Housing Speculation and Entrepreneurship†

Xuan Tian
PBC School of Finance,
Tsinghua University
tianx@pbcsf.tsinghua.edu.cn

Yichu Wang
PBC School of Finance,
Tsinghua University
wangych.18@pbcsf.tsinghua.edu.cn

Current version: Jan 2023

Abstract

We document a speculation channel through which house market booms negatively affect entrepreneurship. To address endogeneity concerns, we exploit plausibly exogenous variation in house prices generated by staggered and unintended policy spillovers in China. We find house market speculation triggered by house booms crowds out entrepreneurship. Reduced labor supply, reduced capital supply, and heightened entry costs do not appear to explain our main findings. The negative effect exhibits in the OECD countries as well. Our paper complements the well-documented collateral channel by offering novel evidence on a previously under-explored adverse consequence of house market booms – their hindrance to entrepreneurship.

Keywords: Housing speculation; entrepreneurship; house purchase restriction policy
JEL code: L26, D84; G28; G51

---

† We thank Zhuo Chen, Grace Hu, Chen Lin, Zhan Shi, Mingzhu Tai, Yue Yuan, Hanyi (Livia) Yi, and seminar participants at Tsinghua University, Fudan University, Nottingham Business School, and 2022 China Finance Annual Meeting for their helpful suggestions and comments. We thank Jinxin Yu and Yuan Sun for their valuable research assistance. We acknowledge financial support from the National Natural Science Foundation of China (Grant No. 71825002, 71790591). We remain responsible for all errors and omissions.

1 Introduction

The real effects of real estate cycles have long drawn intensive attention by both academics and policy makers, especially since the 2008 financial crisis (Mian et al., 2015; Gao et al., 2020). Entrepreneurship, as a key engine of the Schumpeterian economic growth, reflects a direct consequence that house market booms and busts could have on the real economy. Existing studies typically link house prices to entrepreneurship through the well-documented collateral channel (e.g., Adelino et al., 2015; Corradin and Popov, 2015; Schmalz et al., 2017), in the spirit that entrepreneurs could pledge their houses or wealth in hand to extract home equity, and thereby alleviate credit constraints resulted from information asymmetry (Stiglitz and Weiss, 1981). While this positive effect of house price appreciation on business formation is confirmed by subsequent studies, research along this line of inquiry mainly focuses on the role of houses in alleviating credit constraints for a pre-determined group of (potential) entrepreneurs, yet ignores the role of house prices in driving the opportunity cost (or their outside options) and thereby affecting participant incentives of entrepreneurship.[1] In this paper, we examine the effect of house prices on entrepreneurship from a different angle by taking into account the role of real estate cycles that could alter the opportunity cost. Specifically, we hypothesize that house price appreciation induces housing speculation, which significantly alters the opportunity cost of business ownership, and thereby crowds out entrepreneurship.

We particularly consider the participant incentives of entrepreneurship. Conceptually speaking, an individual could either start a new business (i.e., entrepreneurship) or invest in the house market (i.e., speculation). Both are risky investment opportunities, and the tradeoff of their returns determines the realization of entrepreneurship. The above tradeoff is typically trivial when the real estate market is stable. This is because a stable housing market is unlikely to provide high enough speculation returns that materially alter an individual's opportunity cost of starting new businesses, especially for those whose projects could ultimately achieve hits (i.e., going public or being acquired with decent bids), and thereby generate considerably higher expected returns than real estate investment. However, during house booms with fast and dramatic, and even long-lasting, house price increases, speculation in the house market could provide low-risk expected returns that are comparable to, or even higher than, businesses ownership, which significantly increases the opportunity cost of entrepreneurship. As a result, house market booms not only allow quitter-entrepreneurs, probably those with less valuable

[1] Some studies, however, show that the revealed relation is either non-linear or economically limited (summarized by Kerr et al., 2015) using alternative data sets or sample periods of house market booms.

