical strategy detailed above is not without limitations, as follows.

Endogenous savings: It is possible that homeowners’ savings constitute an endogenous response to their level of equity. If so then savings, which affect homeowners’ ability to sell and move, may vary systematically with equity and confound estimates. Because I cannot observe savings, I cannot control for this possibility directly. However, assuming that homeowners’ saving behavior is compensating in nature, meaning that they save more when their equity is low and vice versa, then ignoring homeowners’ endogenous savings would attenuate any observed effect of equity to zero. Thus, any non-zero estimates can be considered to be a lower bound (in absolute value) of the true effect of equity on sales.

Wealth effects: The strategy does not distinguish between different mechanisms through which equity affects sales. One such mechanism is the effect of changes in equity on homeowners’ liquidity constraints, as detailed in section 1.2. Another involves wealth effects that accompanies changes in equity. Analyzing such wealth effects is not trivial (see e.g. Grossman and Laroque (1990)), and is beyond the scope of this paper. Nevertheless, the ultimate finding of this study whereby the effect of equity on home sales occurs within the confined equity range of 70%-100% CLTV, suggests that liquidity constraints rather than wealth effects are the more important mechanism.

General equilibrium effects of equity: The empirical strategy detailed in above estimates only a “partial equilibrium” effect of equity on sales, in the sense that the effect of an individual homeowner’s equity is estimated on his or her sale probability, *all else equal*. The effect of changes in the equity levels of a broader group of homeowners, e.g. higher equity across the board raising all homeowners’ likelihood of mobility, constitute part of the contemporary local economic conditions soaked up by the time \times location fixed effects. Estimating these broader, “general equilibrium” effects of equity requires a different framework than the one presented here, and is beyond the scope of this paper as well.

³¹Note that estimates obtained from a sample limited only to homeowners who have experienced a net *loss* may still be subject to the influence of loss aversion. Suppose for example that loss averse homeowners set their asking price equal to the amount they paid for their home. In this case, homeowners with greater net losses set prices farther above their homes’ market value which reduces their probability of sale more sharply, and this reduction in sales correlates with lower (or more negative) equity, simply because the falling home values that generate the loss also reduce equity.

1.4 Data

This study uses data from King County, Washington, which is the core county of the Seattle metropolitan area. Section 1.4.1 describes the data sources and construction, and provides summary statistics. Section 1.4.3 provides information on the evolution of housing prices in King County and on the Seattle metro area’s relative standing in terms of diminished equity.

1.4.1 Data sources and construction

This study combines data from several sources: the core data are from administrative records kept by the King County assessor and from a matched proprietary dataset on housing loans provided by CoreLogic. In addition to these two main sources, further information is obtained from the Home Mortgage Disclosure Act’s (HMDA) loan application registry, from Washington State’s Voter Registration Database (VRDB), from the US Census. Local housing price indices at the zip code area level are obtained from Zillow.com.

The King County assessor maintains numerous files. The key file used in this study is the real property sales file, which documents the date, dollar amount and identities of the parties involved in every real estate transaction in the county. I shape the file into a panel, such that properties are observed repeatedly over time at a quarterly frequency, and each quarter every property’s sale or lack thereof is indicated. Many sales - almost 42% - are attributed reasons such as placement in a trust, foreclosure, divorce settlement or gift transfer. I focus on the remaining 58% of sales for which a sale reason is not provided, and consider them to have been conducted at arm’s length.³² Additional county assessor files provide information on property characteristics such as their address, year of construction, square footage and bedroom and bathroom count.

The CoreLogic data provide information on housing loans in the county, and they allow me to observe properties’ estimated CLTV ratios. The data consist of a sequence of quarterly extracts from CoreLogic’s database, each of which is a snapshot of the open liens on properties in King County on the last day of the quarter. Approximately 75% of observations in the county assessor data are successfully matched with the corresponding CoreLogic records.³³ Every quarterly observation in the data is associated with up to four loans, and among the characteristics observed for each loan are its dollar amount, origination date, term and interest rate type (fixed or adjustable).³⁴ The CoreLogic data also contain estimates of the value of every property each quarter. The estimates are generated by an automatic valuation model (AVM) whose details are proprietary, but which is essentially a sophisticated hedonic regression model whose predictive capability has been honed over time.³⁵ With quarterly

³²A summary list of sale reasons and their frequency is given in appendix Table 1.3.

³³The unmatched observations in the county assessor data reflect either properties that are not associated with a loan, or properties with address discrepancies that - despite some effort - preclude successful matching. Properties without associated loans are likely to comprise the bulk of unmatched observations.

³⁴Only a fraction of a percent of properties are observed to have more than two loans secured against them at any one time, let alone more than four.

³⁵In fact, the estimates are generated by a cascade of several automated valuation models, where the term

Table 1.3: Frequency of Sale by Reason

	Frequency	Percent
None, i.e. arm's length sale	147,868	58.18
Other	23,888	9.40
Placement in trust	14,992	5.90
Foreclosure	14,442	5.68
Establishment of community property	12,109	4.76
Property settlement	10,871	4.28
Estate settlement	10,019	3.94
Partial interest (love, affiliation, gift)	7,216	2.84
Divorce settlement	7,084	2.79
Tenancy partition	3,024	1.19
Correction (refiling)	1,904	0.75
Will-related transfers	434	0.17
Trade	103	0.04
Other settlement	70	0.03
Quit claim deed	62	0.02
Assumption	55	0.02
Easement	17	0.01
Total	254,158	100.00

Notes: the reported sale frequencies are for all observed residential properties over the period 2007-2010, regardless of tenure duration, price at previous sale or owner-occupancy status.

estimates of each property’s value in hand, the information available on the associated loans allows me to estimate each property’s quarterly CLTV ratio, as described in appendix section A.1.1.

In order to construct the predicted CLTV variable, each observation is assigned two levels of the Zillow housing price index for the appropriate zip code area: One recorded the quarter preceding observation - reflecting the level of the index roughly when decision to sell or not was made - and another recorded earlier, when the quarter the property was bought by its current owner. The Zillow housing price index is described in more detail in appendix section A.1.2.

The Home Mortgage Disclosure Act (HMDA) data allow me to observe homeowner characteristics recorded at the time of loan application, including loan applicants’ reported income, race, ethnicity, joint application status and gender. Unfortunately loan-level HMDA data only report location at the Census tract level, so they cannot be perfectly matched with other data sources and I was only able to obtain a 55% success rate.³⁶ The analysis in this study is therefore conducted using aggregate HMDA variables (census tract averages) which can be assigned to all observations, and the loan-level HMDA data are used only for the purpose of reporting summary statistics.

Finally, the State Voter Registration Database (VRDB) provides information on property occupants - as opposed to owners - which I use to infer owner-occupancy. The process of inferring owner-occupancy is described in appendix A.1.3 and is asymmetric in the sense that properties inferred to be owner-occupied are in fact such, whereas properties *not* inferred to be owner-occupied may nevertheless still be owner-occupied. The VRDB data also provide occupants’ date of birth and hence their age.

1.4.2 Sample selection

I restrict the sample in several ways: First, the data are limited to properties inferred to be owner-occupied, as this is the population of interest. Second, the data are limited to properties whose owners’ tenure duration is between 1 and 10 years. Properties with owner tenure duration below 1 year are omitted to avoid capturing the effect of individuals “flipping” properties for a profit. Properties with owner tenure duration above 10 years are omitted because Zillow.com housing price indices only go back only to mid 1996, preventing the construction of the predicted CLTV variable for observations from 2007 with longer tenure duration. Third, the data are limited to properties whose owners do not concurrently own more than 2 properties in order to avoid capturing real estate investors whose sale decisions

“cascade” refers to an algorithm that pre-determines the contingent sequence in which the different models - each with its particular context-dependent pros and cons - is invoked. CoreLogic marketing materials report that automatic valuation has made significant headway over the last decade and is increasingly accepted as a substitute for human appraisal. CoreLogic is a large player in this market and it is a reasonable assumption that its AVM cascade defines the state-of-the-art.

³⁶A key variable that permits matching HMDA data with other sources is the mortgage lender identity. Unfortunately, the CoreLogic data do not include this variable. See, e.g. Stroebel (2012) and Bayer et al. (2011) for examples of previous merges with loan-level HMDA data.

Table 1.4: Summary statistics

	mean	s.d.
	(1)	(2)
Owner characteristics		
Tenure duration (years)	4.83	2.34
Age (years)	45.0	11.6
Real income* ('000s 2012)	139.7	99.5
Local share college grad. (%)	49.7	17.6
Asian* (%)	6.6	
Black* (%)	1.6	
Hispanic* (%)	1.7	
Home characteristics		
Year built	1,966.9	29.2
Square footage	2,062	849
Normalized purchase price ('000s)	382	217
Loan characteristics		
Cltv ratio at purchase (%)	84.3	17.3
Mortgage term (years)	29.0	4.2
Adjustable rate mortgage (%)	26.0	
Going interest rate at purchase (%)	5.72	0.91
N observations	1,451,696	
N properties	107,440	

Notes: For detailed description of the variables see Table 1.2 and the accompanying notes. Variables marked with an asterisk are observed only for sub-sample successfully merged with loan-level HMDA data.

are likely to reflect an entire portfolio of properties. Fourth, one percent of properties with estimated CLTV values in the top and bottom half percentiles are omitted in order to avoid extreme values, which are likely to be erroneous. Finally, only observations not missing any variables necessary for the reported regressions are kept.

In total the data used consist of 1,451,696 complete observations on 107,440 properties and 116,551 property by ownership spells.³⁷ Table 1.4 reports summary statistics.

³⁷Of these, loan-level HMDA data is available for 963,035 quarterly observations on 70,027 properties and 72,112 property by ownership spells.

1.4.3 Housing prices in King County

Figure 1.4 reports the trajectory of the Zillow housing price index (HPI) for King County, and contrasts it with the national and California HPIs. Housing prices in the county peaked only at the end of the second quarter of 2007, a year later than housing prices did nationally. From then on they dropped continuously through the end of the sample period. As can be seen, all three HPIs follow roughly a similar trajectory, but the California HPI is more accentuated - exhibiting a steeper rise and a steeper fall - than the King County and national HPIs.

Figure 1.5 displays local zip code area housing price indices within the county. There is a clear common trend underlying these local HPIs, in line with the national and California HPIs. This is reassuring, because it suggests that the forces driving these housing price indices even at this local level are far removed, and are even external to the Seattle metropolitan area. Nevertheless, there is also a substantial amount of heterogeneity across zip code areas within the county.

Table 1.5 reports the share of mortgaged homeowners who had negative equity in their homes in the second quarter of 2012 for the 30 largest US metro areas. At 37.8%, Seattle is above the national average of 30.9%. Although it is not as hard hit as the non-coastal southwest and Florida MSAs, it has a higher share of negative equity homeowners than most of the other coastal cities.

1.5 Empirical estimates

This section is organized as follows. Section 1.5.1 observes the data unconditionally, and section 1.5.2 reports naive and reduced form estimates using actual and predicted equity, respectively. Section 1.5.3 reports IV estimates. Section 1.5.4 provides evidence on the sensitivity of sale rates to price innovations and section 1.5.5 addresses loss aversion.

1.5.1 Unconditional observation

Unconditional sale rates: Figure 1.6a shows the unconditional annual rate of arm's length home sales by actual CLTV ratio. Observations on the left side are properties with low CLTV ratios, meaning their owners have high equity in them, whereas observations on the right are properties with high CLTV ratios and low homeowner equity. Properties with CLTV ratios above 100% have negative equity. The Figure indicates that the sale rate is increasing for low CLTV ratios, peaks at approximately 65% CLTV, and then declines until roughly 105%, after which it rises again sharply.

The downward slope from roughly 65% to 105% CLTV is consistent with the notion that diminished equity reduces the probability of sale and, indirectly, mobility. In contrast with the focus on negative equity in previous research and in the media, the raw data suggest that the effect of equity on mobility begins well before homeowners reach negative equity levels.

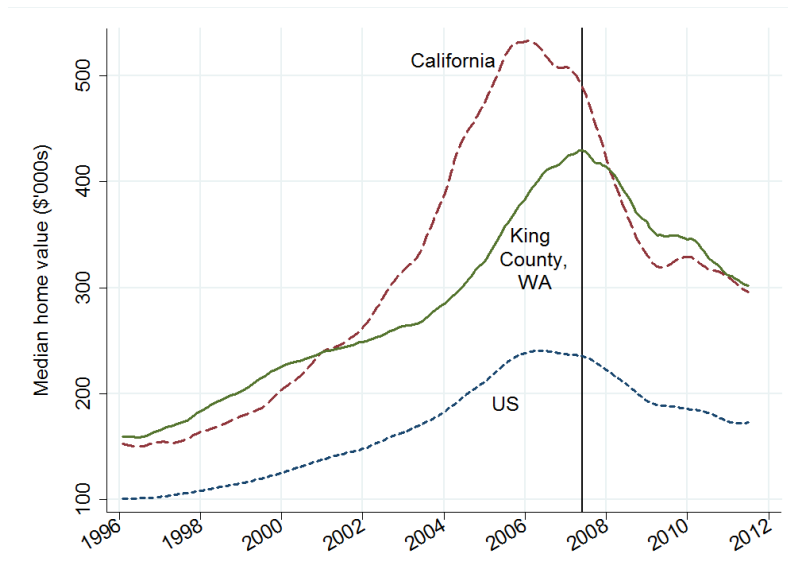


Figure 1.4: The King County, WA, housing price index vs. those of California and US.

Source: Zillow.com.

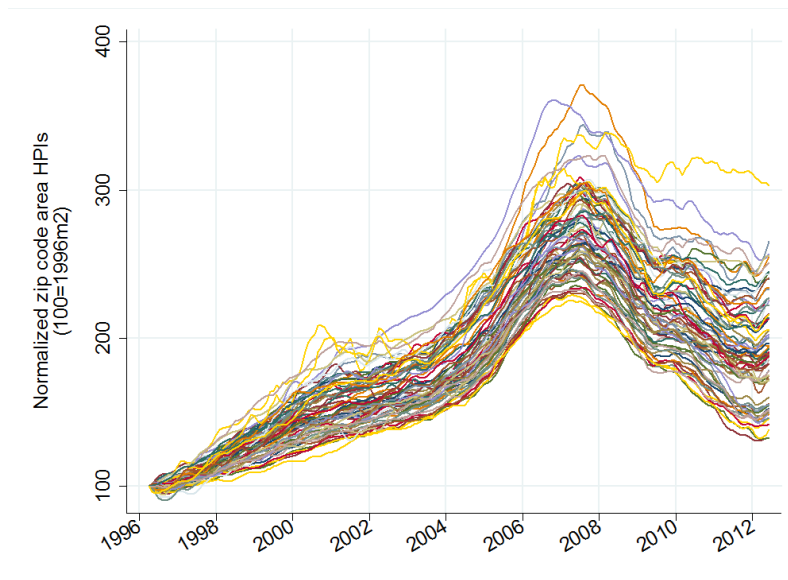


Figure 1.5: 72 zip code area housing price indices within King County. The indices share a clear common trend that aligns with the national HPI, suggesting that they are in fact driven by factors far removed from the zip code areas and from the Seattle metro area in general. Nevertheless, they also exhibit substantial variation within the county.

Source: Zillow.com.

Table 1.5: Percent of mortgaged owner-occupied homes with negative equity

Metro areas:			United States 30.9%		
1	Pittsburgh	15.6%	16	Columbus	33.4%
2	Boston	19.6%	17	San Diego	33.9%
3	San Jose	20.3%	18	Charlotte	36.4%
4	New York	20.7%	19	Seattle	37.8%
5	Philadelphia	25.4%	20	Minneapolis-St. Paul	38.7%
6	Denver	27.1%	21	Chicago	39.2%
7	San Francisco	28.5%	22	Miami-Ft. Lauderdale	43.7%
8	Los Angeles	28.6%	23	Tampa	46.6%
9	Dallas-Ft. Worth	28.9%	24	Detroit	48.3%
10	St. Louis	30.2%	25	Sacramento	49.3%
11	Cincinnati	30.3%	26	Riverside	51.2%
12	Baltimore	30.8%	27	Phoenix	51.6%
13	Washington DC	31.3%	28	Orlando	51.9%
14	Cleveland	32.9%	29	Atlanta	54.4%
15	Portland, Oregon	33.2%	30	Las Vegas	68.5%

Notes: metro Seattle, whose core is King County, is not extreme in terms of its share of properties with negative equity, but it is above the national average and is harder hit than most of the coastal US cities.

Source: Zillow.com, 2012 Q2.

It is intuitive to attribute the sharp upward slope at the right end of the Figure to distressed sales at highly negative levels of equity.³⁸ This upward slope does not have a counterpart in the previous literature, perhaps because distressed sales (and very negative equity) were not as prevalent in the real estate downturn of the late 1980's and early 1990's. The upward slope below 65% CLTV probably reflects systematic differences between homeowners with different CLTV ratios.³⁹

In contrast, Figure 1.6b shows the unconditional rate with respect to predicted CLTV. The same downward slope emerges between 65% and 105% predicted CLTV, but on either side of this interval are two regions that appear flat in comparison to the previous Figure. The upward sloping region on the left hand side of the actual CLTV (1.6a) Figure is a primary cause for concern that estimates using actual CLTV are endogenous. Its elimination in the predicted CLTV Figure (1.6b) is reassuring, and suggests that estimates using predicted

³⁸The sale indicator which I use does not include sales labeled by the county assessor as foreclosures, but it does include short sales. Although some sales are marked as short sales in assessor comments, these markings do not appear to be systematic and there is no reason to believe that they capture all short sales.

³⁹A possible explanation for this slope that can be ruled out is that of dynamic selection, whereby homeowners with higher propensity to sell - e.g. owners of so-called starter homes - are more likely to have sold by the time they would otherwise have obtained low CLTV ratios. However, as shown in appendix section A.2 this upward slope is roughly similar in raw data compiled for homeowners with tenure duration capped at 10, 5 and even just 3 years, suggesting that dynamic selection is not the culprit.

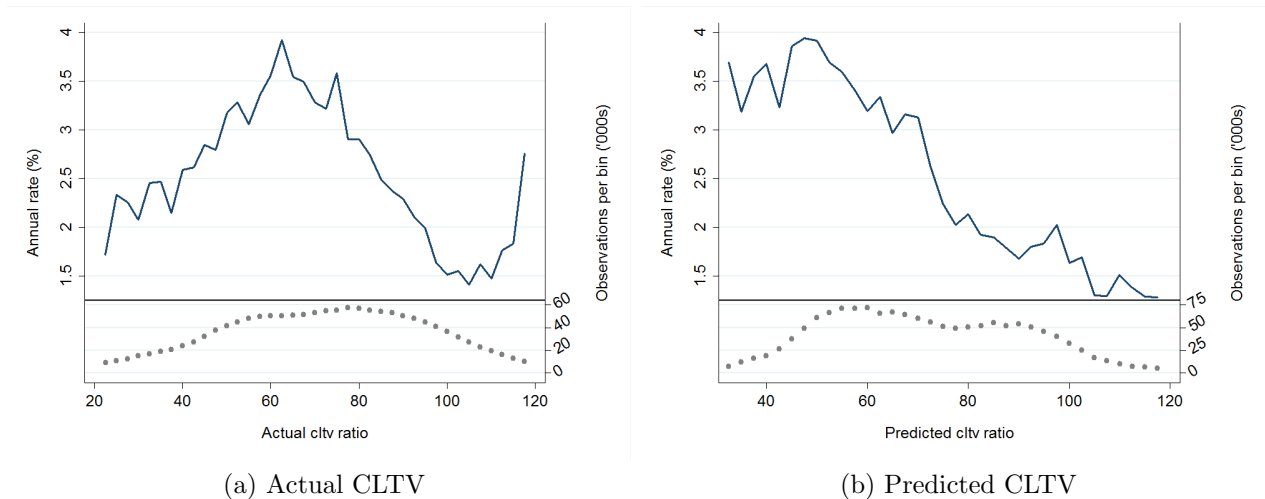


Figure 1.6: Unconditional annual arm’s length sale rates by actual and predicted CLTV ratios. The downward slope from roughly 65% to 105% CLTV in both panels is consistent with the notion that diminished equity reduces the probability of sale and, indirectly, mobility. The upward slope on the right hand side of panel *a* likely reflects short sales, and does not have a counterpart in earlier studies. In contrast, the upward slope on the left hand side of panel *a* raises the concern that actual CLTV is endogenous.

Notes: Bin means are reported for 2.5% CLTV ratio bins. The number of observations per bin refers to property-by-quarter observations.

CLTV can be more safely interpreted to reflect a causal effect of equity. The second upward slope, at right end of the actual CLTV Figure (1.6a), does not raise the same degree of concern with respect to endogeneity, but that it too disappears in the predicted CLTV Figure (1.6b) suggests that it may reflect systematic differences between homeowners with different CLTV ratios as well.

The distribution of actual and predicted CLTV ratios: the lower panels of Figures 1.6a and 1.6b show the distribution of observations over the ranges of actual and predicted CLTV ratios. Casual inspection of these panels is informative. Recall that actual CLTV reflects homeowners’ financial decisions, whereas predicted CLTV does not. Homeowners who made down payments greater (lesser) than 80% have actual CLTV ratios below (above) their predicted CLTV. Comparing the left tail of the Figures, which is thicker in the Figure 1.6a, suggests that homeowners with low actual CLTV probably made down payments greater than the ubiquitous 20%. In addition, homeowners with low predicted CLTV ratios are generally able to engage in additional borrowing against their homes, and in so doing raise their actual CLTV ratio above their predicted one. Such borrowing shifts density to the right in Figure 1.6a relative to Figure 1.6b, and is apparent owing to the roughly bi-modal distribution in both Figures: the modes are similarly located in both Figures, but the right

hand one is denser for actual CLTV than it is for predicted CLTV. Such borrowing is also apparent in the right tail, which is thicker in the Figure 1.6a.

Balancing: Table 1.6 sheds light on observable differences between properties with different actual CLTV ratios. The first three columns report a set of covariate means for subsamples with low, medium and high CLTV ratios, defined as $[30,60)$, $[60,90)$ and $[90,120)$ percent actual CLTV, respectively. Homeowners with higher actual CLTV ratios tend to be younger, poorer and less educated, and to live in smaller, less expensive homes.⁴⁰ Consistent with the being less affluent, they are also more likely to have adjustable rate mortgages (featuring lower monthly payments, all else equal). Homeowners with higher actual CLTV ratios belong to this category because they make smaller down payments and pay off their mortgages more slowly (over longer terms), and also because they tend to have shorter tenure durations, which over the observed period mostly implies having experienced more negative housing prices changes.⁴¹ All in all, homeowners with different actual CLTV ratios are quite dissimilar.⁴²

The last three columns of Table 1.6 report the same covariate means for sub-samples with low, medium and high *predicted* CLTV ratios. With respect to most characteristics, the stark differences between homeowners with different actual CLTV ratios shrink dramatically or even disappear when predicted CLTV is considered instead. That predicted CLTV, which derives only from changes in local housing prices, is substantially more balanced than actual CLTV, suggests that estimates using predicted CLTV can more confidently be regarded to reflect a causal effect of equity.

However, predicted CLTV does correlate with homes' year of construction and with homeowners' tenure duration and age. These correlations are artifacts of the prolonged monotonic rise in housing prices that preceded the current housing crisis, and that generates a positive correlation between low predicted CLTV and any variable that positively increases with tenure duration. As mentioned in section 1.3, controlling for a polynomial of tenure duration helps alleviate concerns of endogeneity that derive from these correlations, and related concerns with respect to dynamic selection are addressed in appendix section A.2 and ultimately relaxed.⁴³

⁴⁰And also in slightly newer homes, perhaps suggesting that they tend to be farther from the city center.

⁴¹Some additional notes with respect to actual CLTV in Table 1.6: (1) while minorities are only a small share of the population in metro Seattle, it is noteworthy that higher actual CLTV ratios are more prevalent among black and Hispanic homeowners, while the opposite is true of Asian homeowners. (2) The higher going interest rates associated with higher CLTV ratios should not be interpreted as evidence that such homeowners receive "worse" loans, because these are going rates determined by the timing of purchase - not the rate actually associated with an observed property's mortgage.