entrepreneurial projects, to enjoy private benefits (e.g., leisure or time to accumulate human capital (Iyigun and Owen, 1998)), but also provide the quitters pecuniary compensation due to house ownership. In fact, quitters could even be more aggressive and engage in housing speculation by pledging their houses in place to invest in the booming real estate market and thereby amplify their total speculative returns.[2] Hence, given that entrepreneurs are by nature with higher risk tolerance (Knight, 1921; Hvide and Panos, 2014), the group of marginal entrepreneurs and that of marginal housing speculators can largely overlap. Moreover, because real estate investment requires comparable amount of disposable capital to that of business creation and is likely subject to financial constraints just as entrepreneurs typically are, house ownership and business ownership are to some extent mutually exclusive, which suggests a crowding-out effect of house prices on entrepreneurship.[3] We term this conjecture as the speculation channel of the link between house prices and entrepreneurship.

Following this logic, we empirically test the effects of house price changes and housing speculation on entrepreneurship in this paper. However, a major empirical challenge is that house prices are likely endogenous with market characteristics, including the dynamics of entrepreneurship, for at least three reasons. First, unobservable local economic fundamentals, such as local demand (Kerr et al., 2015) and economic conditions, correlated with both house markets and entrepreneurship could bias the results, i.e., the omitted variable concern. Second, the causality might run from entrepreneurial activities to the house market, i.e., intensive entrepreneurial activities and subsequent employment expansion could alter local house prices. Third, while speculation could lead to real estate markup, house prices *per se*, potentially resulted from both real demand and gambling motives, can hardly disentangle speculating behavior. Hence, a correlation between house prices and business creation could tell us little about the causal effect of house prices on entrepreneurship.

China's two-decade-long house market booms and staggered countrywide policy interventions that are aimed curtailing real estate speculation at the city level provides an ideal setting to address the identification challenges. China has been witnessing a dramatic house

[2] Plenty of anecdotal evidence shows that this case happens in past stock market booms when people tend to pledge or even sell their houses for engaging into stock market speculation.

[3] We acknowledge that real estate investment and business venturing are not necessarily mutually exclusive. For example, if one can somehow foreknow that an entrepreneurial project can definitely make a home-run such as Facebook or Amazon, there is no doubt that borrowing against her real estate holdings to finance the entrepreneurial project is the first-best choice. However, entrepreneurship means risk-taking, and it is likely that the majority of entrepreneurial projects turn out to be failures. In these cases, failing entrepreneurs can hardly keep their collateralized houses. Hence, in expectation, there are still tradeoffs between home ownership and business ownership. Even if entrepreneurs can first invest in real estate and then use the newly purchased houses as collateral to get bank finance, they are still faced with the choice between investing in businesses and buying more houses. Therefore, house ownership and business ownership are to some extent mutually exclusive.

price boom since 2003, which induces heated speculation on the real estate market due to extrapolative expectation (see Section 2 for detailed institutional background). To prevent the break of uncontainable real estate bubbles and the outbreak of systemic crisis, since 2010, China's central government has prompted a countrywide housing purchase restriction (HPR, hereafter) policy at the city level, yet delegated the detailed decisions (e.g., the timing of implementations) to local governments at the province level. As a result, variation in general considerations and administrative efficiency across these local governments leads to a decade-long process of staggered HPR adoptions, cancellations, and re-implementations in the regulated cities (HPR cities, hereafter).

While the HPR policy squeezes speculation out of local housing markets of regulated cities, it leads to an unintended policy spillover (Deng et al., 2021), i.e., the arrival of out-of-town speculators unexpectedly trigger house price booms and housing speculation in those never-regulated surrounding cities (non-HPR cities, hereafter). Deng et al. (2021) show that cities closer to a regulated city, compared to those farther away, are more likely to experience larger house market booms during the wave of HPR policies in 2016 and 2017. Likewise, using a sample of 42 non-HPR cities' monthly house price information between June 2010 and December 2018 from the WIND's 100-city house price database, we reassure that the pattern of HPR spillover effects on house prices documented by Deng et al. (2021) exists in the 2010-2014 wave of HPR policies as well. Thus, to the extent that uninvited and unexpected arrivals of out-of-town speculation are unlikely to be driven by local entrepreneurship, we rely on these geographical spillovers induced by staggered policy interventions as our main identification strategy. Because this strategy excludes HPR cities, the variation in house prices of non-HPR cities is likely exogenous. It is worth noting that a key advantage of this identification strategy is that there are multiple shocks that affect different cities at exogenously different times, which avoids a common identification difficulty faced by studies with a single shock, namely, the existence of potential omitted variables coinciding with the shock that directly affect entrepreneurship.

We make use of a unique longitudinal survey data at the individual level in our main analyses. The data are rich in terms of its individual-level longitudinal panel structure, wide time span and geographical coverage, and granular information on home ownership (especially multi-property ownership). In addition, we use firm registration data to further explore the effect of housing speculation on the outcome of entrepreneurial activities conditional on entry (i.e., we explore the research question along the intensive margin). We also develop a machine learning algorithm to identify firms' industry classifications based on the mandatory disclosure of business scope of newly registered firms.

We begin our analyses with the baseline regressions using the staggered difference-in-differences (DiD) approach. The main results show that, after the arrival of HPR spillovers, individuals in closer non-HPR cities that are exposed more to the housing shock exhibit a 3.2% lower probability in business ownership compare to those in farther non-HPR cities that are exposed less to the shock.[4] The economical magnitude of our estimation is sizable and roughly comparable to previous studies.

We then undertake a battery of tests to check the internal validity and the robustness of our baseline results. First, while the out-of-town HPR spillover shocks are unlikely driven by local economic fundamentals, especially local entrepreneurship, our findings may still be subject to reverse causality concerns, i.e., heating business dynamics create fortune and thereby could push up house prices and ignite housing speculation in HPR cities, which ultimately induces the HPR policy. To address this concern, we follow Bertrand and Mullainathan (2003) and decompose the key explanatory variable, *HPR spillover*, into a set of time indicators and explore the dynamic effect surrounding the arrival of HPR spillover shocks. The results show that the significantly negative effect appears only in the years after the shocks, but are statistically and economically absent in the years before, which suggests that our baseline results are unlikely driven by reverse causality.

Second, we undertake a variety of robustness checks with alternative variable definitions, alternative sample selection filters, alternative model specifications, various diagnostic tests for the staggered DiD estimation, household-level analyses, and city-level analyses using firm registration data. Our findings continue to hold in these robustness checks. Third, we perform a placebo test by re-estimating the baseline regressions with the subsample of renters (i.e., those who own no house). The results suggest that the HPR spillover shocks have no effect on a renter's probability of business ownership, suggesting that our baseline findings are unlikely capturing the effects of factors that are unrelated to the house market.

Fourth, we conduct another placebo test with the HPR spillover shock years artificially assigned. Specifically, we begin by obtaining an empirical distribution of years when cities implement the HPR policy. Next, we randomize the key explanatory variable and assign the falsified HPR spillover shocks to the treatment group. We then re-estimate the baseline regression with falsified key variable of interest and repeat it 1000 times. This approach

[4] Following the existing literature, we refer to business ownership as entrepreneurship.

maintains the distribution of HPR shock years from our baseline specification, but it disrupts the proper assignment of HPR shock years to cities. Therefore, if an unobservable shock occurs at approximately the same time as the HPR policy, it should still reside in the testing framework, and thus have an opportunity to drive the results. If, however, no such shock exists, then our incorrect assignments of HPR years to cities should weaken our results when we re-estimate the baseline tests. Indeed, the results from the Monte Carlo procedures show that randomly falsified HPR spillover shocks do not relate to individuals' entrepreneurial decisions, suggesting that our baseline findings are unlikely driven by event clustering or sample selection induced by an omitted variable.