⁴²To what degree this holds for the data used in previous studies is uncertain, although Table 2 of Chan (2001) suggests that her sample was qualitatively similar.

⁴³Mid-range homeowners are likelier to have an adjustable rate mortgage than both other groups, reflecting that adjustable rate mortgage originations peaked sharply between 2003 and 2006, and that the general price trajectory since then puts buyers from those years mostly in the 60% to 90% predicted CLTV bin.

Table 1.6: Average observable covariates by CLTV ratio

	Actual CLTV ratio in			Predicted CLTV ratio in		
	[30,60)	[60,90)	[90,120)	[30,60)	[60,90)	[90,120)
	(1)	(2)	(3)	(4)	(5)	(6)
Owner characteristics						
Tenure duration (years)	5.80	4.59	4.03	6.33	4.48	3.45
Age (years)	47.6	44.4	42.3	46.3	44.6	43.8
Real income* ('000s)	147	140	130	141	139	141
Local share college grad. (%)	53.9	50.1	44.4	50.5	50.3	47.7
Asian* (%)	6.9	6.6	6.1	5.9	6.9	7.4
Black* (%)	0.9	1.6	2.4	1.5	1.6	1.6
Hispanic* (%)	1.0	1.8	2.5	1.4	1.8	2.0
Home characteristics						
Year built	1966.0	1967.1	1967.6	1963.0	1967.7	1970.5
Square footage	2,219	2,068	1,864	2,033	2,084	2,061
Normalized purchase price ('000s)	435	381	318	390	386	365
Loan characteristics						
Cltv ratio at purchase (%)	75.7	86.1	93.1	84.1	84.5	83.8
Mortgage term (years)	27.8	29.6	30.1	28.3	29.2	29.5
Adjustable rate mortgage (%)	18.6	28.4	31.4	23.9	28.5	23.2
Going interest rate (%)	5.57	5.68	5.96	5.72	5.69	5.81
N observations	377,331	642,620	360,250	440,528	693,648	307,301
N properties	30,093	49,832	29,823	37,015	51,447	26,812

Notes: Whereas owners, homes and loans differ systematically across *actual* equity levels, they differ far less starkly across *predicted* equity levels, along most dimensions. For detailed description of the variables see Table 1.2 and the accompanying notes. Variables marked with an asterisk are observed only for sub-sample successfully merged with loan-level HMDA data.

1.5.2 Naive and reduced form estimates

I estimate two sequences of linear probability models that are specific forms of (1.6). The first sequence is of the form

$$sale_{it} = \sum_j \beta_j \mathbf{1}\{CLTV_{it} \in [j, j + h)\} + \theta_{lt} + \mathbf{X}_{it}\delta + \epsilon_{it}, \quad (1.8)$$

where each CLTV ratio bin, j , has bandwidth h set to 2.5%. At each step in the sequence I alter the set of covariates \mathbf{X}_{it} , except for one key step in which I add the time by location fixed effects, θ_{lt} , instead. Owing to the way the CLTV bins are set up, this specification is convenient for visual presentation. Naive estimates with respect to actual CLTV are reported in Figure 1.7a, and reduced form estimates with respect to predicted CLTV are reported in Figure 1.7b. In parallel, I estimate another sequence of the form

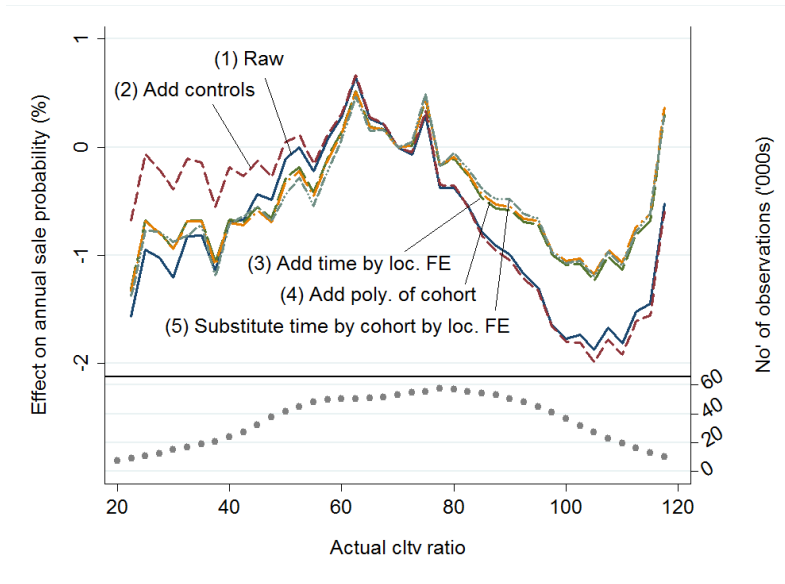
$$sale_{it} = \sum_j \beta_j \mathbf{1}\{CLTV_{it} > j\} + \theta_{lt} + \mathbf{X}_{it}\delta + \epsilon_{it}, \quad (1.9)$$

in which each of the summed $\mathbf{1}\{CLTV_{it} > j\}$ components is an indicator that the CLTV ratio exceeds a cutoff level, j , set at 10% intervals. This setup implies that the β_j coefficients capture the marginal effect of shifting from the $[j - 10, j)$ CLTV bin to the current $[j, j + 10)$ CLTV bin, and is convenient for numerical presentation. Naive estimates with respect to actual CLTV are reported in Table 1.7, and reduced form estimates with respect to predicted CLTV are reported in Table 1.8.⁴⁴

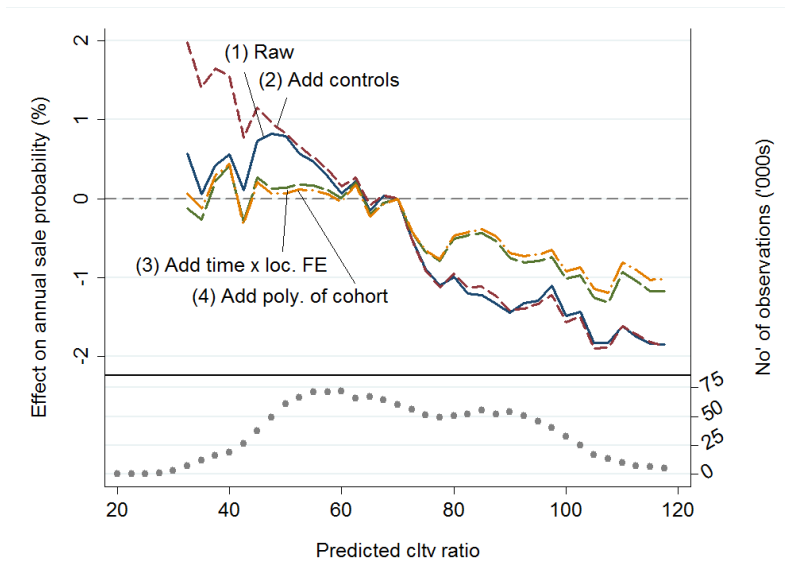
Step 1: In the first step of the sequence I adapt the raw data into a form that is easily comparable to estimates from the subsequent steps of the sequence. The solid lines labeled 1 in Figures 1.7a and 1.7b are identical to the unconditional ones in Figures 1.6a and 1.6b, respectively, other than they have been shifted downwards so that the effect associated with a 70% CLTV ratio is normalized to zero.⁴⁵ The first columns of Table 1.7 and 1.8 report the parallel estimates numerically (from (1.9)), where \mathbf{X}_{it} includes only a constant. The estimates record the same patterns as the Figures.

⁴⁴Standard errors in estimates of (1.9) are clustered at the zip code area level, which serves multiple purposes. First, such clustering allows for arbitrary correlation of errors within observations of the same property (within and between ownership spells), accounting for serial autocorrelation. Second, such clustering allows for arbitrary correlation of errors within observations made in the same time period and the same place, and therefore experiencing the same contemporary local economic conditions. Third, such clustering allows for arbitrary correlation of errors within observations sharing the same cohort of purchase, and therefore experiencing common cohort-related shocks, such as a common set of housing price expectations. Finally, such clustering accounts for spatial and combined spatial and temporal correlation of errors among nearby properties, as long as they are within the same zip code area.

⁴⁵This normalization anticipates that once a set of fixed effects is included in the regression, the estimates of β_j in (1.8) are no longer informative with respect to the absolute level of home sales by CLTV bin, but only with respect to the relative level across different CLTV bins. 70% CLTV proves to be a convenient point for normalization given the results that follow.



(a)



(b)

Figure 1.7: Conditional effects of actual and predicted CLTV ratios on annual sale rates. Estimates with respect to actual and predicted CLTV ratios both indicate a decrease in sale probabilities from roughly 70% to 100% CLTV. However, actual CLTV estimates exhibit implausible non-monotonicity. While the upward slope at the right end of the actual CLTV panel can intuitively be attributed to distressed sales, the upward slope on the left hand side is probably the result of systematic differences between homeowners with different actual CLTV ratios. That these upward slopes do not appear in the predicted CLTV estimates is reassuring with respect to the validity of these estimates' causal interpretation.

Notes: Reported estimates are coefficients of fixed effects for 2.5% CLTV bin indicators from (1.8), whose levels have been normalized such that the effect is zero at 70% CLTV. The steps of the analysis, 1 to 4, correspond to columns 1 to 4 of Table 1.7. The number of observations per bin refers to property-by-quarter observations.

Table 1.7: The effect of actual CLTV ratio on annual sale probability

Actual cltv ratio	(1)	(2)	(3)	(4)	(5)
Over 30%	0.164 (0.149)	-0.050 (0.148)	0.103 (0.145)	0.112 (0.146)	0.105 (0.152)
Over 40%	0.440*** (0.157)	0.078 (0.155)	0.242 (0.157)	0.223 (0.158)	0.319* (0.164)
Over 50%	0.495*** (0.143)	0.239* (0.141)	0.422*** (0.140)	0.428*** (0.141)	0.304** (0.141)
Over 60%	0.408*** (0.114)	0.335*** (0.120)	0.518*** (0.118)	0.540*** (0.118)	0.601*** (0.120)
Over 70%	-0.384*** (0.118)	-0.386*** (0.116)	-0.163 (0.115)	-0.147 (0.115)	-0.099 (0.117)
Over 80%	-0.609*** (0.109)	-0.648*** (0.112)	-0.402*** (0.109)	-0.379*** (0.110)	-0.345*** (0.113)
Over 90%	-0.612*** (0.109)	-0.631*** (0.112)	-0.373*** (0.110)	-0.364*** (0.109)	-0.371*** (0.112)
Over 100%	-0.506*** (0.106)	-0.553*** (0.107)	-0.358*** (0.111)	-0.338*** (0.111)	-0.381*** (0.115)
Over 110%	0.619*** (0.143)	0.618*** (0.143)	0.744*** (0.143)	0.760*** (0.145)	0.766*** (0.150)
Owner/home/loan ctrls		+	+	+	+
Time by Loc. FE			+	+	
Poly. of Cohort				+	
Time by Cohort by Loc. FE					+
N observations			1,451,278		
N properties			107,412		
Avg. outcome level			2.72%		

Notes: this Table corresponds to Figure 1.7a. The dependent variable is $100 \times$ annualized arm's length sale indicator and reported coefficients are marginal effects, measured in percentage points, so shifting from 39% to 41% or to 51% actual CLTV raises the annual probability of sale estimated in column 1 by 0.440 or by 0.935 ($= 0.440 + 0.495$) percentage points, respectively. Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. For the list and description of included home, owner and loan characteristics see Table 1.2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 1.8: The effect of predicted CLTV ratio on annual sale probability

Predicted cltv ratio	(1)	(2)	(3)	(4)
Over 30%	-0.934 (0.898)	-0.919 (0.890)	0.347 (1.147)	-0.071 (1.173)
Over 40%	0.337 (0.206)	-0.517** (0.228)	0.250 (0.253)	-0.087 (0.366)
Over 50%	-0.096 (0.136)	-0.468*** (0.137)	0.075 (0.146)	0.082 (0.176)
Over 60%	-0.479*** (0.107)	-0.474*** (0.109)	-0.123 (0.120)	-0.020 (0.160)
Over 70%	-0.633*** (0.100)	-0.686*** (0.100)	-0.405*** (0.104)	-0.302*** (0.106)
Over 80%	-0.599*** (0.096)	-0.492*** (0.098)	0.005 (0.111)	0.083 (0.115)
Over 90%	-0.112 (0.078)	-0.245*** (0.082)	-0.272*** (0.104)	-0.198* (0.110)
Over 100%	-0.288*** (0.106)	-0.316*** (0.098)	-0.310*** (0.089)	-0.244*** (0.094)
Over 110%	-0.192 (0.165)	-0.069 (0.162)	0.127 (0.175)	0.213 (0.171)
Owner/home/loan ctrls		+	+	+
Time by Loc. FE			+	+
Poly. of Cohort				+
N observations		1,451,278		
N properties		107,412		
Avg. outcome level		2.72%		

Notes: this Table corresponds to Figure 1.7b. The dependent variable is $100 \times$ annualized arm's length sale indicator and reported coefficients are marginal effects, measured in percentage points, so shifting from 59% to 61% or to 71% predicted CLTV raises the annual probability of sale estimated in column 1 by -0.479 or by -1.112 ($= -0.479 - 0.633$) percentage points, respectively. Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. For the list and description of included home, owner and loan characteristics see Table 1.2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

Step 2: In the second step of the sequence I condition the estimates on a broad set of observed owner, home and loan characteristics. In particular, these controls include a cubic of properties' initial CLTV ratios, reflecting the initial down payment, and a cubic of their owners' tenure duration, which controls for a so-called "seasoning effect" (Caplin et al., 1997) whereby owners propensity to sell evolves over time. The cubic of tenure duration also controls for the average degree of dynamic selection over time.

The dashed lines labeled 2 in Figures 1.7a and 1.7b report the effect associated with CLTV conditional on these controls. From roughly 50% CLTV up these lines track the average effect associated with CLTV in the raw data in both Figures, suggesting that non-random selection into CLTV bins above 50% CLTV is not of great concern inasmuch as the controls provide a good indication with respect to confounding unobservables. Below 50% CLTV, however, the controls influence the effect associated with CLTV, suggesting that non-random selection into CLTV bins is a concern. Unreported estimates conditioning on subsets of the controls indicate that the shift in estimates from step 1 to step 2 is driven primarily by homeowners' initial down payments.

Step 3: In this step I condition estimates on a set of quarter by zip code area fixed effects. The estimates from this step appear as long-dashed lines labeled 3 in Figures 1.7a and 1.7b. Conditioning on time by location cells causes a counter-clockwise rotation of the estimates. Low CLTV ratios - actual and predicted - tend to be observed at times and places in which the housing market was "hot", with both prices and transaction volumes high. Accounting for this tendency reduces the effect associated with low CLTV ratios, generating a downward shift on the left hand side of the Figures - from the lines labeled 2 to the lines labeled 3. Conversely, higher CLTV ratios tend to be observed at times and places in which the housing market was "cold", with both prices and transaction volumes low. Accounting for this raises the effect associated with high CLTV ratios, generating an equally substantial - but opposite - upward shift on the right hand side of the Figure.

Conditioning estimates on the time by location cell of observation reduces the estimated magnitude of the effect between 70% and 100% CLTV by half, from roughly 2 percentage points slightly less than 1. Given that the average annual sale rate at 70% actual (predicted) CLTV in the observed sample is only 3.3% (3.1%) per year, this difference is quite substantial. The third column of Tables 1.7 and 1.8 report numerical estimates corresponding to this step of the analysis. The marginal effect of each CLTV bin is estimated to be higher than in the previous column, consistent with a counter-clockwise rotation.

Step 4: In this step I account for the properties' cohort of purchase by including a cubic of the time of purchase in \mathbf{X}_{it} . The importance of this step is in accounting for systematic differences in the pool of buyers at different times. The perforated lines labeled 4 in Figures 1.7a and 1.7b report estimates of this step. These lines hardly differ from those of the previous step, implying that conditional on all previous controls, accounting for factors correlated with time of purchase barely influences the estimates.

A closely related issue is dynamic selection, whereby the composition of the pool of

homeowners in a cohort systematically changes over time as homeowners sell their homes and leave the cohort selectively. Appendix section A.2 provides evidence that dynamic selection is unlikely to be driving the observed results.

Step 5: In a final step of the sequence, I replace the set of time by location fixed effects, θ_{lt} , with a fully saturated three way set of time by cohort by location fixed effects, ψ_{lct} . This step can only be performed with respect to actual CLTV, because it eliminates all variation in predicted CLTV by construction. The double-perforated line labeled 5 in Figure 1.7a and column 5 of Table 1.7 report the estimates, which are almost indistinguishable from those of the previous two steps.

The purpose of this step is to shed light on the source of identifying variation driving the results with respect to actual CLTV. This result implies that, in contrast to the results obtained using predicted CLTV, those obtained using actual CLTV are *not* driven by changes in aggregate local housing prices. Rather, they are driven primarily by homeowners' financial decisions (as well as by property-level home values within time by cohort by location cells). Given that variation derived from homeowners' financial decisions is likely to be endogenous, the fact that the results using actual CLTV persists through this step casts doubt on their causal nature and highlights the advantage of predicted CLTV as in reflecting more plausibly exogenous variation.

1.5.3 Instrumental variable estimates

This section reports instrumental variable estimates of the effect of equity on home sales, using predicted equity as an instrument for actual CLTV. In order to allow for a non-linear effect of equity I estimate the following specification that includes a cubic of actual CLTV, and I use a cubic of predicted CLTV as the set of instruments.⁴⁶

$$sale_{it} = \beta_1 CLTV_{it} + \beta_2 CLTV_{it}^2 + \beta_3 CLTV_{it}^3 + \theta_{lt} + \mathbf{X}_{it}\delta + \epsilon_{it}. \quad (1.10)$$

The set of controls includes the full set of owner, home and loan characteristics, a set of quarter by zip code area fixed effects and a cubic of the cohort of purchase. Estimates are reported in Table 1.9. As the results using a polynomial of CLTV may differ from earlier results using CLTV bin indicators, I report naive OLS estimates of specification (1.10) in column 1 and reduced form estimates in column 2. Column 3 reports the IV estimates.

The naive OLS estimates in column 1 are similar to those in Table 1.7. The effect associated with higher actual CLTV ratios is positive at first, peaking around 65% CLTV, and then falls until a CLTV ratio of approximately 100%. The sharp upward slope observed for homeowners with negative equity at the right end of Figure 1.7a does show up in the estimate, but not with sufficient precision to be statistically significant at standard levels. Similarly, the reduced form estimates in column 2 are similar to those obtained in Table 1.8,

⁴⁶A cubic is the lowest-order polynomial capable of capturing the rough shape of the curve in Figure 1.7a. Results using higher-order polynomials do not differ substantially.

Table 1.9: IV estimates of the effect of CLTV ratio on home sale rates

	OLS	Reduced form	IV
Effect of shifting cltv	(1)	(2)	(3)
From 30% to 40%	0.324*** (0.060)	0.180 (0.503)	1.069 (1.464)
From 40% to 50%	0.512*** (0.048)	-0.014 (0.290)	0.315 (0.524)
From 50% to 60%	0.402*** (0.045)	-0.147 (0.189)	-0.248 (0.460)
From 60% to 70%	0.125*** (0.035)	-0.225 (0.142)	-0.618 (0.491)
From 70% to 80%	-0.189*** (0.031)	-0.253** (0.105)	-0.792** (0.376)
From 80% to 90%	-0.409*** (0.036)	-0.237*** (0.074)	-0.768*** (0.268)
From 90% to 100%	-0.405*** (0.038)	-0.183*** (0.061)	-0.545** (0.239)
From 100% to 110%	-0.047 (0.049)	-0.097 (0.065)	-0.118 (0.190)
From 110% to 120%	0.797*** (0.111)	0.016 (0.127)	0.513 (0.705)
Owner/home/loan ctrls	+	+	+
Time by Loc. FE	+	+	+
Poly. of Cohort	+	+	+
N observations		1,451,696	
N properties		107,440	
Avg. outcome level (percent)		2.72%	

Notes: The dependent variable is $100 \times$ annualized arm's length sale indicator. The explanatory variables of interest form a quartic of actual CLTV, and are instrumented by a quartic of predicted CLTV. Reported coefficients are specific marginal effects, measured in percentage points. Angrist-Pischke F-statistics (p-values) for the CLTV, CLTV², CLTV³ and CLTV⁴ first stage regressions are 59.00 (0), 12.31 (< 0.001), 42.21 (0) and 61.16 (0), respectively, and the partial R² for each of the first-stage regressions is 0.026, 0.036, 0.044 and 0.048, respectively. Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. For the list and description of included home, owner and loan characteristics see Table 1.2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

in that the predicted CLTV range below 70% CLTV is essentially flat, followed by a steep, statistically significant drop from 70% to 100% predicted CLTV, and then a flat region above 100% predicted CLTV.

The IV estimates are useful for obtaining the magnitude of the effect of predicted CLTV on sale probabilities in terms of actual CLTV.⁴⁷ The marginal effect of shifting from an actual CLTV ratio just below 70% to just over 100% is approximately a 2.1 percentage point drop. Because the annual rate of arm's length sales observed in the sample below 70% CLTV is roughly 3.07%, a 2.1% drop is quite dramatic and means that a shift to 100% CLTV almost completely eliminates arm's length home sales.⁴⁸ As in the naive OLS estimate, the upward slope for homeowners with negative equity does show up, but it can not be significantly distinguished from zero.

The OLS bias

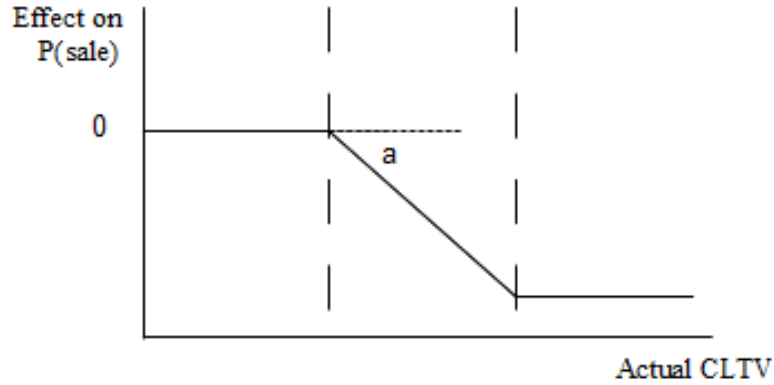
Focusing on the coefficients in the 70% to 100% CLTV range, note that the IV estimates are greater in (absolute) magnitude than the naive OLS estimates, indicating that the latter are biased towards zero. Figures 1.8a-1.8c illustrate the situation in a stylized manner. Figure 1.8a depicts the true effect of equity on the probability of home sale, reflecting the reduced form and IV results with respect to the range and magnitude of the effect, respectively. The true effect is confined to the 70% and 100% CLTV range, in which its slope is steep and negative. If the sum of confounding effects is *positive* and has a moderate slope, as depicted in Figure 1.8b, then the super-position of the true and confounding effects yields a non-monotonic pattern resembling the naive OLS estimates. In this non-monotonic pattern shown in Figure 1.8c sales increase in CLTV below 70%, then decrease between 70% and 100% CLTV - but with a more moderate slope than the true effect - and then increase again above 100%. Thus, a positive sum of confounding effects can explain the bias of the naive OLS estimates towards zero in the 70% to 100% CLTV range. But why might the sum of confounding effects associated with equity be positive?