Fifth, because our identification strategy naturally predicts that the negative effect of house prices on entrepreneurship diminishes after the cancellation of HPR policies (i.e., the reversal shocks), we look at the spillover effect of HPR cancellations on entrepreneurship. By estimating year-by-year spillover differences around HPR cancellations with the control variables and fixed effects the same as those in our baseline estimation, we find that the effect of HPR cancellations exhibit a symmetric pattern compared to what we show in the reverse causality test, i.e., the negative spillover effect is significant before the HPR cancellations but is insignificant and close to zero after the cancellations. Once again, this finding is consistent with our conjecture and further strengthens the identification.

In summary, all the empirical attempts above produce evidence that is consistent with the speculation channel, i.e., house market booms induced by the spillover of HPRs of nearby cities negatively affect entrepreneurship. While each piece of evidence might be subject to alternative interpretations, these pieces of evidence collectively are difficult to reconcile with specific alternative arguments. Hence, these identification attempts suggest that the negative relation between housing speculation and entrepreneurship is likely causal.

Next, we undertake more tests to further explore the speculation channel. We examine the heterogeneity of our baseline results with respect to multi-property ownership. Compared to single-property owners and renters, multi-property owners are more likely to be housing speculators (Gao et al., 2020). Consistent with this conjecture, we first show that multi-property owners have significantly higher expectations on future house price appreciation, easier access to external finance, and stronger preferences for risk-taking. We then interact an individual's multi-property ownership status with the key variable of interest, *HPR spillover*. We find the effect of house prices on entrepreneurship is more pronounced for multi-property owners. Likewise, we find the effect is more accentuated for those with high expectations on local house price changes in the near future. The evidence suggests that the crowding-out effect of housing speculation is accentuated for individuals with a high propensity to be housing speculators.

While our results so far point to a housing speculation channel underlying the negative relation between house prices and entrepreneurship, three alternative interpretations could still explain the results: reduced labor supply caused by the wealth effect (i.e., the labor supply argument), reduced capital supply in entrepreneurial finance (i.e., the capital supply argument), and heightened entry costs (i.e., the entry cost argument). The first explanation argues that multi-property owners can experience a larger wealth effect due to the arrival of house booms and thus are more likely to reduce their labor supply (Li et al., 2020), including entrepreneurial labor supply; the second explanation argues that house booms could alter the participant incentives of entrepreneurial finance providers and drive away capital supply from entrepreneurship to the real estate market, which would lead to a negative effect on entrepreneurship; the last explanation argues that house booms can increase the entry cost of starting new businesses, such as workplace rent, and consequently crowd out entrepreneurship.

To address the labor supply argument, we first conduct a heterogeneous test based on household wealth captured by the value of household net assets. We do not find that the negative effect of house prices on entrepreneurship is altered by household wealth levels, standing against the reduced labor supply caused by the wealth effect argument. Then we include the household wealth proxy and its interaction term with the key explanatory variable into the heterogeneity tests based on multi-property ownership status and house price expectations. The results show that the negative effect of *HPR spillover* is still more pronounced for multi-property owners and those with high expectations of house appreciation, and remains largely unchanged even after controlling for household wealth levels. This finding suggests that the reduction in labor supply caused by the wealth effect cannot explain away the speculation channel. In addition, we replace the entrepreneurship dependent variable with two measures on an individual's labor supply choice, i.e., the *Work-or-quit* dummy and the logarithm of one's wage from being employed, and re-estimate the baseline model as well as the heterogeneity tests, to explore whether our findings reflect reduced labor supply instead of deterred entrepreneurship. On contrary to the labor supply argument, with the two labor supply proxies as dependent variables, we could no longer find any significant result similar to what we get with the entrepreneurship proxy as the dependent variable. This finding helps further rule out the labor supply argument.