There could be any number of confounding effects that generate a positive slope. An example mentioned earlier is that greater household prosperity may induce both larger down payments and lower propensities to move, thereby biasing OLS regressions of sales on CLTV upward (similarly, greater prosperity may induce shorter mortgage terms, to the same effect). Another example mentioned earlier involves renovations: As long as renovations which increase the value of a home and decrease CLTV tend to coincide with intentions to stay in the home for an extended period, they give rise to a positive slope.⁴⁹ A separate source of bias could be measurement error in the (actual) CLTV variable, which may arise for multiple

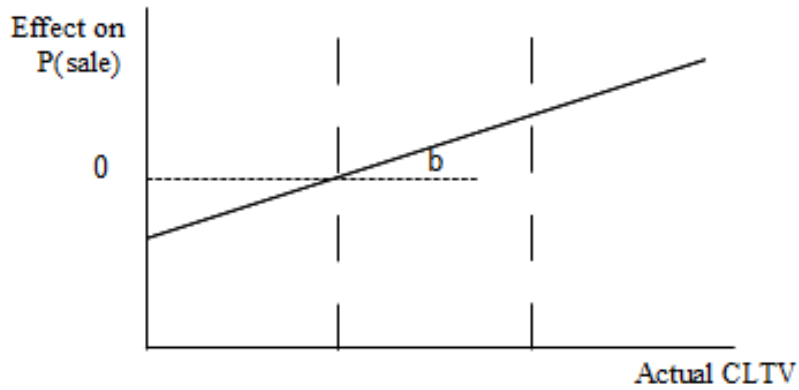
⁴⁷The IV estimates are larger in (absolute) magnitude than the reduced form estimates because every unit of predicted CLTV shifts actual CLTV only by a fraction of a unit, as indicated (unconditionally and linearly) in Figure 1.3. For a textbook example see Cameron and Trivedi (2005) section 4.8.3 (page 98).

⁴⁸At its highest, in 2007, the annual rate of arm's length sales observed in the sample below 70% CLTV is roughly 4.5%, and at its lowest, in 2011, it is roughly 2.0%. The observed 2.1% drop is substantial compared to either of these figures.

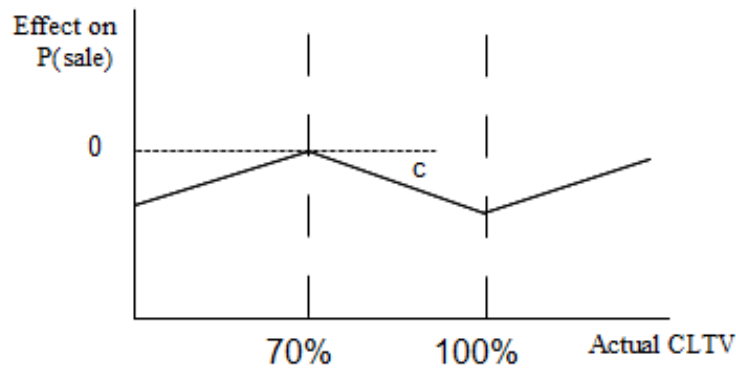
⁴⁹Renovations may show up in the data if they are of a type that is recorded by the county assessor's office, and if they the recorded variables are reflected in CoreLogic's AVM.



(a) The causal effect of equity. This is the effect reflected by the IV estimates.



(b) The sum of confounding effects associated with equity.



(c) The super-position of 1.8a and 1.8b given $|a| > |b|$. This is the confounded “effect” reflected by the naive OLS estimates.

Figure 1.8: Stylized illustration of the relationship between naive OLS and IV estimates. Figure 1.8a depicts the true effect of equity on the probability of home sale, reflecting the reduced form and IV results with respect to the range and magnitude of the effect, respectively. Figure 1.8b portrays the sum of confounding effects as having a moderate positive slope. The super-position of Figures 1.8a and 1.8b yields a non-monotonic pattern resembling the naive OLS estimates, shown in Figure 1.8c. 34

reasons. For instance, as I do not observe payment histories directly, I assume that payments towards principal are paid on time and set according to the standard (fixed payment) amortization formula. Neither of these assumptions is likely to be true in general, attenuating the naive OLS estimates to zero.⁵⁰

The “better homes” hypothesis: Perhaps the most important confounding effect stems from the distinction between aggregate and individual changes in home values (recall that both types of changes are reflected in actual CLTV, but only aggregate changes are reflected in predicted CLTV). Suppose certain homes within a quarter \times zipcode cell are, on average, more appealing to incumbent and potential residents - e.g. because of their layout or precise location - in a way that is unobservable to the econometrician. The incumbent residents of such “better homes” will be less likely to sell them, so compared to observably similar homes they will have a lower sale rate, or lesser “churn”. At the same time, the demand for “better homes” will be more robust, in the sense that when the demand for housing falls, these will be the last homes in which buyers lose interest. Thus, when housing prices fall “better homes” will lose less of their value - thereby maintaining lower CLTV ratios - than the average observably similar home. Combining these two observations implies that “better homes” will tend to have both lower sale rates (less “churn”) and lower CLTV ratios in periods when housing prices are falling, generating the upward slope shown in Figure 1.8b. A simulation and empirical evidence reported in appendix section A.3 indicate that the “better homes” hypothesis is likely to explain at least part of the OLS bias.

1.5.4 Sensitivity of sales to housing price innovations

In this section I ask whether the sale probability of properties with low or negative equity is more sensitive to the latest changes in local housing prices. If the observed reduction in the home sale rates from 70% to 100% CLTV is caused by insufficient equity then a drop in housing prices should reduce the sale probability *around this range* by further eroding equity, and an increase in prices should do the opposite. There is no clear reason why housing price innovations should affect sale probabilities *differentially* by CLTV ratio, unless insufficient equity is an issue.

To estimate the sensitivity of sale probabilities with respect to housing price innovations I estimate a sequence of linear probability models of the following form

$$\begin{aligned} sale_{it} = & \alpha(P_{i,t} - P_{i,t-4}) + \sum_j \beta_j \mathbf{1}\{CLTV_{i,t-4} > j\} \\ & + \sum_j \gamma_j [(P_{i,t} - P_{i,t-4}) \mathbf{1}\{CLTV_{i,t-4} > j\}] + \theta_{it} + \mathbf{X}_{it} \delta + \epsilon_{it}. \end{aligned} \quad (1.11)$$

This specification is similar to that in (1.9), but it contains CLTV ratios that are lagged by one year (expressed as four quarters). In addition, it includes the change in the value of property i over the last year, $(P_{i,t} - P_{i,t-4})$, as a covariate, and a set of interaction

⁵⁰Attenuation to zero amounts to an upward bias in the 70% to 100% CLTV range, but attenuation alone is not sufficient to produce the slope shown in Figure 1.8b.

terms between this price innovation and the one-year lagged CLTV ratios. The γ s are the coefficients of interest.

Table 1.10 reports estimates of (1.11) for specifications that parallel steps 1 to 5 of the previous analysis. All four specifications reveal a similar picture: the sensitivity of sale probabilities with respect to price innovations first become more positive at lagged CLTV ratios of roughly 70%, and then become more negative above 100%. This finding means that in the window ranging from roughly 70% CLTV to 100%, price innovations have a more positive effect on the likelihood of sale than they do at outside this range, consistent with a causal effect of equity.

Note that the negative interaction effect above 100% CLTV is even greater in magnitude than the sum of the positive effects at 60% and 70% CLTV. This result is consistent with the possibility that most sales observed to occur with negative equity are short sales, and it reveals that the likelihood of such sales is negatively correlated with price innovations, i.e. short sales are more common when prices are falling than when they are rising.

1.5.5 Loss Aversion

Genesove and Mayer (2001) show that sellers who have lost in home value tend to set higher asking prices and are therefore less likely to succeed in selling their homes. The fact that falling values mechanically reduce equity suggests that low equity may non-causally correlate with depressed home sales simply because of loss aversion. However, loss aversion should not affect the sale probability of sellers who experience both falling home values and net gain in equity. Thus, to ensure that my results are not driven by this alternative explanation, I re-estimate the models above on the sub-sample of owners who experience a net gain in equity.⁵¹

Figures 1.9a and 1.9b summarize the analysis within the gain-only sample, and their resemblance to the corresponding Figures for the full sample (1.7a and 1.7b) is striking. Corresponding estimates are reported in Table 1.11 for specifications that include controls for the full set of owner, home and loan characteristics as well as quarter by zip code area fixed effects and a cubic of the time of purchase (for completeness, Table 1.12 reports the parallel estimates for the Loss only sample). Column 1 reports naive OLS estimates using actual CLTV which reflect the non-monotonic pattern in Figure 1.9a.⁵² Column 2 reports reduced form estimates using predicted CLTV, omitting the ranges below 30% and above 90% predicted CLTV which have few observations. A predicted CLTV ratio in excess of 70% reduces the probability of sale by 0.4 percentage points in this sample, a slightly larger effect

⁵¹Net loss and net gain refer to a home's current value minus its purchase price being negative or positive, respectively. A homeowner who first experiences a gain in home value and then experiences a greater loss is considered to experience only *net* loss and not *net* gain, even though he has experienced both gain and loss during his tenure.

⁵²The upward slope at high negative equity levels that emerges in the full sample does not emerge as sharply in Figure 1.9a, and does not show up at all in Table 1.11. Estimates obtained from an opposite loss-only sample, as well as the results in Table 1.10, indicate that the increase in sales (short sales) at high negative equity levels is driven by homeowners experiencing net loss.

Table 1.10: Sensitivity of annual sale rates to price innovations by lagged actual CLTV

1 yr lagged CLTV ratio × year-on-year Δ value ($\times 10,000$)	(1)	(2)	(3)	(4)	(5)
Over 30%	0.019 (0.037)	0.019 (0.037)	-0.005 (0.036)	-0.005 (0.036)	0.002 (0.037)
Over 40%	-0.062** (0.031)	-0.065** (0.031)	-0.077** (0.032)	-0.081** (0.032)	-0.088*** (0.031)
Over 50%	0.044 (0.033)	0.035 (0.033)	0.017 (0.033)	0.013 (0.034)	0.007 (0.036)
Over 60%	0.099*** (0.033)	0.103*** (0.033)	0.080** (0.032)	0.079** (0.032)	0.070** (0.032)
Over 70%	0.046 (0.032)	0.062* (0.032)	0.064** (0.032)	0.064** (0.032)	0.083** (0.032)
Over 80%	-0.008 (0.035)	0.009 (0.035)	0.019 (0.035)	0.019 (0.035)	0.022 (0.036)
Over 90%	-0.111** (0.048)	-0.088* (0.048)	-0.058 (0.049)	-0.058 (0.049)	-0.026 (0.052)
Over 100%	-0.268*** (0.098)	-0.282*** (0.098)	-0.204** (0.097)	-0.204** (0.097)	-0.226** (0.101)
Over 110%	0.132 (0.138)	0.088 (0.139)	0.119 (0.139)	0.125 (0.139)	0.100 (0.146)
1 yr lagged CLTV ratio bins YoY change in value	+	+	+	+	+
Owner/home/loan ctrls		+	+	+	+
Time by Loc. FE			+	+	
Poly. of Cohort				+	
Time by Cohort by Loc. FE					+
N observations			1,451,278		
N properties			107,412		
Avg. outcome level			2.72%		
Median home value ('000s)			422		

Notes: The dependent variable is $100 \times$ annualized arm's length sale indicator. Reported coefficients are marginal effects, measured in percentage points. For example: the coefficient for "over 60%" in column 1 indicates that if actual CLTV is greater than 60%, a \$10,000 year-on-year increase in value raises the annual prob. of sale by 0.099 percentage points. Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. For the list and description of included home, owner and loan characteristics see Table 1.2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

than in the full sample.⁵³ Columns 3, 4 and 5 of the Table report separate naive and reduced form estimates, and of course IV estimates, using a third order polynomial of CLTV. These results, too, are qualitatively similar to the corresponding full sample results (Table 1.9), although they are substantially less precise.⁵⁴

The results for the gain-only sample allow me to rule out that the full sample results are driven purely by loss aversion, especially considering that roughly 60% of the full sample also belongs to the gain-only sample. However, I cannot rule out that the effect of equity on mobility estimated in the full sample is *at least partially* driven by loss aversion. An observed correlation between low equity and reduced sale rates within a similarly defined *loss-only* sample can still reflect loss aversion, as can differences between gain-only and loss-only samples.⁵⁵

1.6 Conclusion

In this study, I estimate the effect of owner-occupant equity on the probability of home sale and, indirectly, on mobility. Most importantly, I establish the causal nature of this empirical relationship. I do so primarily by exploiting plausibly exogenous variation in equity that stems only from changes in local housing price indices, and by comparing properties observed in the same time and location, thereby controlling for a wide variety of confounding time- and location-varying economic conditions. In addition, I bring the literature on the equity-mobility nexus up to date, using a novel combination of proprietary and administrative data to provide estimates from the current housing crisis.

I find that sale probabilities decline well before homeowners reach negative equity levels, suggesting that the share of homeowners who have negative equity substantially understates the share of homeowners whose mobility is impaired by insufficient equity. More importantly, my findings indicate that policy attempting to influence the housing market or the economy at large by manipulating homeowners' equity - e.g. by some means of reducing mortgage principal - is likely to be most effective in inducing mobility if it targets homeowners whose equity ranges from 70% to 100% CLTV. Admittedly, targeting this group of homeowners rather than those who are deeper "underwater" may not be an attractive policy, but recognizing the range in which equity shifts the probability of sale and mobility can nevertheless help shape more effective and informed policy. Increasing mortgaged homeowners' equity across the board by generating mild inflation, for example, is likely to induce

⁵³The continued decline above 90% predicted CLTV that is observed in the full sample cannot be observed in the gain-only sample, as very few homeowners experiencing gains have such high predicted CLTV ratios (many have actual CLTV ratios above 90% due to their financial decisions, but not due to changes in price).

⁵⁴The intuition behind this loss of precision is that restricting the sample only to homeowners with net gain substantially reduces variation in predicted CLTV, but does not affect the amount of variation in other determinants of actual equity, such as down payment amounts. Consequently predicted CLTV is not as good a predictor of actual CLTV in the gain-only sample as it is in the full sample, implying a weaker first stage and less precise IV estimates. This problem is even more severe in the loss-only sample, in which IV results are too imprecise to be informative.

⁵⁵See footnote 31.

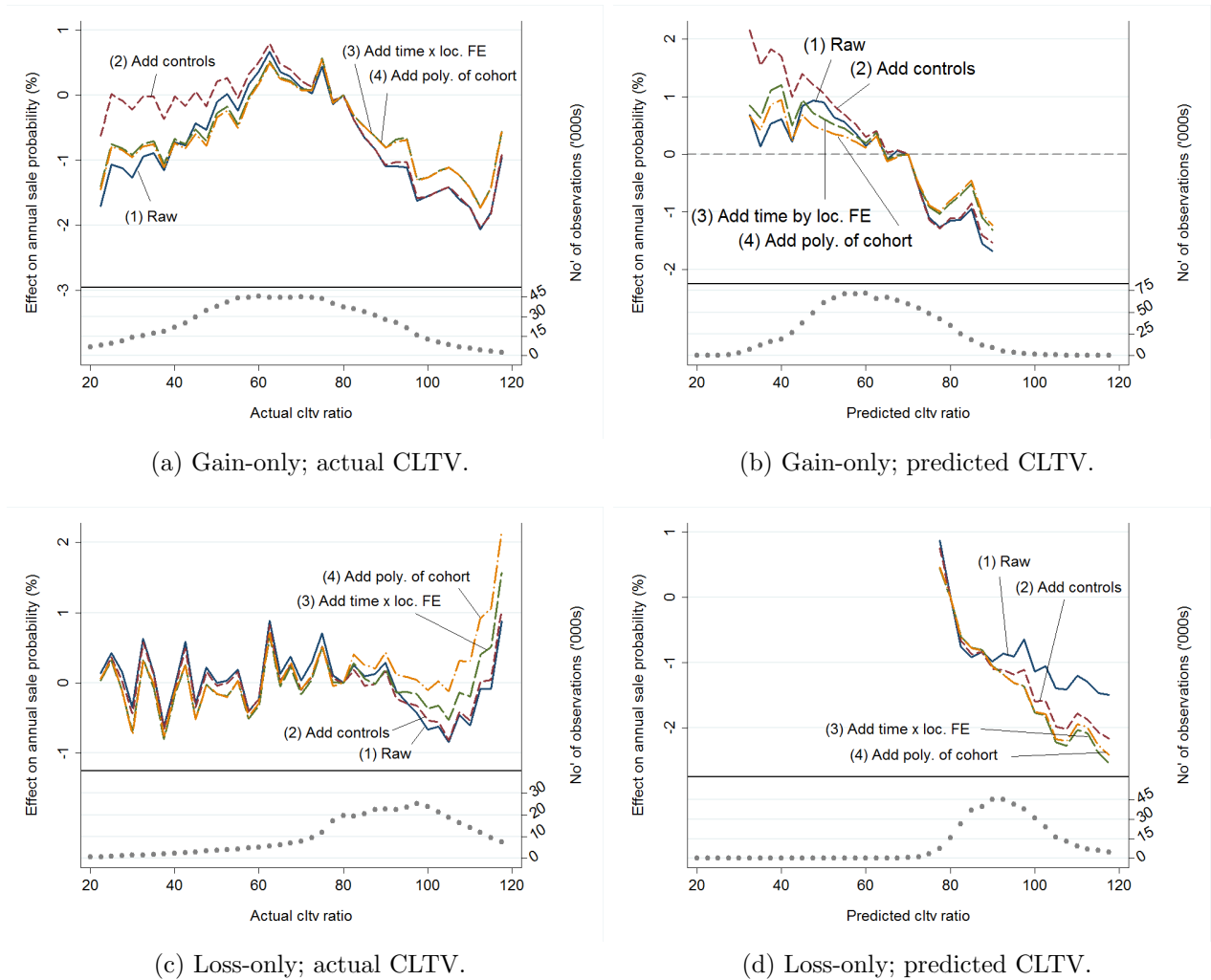


Figure 1.9: Gain-only and loss-only sample results. The gain-only results in panels *a* and *b* are remarkably similar to those in the full sample (Figures 1.7a and 1.7b), and they are not subject to the influence of loss aversion. Panel *d* indicates that in the loss-only sample sale rates increase (decrease) with predicted equity (CLTV), which is consistent with equity affecting sales through the liquidity constraint mechanism, but not through a wealth effect. The contrast between the loss-only result in panel *c* - in which sale rates do not decrease with actual CLTV (when fully conditioned in line (4)) - and those in panel *d*, stems from the fact that predicted and actual CLTV ratios are only weakly correlated within the loss-only sample (confining the sample only to properties with net loss substantially reduces variation in predicted equity without reducing variation in other determinants of actual equity, such as down payment amounts, with the result that predicted equity explains a much smaller share of variation in actual equity within this sample).

Table 1.11: Estimates for gain-only sample

Cltv ratio	cltv bins		Effect of shifting cltv	3^{rd} -order polynomial of cltv		
	actual	predicted		OLS	Reduced form	IV
	(1)	(2)		(3)	(4)	(5)
Over 40%	0.171 (0.167)	-0.124 (0.367)	From 30% to 40%	0.442*** (0.085)	0.264 (0.579)	2.630* (1.355)
Over 50%	0.441*** (0.152)	0.008 (0.181)	From 40% to 50%	0.297*** (0.056)	-0.030 (0.373)	0.806 (0.772)
Over 60%	0.563*** (0.127)	0.004 (0.160)	From 50% to 60%	0.150*** (0.050)	-0.230 (0.237)	-0.486 (0.529)
Over 70%	-0.131 (0.128)	-0.412*** (0.117)	From 60% to 70%	0.003 (0.049)	-0.335* (0.173)	-1.247** (0.525)
Over 80%	-0.493*** (0.148)	0.000 (0.127)	From 70% to 80%	-0.145*** (0.045)	-0.346** (0.156)	-1.476*** (0.549)
Over 90%	-0.501*** (0.130)		From 80% to 90%	-0.293*** (0.038)	-0.263* (0.150)	-1.174* (0.637)
Over 100%	-0.359** (0.176)		From 90% to 100%	-0.443*** (0.052)	-0.085 (0.164)	-0.340 (0.985)
Over 110%	-0.021 (0.248)		From 100% to 110%	-0.593*** (0.096)	0.188 (0.246)	1.025 (1.643)
Owner, home and loan ctrls	+	+	Owner, home and loan ctrls	+	+	+
Time by Loc. FE	+	+	Time by Loc. FE	+	+	+
Poly. of Cohort	+	+	Poly. of Cohort	+	+	+
N observations	1,028,770		N observations	1,028,770		
N properties	90,530		N properties	90,530		
Avg. outcome level	3.03%		Avg. outcome level	3.03%		

Notes: the dependent variable is $100 \times$ annualized arm's length sale indicator. This Table reports results for the gain-only sample, which includes all properties in the full sample whose estimated value at the time of observation exceeds the purchase price. The results correspond to full-sample results as follows: columns 1 and 2 correspond to the fourth columns of Tables 1.7 and 1.8, respectively, and columns 3, 4 and 5 correspond to the columns of Table 1.9 - see notes in the corresponding full-sample Tables. With respect to column 5, Angrist-Pischke F-statistics (p-values) for the CLTV, CLTV² and CLTV³ first stage regressions are 335.28 (0), 208.77 (0) and 132.59 (0), respectively, and the partial R² for each of the first-stage regressions is 0.0017, 0.0034 and 0.0045 respectively. These low partial R² values make the estimates substantially less precise than the full sample ones - see footnote 54 for the intuition. For the list and description of included home, owner and loan characteristics see Table 1.2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

Table 1.12: Estimates for loss-only sample

Cltv ratio	cltv bins		Effect of shifting cltv	3 rd -order polynomial of cltv		
	actual	predicted		OLS	Reduced form	IV
	(1)	(2)		(3)	(4)	(5)
Over 40%	0.291 (0.378)		From 30% to 40%	0.647*** (0.192)	-13.067*** (4.895)	55.943 (43.366)
Over 50%	-0.031 (0.324)		From 40% to 50%	0.307** (0.128)	-9.954*** (3.566)	32.480 (27.303)
Over 60%	0.467 (0.288)		From 50% to 60%	0.076 (0.084)	-7.267*** (2.449)	13.861 (14.429)
Over 70%	-0.000 (0.241)		From 60% to 70%	-0.047 (0.061)	-5.007*** (1.545)	0.085 (5.187)
Over 80%	0.130 (0.187)	-2.141*** (0.558)	From 70% to 80%	-0.061 (0.053)	-3.174*** (0.856)	-8.846** (4.114)
Over 90%	-0.030 (0.165)	-0.166 (0.147)	From 80% to 90%	0.033 (0.050)	-1.767*** (0.391)	-12.935* (7.140)
Over 100%	-0.139 (0.142)	-0.412*** (0.147)	From 90% to 100%	0.235*** (0.055)	-0.788*** (0.178)	-12.179 (7.691)
Over 110%	1.362*** (0.160)	0.116 (0.197)	From 100% to 110%	0.546*** (0.079)	-0.235* (0.136)	-6.581 (5.374)
Owner, home and loan ctrls	+	+	Owner, home and loan ctrls	+	+	+
Time by Loc. FE	+	+	Time by Loc. FE	+	+	+
Poly. of Cohort	+	+	Poly. of Cohort	+	+	+
N observations	422,926		N observations	422,926		
N properties	46,865		N properties	46,865		
Avg. outcome level	1.95%		Avg. outcome level	1.95%		

Notes: the dependent variable is $100 \times$ annualized arm's length sale indicator. This Table reports results for the gain-only sample, which includes all properties in the full sample whose estimated value at the time of observation exceeds the purchase price. The results correspond to full-sample results as follows: columns 1 and 2 correspond to the fourth columns of Tables 1.7 and 1.8, respectively, and columns 3, 4 and 5 correspond to the columns of Table 1.9 - see notes in the corresponding full-sample Tables. With respect to column 5, Angrist-Pischke F-statistics (p-values) for the CLTV, CLTV² and CLTV³ first stage regressions are 8.00 (0.006), 11.21 (0.001) and 24.08 (0), respectively, and the partial R² for each of the first-stage regressions is 0.0248, 0.0318 and 0.0346 respectively. The low F-statistics imply that the estimates are not informative, but I report them for the sake of completeness - see footnote 54 for the intuition. For the list and description of included home, owner and loan characteristics see Table 1.2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

mobility among homeowners in the 70% to 100% CLTV range, whereas principal reductions for deeply “underwater” homeowners may fail to induce any mobility whatsoever.

Abstracting from policy and the present context, the theoretical link put forth in Stein (1995) whereby owner-occupant mobility hinges upon sufficient equity and its relation to the positive price-volume correlation in housing markets form a basic tenet of our understanding of housing markets and their cycles. This study underpins our understanding by providing evidence that the corresponding *empirical* relationship between equity and mobility is causal. By shedding light on a causal effect of price on trading volume, this study contributes to the literature disentangling the multiple causal relationships driving the positive price-volume correlation in housing markets. Establishing both theoretical and causal empirical relationships running in the opposite direction, from volume to prices, could provide a foundation for a more complete, dynamic understanding of housing markets and their cycles, and may be a fruitful avenue for future research.

Chapter 2

Re-Interpreting Burdett and Mortensen: Equilibrium Wage Dispersion with Balanced Matching

2.1 Introduction

The fundamental result in Burdett and Mortensen’s celebrated model is that even in the absence of worker or firm heterogeneity, search frictions alone give rise to equilibrium wage dispersion.¹ In a nutshell, the mechanics of worker flows are such that firms posting higher wages recruit workers more frequently and are quit upon less often. This state of affairs results in high wage firms retaining a larger steady-state workforce, thereby giving rise to an iso-profit wage-workforce schedule along which firms can disperse.² An immediate implication is that firms posting higher wages employ more labor. This consequence is especially important because it means that firms face upward sloping labor supply curves, i.e. they have monopsony power. This insight theoretically underpins the competitive monopsony literature.³

The devil, of course, is in the details. Burdett and Mortensen assume that workers randomly search among employers and assign each firm equal odds of being encountered, irrespective of firm size. This regime is known as *random matching* (with firms), and it generates some troubling implications. Burdett and Vishwanath (1988) point out that random matching with firms has the bizarre implication that by spitting itself in two, a firm can increase its recruitment rate. Less abstractly, Kuhn (2004) lets the reader judge random matching with firms based on the implication that “regardless of its size, every firm - from the local bakery to Microsoft - receives the same absolute number of job applications per period.”⁴ At the opposite end of the spectrum is the alternative of *balanced matching* (with

¹Burdett and Mortensen (1998).

²That firms *necessarily* disperse along an iso-profit wage-workforce schedule in equilibrium is shown in the most memorable step of Burdett and Mortensen’s proof. Manning (2011) questions how accurately this step captures the real process leading to dispersion of the wage offer distribution, because it relies on workers switching jobs for the sake of infinitesimal gains.

³Ashenfelter et al. (2010) and Manning (2011) provide brief and extensive overviews of this literature, respectively, and Manning (2003) is the definitive volume on the topic.

⁴Kuhn (2004), page 375.

firms), in which a firm is encountered with probability proportional to its workforce. Burdett and Vishwanath (1988), who coined the terms for these two matching regimes, were also the first to show that, under balanced matching, equilibrium wage dispersion degenerates into a single mass point at the competitive wage.⁵ Balanced matching with firms eliminates pure equilibrium wage dispersion.

The relative importance of random versus balanced matching with firms in actual labor markets is crucial in determining the amount of monopsony power firms can muster, and hence also in determining the empirical relevance of competitive monopsony. For this reason Manning (2003) discusses the matter carefully.⁶ He argues that some methods of recruitment, such as using existing employees' word-of-mouth, are more suggestive of balanced matching with firms than are other methods, such as the use of public employment services or advertisement. He then provides empirical evidence that methods tending less strongly towards balanced matching with firms are quite commonly used, which is compelling because it suggests that at the very least a complete absence of monopsony power is unlikely. Nevertheless, it remains easy to argue that just about any recruitment method sends more job seekers to the Microsofts of the world than to local bakeries, raising doubt with respect to the extent of firms' monopsony power. The ground on which competitive monopsony has been standing is shaky.

This paper attempts to secure competitive monopsony on a more solid foundation. A modified version of Burdett and Mortensen's model is proposed, in which pure equilibrium wage dispersion arises under a form of balanced matching with firms. Instead of workers and firms seeking each other out, workers in the model seek jobs and firms seek workers to fill jobs. I assume random matching between workers and jobs (not firms!), with two consequences: first, doing so allows me to formulate a model that follows the lines of Burdett and Mortensen's original very closely. In particular, the modified model retains the key result of pure equilibrium wage dispersion and it continues to imply that employers face an upward sloping labor supply curve, indicating monopsony power. Second, random matching with jobs implies that the rate at which a firm is contacted is proportional to its measure of jobs, which amounts to a form of balanced matching with firms. Firms wield monopsony power even when matching is balanced.

Wording the statement that random matching with jobs amounts to *a form of* balanced matching marks a nuance in the way balanced matching is defined here. Burdett and Vishwanath (1988) define balanced matching as holding when the probability that a worker contacts a given firm, conditional on contacting some firm, equals the number employed by that firm divided by the total employed labor force. Random matching with jobs implies a form of balanced matching with firms that is closely related, but distinct. This form of balanced matching holds when the probability that a worker contacts a given firm, conditional on contacting some firm, equals the number of jobs at that firm divided by the total number of jobs in the labor market. Thus, I define balanced matching with firms in terms of the

⁵The breakdown of equilibrium wage dispersion under balanced matching also arises in Mortensen and Vishwanath (1991), Manning (1993) and Mortensen and Vishwanath (1994). This result is reviewed in section 10.3 of Manning (2003).

⁶See sections 10.3 and 10.4 of Manning (2003).

number of jobs at the firms, rather than the number of workers they employ.

Clearly, these definitions of balanced matching refer to different statistics. It is also clear, however, that both definitions are consistent with the casual observation that larger firms receive greater numbers of applications per period than small firms do. In this respect both definitions are similarly apt. Unless one insists that the matching regime is balanced precisely with respect to the number of workers employed and not the number of jobs at a firm, the argument that Microsoft attracts more job seekers than the local bakery no longer poses a challenge. Moreover, when matching is balanced in terms of jobs then a firm's recruitment rate is not affected by the act of splitting the firm, so the bizarre implication pointed out by Burdett and Vishwanath (1988) is eliminated.

The proposed model closely resembles Burdett and Mortensen's original. The labor market comprises two continuums, one of workers and one of jobs. The mechanics of worker flows are such that jobs associated with higher posted wages recruit workers more frequently, and are quit upon less often. This state of affairs results in high wage jobs maintaining lower steady-state vacancy rates, thereby giving rise to an iso-profit wage-vacancy rate schedule along which firms can disperse their jobs.

Some appealing results carry over from the original model. Raising wages speeds up recruitment. By increasing the wage posted for a vacant job, an employer increases the probability that an encountered worker will forego his prior job in favor of its offer. Raising wages reduces worker turnover, too. By increasing the wage posted for an occupied job, an employer reduces the probability of the incumbent worker quitting in favor of a higher paying alternative.

Note that both results are stated in terms of jobs, not firms. A firm in the model is no more than an arbitrary grouping of jobs. That any two jobs belong to the same firm does not prevent that firm from posting different wages for them, because the model is essentially blind as to which firm a job belongs to.⁷ Whether a firm in the model encompasses one job or seventeen remains arbitrary, too, and this matter is worth dwelling upon.

Most wage-posting models implicitly assume that in equilibrium, steady-state firm size is determined by labor supply alone, or equivalently that all firms have excess demand for labor.⁸ This implicit assumption stems from the simplifying assumption that (conditional on worker and firm characteristics) the marginal revenue product of labor is fixed. Because firms post wages below the marginal revenue product of labor, their labor demand at the

⁷The model makes no prediction with respect to within-firm wage distributions, but it is worth noting that wage-posting models in which it is assumed that firms post a single wage for all employees (such as Burdett and Mortensen (1998)) make no such prediction either. Some wage-posting models do make predictions with respect to within-firm wage distributions. An example of such a model is Postel-Vinay and Robin (2002), in which firms can make counter-offers to their employees when the latter encounter alternative employers.

⁸This is the state of affairs, for example, in Burdett and Mortensen (1998), as well as Mortensen (1998), Mortensen (1999), Bontemps et al. (2000), Postel-Vinay and Robin (2002), Burdett and Coles (2003) and Cahuc et al. (2006), as well as in most of the models presented in Manning (2003). Exceptions to this rule are the wage-posting models containing some degree of balanced sampling: Burdett and Vishwanath (1988), Mortensen and Vishwanath (1991), Manning (1993) and Mortensen and Vishwanath (1994). In these papers decreasing marginal returns to labor are assumed and the size of some (or all) firms is determined by labor demand.

posted wages is infinite, leaving finite labor supply to determine firm size. In this paper it is assumed that firm size, measured by the number of jobs available in a firm, is determined solely by labor demand. Specifically, it is assumed that the marginal revenue product of labor in each firm is fixed until an arbitrary firm-specific level of employment is reached, after which it drops to zero. That a firm's labor demand is satisfied after employment hits a firm-specific cap reflects, in crude approximation fashion, the role of non-labor inputs as well as conditions in the firm's output market, both of which are taken to be exogenous. That each firm's measure of jobs is exogenous means that the model makes no prediction with respect to firm size (by either measure thereof - jobs or employment).

Even though firm size is undetermined in the model, firms still face an upward-sloping labor supply. To see this note that while fixing the number of jobs in the firm, raising wages reduces the firm's steady-state vacancy rate. As employment is simply the measure of occupied or non-vacant jobs, raising wages must increase the firm's steady-state number of employees, facing the firm with upward-sloping labor supply. This point is crucial because it is what tells us that even with balanced matching, firms wield monopsony power.

The model is presented in the next section, followed by a brief conclusion.

2.2 Model

2.2.1 Setup:

The labor market consists of a continuum of workers with measure m and a continuum of jobs with measure normalized to 1. Certain mutually exclusive sets of jobs comprise firms, and firms maximize the steady-state flow of profits from their constituent jobs. No single firm encompasses sufficient jobs to exceed measure 0, and without loss of generality it is assumed that each firm is comprised of a single job. Workers can be employed or unemployed and jobs can be occupied or vacant.⁹

All jobs are identical outside of a fixed wage flow, w , which is posted by the employer at the onset. The wage-offer distribution is denoted F . The flow utility of an employed worker is his wage w , and the flow utility of an unemployed worker is normalized to 0.¹⁰ All workers are identical, they discount the future exponentially at rate r , and they yield their employer a revenue product flow of $p \in (0, \infty)$ when employed in any available job.

Workers conduct on-the-job search, encountering jobs according to a Poisson arrival process with intensity λ . I focus on the steady-state only, and for simplicity I assume that the

⁹The notion of a vacant job may be intuitive, but it is elusive to define and its use in an empirical context poses a challenge - see e.g. the discussion in section 10.1 of Manning (2003). Nevertheless, vacancies are common currency in labor economics: they are a key ingredient in matching models along the lines of Mortensen and Pissarides (1994), and in light of Mortensen (1998) one could argue that they are implicit in Burdett and Mortensen (1998) as well.

¹⁰Implicitly, the wage offered by a job is the pecuniary equivalent of the flow utility from all of the job's characteristics, whether these characteristics are pecuniary or not. This point is inconsequential in theory, but requires careful attention in any empirical context.

arrival intensity is independent of a worker's employment status.¹¹ Upon encountering a job workers discover its wage, which is a draw from F . If the job is vacant, it is offered to the worker, who decides whether or not to accept it based on an acceptance rule, $A : \mathbb{R}^2 \rightarrow \{0, 1\}$, whose arguments are the offered wage and the worker's current wage. Job-worker matches separate exogenously at a positive rate, δ , and it is assumed that $\lambda/\delta \in (0, \infty)$.

Whereas the wage-offer distribution F is the distribution of wages among jobs, the distinct distribution of wages among employed workers is denoted by G . The relationship between these two distributions is simple: *the number of workers employed at any wage is equal to the number of non-vacant jobs offering that wage*. The following identity formally characterizes their relationship, and is the key assumption distinguishing the model from Burdett and Mortensen's original.

$$m(1 - u(t|A, F))G(w, t) \equiv F(w, t) - \int_{-\infty}^w v(\tilde{w}, t|A, F)dF(\tilde{w}) \quad \forall w. \quad (2.1)$$

Here, $u(t|A, F)$ is the unemployment rate at time t (conditional on the behavior of workers and firms, as summarized by A and F), and the LHS is the measure at that time of all employed workers earning a wage less than or equal to w . The RHS is the measure at time t of all non-vacant jobs offering a wage less than or equal to w . In particular, $F(w, t)$ is the measure of all jobs - vacant or not - offering wages less than or equal to w at time t , and the integral on the RHS is the contemporary measure of all vacant jobs offering wages less than or equal to w . $v(w, t|A, F)$ denotes the share of vacant jobs among jobs offering wage w at time t , given that workers follow an acceptance rule A and that firm post wages distributed F .

“Non-vacant jobs” and “occupied jobs” are synonymous, reflected by

$$o(w, t|A, F) \equiv 1 - v(w, t|A, F), \quad (2.2)$$

where $o(w, t|A, F)$ is the occupancy rate of jobs offering wage w at time t , given A and F . Substituting occupancy for non-vacancy in the identity (2.1) using (2.2) yields

$$m(1 - u(t|A, F))G(w, t) \equiv \int_{-\infty}^w o(\tilde{w}, t|A, F)dF(\tilde{w}) \quad \forall w \quad (2.3)$$

as an equivalent form.¹² Plugging $w \rightarrow \infty$ into (2.1) and (2.3) yields

$$m(1 - u(t|A, F)) = 1 - v(t|A, F) \equiv o(t|A, F), \quad (2.4)$$

¹¹The consequences of this simplification are addressed in a subsequent footnote, below.

¹²The occupancy rate of jobs offering wage w plays a similar role in this model to that of firm size (measured by employment) of firms offering wage w in Burdett and Mortensen's original ($l(w|R, F)$ therein, where workers' reservation wage R comprises their acceptance rule).

where $v(t|A, F) \equiv \int v(w, t|A, F)dF(w)$ and $o(t|A, F) \equiv \int o(w, t|A, F)dF(w)$ are the vacancy and occupancy rates for all jobs.

The setup described comprises a wage-posting game. Each worker's strategy space is the set of acceptance rules and his (instantaneous) payoff at every instant is his value flow. Each firm's strategy space is \mathbb{R} , from which it selects a wage, and its payoff is the steady-state flow of profits from its constituent job. An equilibrium of the game consists of an acceptance rule for each worker and a posted wage for each job. As in Burdett and Mortensen's original model the game has a unique equilibrium, in which all workers apply a common simple acceptance rule and firms post wages along a continuous, non-degenerate distribution.

2.2.2 Preliminaries:

Acceptance Rule:

A worker's steady-state value of employment at wage w is¹³

$$J(w) = \frac{1}{r + \lambda + \delta} \left[w + \lambda \int v(\tilde{w}|A, F) \max\{J(w), J(\tilde{w})\}dF(\tilde{w}) + \delta J(0) \right], \quad (2.5)$$

where the first bracketed term reflects the instantaneous utility flow from wage, the second the option value of switching to a higher paying job, and the third the possibility of exogenous separation. Note that encountering a higher paying job does not imply switching to one. When a worker encounters a non-vacant job he discovers its posted wage, but is only offered the job if it is vacant - hence the role of $v(\tilde{w}|A, F)$ in the expression. As all workers are identical, firms have nothing to gain by replacing the incumbent worker with a fresh hire.

The steady-state value of unemployment, J^0 , is

$$J^0 \equiv J(0) = \frac{1}{r + \lambda} \int v(\tilde{w}|A, F) \max\{J(0), J(\tilde{w})\}dF(\tilde{w}). \quad (2.6)$$

Whereas the $J(w)$ is strictly increasing in w , J^0 is independent of w , so 0 is the unique reservation wage such that $J(w) > J^0$ if and only if $w > 0$.¹⁴ The acceptance rule selected by all workers is thus

¹³To derive this equation, define $J^0 \equiv E_T[e^{-rT} \int v(\tilde{w}|A, F) \max\{J^0, J^1(\tilde{w})\}dF(\tilde{w})]$, where T is the time elapsed until the next job offer arrives and is exponentially distributed with intensity λ (i.e. the density of T is $\lambda e^{-\lambda T}$) and \tilde{w} is the wage drawn from F when the first job offer arrives. E_T denotes an expectation with respect to the random variable T . A useful primer on deriving Bellman equations of this type and on Poisson processes can be found in the appendices of Zenou (2009).

¹⁴That $J(w)$ is increasing in w is intuitive, as it raises the instantaneous flow utility without affecting the worker's opportunities, and it is often taken as a standard result without proof. Showing it formally is not straightforward, but can be achieved - for example - by applying proposition 3 of Smith and McCardle (2002) to the property of monotonicity.

$$A(\hat{w}|w) = \mathbf{1} \{ \hat{w} > \max \{ w, 0 \} \},$$

i.e. “accept any job with a positive wage if unemployed, and accept any job associated with a wage increase otherwise.” For notational brevity, conditioning on workers’ acceptance rule A is omitted in what follows.¹⁵

Steady-State Unemployment and Vacancy Rates:

The steady-state unemployment, vacancy and occupancy rates are fully determined by the exogenous parameters m , λ and δ . To see this, note that the steady-state flows in and out of unemployment are $m(1 - u) \cdot \delta$ and $mu \cdot \lambda v$, respectively, where the conditioning of u and v on F has already been omitted. Equating the two and using $m(1 - u) = 1 - v$ from (2.4) yields

$$\frac{1 - u}{u} = \frac{\lambda}{\delta} (1 - m(1 - u)). \quad (2.7)$$

That the LHS is decreasing in u and the RHS is increasing in it implies the existence of a unique steady-state unemployment rate u , which depends only on m , λ and δ , and (2.4) implies the same for the steady-state vacancy and occupancy rates, v and o .

Note that while this argument renders v and o independent of F , the steady-state vacancy and occupancy rates for jobs offering a specified wage ($v(w|F)$ and $o(w|F)$) may still depend on F .

2.2.3 Steady-State Job Occupancy:

Given an initial allocation of workers to firms and their derived acceptance rule, the number of employed workers receiving a wage no greater than w at time t , $m(1 - u(t))G(w, t)$, can be calculated. Its time derivative can be written as

$$\frac{d}{dt} (m(1 - u(t))G(w, t)) = \lambda \int_0^w v(\tilde{w}|F) dF(\tilde{w}) mu(t) - \left[\delta + \lambda \int_w^\infty v(\tilde{w}|F) dF(\tilde{w}) \right] m(1 - u(t))G(w, t).$$

Consequently, the steady state distribution of wages earned by employed workers is

$$G(w) = \frac{kv_0^w}{1 + kv_w^\infty} \cdot \frac{u}{1 - u}, \quad (2.8)$$

where $k \equiv \lambda/\delta$ and $v_a^b \equiv \int_a^b v(\tilde{w}|F) dF(\tilde{w})$ is a convenient abbreviated notation.

¹⁵That workers’ reservation wage is not greater than 0 is driven purely by the simplifying assumption that employed and unemployed workers face the same job offer arrival rate. In contrast, Burdett and Mortensen (1998) find that the reservation wage, R , is such that $R > b \geq 0$, where b is the non-negative flow value of unemployment. This discrepancy arises from their assumption that job opportunities arrive more frequently during unemployment whereas I have assumed a rate of arrival which is independent of state. In their model workers refuse an offer if the wage does not compensate for the reduced rate of the future flow of offers, whereas here accepting a job implies no such reduction, so any job offer with positive wage is accepted.

The steady state occupancy rate of jobs offering wage w can be expressed as

$$o(w|F) = \lim_{\epsilon \rightarrow 0} \frac{G(w) - G(w - \epsilon)}{F(w) - F(w - \epsilon)} m(1 - u). \quad (2.9)$$

Following Burdett and Mortensen (1998), define the fraction, or mass, of jobs offering wage w as

$$\nu(w) \equiv F(w) - F(w^-) = \lim_{\epsilon \rightarrow 0} F(w) - F(w - \epsilon),$$

and note that together with $u = (1 + kv)^{-1}$ from (2.4), (2.8) implies¹⁶

$$\lim_{\epsilon \rightarrow 0} G(w) - G(w - \epsilon) = \frac{uk}{1 - u} \left(\frac{v_0^w}{1 + kv_w^\infty} - \frac{v_0^{w^-}}{1 + kv_{w^-}^\infty} \right) = \frac{k\nu(w)v(w|F)}{(1 - u)(1 + kv_w^\infty)(1 + kv_{w^-}^\infty)}. \quad (2.10)$$

Substituting the above into (2.9) and re-arranging yields

$$o(w|F) = \frac{mk}{(1 + kv_w^\infty)(1 + kv_{w^-}^\infty) + mk} \in (0, 1), \quad (2.11)$$

Combined with (2.10), equation (2.11) implies that G is continuous if and only if F is continuous (i.e. $\nu(w) = 0 \quad \forall w$).¹⁷

$o(w|R, F)$ is the steady state occupancy rate of jobs offering wage w , and (2.11) implies the following properties:

1. $o(w|R, F)$ is increasing in w .
2. $o(w|R, F)$ is strictly increasing in w on the support of F and is a constant on any connected interval off the support of F .
3. $o(w|R, F)$ is continuous if and only if F is continuous (and hence G , too).¹⁸

¹⁶The second equality is derived by obtaining a common denominator as follows $\frac{v_0^w}{1 + kv_w^\infty} - \frac{v_0^{w^-}}{1 + kv_{w^-}^\infty} = \frac{v_0^w(1 + kv_{w^-}^\infty) - v_0^{w^-}(1 + kv_w^\infty)}{(1 + kv_w^\infty)(1 + kv_{w^-}^\infty)}$ and applying the identity $ab - cd = (a - c)b - c(d - b)$ to yield that $v_0^w(1 + kv_{w^-}^\infty) - v_0^{w^-}(1 + kv_w^\infty) = (v_0^w - v_0^{w^-})(1 + kv_{w^-}^\infty) - v_0^{w^-}k(v_w^\infty - v_{w^-}^\infty) = \nu(w)v(w|F)(1 + k(v_0^{w^-} + v_{w^-}^\infty)) = \nu(w)v(w|F)(1 + kv)$, where $v \equiv v_0^\infty$.

¹⁷To see this, note first that if F is continuous then $\nu(w) = 0$ for all w , so by (2.10) G must be continuous. In the other direction, if $G(w)$ is continuous then the RHS of (2.10) must equal zero, but because $k \in (0, \infty)$, $v(w|R, F) \equiv 1 - o(w|R, F) \in (0, 1)$ and the denominator is necessarily finite ($u \in (0, 1)$, $k \in (0, \infty)$ and $v_a^b \in (0, 1) \quad \forall a, b$) this implies that $\nu(w) = 0$, so F must be continuous.

¹⁸The link between the continuity of v_a^b and that of F and G is provided by (2.1), which can be re-written in steady-state using the v_a^b notation as $v_0^w \equiv F(w) - m(1 - u)G(w)$.

2.2.4 Equilibrium Wage Dispersion:

For each job, firms post a wage so as to maximize their steady-state flow of profits from that job,

$$\pi = \max_w (p - w) \cdot o(w|F). \quad (2.12)$$

In equilibrium F must be such that

$$\begin{aligned} (p - w) \cdot o(w|F) &= \pi \text{ for all } w \text{ on support of } F \\ (p - w) \cdot o(w|F) &\leq \pi \text{ otherwise.} \end{aligned} \quad (2.13)$$

Denote the infimum and supremum of the support of an equilibrium F (supposing that one exists) by \underline{w} and \bar{w} . Note that no employer will offer a wage $w < 0$ because he would have a permanent vacancy which is costly to maintain, so I consider only $\underline{w} \geq 0$.