To address the capital supply argument, we use a city's level of venture capital investment activities (*VC inv. score*) constructed by Dai et al. (2021). We first run city-level regression that is analogous with our baseline model but replace the dependent variable with *VC inv. score*. While our empirical model captures a significant city-level effect of *HPR spillover* on the number of new businesses (reported in robustness checks), we do not find significant relation between the uninvited and unexpected house price changes in a city and the city's level of VC investment activities. We then conduct heterogeneity tests at the individual level based on *VC inv. score*, and find that the interaction term between *HPR spillover* and *VC inv. score* is economically close to zero and statistically insignificant. The two sets of tests collectively suggest that reduced capital supply in entrepreneurial finance (measured by the level of VC investment activities) unlikely explains our main results. While our housing speculation argument actually does not deny the possibility of changes in entrepreneurial investor's participant incentives, we do not observe evidence that this capital supply argument could explain away the housing speculation channel.

To address the entry cost argument that predicts the negative effect of house prices on entrepreneurship is more pronounced for individuals with more financial constraints, we undertake heterogeneity tests with alternative indicators of individuals' levels of financial constraints that could help differentiate the speculation channel from the entry cost argument. The first set of tests examines the heterogeneous effect based on a respondent's answers to her subjective attitudes. We find that the crowding-out effect is more pronounced for individuals who have higher income, have more positive sense of wealth, are more confident to their future, do not suffer from inequality, and are less educated. To the extent that speculators are likely those who are wealthy, optimistic, and less educated (and thus more subject to speculative impulses), all these pieces of evidence collectively are consistent with the speculation channel and contradicting to the entry cost argument.

The second set of tests explores how a city's supply of disposable liquidity alters the baseline results. The rationality behind is that cities with higher availability of disposable finance are more subject to housing speculation, given the geographical segmentation of financial markets across cities in China (Huang et al., 2020). We find that the negative effect of house prices on entrepreneurship is more pronounced if the individual is in a city with higher income inequality (and thus more disposable liquidity due to limited marginal propensity of consumption), with larger amount of citizen deposit, and with higher supply of bank loans. These observations suggest that entrepreneurial activities are more sensitive to the crowding-out effect of housing speculation in cities with higher availability of disposable finance. Again, the entry cost argument cannot reconcile with these findings. Further evidence with local GDP, fiscal expenditures, and unemployment controlled suggests that the heterogeneous effect of HPR spillovers cannot be explained by local economic conditions.

In the final part of the paper, we undertake two additional tests. First, we push the research question along the intensive margin, i.e., given startups are founded, what is the relation between housing speculation and entrepreneurial outcomes, using firm registration data? We develop a machine learning approach to identify firms' industry classifications based on their business scopes. We run baseline regressions at both the firm level and the city level. Though we find that HPR spillovers have no effect on firms' sizes at creation conditional on entry (which, once again, contradicts to the entry cost argument), the startups exhibit higher survival rates. Recalling our earlier findings that the baseline results are more pronounced for individuals with a lower level of education, these findings together suggest that housing speculation is more likely to serve as a screening mechanism that screens out "low-quality" and "oscillating" entrepreneurs. Housing speculation, however, is unlikely associated with financial constraints that hinder entrepreneurship, which is consistent to our conjecture that house market booms could give rise to the opportunity cost of capital and thereby crowd out entrepreneurship. The results are also complementary to the existing findings on the collateral channel, e.g., Jensen et al. (2021), in the sense that the housing collateral supports high-quality entrepreneurs while the speculation channel is likely to screen those low-quality entrepreneurs out.