The following argument rules out continuous wage offer distributions, and is worded as closely as possible to Burdett and Mortensen's original text. $o(w|F)$ is discontinuous at $w = \hat{w}$ if and only if \hat{w} is a mass point of F and $\hat{w} \geq 0$. This implies that any employer offering a wage slightly greater than \hat{w} , a mass point where $0 \leq \hat{w} < p$, has a significantly larger steady state occupancy rate for the job and only a slightly smaller profit per unit of occupancy¹⁹ than an employer offering \hat{w} , as $(p - w)$ is continuous in w .²⁰ Hence, any wage just above \hat{w} yields a greater profit. If there were a mass of F at $\hat{w} \geq p$, all jobs offering such a wage yield non-positive profit. However, any job offering a wage slightly lower than p yields a strictly positive profit as it maintains a positive steady state occupancy rate. In short, offering a wage equal to a mass point \hat{w} cannot be profit maximizing in the sense of (2.12).

As non-continuous offer distributions have been ruled out, (2.11) implies that for \underline{w} ,

$$o(\underline{w}|F) = \frac{mk}{(1 + kv)^2 + mk} \in (0, 1) \quad (2.14)$$

where the equality follows from $v_{\underline{w}}^{\infty} = v$ by the definition of \underline{w} . Thus $o(\underline{w}|F)$ depends only on the parameters m , λ and δ (as do u and v) and is independent of F . Note that the occupancy rate at wage $w = \underline{w}$ is independent of \underline{w} , too, implying that the lowest paying job in the market will yield the maximum profit flow if and only if $\underline{w} = 0$.

In equilibrium, every offer must yield the same steady state profit, which equals

$$\pi = p \cdot \frac{mk}{(1 + kv)^2 + mk} = (p - w) \cdot o(w|F) \quad (2.15)$$

¹⁹For concreteness, "a unit of occupancy" can be measured, e.g., in terms of days that a job is non-vacant per year.

²⁰Algebraically, if $\hat{w}^+ \equiv \lim_{\epsilon \rightarrow 0} \hat{w} + \epsilon$ then $(p - \hat{w}^+) \cdot o(\hat{w}^+|F) - (p - \hat{w}) \cdot o(\hat{w}|F) = (p - \hat{w}) \cdot (o(\hat{w}^+|F) - o(\hat{w}|F)) - (\hat{w}^+ - \hat{w}) \cdot o(\hat{w}^+|F) > 0$ for $\hat{w} < p$, because $o(\hat{w}^+|F) - o(\hat{w}|F) > 0$ if \hat{w} is a mass point, whereas $\hat{w}^+ - \hat{w} \rightarrow 0$ and $o(\hat{w}^+|F) \geq 0$.

for all w on the support of F , yielding²¹

$$o(w|F) = \frac{p}{p-w} \cdot \frac{mk}{(1+kv)^2 + mk}. \quad (2.16)$$

Equating (2.11) with (2.16) and re-arranging yields

$$\frac{p}{p-w} = \frac{(1+kv)^2 + k}{(1+kv_w^\infty)^2 + k} \in (0, 1), \quad (2.17)$$

which, given the parameters λ , δ and m , uniquely determines $v_w^\infty \equiv 1 - v_0^w$ for all w on the support of F (recall from section (2.2.2) that both u and v are fully determined by these three parameters, too). Substituting (2.8) into (2.1) yields

$$F(w) = v_0^w \frac{1 + k(mu + v_w^\infty)}{1 + kv_w^\infty}, \quad (2.18)$$

implying that F is unique.²²

As $v_{\bar{w}}^\infty$ must equal 0, it follows from (2.17) that

$$p - \bar{w} = p \cdot \frac{1+k}{(1+kv)^2 + k} > 0. \quad (2.19)$$

As the occupancy rate and profits per unit of occupancy are both positive for a firm offering wage \bar{w} , and equilibrium profit flows are equal for all firms offering wage on the support of F , it must be that $\pi > 0$ for all firms.²³

Finally, to complete the proof that the acceptance rule obtained in section (2.2.2) coupled with the wage offer distribution F constitute the unique equilibrium of the wage posting game, it must be shown that no wage off the support of F yields an employer a profit flow greater than π . A job offering wage $w < 0$ will have a vacancy rate of 1, and so yields zero profits. A job offering wage $w > \bar{w}$ has the same vacancy rate as a job offering \bar{w} , because $v(w|F)$ is constant on any connected interval off the support of F , but has a lower profit flow because $w > \bar{w}$.

²¹(2.16) implies that $o(w|F) \equiv o(w)$ and $v(w|F) \equiv v(w)$, i.e. that the occupancy and vacancy rates at each wage level w do not depend on F , but only on the exogenous parameters m , λ, δ and p . However, this does *not* imply that their integral v_0^w (over strictly less than the complete support of F) is independent of F , because $v_0^w \equiv \int_0^w v(\tilde{w}|R)dF(\tilde{w})$.

²²That $F(\bar{w}) = 1$ and $v_{\bar{w}}^\infty = 0$ implies that $v = (1 + mku)^{-1}$.

²³In line with footnote 6 of Burdett and Mortensen (1998), at this stage one can endogenize the measure of firms by assuming the existence of a positive fixed cost $c > 0$ and invoking free entry of firms, so that $\pi = p \cdot \frac{mk}{(1+kv)^2 + mk} - c = 0$.

2.3 Conclusion

A known shortcoming of Burdett and Mortensen's result of pure equilibrium wage dispersion is that it hinges upon random matching, whereby workers are equally likely to encounter any firm, regardless of its size. The casual observation that larger employers attract more job applicants than small ones contrasts starkly with random matching. On the other hand it is well known that with balanced matching, whereby a worker's probability of encountering a firm is proportional to its employment, equilibrium wage dispersion degenerates to a single mass at the competitive wage. The unrealistic nature of random matching with firms, coupled with the lack of wage dispersion under balanced matching, casts doubt on the empirical relevance of Burdett and Mortensen's result. The stakes are raised further by the fact that Burdett and Mortensen's result provides the theoretical underpinning for the competitive monopsony literature.

This paper presents a modification of Burdett and Mortensen's model, in which the key result of pure equilibrium wage dispersion holds with balanced matching. The matching technology is balanced in the sense that workers encounter every firm with probability that is proportional to the number of jobs at the firm (but not with employment at the firm). The key insight is that instead of addressing the matching of workers and firms, a variation of the original model can be used to address the matching of workers and jobs, and that random matching with jobs amounts to a form of balanced matching with firms. In particular, random matching with jobs implies that a worker's probability of encountering a firm is proportional to the number of jobs at that firm (which differs from the number of employees). The upshot, however, is that the casual observation that larger employers attract more job applicants than small ones is *not* at odds with balanced matching in terms of jobs, so this observation no longer challenges the empirical relevance of Burdett and Mortensen's result. This paper removes an otherwise persistent source of doubt in the empirical relevance of pure equilibrium wage dispersion, and by doing so provides a more solid theoretical foundation for the competitive monopsony literature.

Chapter 3

The Unraveling of Prolonged Stability: The Fall of the Old Kingdom in Ancient Egypt

3.1 Introduction

The Old Kingdom was the first great age of Ancient Egypt. Together with the preceding Early Dynastic Period it spanned most of the third millennium BCE and saw the construction of such grand monumental architecture as the Great Pyramids of Giza and the Sphinx.¹ The Pharaoh's of Egypt were absolute rulers, believed to be deities, and the period was one of extraordinary stability and uniformity, not just politically but also in terms of cultural, social and economic development. Remarkably little change took place over this immense stretch of time, compared with other historical periods. In the absence of any marked external events affecting Egypt, the Old Kingdom's downfall was the outcome of an internal process that unfolded over centuries, whereby the balance of power gradually shifted from the royal court to an emerging provincial elite. This narrative raises two questions: What underlying currents caused the balance of power to shift, and why did the Pharaoh's not reverse the shift unilaterally, given their absolute power?

I maintain that the process was born out of the Egyptian state's policy. The emergence of the provincial elite is closely related to the relationship between the royal court, i.e. the state, and its local tax-collecting proxies. Local administrators fall among the proxies and so do the emerging provincial elite, either directly or indirectly through the elite's affiliation with the local administrators. In order to efficiently guarantee the proxies' compliance, the court needed to permit them to keep a certain share of the taxes they collected. The size of this share simultaneously depended on the existing balance of power and affected the future balance of power, yielding a dynamic system. Viewing the balance of power in terms of the court's and the proxies' resources implies that the state could control the balance of power by carefully managing its expenditure.²

¹The two periods jointly span the years 3000 to 2181 BCE. The distinction between them is essentially a scholarly one, without any real discontinuity. See Malek (2000) page 89.

²The link between the balance of power and the ratio of the court's and the proxies' resources is formalized in the model that follows by way of a *fighting function*.

Despite the simplicity of the system, the chain of causality was virtually impossible for the royal court to perceive, because the ultimate consequences of .³ Consequently, I maintain that the royal court continuously used more resources than required to keep the balance of power stable, setting the economy on a divergent path that empowered the provincial elite at the state's expense. The underlying strategic interaction between the royal court and its local proxies was present from the onset of the Early Dynastic Period, when the former first levied taxes from outside its immediate vicinity, and many centuries before a provincial elite ever emerged. Moreover, while my analysis in what follows is specific to Ancient Egypt and to the era in question, the model that I have implicitly referred to is more widely applicable.

The gradual nature of the Old Kingdom's downfall and its roots in political developments are well described in the literature.⁴ The authority of the royal court is said to have declined in the later years of the Old Kingdom as local administrators and the associated provincial elite gained power and increasingly rivaled the state. This trend occurred because an increasing amount of resources was channeled to these parties by the state, leading to its relative impoverishment and hence its weakening. Additional flows of resources, earmarked for the upkeep of the temples and tombs which are the hallmark of Ancient Egyptian religion only increased the strain on the state's resources.⁵

From such descriptions it appears that once the shift in the balance of power had progressed sufficiently, it was quite apparent for a very long period of time. How is it possible then, that not a single Pharaoh attempted to divert the flow of resources back to the state? Surely an absolute king (and a deity, no less!) could issue an immediate decree on the matter?⁶ For the descriptions in the literature to comprise a more robust theory of the fall of the Old Kingdom, they ought to provide a convincing answer to this question, which I address via the model implicitly outlined above.⁷

In the model, the state incentivizes proxies to act as its loyal agents by allowing them to maintain a share of the tax revenue they collect. It sets the level of this share such that the proxies are indifferent between complying with the state and confronting it militarily. Thus, the proxies' compliance is guaranteed and with it the royal court's income, and military conflict between the state and its proxies is kept off of the equilibrium path. In this scenario,

³This statement is developed in more depth in section 3.5.

⁴See Grimal (1992) pages 88-93 and also Malek (2000) pages 114-117.

⁵Ibid.

⁶The flow of resources away from the royal court was accompanied by immaterial flows of power, too. Egypt was divided into local districts called *nomes* that were run by state-appointed administrators called *nomarchs*. Grimal (1992) (page 92) and Seidlmayer (2000) (pages 120-121) mention that during the 5th and 6th Dynasties (the last two Dynasties of the Old Kingdom) hereditary transfer of the nomarch position gradually replaced the previous method of appointment by the state, allowing nomarchs to establish a stronger hold over their districts. Eradicating this norm seems like an easy enough task for a Pharaoh, so why was it not abolished?

⁷Absolute religious ideology may explain the reticence to pull resources away from some uses, such as temples and funerary cults, but it can hardly explain the whole picture. Moreover, from a cynic's point of view it is hard to believe that a powerful Pharaoh would voluntarily permit religious ideology to develop in a way that sheds him of power, let alone entire Dynasties of Pharaohs surrounded by intelligent, savvy advisors.

the royal court cannot, at least in the short-run, decrease the share of resources it affords the local proxies.

In the long run, however, the royal court can influence the proxy's share via its policy. In particular, it can control the state's expenses so as to alter the ratio of resources held by the royal court and its proxies, which implies the balance of power.

Nevertheless, it is highly unlikely that members of the royal court were historically aware of the dynamic system governing the balance of power, and therefore I pay careful attention to distinguishing which decisions can be modeled by optimizing behavior and which cannot. As discussed in more detail in section 3.5, the Pharaohs and their aides were almost certainly incapable of detecting the long-term dynamic relationship between state expenditure and the subsequent share of resources afforded to proxies. Due to this inability, I argue that the state failed to control its expenses so as to maintain a stable balance of power between itself and its proxies, and that the Egyptian economy consequently followed a divergent path, whereby the state was impoverished as it afforded the provincial elite an ever-increasing share of its tax revenue. Finally, I argue that the ratio of resources was such that the state could no longer credibly threaten to confront the proxies militarily, and it was effectively rendered powerless.

The period following the Old Kingdom, known as the First Intermediate Period, lasted for almost two centuries and is often perceived as a dark age. During this period the state dissipated into several polities and the country was struck by strife and bouts of famine. Very little is known with any certainty about the nature of the transition into the First Intermediate Period.⁸

Throughout the periods involved, agricultural output in Egypt depended heavily on local administrators, who were responsible for developing and maintaining local irrigation systems composed of floodwater dams and canals.⁹ The state, on the other hand, managed the storage of grain. The main determinant of agricultural output was the level of annual Nile floods that served to irrigate the land, and which was inherently volatile. By storing grain the state could smooth the food supply over time so as to cushion the hazardous consequences of this volatility, and it could do the same over space by transferring food from areas of plenty to areas of shortage.¹⁰ The state also provided Egypt with external security.¹¹

In the absence of the state the food supply could not be smoothed over space, and the capacity to smooth it over time was at best on par with what it had been earlier. Thus, it is no mystery how the collapse of the Old Kingdom may have led to famine, while internal

⁸Despite the certainty with which Bell (1971) states that Egypt fell into anarchy with seeming suddenness, it is uncertain whether the deterioration of the state was gradual or if abrupt events occurred. See Grimal (1992) pages 137-140, and Seidlmayer (2000), pages 119-122.

⁹see Malek (2000) page 102, Butzer (1976), page 43, Butzer (1984), page 104, and Hassan (1997), page 56, as well as others.

¹⁰Although its origin is likely to be from a later period (see Knohl (2008)), the biblical story of Joseph and Pharaoh's dreams is a perfect example of the state preventing famine this way. In the story, Joseph oversees the royal bureaucracy as it stocks grain over seven years of plenty and then cushions the blow of the seven lean years that followed (Genesis, cap. 41). The state's role in inter-temporally smoothing the food supply is commonly acknowledged.

¹¹See Malek (2000), page 102.

strife may have preceded the famine, followed it, or both.¹²

Section 3.2 provides a set of stylized facts that portray the era in question. A model introduced in section 3.3 is then used in section 3.4 to analyze the stylized facts. Section 3.5 provides justification for the way various agents' decisions are modeled and section 3.6 concludes.

3.2 Stylized Facts

The facts:

- 1) *Beginning with the formation of the state, the region from which taxes were levied gradually expanded and the tax collection system solidified. The process came to an end roughly during the 3rd Dynasty when the tax system encompassed all of Egypt. Most of the growth in the state's income occurred up until the 3rd Dynasty with more modest growth, if any, taking place afterwards.*
- 2) *The peak of state expenditure occurred during the 4th Dynasty, and is marked by the construction of the Great Pyramids at Giza. During the 5th and 6th Dynasties state expenditure was substantially reduced.*
- 3) *The provincial elite was negligible in size and wealth before the 5th Dynasty. During the 5th and 6th Dynasties it grew substantially larger and wealthier, at the expense of the state.*
- 4) *From the end of the 6th Dynasty the state lost de-facto authority over the country, although not necessarily by way of any military conflict.*

Basis for the facts:

- 1) *Beginning with the formation of the state, the region from which taxes were levied gradually expanded and the tax collection system solidified. The process came to an end roughly during the 3rd Dynasty when the tax system encompassed all of Egypt. Most of the growth in the state's income occurred up until the 3rd Dynasty with more modest growth, if any, taking place afterwards.*

The unification took place roughly during the first century of the third millennium BCE, marking the beginning of the Early Dynastic Period. The Old Kingdom refers to the period from 2686 BCE to circa 2150 BCE, and is formally distinguished from the Early Dynastic

¹²An alternative theory explaining the fall of the Old Kingdom is that of Bell (1971), who hypothesizes that collapse occurred due to a catastrophic sequence of unusually low Nile floods, which could not be mitigated by grain storage (For a thorough description of Nile floodplain agriculture and the ramifications of too low or too high a flood, see Butzer (1976), pages 39-56). Implicitly, the low Nile floods induced famine, which in turn brought about strife. However, the only relevant evidence of unusually low Nile floods around the time of the collapse is comprised of inferences from contemporary texts describing famine and strife (there is much physical evidence that Nile floods were lower during the later *centuries* of the Old Kingdom. See Butzer (1976), pages 30-33). While there is no ruling out the climatic theory of Bell (1971), I agree with others who argue that it is redundant in explaining the fall of the Old Kingdom. See Butzer (1984), pages 101 and 106, and Seidlmayer (2000), page 129.

Period by the reign of King Djoser (3rd Dynasty), whose construction projects foreshadowed the subsequent wonders of the 4th Dynasty.¹³ Referring to the “enormous control exercised by the Crown” with respect to the step-pyramid complex of King Djoser, Bard (2000) writes the following:¹⁴

“...such power must have been developed incrementally throughout the 1st and 2nd Dynasties, following the unification of the large territorial state... The Early Dynastic Period was a time of consolidation of the enormous gains of unification... when a state bureaucracy was successfully organized and expanded to bring the entire country under royal control. This was done through taxation, to support the Crown and its projects on a grand scale...”

The “enormous gains of unification” to which Bard (2000) refers are the focus of Allen (1997), who argues the following:

“Successful states in the ancient world depended on the ability of elites to extract a surplus from farmers and other producers. This ability was greatest when the population was immobile. The success of the Pharaohs was due to the geography of Egypt - the deserts bordering the Nile meant that habitation was confined to the valley. Farmers could flee tax or rent collectors only along the river. The population control problem was, thus, simpler than elsewhere and was the reason a unified state was created and lasted for millennia.”¹⁵

During the formative period of the tax system that preceded the 3rd Dynasty the state’s income grew both on the extensive margin, due to expansion of the taxed region, and on the intensive margin due to more thorough extraction. From the 3rd Dynasty onwards, increasing the state’s income was possible only on the intensive margin.¹⁶ Therefore, and in light of the quotes above, it is reasonable to assume that most of the growth in the state’s income occurred up until the 3rd Dynasty, with more only more modest growth, if any, taking place afterwards.

2) *The peak of state expenditure occurred during the 4th Dynasty, and is marked by the construction of the Great Pyramids at Giza. During the 5th and 6th Dynasties state expenditure was substantially reduced.*

This stylized fact is inferred from viewing the construction of monumental architecture, and in particular of royal tombs, as a proxy for state expenditure. There are no marked periods of increased warfare during the Early Dynastic Period or the Old Kingdom, nor do

¹³King Djoser is known for constructing the first step-pyramid which provides the link between the later, true pyramids, and the preceding mastaba tombs. The given classification into periods was imposed by 19th century historians and was not accompanied by discontinuous changes on the ground - see Malek (2000), page 89.

¹⁴See Bard (2000), page 87.

¹⁵See Allen (1997), abstract. His argument is related to the “circumscription theory” of Bard and Carneiro (1989).

¹⁶Explaining what limited the geographic expansion of the Egyptian state is beyond the scope of this paper.

there appear to be any other grounds to suspect that the share of monumental construction in the total state expenditure fluctuated significantly during the period, so at least as a first pass this seems to be a valid proxy.

The step-pyramid of King Djoser, in the 3rd Dynasty, overshadowed all royal construction that preceded it and was itself overshadowed by the Great Pyramids built by the 4th Dynasty kings. The later pyramids of the 5th and 6th Dynasties are significantly smaller.¹⁷ Referring to the reign of the King Pepi II (6th Dynasty), the last of the Old Kingdom's kings, Kemp (1983) states that "the ability of the court to build on a truly monumental scale seems to have gone altogether".¹⁸

A potential objection to this argument would be that the size of a king's pyramid is a function of how long he reigned. Kings Khufu, Khefren and Menkaura (4th Dynasty), builders of the three great pyramids at Giza, ruled for 23, 26 and 18 years, respectively. While some of the 5th and 6th dynasty kings were not so lucky, two of the 5th Dynasty kings ruled for over 30 years and four of them ruled for over 20 years. During the 6th dynasty three kings ruled in excess of 30 years.

Another potential objection is that construction technology improved due to learning over time, so that a larger tomb at a later date could imply a smaller expenditure than a smaller one built earlier. While this argument may support the claim that state expenditure was greater at the time of King Djoser than it was at the height of the 4th Dynasty, it only supports further the claim that state expenditure was lower during the 5th and 6th Dynasties than it was at the height of the 4th.

3) *The provincial elite was negligible in size and wealth before the 5th Dynasty. During the 5th and 6th Dynasties it grew substantially larger and wealthier, at the expense of the state.*

According to Seidlmayer (2000), "[u]ntil well into the 5th Dynasty, nothing of the cultural achievements that attest to the grandeur of the Old Kingdom was to be seen outside the Memphite region" (Memphis being the capital of the Old Kingdom). He continues:¹⁹

"However, a profound change in the system began to appear in the 5th Dynasty and was completely in place by the end of the 6th... Originally, economic resources were concentrated at the royal residence and redistributed to the beneficiaries by the central administration. Now, however, the nobles residing in the provinces were able to gain direct access to the products of the country... The provincial aristocracy was eager to ensure that its way of life was on a par with the style of the royal court. This is evident in the decorated monumental tombs that began to appear in the cemeteries of the regional centres throughout the country... These tombs, however, are only the tip of the iceberg; in fact, the various provincial elites and their staff acted as separate centres within the political organization, sustaining specialist professionals and keeping a growing amount of

¹⁷For a sequential list of pyramid base sizes by king and dynasty, see Grimal (1992) pages 116-119.

¹⁸See Kemp (1983), page 112.

¹⁹Seidlmayer (2000), pages 120-121. See also Kemp (2006), page 309. For a reflection of the growing wealth of the provincial elite on the relative importance of local deities, see Malek (2000), page 111.

local produce for use within the provinces themselves (rather than allowing it to be exploited by the royal court)..."

The process whereby the provincial elite gained power was a very gradual one, and was a result of the state's financial policy. Malek (2000) describes how state officials were typically remunerated in the Old Kingdom:²⁰

These officials were remunerated for their services in several different ways, but the most significant was an *ex officio* lease of state (royal) land, usually estates settled with their cultivators. Such estates produced practically all that their personnel needed... and the *ex officio* remuneration was their surplus produce. This land reverted, at least in theory, to the king after the official's term of office expired and so could be assigned as remuneration of another official. In an economic system that did not know money it was a very effective way of paying salaries of officials, but it also represented a significant erosion of the king's resources."

The remuneration of officials was not limited to the realm of material compensation. Grimal (1992) describes a great number of new official titles appearing during the 5th and 6th Dynasties that bore no real meaning and were clearly granted to satisfy political needs.²¹

An even greater drain on the royal treasury than ex-officio remunerations were pious foundations. The latter were funds, usually established by a donation of property. Their role was to guarantee perpetual maintenance of religious sites and the cults that accompanied them, such as temples, the tombs of royalty and the tombs of growing numbers of other wealthy individuals.²² Referring to the growing number of funerary endowments of provincial officials for which the state was responsible, Grimal (1992) states that:²³

"...this principle... contained the seeds of the state's destruction, in that it favored the dissemination of wealth and the gradual - and ultimately irreversible - impoverishment of the king. The profits enjoyed by the recipients of these concessions acted as a drain on the economy, since they effectively lay outside the redistribution network provided by the state. But even this was not the most important effect. The most serious problem was the social mechanism that was created by these concessions: ...the recipients attempted to acquire not only wealth but also the prerogatives associated with royal property."

The historians' quotations presented clearly identify the symptoms of the process by which the state's power diminished in terms of economic resources. The contribution of this paper is in analyzing the mechanism causing these symptoms.

4) *From the end of the 6th Dynasty the state lost de-facto authority over the country, although not necessarily by way of any military conflict.*

²⁰Malek (2000), pages 104-105.

²¹See Grimal (1992), page 90, as well as Kemp (1983), page 80.

²²See Kemp (1983), page 85.

²³See Grimal (1992), pages 92-93.

Describing the end of the Old Kingdom following the 6th Dynasty, Malek (2000) is very straightforward: “Centralized government all but ceased to exist, and the advantages of a unified state were lost.”²⁴

There is no material or textual evidence that the end of the Old Kingdom came about through military conflict. Rather, it may have been the conclusion of a gradual process in which the royal court’s power diminished and it fell into political irrelevance for much of the country. There appears to be no archaeological evidence for a social revolution or a civil war at the end of the Old Kingdom.²⁵

Whether or not the remnants of the Old Kingdom at the end of the 6th Dynasty were destroyed in armed conflict is inconsequential. Both possibilities will prove to be consistent with the theory presented below.

The following section presents a model with which these stylized facts can be analyzed.

3.3 Model

Consider the king, the royal court and the state to be interchangeable terms describing one player, whose marginal utility from an additional unit of the single existing type of good, grain, is always positive.

The royal court extracts surplus product from the economy by way of local proxies. The proxies may be nomarchs (administrators of local districts called *nomes*), mayors, temple priests, combinations of these, or perhaps bearers of different posts or informal roles.

Consider then a proxy for the state, i , who presides over a domain with product Y_i and extracts its surplus. He transfers to the state all but a share of the extracted surplus, which he retains for himself. This share is referred to as the *proxy share* and is labeled M_i . His transfer combines with that of other proxies to comprise the state’s tax revenue.²⁶

The proxy’s share is retained under the auspices of the state, either formally or informally by way of the court turning a blind eye. The product of ex-officio remunerations and pious foundations falls under the former category, whereas embezzlement of tax revenue by local officials falls under the latter.

The proxy’s share is an essential incentive for him to fulfil his duty. If he complies with his role by transferring all but his share of surplus to the state he retains:

$$M_i \cdot \tau Y_i \tag{3.1}$$

where τ is the share of product that is extracted as surplus, i.e. the tax rate. According to Brewer and Teeter (2007) it appears that approximately a tithe of the total harvest was appropriated as tax.²⁷ For simplicity, assume that tax is either levied at a fixed rate, τ , or

²⁴See Malek (2000), page 117.

²⁵See Seidlmayer (2000), page 140.

²⁶The royal court may extract some surplus directly from its vicinity. This is inconsequential to the analysis.

²⁷Brewer and Teeter (2007), pages 94-96.

not levied at all. Consequently, the state's tax revenue is:

$$\sum_{j \in J} (1 - M_j) \cdot \tau Y_j \quad (3.2)$$

where J is the set of all proxies.

Alternatively, if the proxy chooses to revolt by not transferring tax revenue he may retain all, some or none of the extracted surplus from his domain. The amount which he retains is a random variable that depends on whether or not the state chooses to confront him and on the outcome and the cost of such a confrontation if it takes place.²⁸ Designate $P_i \in [0, \infty)$ to be the share of extracted surplus retained by the proxy if he revolts, so that in a particular realization he retains:

$$P_i \cdot \tau Y_i$$

and in expectancy he retains:

$$E(P_i) \cdot \tau Y_i \quad (3.3)$$

P_i can be greater than one, because a military confrontation may potentially result in grand victory for the proxy, where he is left with more resources than he extracted from his domain to begin with. On the other hand, it is also possible that $P_i = 0$ if the state crushes the proxy.

Regardless of whether a military confrontation is realized, assume that $E(P_i)$ is a positive function of the proxy's potential military ability versus the royal court, and refer to it as a *fighting function*.²⁹ Assume further that the ratio of resources available to the parties at the time of confrontation is a valid proxy for their relative, potential military ability. Doing so is especially appealing in the context of ancient Egypt, in which military force was generally hired and even the state itself never maintained a standing army.³⁰

Should the proxy choose to revolt, the resources available to him are those he extracts from his domain in the current period, τY_i , and whatever remains of his share from previous periods, which I shall assume to be zero for simplicity.³¹ The royal court's resources are the current period's tax revenue save that of the revolting proxy, and what remains of the previous period's tax revenue. The length of time periods is normalized so that resources

²⁸The cost of conflict is in terms of the grain used to hire military force.

²⁹To be more precise: let x be a parameter positively describing the proxy's military ability versus the royal court such that $P_i(x) \in [0, 1]$ is a random variable with cdf $F(P_i|x)$. The assumption can be re-stated as $F(P_i|x') \leq F(P_i|x'') \quad \forall x' \geq x''$ (first-order stochastic dominance), which implies $\frac{\partial}{\partial x} E(P_i(x)) \geq 0$.

³⁰See Brewer and Teeter (2007), page 74. Also, according to Seidlmayer (2000) much of Egypt's external warfare was conducted by local administrators who hired temporary force. See page 130.

³¹Without this simplifying assumption, the proxy's resources would depend positively on the proxy's share in the previous period, lending further support to the results developed below.

can be stored for no more than one period.³² Thus:

$$E(P_{i,t}) = E\left(P\left(\frac{X_{i,t}}{X_{c,t}}; Z_t\right)\right) \quad \text{s.t.} \quad E(P'(\cdot|Z)) > 0 \quad (3.4)$$

where Z_t is a vector of all exogenous parameters affecting the outcome at time t , such as the location and characteristics of potential sites of battle, the parties' military and political strategies, the players' temperaments, talents and so forth. $X_{i,t}$ and $X_{c,t}$ are the resources of the proxy and the court in period t , respectively, conditional on the proxy choosing to revolt:

$$X_{i,t} = \tau Y_i \quad (3.5)$$

$$X_{c,t} = \sum_{j \in J} [(1 - M_{j,t}) + (1 - \alpha_{t-1})(1 - M_{j,t-1})] \tau Y_j - (1 - M_{i,t}) \tau Y_i \quad (3.6)$$

where $(1 - \alpha_{t-1}) \in (0, 1)$ is the proportion of resources that remain stored from the previous period and is controlled by the state in period $t - 1$.

As will become evident, the results of the model depend crucially on the dynamic link between the royal court's present resources, $X_{c,t}$, and the proxy's share in the previous period, $M_{i,t-1}$. This link is intact as long as the court's resources are partly comprised of the previous period's tax revenue that has been stored by the state. Therefore, I shall assume $(1 - \alpha_t) > 0$ for all t and express $X_{c,t}$ as a function of the lagged proxy's share, $M_{i,t-1}$:

$$X_{c,t}(M_{i,t-1}) = K - (1 - \alpha_{t-1}) M_{i,t-1} \tau Y_i$$

where K follows from (3.6).

Note that this assumption is far from arbitrary and is most likely to have held in reality. To see this, suppose that the royal court's resources depended entirely on the present period's tax revenue. If this were the case, then any coalition of proxies which comprised a sufficient proportion of the total could have revolted with reasonable odds of success at any moment. This would have been an inherently unstable situation and the state could not have persisted over any length of time under these circumstances, let alone for the many centuries that it did (during which it relied heavily on proxies).³³

³²This choice of normalization makes the model more tractable, and in particular more amenable to graphic presentation. Note that a normalized period may be longer than the intervals at which tax revenue is transferred to the royal court (these were probably annual in ancient Egypt). If this is the case then the resources actually in the hands of a revolting proxy are only the part of the resources extracted from his domain in the latest tax transfer interval, which he has not yet transferred to the court and can be expressed as $\frac{1}{\gamma} \tau Y_i$, where γ is the number of tax transfer intervals in one normalized period (the storage lifetime of grain). Carrying the coefficient $\frac{1}{\gamma}$ through all of what follows is inconsequential to the analysis, so I assume $\gamma = 1$ throughout even though this is unrealistic.

³³This line of thought also sheds light on the historical role of granaries, where resources were stored in grain. Prior to the introduction of currency and perhaps even thereafter, the state's motivation for maintaining granaries went well beyond their role in buffering the food supply from shocks. Granaries were a crucial vehicle for the power and stability of the state. In their absence, the state could not have existed

For the royal court to assure the proxy's compliance, it must set the proxy's share so that by complying he will dynamically maximize his utility. To eliminate unnecessary complexity, I make two simplifying assumptions:

1) *The proxy is assumed to be risk-neutral.* To justify this, note that if the proxy is risk-averse $P_i(\cdot)$ can be replaced with some function $\tilde{P}_i(\cdot) \in [0, \infty)$ that is also strictly increasing and which is referred to as the *risk-adjusted fighting function*. A formal proof is provided in the appendix.³⁴

2) *The proxy is assumed to be myopic.* In addition, I assume that changes over time in the proxy's share are very gradual if he does not revolt, and are therefore negligible in his foreseeable horizon. It is rather tedious to show that the models with myopic and non-myopic proxies are equivalent, so I relegate this to the appendix, too.³⁵

Given these assumptions, the royal court's task is to maintain the following condition, which states that the proxy's present utility is at least as great as his present expected utility from revolting.

$$u_i|_{\text{comply}} \geq E[u_i|_{\text{revolt}}]$$

$u(\cdot)$ is a utility function representing the proxy's preferences over resources in the current period, and is assumed to exist and to be "well behaved".³⁶

Because the state always has positive marginal utility from resources, it will ensure that this condition binds. Normalizing $u(0)$ to zero, this condition yields:

$$\begin{aligned} u_i(M_{i,t}\tau Y_i) &= u_i(E(P_{i,t})\tau Y_i) \\ \Rightarrow M_{i,t} &\equiv E(P_{i,t}) \end{aligned} \tag{3.7}$$

If the proxy's share, $M_{i,t}$, were smaller he would choose to revolt and if it were larger the royal court would reduce it, keeping more for itself. (3.7) is an equivalence because $M_{i,t}$ is constantly adjusted by the state so as to maintain it.

Substituting (3.4) into (3.7) we obtain the dynamic path of M_i :

$$M_{i,t+1} = E\left(P\left(\frac{\tau Y_i}{K - (1 - \alpha_t)M_{i,t}\tau Y_i}; Z_{t+1}\right)\right) \tag{3.8}$$

Shortly, we shall proceed to examine the dynamics of the proxy's share graphically, however there are some further issues that must be addressed before doing so.

First, notice that until now I have implicitly modeled the state's strategy as an automatic response to that of the proxy's: it provides him with his proxy's share if he complies and transfers tax revenue to the state and confronts him otherwise. This strategy, however, is not

for any length of time on a large geographic scale (one that required reliance on proxies for tax collection).

³⁴See appendix section B.1

³⁵See appendix section B.2.

³⁶That is: $\frac{\partial u(x)}{\partial x} > 0$ and $\frac{\partial^2 u(x)}{\partial x^2} < 0 \quad \forall x$.

necessarily applied. When confronting a revolting proxy, the state takes upon itself a risk of loss in the battlefield that may leave it with fewer resources than if it were to acquiesce to the proxy. Formally, if the state chooses to acquiesce it obtains:

$$X_{c,t}|_{\text{acquiesce}} = \sum_{j \neq i} (1 - M_{j,t}) \tau Y_j \quad (3.9)$$

in the current period, whereas if it confronts the proxy then in expectancy it maintains:

$$E(X_{c,t}|_{\text{confront}}) = E(Q_{c,t} \cdot \sum_{j \in J} (1 - M_{j,t}) \tau Y_j) \quad (3.10)$$

where $Q_{c,t}$ is a fighting function viewed from the royal court's perspective, i.e. whose argument is the inverse resource ratio, $\frac{X_{c,t}}{X_{i,t}}$.³⁷ Thus, given a rival with sufficient resources, the state prefers to acquiesce to a revolt rather than confront it. Equating (3.9) and (3.10) yields a unique proxy share, \bar{M}_i , beneath which the state confronts a revolting proxy and above which it acquiesces.

If the proxy's share exceeds \bar{M}_i then he ceases to transfer tax revenue to the state, knowing that no military confrontation will be required of him. This effectively means that the proxy's domain becomes independent of the state.

Second, so far I have considered only a single proxy, i , however there is no barring the cooperation of several proxies in revolting against the state. Clearly, several proxies would stand a better chance of success if they revolted simultaneously. This raises the question whether the state determines each proxy's share based on his potential to rebel single-handedly, or whether it takes into account his potential role in a larger coalition. In order to avoid this matter, let us assume that the royal court can correctly identify coalitions that are most likely to revolt and predict their internal division of resources, and that it determines each proxy's share accordingly. This amounts to giving the members of the royal court credit as savvy political players, who know the intricate human material of their proxies and can establish the correct payoffs to efficiently guarantee their compliance. Moreover, historically it must have been the case that the royal court was proficient at determining proxy shares that successfully guaranteed compliance, or else it would not have lasted the extensive period that it did. Accordingly, let subscript i denote any relevant set of proxies acting as a single player, rather than just any single proxy, and let us continue referring to it as the proxy's share.

Let us proceed to examine the dynamics of the proxy's share by graphing $M_{i,t}$ as a function of $M_{i,t-1}$. Recalling the assumption that the fighting function is strictly increasing, the dynamic curve to the left of \bar{M}_i in figure 3.1 must be strictly increasing, too, because

³⁷The fighting function $Q_{c,t}$ differs from $1 - P_{i,t}$ in two respects. First, if either the state, the proxy or both are not risk-neutral then their risk attitudes are incorporated and the two are distinct. Second, $P_{i,t}$ is "truncated" at zero for instances where the proxy is absolutely defeated (supposing some latent, underlying fighting function). Consequently, $1 - P_{i,t}$ may take on negative values and never exceeds 1, whereas $Q_{i,t} \in [0, \infty)$, similar to $P_{i,t}$.

the state's resources in period $t + 1$ diminish when $M_{i,t}$ increases. The linear presentation of the dynamic curve to the left of \bar{M}_i is merely a simplification, and there may in fact be multiple equilibria along the depicted stretch. We shall assume, however, that there exists at least one stable equilibrium with $M < \bar{M}_i$, on the grounds that the Old Kingdom state held on to power for more than a brief moment.

Thus, there are at least two stable equilibria in the model: one or more *low equilibria* to the left of \bar{M}_i and a *high equilibrium* where $M = 1$. When a low equilibrium prevails tax revenue is transferred to the state by the proxies and it maintains its hegemony over the country. In contrast, when the high equilibrium prevails no tax revenue is transferred to the state, and it is in fact no longer a state. Rather, it is a royal court stripped of its power. The high equilibrium is an apt description of the state of affairs during the First Intermediate Period, while any low equilibria describes the state of affairs during the Old Kingdom.

Finally, let us address one last aspect of the model: the state's ability to influence the outcome. Even given the state's choice of the proxy's share and its decision whether to acquiesce or confront a revolting proxy, it still has the ability to maneuver between the model's multiple equilibria. The state can alter the position of the dynamic curve via its policy, thereby affecting both the location of stable equilibria (and in some instances their number). There are two categories of policy it can apply: first, it can inadvertently affect the amount of resources it will have in the future by spending more today, as modeled by the parameter α_t . Increased (decreased) spending in period t is represented by a higher (lower) level of α_t , leaving the state with fewer (more) resources to confront a potential revolt in period $t + 1$, and thereby shifting upwards (downwards) the section of the dynamic curve to the left of \bar{M}_i . To see this, recall equations (3.6) and (3.7), and note that $M_{i,t+1}$ can be expressed as follows:

$$M_{i,t+1} = P\left(\frac{\tau Y_i}{K' + (1 - \alpha_t)K''}; Z_{t+1}\right) \quad \forall t \quad (3.11)$$

where K' and K'' follow from equation (3.6).

Secondly, the state can take measures that affect the fighting function via the parameter vector Z , and which may or may not be accompanied by the use of resources and a consequently higher level of α_t . Maintaining a fortress overlooking a proxy's abode would qualify into this category, as would mounting obstacles to the construction of a rebel coalition. I will assume, however, that the effect of changes to Z is limited when they are not accompanied by the use of any resources.

3.4 Analysis

The hypothesis brought forth in this paper is that the state's policy, reflected by the control variable α , was such that the equilibrium proxy's share increased gradually until it eventually reached the level \bar{M}_i . At that point the state ceased to confront revolting proxies and there occurred a shift to the high equilibrium, marking the end of the Old Kingdom and the onset

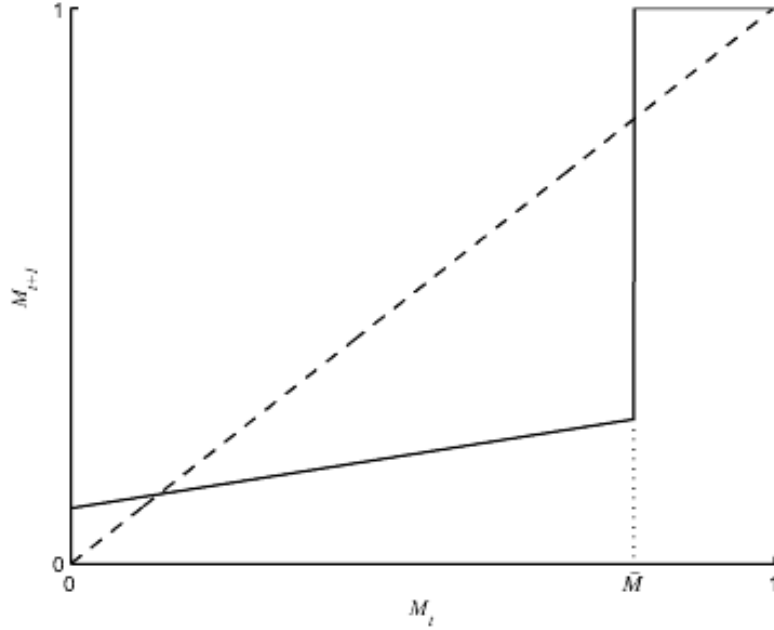


Figure 3.1: Dynamic path of the proxy's share.

of the First Intermediate Period.

In what follows I present this hypothesis more carefully and in greater detail, incorporating each of the stylized facts presented earlier into the framework of the model. To do so, I track the likely value over time of four variables, from which the path of α over time can be reconstructed:³⁸

- 1) GI : the state's gross income.
- 2) M : the proxy's share (recall the issue of proxy coalitions, addressed above).
- 3) NI : the state's net income, which refers to the income remaining after deducting the proxy's share and is equal to $GI - M$.

4) E : the state's expenditure of present-period income. To understand the qualifier "present-period", recall from equation (3.6) that α_t refers to the share of the tax revenue received in period t that is spent in the same period, leaving behind a share $1 - \alpha_t$ for use in period $t + 1$. Thus, $E = \alpha \cdot NI$.³⁹

Having described the paths of these variables, it is then straightforward to reconstruct α as:

$$\alpha = \frac{E}{NI} = \frac{E}{GI - M} \quad (3.12)$$

³⁸In what follows: $TI = \sum_{j \in J} \tau Y_j$; $M = \sum_{j \in J} M_j \tau Y_j$; $NI = \sum_{j \in J} (1 - M_j) \tau Y_j$.

³⁹An implicit assumption made here is that each period the state first spent resources remaining from the previous period, before they went to waste, and only then turned to using resources collected in the present period. It would be noticeably wasteful for the state to do otherwise.

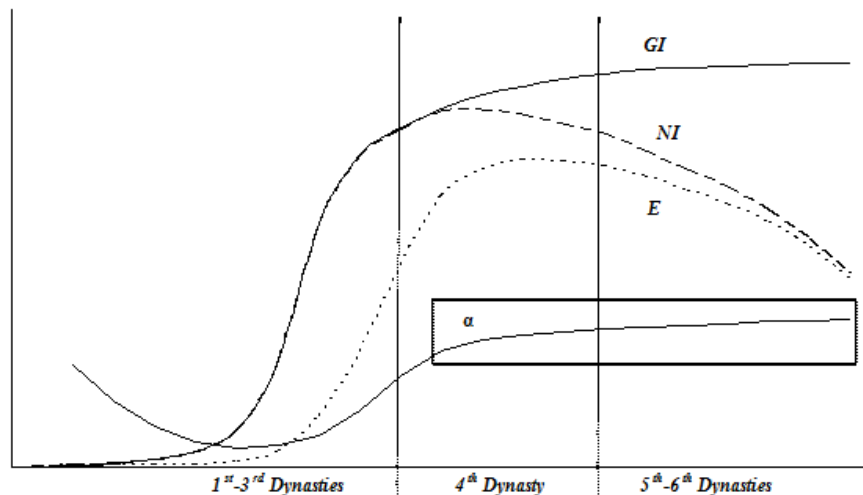


Figure 3.2: The development of GI, NI, E and α during the Early Dynastic Period and the Old Kingdom. Note that the from the 4th Dynasty onwards, the level of α steadily increases as it converges to 1.

Figure 3.2 is useful in illustrating the analysis that follows.

1) *Beginning with the formation of the state, the region from which taxes were levied gradually expanded and the tax collection system solidified. The process came to an end roughly during the 3rd Dynasty when the tax system encompassed all of Egypt. Most of the growth in the state's income occurred up until the 3rd Dynasty with more modest growth, if any, taking place afterwards.*

First of all, note that a new proxy introduced in period t has a previous-period proxy share of $M_{i,t-1} = 0$. Therefore, with all else equal, his share necessarily converges to the lowest existing equilibrium, as indicated by figure 3.1.

Next, consider the dynamics of the proxy's share when the amount of surplus extracted by the state is increasing. Equation (3.7) is key so it is repeated here:

$$M_{i,t+1} = E\left(P\left(\frac{X_{i,t}}{X_{c,t}}; Z_{t+1}\right)\right)$$

This stylized fact is essentially an observation on GI . As long as the growth in surplus extraction was more or less evenly distributed among proxies the resource ratio for each individual proxy, $X_{i,t}/X_{c,t}$, is likely to have been constant.⁴⁰ Inasmuch as the increase in

⁴⁰The Egyptian state was subdivided into administrative units of roughly similar size called *nomes*, of which there were finally several dozen. Thus, momentarily equating proxies with nomarchs (administrators of local districts called *nomes*), even if the expansion of surplus extraction was concentrated in the domains of a small number of nomarchs during some period, it is unlikely that this was the case throughout any

surplus extraction involved the introduction of new proxies, it would have yielded reductions in individual proxies' resource ratios. In figure 3.1, any increase in $X_{c,t}$ given an existing level of $M_{i,t}$ induces a downward shift of the dynamic curve, leading to a drop in the levels of all low equilibria. Thus, it is likely that while the state's extraction of surplus expanded from its inception through the 3rd Dynasty, the proxy's share did not grow and probably even decreased.

The state's gross income, GI , increased during this period. The lack of increase in the proxy's share, M , established above implies that the state's net income, NI , followed gross income very closely. While the state's present-period expenditure, E , was rising, it is unclear whether or not it was rising relative to net income, so the path of α while the state's surplus extraction system was expanding is unclear (recall $\alpha = E/NI$).⁴¹

2) *The peak of state expenditure occurred during the 4th Dynasty, and is marked by the construction of the Great Pyramids at Giza. During the 5th and 6th Dynasties state expenditure was substantially reduced.*

This stylized fact is essentially an observation on E . The state's present-period expenditure peaked during the 4th Dynasty and dropped thereafter. In combination with the next stylized fact, it is possible to map out the path of alpha through the end of the 6th Dynasty, so let us proceed.

3) *The provincial elite was negligible in size and wealth before the 5th Dynasty. During the 5th and 6th Dynasties it grew substantially larger and wealthier, at the expense of the state.*

This stylized fact, on the other hand, is essentially an observation on M . The proxy's share was very small until roughly the 5th Dynasty, after which it steadily grew larger. For this to have been the case, it is necessary for α to have been increasing, thereby raising the dynamic curve in figure 3.1 and shifting any low equilibrium to a higher level of M .⁴²

But is such an increase in α plausible? The proxy's share was negligible until the 3rd Dynasty (see above), so $\alpha = E/NI \simeq E/GI$. For α to have increased, it must have been the case that the state's present-period expenditure, E , increased more quickly than the state's income, GI . Recall from the first stylized fact that the increase in the state's gross income, GI , slowed substantially after the 3rd Dynasty and possibly even stopped. The emergence of a provincial elite was therefore the result of state expenditure, E , whose growth after the 3rd Dynasty did not slow as quickly as gross income.

The mutually unsustainable paths of state expenditure and income eventually induced the state to curb its expenditure, *albeit not enough to curb the growth of α* .⁴³ The curbing of

length of time.

⁴¹Of course, one could assert the path of α more accurately by making educated assumptions about the state's spending behavior. Also, given that the state continuously wielded power throughout this period, a sharp increase in α of the kind that would induce a shift to the high equilibrium can be ruled out.

⁴²It is inconsequential whether the increase in M occurred through a gradual shift of a realized low equilibrium or through a jump to some higher low equilibrium in a set of multiple equilibria.

⁴³Income here refers to both GI and NI , which diverged now that the proxy's share, M , was no longer negligible.

state expenditure brought the path of E to a peak during the 4th Dynasty. The growth of α , however, was only slowed but not stopped and so the proxy's share continued to expand so that it became necessary to curb E even further. This occurred continuously, sending state expenditure on a decreasing path well below the 4th Dynasty peak and raising the proxy's share to unprecedented levels. Eventually, it hit the critical level \bar{M} at which the state ceased to confront revolting proxies. The result was stylized fact number four.

4) *From the end of the 6th Dynasty the state lost de-facto authority over the country, although not necessarily by way of any military conflict.*

Earlier, I stated that whether or not the remnants of the Old Kingdom at the end of the 6th Dynasty were destroyed in armed conflict is inconsequential, and both possibilities are consistent with the analysis. To clarify this point, note that a risk-averse proxy may choose to comply with the state even when the proxy's share exceeds the critical level \bar{M} , for fear that the state will confront him despite the odds.

Formally, the interaction between the state and the proxy can be modeled as an extensive form game where the proxy moves first and can either comply with the state or revolt. In the event that the proxy revolts, the state can either confront the proxy militarily or acquiesce. When the proxy's share exceeds the critical level, \bar{M} , and the proxy is sufficiently risk-averse with respect to the outcome, (*revolt, acquiesce*) may not qualify as a trembling-hand perfect equilibrium, even if it is a perfect Bayesian equilibrium.

3.5 What Is Optimized And What Is Not:

Before proceeding, note that all of the decisions in the model are made by way of optimization with one exception: the state's expenditure is not optimally controlled. The contrast is sharpest with respect to the state's optimal determination of the proxy's share. This matter is worth dwelling upon.

The grounds for modeling the state's choice of the proxy's share as an optimizing one is that its results were apparent to the state almost immediately. A royal court member setting an insufficient proxy's share would quickly be confronted - possibly in person - either by a proxy withholding tax revenue (i.e. revolting), by a threat to do so, or at least by an expression of discontent on the proxy's part that served as a warning. This type of interaction can be viewed as implicit bargaining, and perhaps the state even bargained with the proxies explicitly. I broadly assume that this feedback allowed the state to gauge the proxies' perception of their relative power as well as their risk attitude, and consequently to set near-optimal proxy's shares. The process was likely one of learning through trial and error on the state's part, which yielded political intuition and successfully guaranteed the proxies' compliance. Therefore, optimal determination of the proxy's share appears to be a reasonable approximation of the state's behavior.

A feedback loop such as this, which was immediate in historical terms, did not exist with respect to the level of state expenditure. The meager feedback that did exist would have looked like this: suppose the state failed to cut back on its expenses in response to a

reduction in its net income - for instance a reduction in net income due to an increase in a certain proxy's share. Such a failure by the state to respond would not be unusual behavior even for a modern government today, let alone for an ancient one. In terms of the model this would constitute an increase in α for that period (recall once again that $\alpha = E/NI$). After the next harvest season, when tax revenue was transferred to the royal court, the state's resources would have dwindled, placing it in an inferior position with respect to the proxies (if not after the following harvest then cumulatively, after a longer stretch of time). The proxies, with or without knowledge of the state's exact fiscal condition, could then drive a harder bargain and subsequently up their proxy's share.

It seems unlikely that any one person would make the connection between the state's initial inaction and the eventual, implicitly harder bargain driven by proxies. In fact, picking up on such a causal link was practically impossible given the cumulative nature of the process, which occurred over several generations, and so I assume that it was never made. In light of this, it does not appear that optimal control over α incorporating perfect knowledge of the model would be a reasonable approximation of the state's behavior.

When state expenditure was eventually and insufficiently curbed, it was likely in a general attempt to economize and does not indicate any perception of the causal link between excess spending at one point in time and dwindled net income later on. Kemp (2006) addresses this matter:⁴⁴

“In revenue and expenditure terms the sum of [the pious foundations'] activities plus a general level of royal expenditure on court life, on large and thus long-term building programmes, and on the military, represented a general 'budget' or balance-sheet for the country. It was probably never seen in quite so abstract a way. But complaints from below of insufficient resources would have signalled to senior officials a degree of imbalance, which they could then have sought to correct.”

It is likely that signals of shortage and a call for economization arose in a very simple manner, when the stocks in the state's granaries were reported.

Given the following quote, also from Kemp (2006), the inference of the causal link by the Egyptian state seems all the more unlikely:⁴⁵

“Within any one channel the procedures [of the Egyptian government] could be remarkably effective (though not efficient) in achieving a given target, such as quarry, transport, and erect a colossus of a particular size. This is where bureaucratic talents flourished. But we will look in vain for evidence of conscious integration of the individual parts into a general scheme of management.”

⁴⁴See Kemp (2006), page 236.

⁴⁵See Kemp (2006), page 235.

3.6 Conclusion

The prolonged stability of the Early Dynastic Period and the Old Kingdom came to an end when local administrators and the affiliated provincial elite grew too powerful for the royal court to credibly threaten them militarily, rendering the state powerless. This outcome was the culmination of a long and gradual process, rooted in the relationship between the royal court and its tax-collecting proxies, who directly or indirectly comprised the provincial elite. By making the minimal assumption of a positive relation between the two parties' ratio of resources on the one hand and the outcome of a military conflict between them on the other, I characterize the dynamics of the balance of power. The formal model yields that in the short run the state could not deviate from its policy by reducing the flow of resources from its treasury to the provincial elite, because doing so would fail to guarantee the proxies' compliance with their role. This result fills a gap in the existing literature, which does not provide a robust explanation why the Pharaohs did not exercise their power to immediately halt or even slow the flow of resources to the provincial elite.

In the long term, the state could potentially have prevented its fall by carefully manipulating its expenses, however the nature of the dynamics governing the balance of power was essentially unobservable to the royal court. Failing to understand the link between state expenditure and the balance of power, the royal court failed to maintain a stable ratio of resources with its proxies. Thereby, an ever increasing flow of resources from the state to the provincial elite was required to guarantee the latter's cooperation with the state, and eventually eroded the royal court's power beyond the threshold of making a credible military threat. A critical issue here is distinguishing between policy decisions that can be modeled by optimizing choices and decisions that cannot.

The analysis presented in section 3.4 is specific to Ancient Egypt and to the era in question, however the model to which it applies is more general, and raises the possibility of wider applicability. The relationship between central government and tax-collecting proxies modeled here is a general one and may be widely applicable to ancient periods as well as to more recent ones.

References

- Aaronson, D. and J. Davis (2011). How much has house lock affected labor mobility and the unemployment rate? *Chicago Fed Letter* (Sep).
- Akerlof, G. A., A. K. Rose, J. L. Yellen, L. Ball, and R. E. Hall (1988). Job Switching and Job Satisfaction in the U.S. Labor Market. *Brookings Papers on Economic Activity* 1988(2), 495–594.
- Allen, R. (1997). Agriculture and the Origins of the State in Ancient Egypt. *Explorations in Economic History* 34, 135–154.
- Andrew, M. and G. Meen (2003). House Price Appreciation, Transactions and Structural Change in the British Housing Market: A Macroeconomic Perspective. *Real Estate Economics* 31(1), 99–116.
- Anenberg, E. (2011). Loss aversion, equity constraints and seller behavior in the real estate market. *Regional Science and Urban Economics* 41(1), 67 – 76.
- Anenberg, E. (2012). Information Frictions and Housing Market Dynamics. *Finance and Economics Discussion Series 2012-48. Board of Governors of the Federal Reserve System (U.S.)*.
- Ashenfelter, O., H. Farber, and M. R. Ransom (2010). Modern Models of Monopsony in Labor Markets: A Brief Survey. Working Papers 1223, Princeton University, Department of Economics, Industrial Relations Section.
- Bard, K. (2000). The Emergence of the Egyptian State. In I. Shaw (Ed.), *The Oxford history of Ancient Egypt*, Chapter 4, pp. 61–88. Oxford: Oxford University Press.
- Bard, K. and R. Carneiro (1989). Patterns of Predynastic Settlement Location, Social Evolution, and the Circumscription Theory. *Societes Urbaines en Egypte et au Soudan, Cahiers de Recherches de l'Institut de Papyrologie et d'Egyptologie de Lille* 11, 15–23.
- Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2011, July). A Dynamic Model of Demand for Houses and Neighborhoods. Working Paper 17250, National Bureau of Economic Research.

- Bell, B. (1971). The Dark Ages in Ancient History. I. The First Dark Age in Egypt. *American Journal of Archaeology* 75(1), 1–26.
- Benito, A. (2006). The down-payment constraint and UK housing market: Does the theory fit the facts? *Journal of Housing Economics* 15(1), 1 – 20.
- Berkovec, J. A. and J. L. Goodman (1996). Turnover as a Measure of Demand for Existing Homes. *Real Estate Economics* 24(4), 421–440.
- Blanchard, O. J. and L. Katz (1992). Regional Evolutions. *Brookings Papers on Economic Activity* 1992(1), pp. 1–75.
- Bontemps, C., J.-M. Robin, and G. J. Van Den Berg (2000). Equilibrium Search with Continuous Productivity Dispersion: Theory and Nonparametric Estimation. *International Economic Review* 41(2), 305–358.
- Brewer, D. and E. Teeter (2007). *Egypt and the Egyptians*. Cambridge: Cambridge University Press.
- Burdett, K. and M. Coles (2003). Equilibrium Wage-Tenure Contracts. *Econometrica* 71(5), 1377–1404.
- Burdett, K. and D. T. Mortensen (1998). Wage Differentials, Employer Size, and Unemployment. *International Economic Review* 39(2), 257–273.
- Burdett, K. and T. Vishwanath (1988). Balanced Matching and Labor Market Equilibrium. *Journal of Political Economy* 96(5), pp. 1048–1065.
- Butzer, K. (1976). *Early Hydraulic Civilization in Egypt: A Study in Cultural Ecology*. Chicago: The University of Chicago Press.
- Butzer, K. (1984). Long-Term Nile Flood Variation and Political Discontinuities in Pharaonic Egypt. In Clark, J. And Brandt, S. (eds.) *From Hunters to Farmers: the Causes and Consequences of Food Production in Africa*, Berkeley: University of California Press, pages 102-112.
- Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006). Wage Bargaining with On-the-Job Search: Theory and Evidence. *Econometrica* 74(2), 323–364.
- Cameron, C. and P. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Caplin, A., C. Freeman, and J. Tracy (1997, November). Collateral Damage: Refinancing Constraints and Regional Recessions. *Journal of Money, Credit and Banking* 29(4), 496–516.
- Case, B., H. O. Pollakowski, and S. M. Wachter (1997). Frequency of Transaction and House Price Modeling. *The Journal of Real Estate Finance and Economics* 14(1-2), 173–87.

- Case, K. E., R. J. Shiller, and A. Thompson (2012, September). What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets. Working Paper 18400, National Bureau of Economic Research.
- Chan, S. (May 2001). Spatial Lock-in: Do Falling House Prices Constrain Residential Mobility? *Journal of Urban Economics* 49, 567–586.
- Clayton, J., N. Miller, and L. Peng (2010). Price-volume Correlation in the Housing Market: Causality and Co-movements. *The Journal of Real Estate Finance and Economics* 40, 14–40.
- Coulson, N. E. and P. L. Grieco (2013). Mobility and mortgages: Evidence from the PSID. *Regional Science and Urban Economics* 43(1), 1 – 7.
- Donovan, C. and C. Schnure (2011). Locked in the House: Do Underwater Mortgages Reduce Labor Market Mobility? Unpublished manuscript.
- Engelhardt, G. V. (2003). Nominal loss aversion, housing equity constraints, and household mobility: evidence from the United States. *Journal of Urban Economics* 53(1), 171 – 195.
- Estevao, M. and E. Tsounta (2011). Has the Great Recession raised US structural unemployment? *IMF Working Papers*, 1–46.
- Farnham, M., L. Schmidt, and P. Sevak (2011, May). House Prices and Marital Stability. *American Economic Review* 101(3), 615–19.
- Ferreira, F., J. Gyourko, and J. Tracy (2010). Housing busts and household mobility. *Journal of Urban Economics* 68(1), 34 – 45.
- Ferreira, F., J. Gyourko, and J. Tracy (2011, September). Housing Busts and Household Mobility: An Update. Working Paper 17405, National Bureau of Economic Research.
- Ferreira, F., J. Gyourko, and J. Tracy (2012). Housing busts and household mobility: an update. *Economic Policy Review* (Nov), 1–15.
- Geanakoplos, J. (2010). The Leverage Cycle. *NBER Macroeconomics Annual* 24(1), pp. 1–66.
- Genesove, D. and C. Mayer (2001). Loss Aversion and Seller Behavior: Evidence from the Housing Market. *Quarterly Journal of Economics* 116(4), 1233–1260.
- Grimal, N. (1992). *A History of Ancient Egypt*. Oxford: Blackwell.
- Grossman, S. and G. Laroque (1990). Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods. *Econometrica* 58, 25–51.
- Hassan, F. (1997). The Dynamics of a Riverine Civilization: A Geoarchaeological Perspective on the Nile Valley, Egypt. *World Archaeology* 29(1), 51–74.

- Henley, A. (1998). Residential Mobility, Housing Equity and the Labour Market. *The Economic Journal* 108(447), pp. 414–427.
- Hort, K. (2000). Prices and turnover in the market for owner-occupied homes. *Regional Science and Urban Economics* 30(1), 99 – 119.
- Karahan, F. and S. Rhee (2011). Geographical Reallocation and Unemployment during the Great Recession: The Role of the Housing Bust.
- Kemp, B. (1983). Old Kingdom, Middle Kingdom and Second Intermediate Period. In O. D. Trigger B., Kemp B. and L. A. (Eds.), *Ancient Egypt: a Social History*, Chapter 2, pp. 71–182. Cambridge: Cambridge University Press.
- Kemp, B. (2006). *Ancient Egypt: Anatomy of a Civilization*. London: Routledge.
- Knohl, Y. (2008). *The Bible's Genetic Code (Hebrew)*. Or-Yehudah: Dvir.
- Kothari, S., I. Saporta-Eksten, and E. Yu (2012). The (Un)importance of Mobility in the Great Recession. Stanford University, unpublished manuscript.
- Kuhn, P. (2004). Is monopsony the right way to model labor markets? A review of Alan Manning's monopsony in motion. *International Journal of the Economics of Business* 11(3), 369 – 378.
- Leung, C. K. Y., G. C. K. Lau, and Y. C. F. Leong (2002). Testing Alternative Theories of the Property Price-Trading Volume Correlation. *Journal of Real Estate Research* 23(3), 253–264.
- Malek, J. (2000). The Old Kingdom. In I. Shaw (Ed.), *The Oxford history of Ancient Egypt*, Chapter 5, pp. 89–117. Oxford: Oxford University Press.
- Manning, A. (1993). Endogenous Labour Market Segmentation in a Matching Model. *Centre for Economic Performance, Discussion Papers No' 126*.
- Manning, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton: Princeton University Press.
- Manning, A. (2011). Imperfect Competition in the Labor Market. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4B of *Handbook of Labor Economics*, pp. 973 – 1041. Elsevier.
- Modestino, A. S. and J. Dennett (2012). Are American homeowners locked into their houses? The impact of housing market conditions on state-to-state migration. *Regional Science and Urban Economics*.
- Mortensen, D. T. (1998). Equilibrium Unemployment with Wage Posting: Burdett-Mortensen Meet Pissarides. Working papers, Centre for Labour Market and Social Research, Denmark-.

- Mortensen, D. T. (1999). Modeling Matched Job-Worker Flows. Econometric Society World Congress 2000 Contributed Papers 1493, Econometric Society.
- Mortensen, D. T. and C. A. Pissarides (1994, July). Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies* 61(3), 397–415.
- Mortensen, D. T. and T. Vishwanath (1991). Information Sources and Equilibrium Wage Outcomes. *Northwestern University, Department of Economics Discussion Papers No' 948*.
- Mortensen, D. T. and T. Vishwanath (1994). Personal contacts and earnings : It is who you know! *Labour Economics* 1(2), 187 – 201.
- Nenov, P. (2012). Labor Market and Regional Reallocation Effects of Housing Busts. Unpublished manuscript.
- Ortalo-Magne, F. and S. Rady (2004). Housing transactions and macroeconomic fluctuations: a case study of England and Wales. *Journal of Housing Economics* 13(4), 287 – 303.
- Ortalo-Magne, F. and S. Rady (2006). Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints*. *Review of Economic Studies* 73(2), 459–485.
- Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium Wage Dispersion with Worker and Employer Heterogeneity. *Econometrica* 70(6), 2295–2350.
- Sahin, A., J. Song, G. Topa, and G. L. Violante (2012, August). Mismatch Unemployment. Working Paper 18265, National Bureau of Economic Research.
- Schulhofer-Wohl, S. (2011, January). Negative Equity Does Not Reduce Homeowners' Mobility. Working Paper 16701, National Bureau of Economic Research.
- Seidlmayer, S. (2000). The First Intermediate Period. In I. Shaw (Ed.), *The Oxford history of Ancient Egypt*, Chapter 6, pp. 118–147. Oxford: Oxford University Press.
- Smith, J. E. and K. F. McCardle (2002). Structural Properties of Stochastic Dynamic Programs. *Operations Research* 50(5), pp. 796–809.
- Stein, J. C. (1995, May). Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects. *The Quarterly Journal of Economics* 110(2), 379–406.
- Sterk, V. (2010, October). Home Equity, Mobility, and Macroeconomic Fluctuations. DNB Working Papers 265, Netherlands Central Bank, Research Department.
- Stroebel, J. (2012). The impact of asymmetric information about collateral values in mortgage lending. Unpublished manuscript.
- Valletta, R. G. (2013). House Lock and Structural Unemployment. *Labour Economics*.
- Zenou, Y. (2009). *Urban Labor Economics*. New York: Cambridge University Press.

Appendix A

Appendix to Chapter 1

A.1 Data

A.1.1 Measuring equity

By convention, home equity is most often measured inversely, in terms of a property’s combined loan-to-value ratio (CLTV), and this is the measure of equity used in (1.5). The CLTV ratio of property i at time t is the sum of principal owed at that time on all loans secured by the property, divided by the property’s contemporary value.¹ Formally,

$$CLTV_{it} \equiv \frac{\sum_{j \in J(i,t)} loan_{jt}}{value_{it}}, \quad (\text{A.1})$$

where $J(i, t)$ is the set of loans for which property i serves as collateral and $loan_{jt}$ is the amount of principal owed on loan j .² The CLTV ratio is inversely related to the property owner’s share of equity, which is simply

$$\% \text{ equity}_{it} \equiv \frac{value_{it} - \sum_{j \in J(i,t)} loan_{jt}}{value_{it}} \equiv 1 - CLTV_{it}. \quad (\text{A.2})$$

The CLTV ratio used in the paper is an estimate. First, because property values can never be observed directly, $value_{it}$ is taken to be CoreLogic’s corresponding Automated Valuation Model (AVM) estimate, described in the main text.³ Second, only the initial

¹Each loan secured against a property has its own loan-to-value (ltv) ratio. The word “combined” simply reflects that in instances in which there is more than one loan, the ratio involves their sum.

²The primary loan secured against a property is typically a mortgage obtained when the property is purchased, but a sizeable share of mortgaged properties - 40% in my sample - are observed serving as collateral for further borrowing, either since the time of purchase (“piggy back” loans) or from later on. Additional borrowing comes in many shapes and forms, including secondary mortgages, refinancing (often involving cash-in or cash-out), home equity loans or lines of credit, and various other methods means of borrowing against one’s home. A property’s combined loan-to-value (CLTV) ratio is simple the sum of loan-to-value (ltv) ratios for any individual loans secured against the property.

³Even at times when a property is sold, one could deliberate whether the sale price reflects the property’s

balance of the loans secured against a property is observed, not their outstanding balance once payments have been made towards principal, as CoreLogic does not observe these payments. I estimate payments towards principal in a very crude way, assuming standard (fixed payment) amortization over the observed term of the loan and given the going interest rate associated with the loan by CoreLogic.⁴

A.1.2 The Zillow housing price index

The Zillow housing price index is hedonic in the sense that it relies on estimating home values based on their observable characteristics, rather than relying on repeat sales to account for all fixed observable and unobservable home attributes. However, this index is not simply a normalized sequence of time fixed effects estimated in a hedonic regression. In order to minimize the selection bias inherent in hedonic regressions because the set of properties selling in a given period is non-random, Zillow hedonically estimates the value of each and every home in every period. The Zillow housing index value for a set of homes in a certain period is then taken to be the median home value for that set in that period.⁵

A.1.3 Inferring owner-occupancy status

I infer that a property was owner-occupied in a given quarter if at least one contemporary occupant shared a last name with a contemporary owner. Thus, I broadly interpret owner-occupancy to include relatives of owners (as well as occupants who by coincidence share a last name with an owner). I observe the occupants of a property by matching each property-by-quarter observation with contemporary voter registry records by address.⁶ Registering to vote is thus used as a proxy for occupancy. Occupants who do not register to vote, or prefer to use another permanent address, e.g. of a parent's home, remain unobserved. Within property-by-ownership spells, I take a property's owner-occupancy status to be that of the last period of the ownership spell. This way, occupants who take their time about updating their voter registry after moving are still captured as occupants, as long as they update their voters registry either before the last period of their ownership spell or before the last period observed (2011Q4). This method of inferring owner-occupancy is asymmetric, in the sense that properties *not* inferred to be owner-occupied may in fact still be owner-occupied, e.g. if their owner-occupants failed to update their voter registration records.

value in a way that generalizes beyond the buyer and seller's agreed upon price.

⁴For a loan with principal P , an N period term, an interest rate of r per period and a fixed per period payment c , the balance after n periods is $(1+r)^n P \cdot [1 - \frac{1-(1+r)^{-n}}{1-(1+r)^{-N}}]$.

⁵Additional information on Zillow's indexing methodology is available online at the time of writing, at: <http://www.zillowblog.com/research/2012/01/21/zillow-home-value-index-methodology>.

⁶Roughly 82% of voter registry records are successfully matched with properties; the unmatched voter registry records primarily reflect condos with idiosyncratically recorded numbering, and apartments for which the county assessor record encompasses the entire multi-unit property.

A.2 Dynamic selection

Dynamic selection refers to the process whereby the composition of the pool of homeowners in a cohort systematically changes over time as homeowners sell their homes and exit the cohort selectively. A simple way of thinking about the problem is to suppose that homeowners belong to one of two types: owners of starter homes who have a high propensity to move and owners of permanent homes who have a low propensity to move. As a purchase cohort’s tenure duration grows longer its relative share of starter-home owners dwindles. This “weeding out” process is of no concern inasmuch as it is identical across cohorts, because then it gets soaked up by the flexible tenure duration control included in the regressions. However if the “weeding out” process is differential across cohorts, i.e. if starter-home owners are “weeded out” at a quicker pace in some cohorts than in others because of conditions that facilitate or hinder mobility, then the flexible tenure duration control is inadequate and captures only the average effect of dynamic selection across all cohorts.⁷

Dynamic selection is a potential concern if increasing home values raise equity and in so doing facilitate mobility, because then cohorts that experience greater increases in home values during their tenure are likely to experience a faster pace of dynamic selection. Homeowners remaining in such cohorts long enough to have low CLTV ratios are likely to have a relatively low propensity to move as well, biasing estimates of the effect of equity (CLTV) on mobility downward (upward).

To gauge the extent of dynamic selection I compare the rates of home sale by CLTV ratio for three sets of homeowners. The first set is the full sample used in this study, which consists of properties whose owners have tenure duration of 1 to 10 years. The second and third are subsets whose tenure duration is within 1 to 5 years and 1 to 3 years, respectively. If differences in sale rates across equity levels are driven by dynamic selection then sale rates should increase (decrease) with equity the most (least) for the 1 to 3 year group, and least (most) for the 1 to 10 year group.

Figure A.2.1a shows the unconditional home sale rates for these three groups with respect to actual CLTV. The three groups do not differ substantially, suggesting that dynamic selection does not play an important role in shaping the relationship between actual equity and sale rates. Figure A.2.1b shows the corresponding home sale rates conditional on the full set of owner, home and loan characteristics - and in particular a cubic of tenure duration - as well as quarter by zip code area fixed effects and a cubic of purchase cohort, and does not suggest an important role for dynamic selection either. Figure A.2.1c, on the other hand, shows the unconditional home sale rates for these three groups by *predicted* CLTV. Below 80% predicted CLTV the sale rate increases the most with equity for the 1 to 3 year group, then for the 1 to 5 year group and least of all for the 1 to 10 year group, which suggests that dynamic selection is taking place. However, once the home sale rates are fully conditioned in Figure A.2.1d (as in Figure A.2.1b) the effect of equity on sale rates appears to be similar for the different tenure duration groups, suggesting that the controls adequately account

⁷Dynamic selection is differential across cohorts if, for example, factors influencing mobility interact multiplicatively as they do in a mixed proportional hazard model.

for dynamic selection. Overall, these results alleviates concerns with respect to dynamic selection.⁸

A.3 The “better homes” hypothesis

This appendix addresses the “better homes” hypothesis outlined in section 1.5.3. The hypothesis is that relative to observably similar homes, certain “better homes” tend to sell less frequently and to incur smaller losses in value when aggregate housing prices fall. If both traits coincide this generates a positive correlation between homes sales and CLTV ratios - which rise in proportion to loss in home value - causing naive OLS regressions of home sales on CLTV ratios to be biased upwards.

The observed empirical relationship between sale probability and actual CLTV is shown in figure 1.6a and is stylized in figure 1.8c. The argument put forth in section 1.5.3 is that this pattern emerges from the superposition of a causal effect of CLTV on the probability of sale (from 70% to 100% CLTV) on one hand and confounding factors that bias the OLS estimate of the effect upward over the entire CLTV range on the other. These confounding factors include the “better homes” hypothesis. To illustrate how the observed pattern emerges I generate data and run a simulation in which:

- Aggregate housing prices fall.
- A higher CLTV ratio reduces the probability of sale over the 70% to 100% CLTV range for all homes.
- All homes are valued equally at the onset, but half of the homes are “better” and therefore, compared to the remaining “worse” homes, they sell less frequently and incur smaller losses in value when aggregate housing prices fall.
- The owner of “better” homes are assumed to have made somewhat larger initial down payments.⁹

Figure A.3.1 visualizes the results of the simulation.¹⁰ The solid lines in the upper panel reflect the annual sale probability of “better” and “worse” homes, respectively, and the

⁸Moreover, note that the direction in which dynamic selection potentially biases results works *against* finding a positive effect of equity on home sales. Therefore if one remains concerned about dynamic selection despite the results shown here, the implication is that estimates of the effect of equity on mobility reported in this paper are in fact lower bounds.

⁹As shown shortly, this assumption is supported by empirical evidence. This assumption is not strictly necessary, but it helps the simulation visually mimic the relationship between CLTV and annual sale probability observed in the raw data, as seen in Figure 1.6a (particularly at low levels of CLTV).

¹⁰In more detail: The data consist of 20 quarterly observations on 1000 simulated properties, half of which are “better” and half of which are “worse”. All homes are valued equally when the simulation begins, but the value of “better” and “worse” homes evolves as 0.9 and 1.1 times the change in the housing pricing index (HPI), respectively, and the HPI is falls by 2% each quarter. Thus, “better” homes lose less value

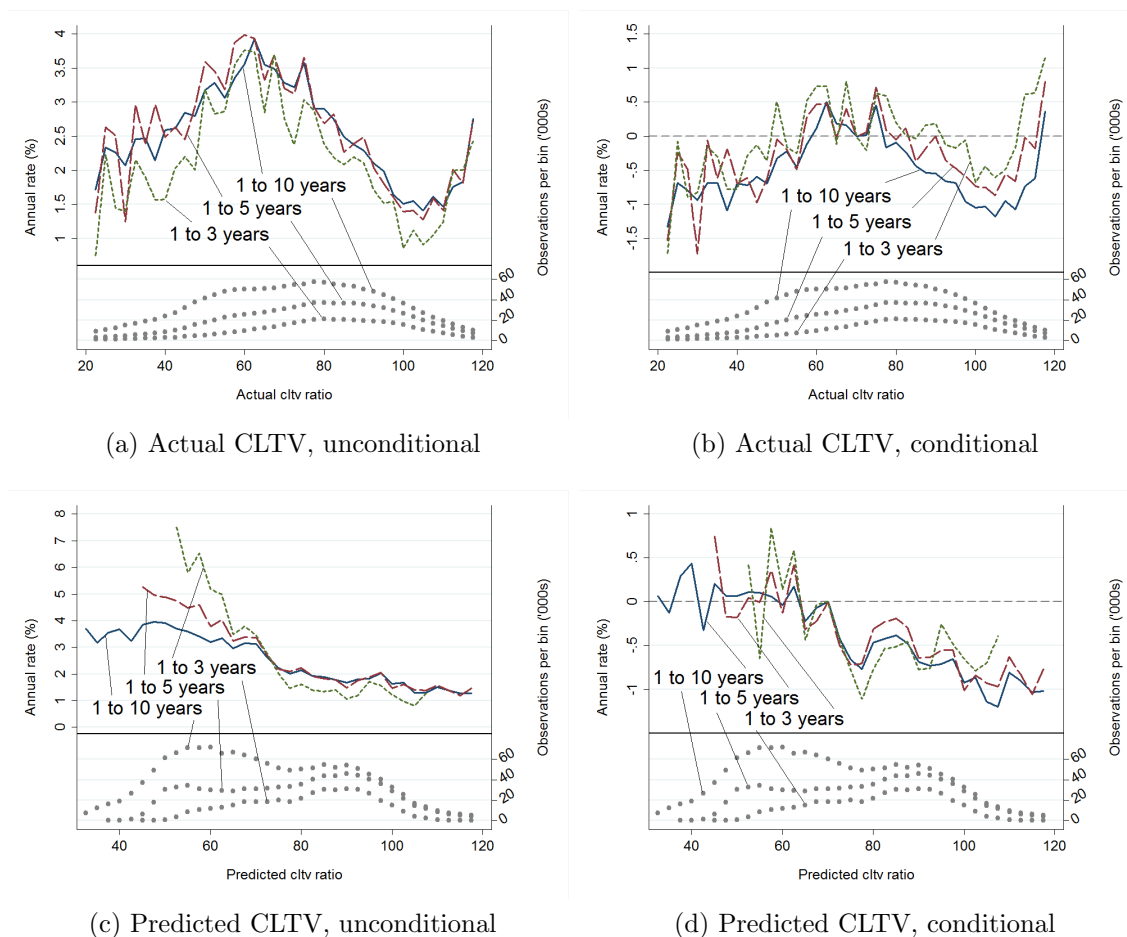


Figure A.2.1: The effect of equity on sales by actual and predicted CLTV and by tenure duration group. Panels *a* and *b* (*c* and *d*) report unconditional sale rates and conditional effects on sale rates by actual (predicted) CLTV ratio for the full sample, which contains homeowners with tenure duration of 1 to 10 years, and for subsets with tenure duration of 1 to 5 years and of 1 to 3 years. Conditioning refers to controlling for the full set of owner, home and loan characteristics (including a cubic of tenure duration) as well as quarter by zip code area fixed effects and a cubic of purchase cohort - see Table 1.2 for a detailed account of the included controls. If dynamic selection is taking place then an increase (decrease) in equity (CLTV) should correspond to a greater increase in sale rates for groups with lower tenure duration. Dynamic selection does not appear to be taking place along the dimension of actual CLTV. Dynamic selection does appear to take place along the predicted CLTV dimension, but it is adequately accounted for by conditioning the estimates on the above set of controls.

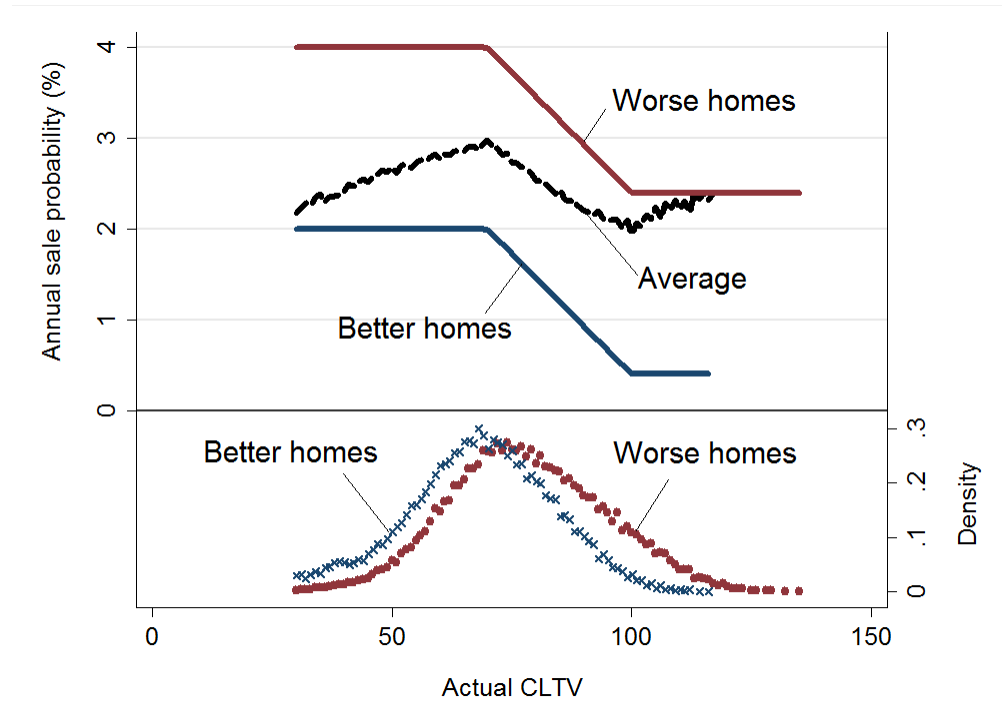


Figure A.3.1: Data generated under the assumptions of the “better homes” hypothesis mimic the relationship between the CLTV ratio and the annual probability of sale observed in the raw data, as in Figure 1.6a. The sale probability of all properties is reduced causally and similarly when CLTV increases from 70% to 100%, but “better” homes are less likely to sell across the board. Because “better” homes lose less value than “worse” homes when aggregate housing prices fall (and because their owners tend to make larger down payments), these homes tend to have lower CLTV ratios. The weighted average of homes’ annual sale probabilities at each level of CLTV reflects the *composition* of homes at that level, reflected by an upward slope in the ranges below 70% and above 100% CLTV, and in a downward slope in the 70% to 100% CLTV range that is *moderate* than the underlying causal effect of CLTV. The moderate slope of the average annual sale probability with respect to CLTV in the 70% to 100% CLTV range qualitatively corresponds to the OLS estimates in column 1 of Table 1.9, whereas the steeper slopes of uniquely “better” or “worse” homes over this range correspond to the IV estimates in column 3 of Table 1.9.

as aggregate housing prices fall. “Better” homes have an annual sale probability of 2% whereas “worse” homes have an annual sale probability of 4%, i.e. they sell more frequently. All homes experience a linear reduction in annual sale probability of 1.6% as they transition from 70% to 100% CLTV. In addition, the owners of “better” homes are assumed to have made down payment that is 5 percentage points larger on average. Specifically, down payments are assumed to be distributed lognormal $(\mu, 0.25)$, with $\mu = 42.5\%$ for “better” homes and 37.5% for “worse” (this implies a mass of down payments centered at 40% with a left tail that extends to 0% and a longer right tail. The lognormal distribution captures the asymmetry of the

dashed line reflects weighted average given the density of “better” and “worse” homes at each level of CLTV (shown in the bottom panel). An increase in the CLTV ratio over the 70% to 100% range decreases the sale probability for both types of homes and as per the “better homes” hypothesis the sale rate is lower for “better” homes across the board.

However, the decreasing share of “better” homes at higher CLTV levels generates a composition effect whereby higher CLTV levels appear to raise the annual probability of sale. This composition effect does *not* reflect causality. Outside of the 70% to 100% CLTV range the composition effect is immediately evident in the positive slope of the dashed line, whereas within this range it merely flattens the negative slope of the dashed line compared to the causal effect captured by the solid lines. The dashed line essentially mimics the relationship between CLTV and the annual sale probability seen in the raw data, as in Figure 1.6a. Further, the moderate slope of the dashed line qualitatively corresponds to the naive OLS estimates of the effect of CLTV on the probability of sale in the 70% to 100% CLTV range, as they appear in column 1 of Table 1.9, whereas the steeper slopes of the solid lines correspond to the IV estimates over this range that reflect the *causal* effect of CLTV, as they appear in column 3 of Table 1.9.

But are the assumptions underpinning the “better homes” hypothesis and the simulation true to reality? To answer this question I test among observably similar homes whether those that tend to sell less frequently also tend to incur smaller losses in value when aggregate housing prices fall (and to have had larger down payments). I quantify properties’ tendency to sell more or less frequently by observing their average annual sale rate in the 10 years preceding my sample, from Jan 1st 1997 through December 31st 2006 (omitting properties built after Jan 1st 1997). I then regress these properties’ year-on-year changes in estimated value when housing prices fell during the recent housing crisis on their prior sale rate, while controlling for observables and conducting the estimate within sets of properties experiencing identical aggregate price changes. Specifically, I estimate the regression

$$\Delta P_{i,t,t-4} = \beta SR_i + \mathbf{X}_{it}\delta + \psi_{lct} + \epsilon_{it},$$

where $\Delta P_{i,t,t-4}$ is the percentage year-on-year change in the estimated value of property i between time (quarter) $t - 4$ and t ; the sale rate, SR_i , is the average annual number of sales recorded for property i , built no later than 1996, from Jan 1st 1997 through December 31st 2006; \mathbf{X}_{it} is the full vector of control variables detailed in Table 1.2; ψ_{lct} is a saturated set of zipcode by quarter by time of purchase fixed effects and ϵ_{it} is an error term. The fixed effects ψ_{lct} ensure that the estimate is conducted within sets of properties with identical aggregate house price histories since their last purchase. All properties within such a cell experience - by construction - the same change in the aggregate local housing price index, so identifying variation in housing price changes stems from *changes in individual home values, conditional on aggregate changes in home values*. These fixed effects also eliminate any potential confounding effects generated by any other conditions that vary over time, location or cohort of purchase. The sample is limited to properties observed in quarter by

distribution of down payments, and generates only a negligible density of down payments to the right of 100%.)

Table A.3.1: The Correlation Between Properties' Prior Sale Frequency and the Sensitivity of Their Value to Decreases in the Local Housing Price Index

	Year-on-year change in home value (%)
Average annual sale rate (1997-2006)	-0.0060*** (0.0009)
Owner, home and loan characteristics	+
Zip \times qtr. of obs. \times qtr. of purchase FE	+
N	757,639

Notes: Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. The sample is also limited to properties observed in 2008-2011, in zipcode by quarter cells experiencing year-on-year decreases in the zipcode area HPI, and that were built no later than 1996. For the list and description of included home, owner and loan characteristics see Table 1.2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

zipcode cells experiencing a year-on-year decrease in the zipcode area housing price index. The sample is also limited to properties observed in 2008-2011, so that sales determining a property's sale rate do not coincide with the period of the latest year-on-year change in value. Finally, the sample is confined to properties built no later than 1996 in order to avoid a "mechanical" bias created by newly constructed properties. Such properties have a "mechanically" higher than average sale rate in their first several years of existence because their history necessarily begins with a sale (e.g. a property built one month ago has a misleading average of 12 sales per year), and all else equal they are likely to have higher value than older properties.

The estimate reported in Table A.3.1 indicates that, on average, the value of homes which sold more frequently during 1997-2006 fell more sharply than observably similar homes which experienced the same aggregate changes in housing prices, validating the key assumption behind the "better homes" hypothesis. To assess whether the owners of "better" homes tend to make somewhat larger down payments I estimate a similar regression in which I replace $\Delta P_{i,t,t-4}$ with the initial CLTV ratio from the most recent purchase. The estimate reported in Table A.3.2 indicates that on average the owners of homes which sold more frequently during 1997-2006 had higher initial CLTV ratios, implying that they did in fact tend to make smaller down payments.

Table A.3.2: The Correlation Between Properties' Prior Sale Frequency and Down Payment

	Down Payment (% of purchase price)
Average annual sale rate (1997-2006)	3.136*** (0.805)
Owner, home and loan characteristics	+
Zip \times qtr. of obs. \times qtr. of purchase FE	+
N	757,639

Notes: Sample includes only owner-occupied homes, whose occupants have tenure duration of 1 to 10 years and do not concurrently own more than two properties. The sample is also limited to properties observed in 2008-2011, in zipcode by quarter cells experiencing year-on-year decreases in the zipcode area HPI, and that were built no later than 1996. For the list and description of included home, owner and loan characteristics see Table 1.2 and the accompanying notes. Standard errors clustered in 72 zip code areas. One, two and three asterisks reflect statistical significance at the 10, 5 and 1 percent levels, respectively.

Appendix B

Appendix to Chapter 3

B.1 A risk-neutral proxy as a simplifying assumption

Suppose the proxy is strictly risk-averse, so his current-period utility function, $u(\cdot)$, is continuous, strictly concave and strictly increasing. By (3.3), (3.4) and Jensen's inequality:

$$P_i(\cdot) \cdot u[\tau Y_i] \leq u[P_i(\cdot) \cdot \tau Y_i]$$

Because $u(\cdot)$ is strictly increasing, there exists a function $\tilde{P}_i : \mathfrak{R} \rightarrow [0, 1]$ referred to as the *risk-adjusted fighting function*, s.t.:

$$P_i(\cdot) \cdot u[\tau Y_i] \equiv u[\tilde{P}_i(\cdot) \cdot \tau Y_i] \quad (\text{B.1})$$

Note that when $P(\cdot) = 1$ then $\tilde{P}(\cdot) = 1$ and recall that $u(0) = 0$ by normalization, so that when $P(\cdot) = 0$ then $\tilde{P}(\cdot) = 0$. Otherwise $\tilde{P}(\cdot) \in (0, P(\cdot))$. Because equation (B.1) is an equivalence it will hold for the derivatives of both sides, yielding:

$$\begin{aligned} \frac{\partial}{\partial(\cdot)}[u(\tilde{P}_i(\cdot) \cdot \tau Y_i)] &\equiv \frac{\partial}{\partial(\cdot)}[P_i(\cdot) \cdot u(\tau Y_i)] \\ \Rightarrow \tilde{P}'_i(\cdot) &= P'_i(\cdot) \cdot \frac{u(\tau Y_i)}{\tau Y_i \cdot u'(\tilde{P}_i(\cdot) \cdot \tau Y_i)} > 0 \quad \forall(\cdot) \end{aligned}$$

where the strict inequality follows from (3.4) and from the characterization and normalization of $u(\cdot)$. Thus, $\tilde{P}_i(\cdot)$ is strictly increasing.

B.2 A myopic proxy as a simplifying assumption

Suppose the proxy is not myopic, but that all other assumptions made in the main text remain applicable. In particular, assume the proxy's inter-temporal utility function is of the

“standard”, time-consistent form:

$$U_{i,t} = \sum_{s=0}^{\infty} \beta^s u_{i,t+s}; \quad \beta \in [0, 1)$$

Every period the proxy maximizes $E(U_{i,t})$ over the policy set $\{comply_t, revolt_t\}$. It is useful to express the proxy’s optimal policy in terms of a value function, V :

$$\max\{E(U_{i,t})\} \equiv V(M_{i,t}, Y_i) \equiv \max\{V_{comply_t}, V_{revolt_t}\}$$

where:¹

$$V_{comply_t} \equiv u(M_{i,t}\tau Y_i) + \beta V(M_{i,t+1}|comply_t, Y_i) \tag{B.2}$$

$$V_{revolt_t} \equiv E[u((P_{i,t}|M_{i,t-1}) \cdot \tau Y_i)] + \beta[P^- \cdot \underline{U} + P^+ \cdot \bar{U}] \tag{B.3}$$

where:

$$P^- \equiv Prob[(P_{i,t}|M_{i,t-1}) = 0] = Prob(\text{Revolt fails});$$

$$P^+ \equiv Prob[(P_{i,t}|M_{i,t-1}) > 0] = Prob(\text{Revolt succeeds});$$

\underline{U} is the proxy’s long-run utility if he is crushed by the state and \bar{U} is his long-run utility if he successfully revolts. Naturally, I assume $\underline{U} < \bar{U}$. It is also reasonable to assume that \bar{U} depends positively on the amount of resources the proxy retains after revolting, seeing as he may need to confront the state militarily later on. The last assumption, however, is inconsequential and so I omit it in what follows.

In every period the state provides the proxy with the smallest share, $M_{i,t}$, such that the proxy chooses to comply (assume that when indifferent, the proxy complies), as long as the proxy’s share is below the critical level, \bar{M} . Hence, in every such period $M_{i,t}$ is set so that:

$$V(M_{i,t}, Y_i) = V_{comply_t} = V_{revolt_t} \tag{B.4}$$

Note from equation (B.3) that V_{revolt_t} does not depend on $M_{i,t}$. This independence occurs because once the proxy chooses to revolt in period t , his prospects thereafter are not affected by the counterfactual proxy’s share that he would have received had he chosen to comply.

Now, in order to simplify the situation, I assume that changes over time in the proxy’s share were very gradual if a proxy did not revolt, sufficiently so for them to be negligible

¹Note that the additive form of V_{revolt_t} below does not indicate risk neutrality, because $V(\cdot, \cdot)$ can be adapted to incorporate risk-aversion similarly to $P(\cdot)$, as shown in appendix section B.1.

within the proxy's foreseeable horizon. This implies:

$$M_{i,t+1}|comply_t \simeq M_{i,t}$$

Equations (B.2) and (B.4) now yield:

$$V_{comply_t} \simeq \frac{u(M_{i,t}\tau Y_i)}{1 - \beta} \quad (\text{B.5})$$

Substituting equation (B.5) into (B.2) and using these to re-write condition (B.4) gives:

$$\frac{u(M_{i,t}\tau Y_i)}{1 - \beta} \simeq E[u((P_{i,t}|M_{i,t-1}) \cdot \tau Y_i)] + \beta \cdot [P^- \cdot \underline{U} + P^+ \cdot \bar{U}]$$

Finally, applying the simplifying assumption of a risk-neutral proxy and re-arranging yields an equation for the dynamic path of $M_{i,t}$ when proxies are non-myopic:

$$M_{i,t} \simeq (1 - \beta) \cdot (P_{i,t}|M_{i,t-1}) + \frac{\beta \cdot (1 - \beta)}{u(\tau Y_i)} \cdot [P^- \cdot \underline{U} + P^+ \cdot \bar{U}] \quad (\text{B.6})$$

In light of equations (3.4) and (3.6), $P_{i,t}$ is increasing in $M_{i,t-1}$. Therefore, given the dynamic path, equation (B.6) and the definitions of P^- and P^+ , $M_{i,t+1}$ is unambiguously increasing in $M_{i,t}$ when the two are below \bar{M}_i . This state of affairs is qualitatively similar to that when proxies are myopic.