Second, we check the external validity of the speculation channel using cross-country data from the OECD countries. The OECD group has been experiencing a widespread real estate boom resulted from low interest rates since 2015. The international comparison shows that the 2015-2019 changes in entrepreneurship index are in general negatively and concavely correlated with house price growth. Furthermore, we dichotomize the sample with the median housing growth rate. The subsample analysis suggests that, on one hand, the relation between house price growth and entrepreneurship is moderately positive in the countries with stabler house markets (i.e., housing appreciation below median), such as the U.S., France, Italy, and Japan, which is consistent with the collateral channel. On the other hand, the relation between house price growth and entrepreneurship is negative in the countries with relatively booming house markets (i.e., housing appreciation above median), such as Norway, Hungary, Portuguese, and Turkey, which is more in line with the speculation channel. The cross-country evidence helps the generalization of our main results. It also suggests that the market-contingent speculating behavior in the real estate market could "mask" the positive relation predicted by the collateral channel as well, especially during house market booms, which provides a possible explanation that helps reconcile the seemingly mixed findings in the existing literature summarized by Kerr et al. (2015).

Our study contributes to two strands of literature. First, it adds to the literature on house prices and entrepreneurship. Existing literature explains the effects of house prices on entrepreneurial activities with the collateral channel (e.g., Adelino et al., 2015; Corradin and Popov, 2015; Schmalz et al., 2017). The rationality behind is that, since financial constraint is a major barrier against starting new ventures, house price appreciation gives rise to home equity and thus help alleviate this friction. However, some studies, such as Hurst and Lusardi (2004), Kerr et al. (2015), and Jensen et al. (2021), among others, collectively find that the effect of house money on business entry is either non-linear or economically limited with alternative data sources or alternative sample periods. By focusing on the role of house prices in driving the opportunity cost of starting new businesses, our study provides a new angle on the relation between house markets and entrepreneurship and examines the speculation channel, which is previously under-explored. We note that the speculation channel is by no means on contrary, but rather complementary, to the stream of entrepreneurship literature on the collateral channel, because we find the relation between house prices and entrepreneurship depends on the degree of house price appreciation in a market.

Second, our paper is related to the literature on economic distortions of house market booms and busts. This topic draws particularly intensive attention in the recent post-crisis decade. Existing literature has explored the effects of house market booms from various perspectives, such as household leverage (Mian and Sufi, 2011), educational choices (Charles et al., 2015), labor market (Charles et al., 2019; Li et al, 2020), firm behavior (Bahaj et al., 2020), corporate innovation (Mao, 2021), welfare of home owners and renters (Favilukis and Van Nieuwerburgh, 2021), and banking credit (Mian and Sufi, 2022). As a seminal study, Chaney et al. (2012) show that housing appreciation could enhance corporate investment via the collateral channel. Compared to the existing literature that mainly documents an expansionary effect of booming house prices on the real economy, some recent studies tend to underline negative relations. For example, Gao et al. (2020) show that the effect of housing speculation during booming periods can propagate to the real economy during the bust, and thus account for the decline in construction and local retail and service sectors during busting periods. More directly, Chakraborty et al. (2018) find that the rising proportion of mortgage loans in a booming real estate market crowds out commercial lending, and thus house market booms have negative spillover to the real economy. Likewise, Chen et al. (2017) find that firms in China extract resources from their core business to engage into real estate speculation. We contribute the debate by supplying micro evidence for a crowding-out effect, i.e., the speculation channel, of house market booms on entrepreneurship.

Our paper is related to two recent studies. Li and Wu (2014) use a cross-sectional data in 2010 and find evidence consistent with ours. Han et al. (2020) also use a cross-sectional data in 2017 and document a negative association between house price risk and entrepreneurial activities. Our paper is distinct from them in three important ways. First, we make use of a staggered quasi-experimental setting to establish the causal relation between house prices and entrepreneurship, while the two papers are agnostic on the identification issue. Second, we use longitudinal panel data covering 10 years to explore the speculation channel, which allows us to observe the dynamics of home ownership and business creation. Third, using cross-country data, we are able to reconcile the seemingly mixed observations on the relation between house prices and entrepreneurship, pointing out that the relation depends on a country's housing market booming level.

The remainder of the paper is organized as follows. Section 2 discusses the institutional background of China's HPR policy. Section 3 describes sample selection and the identification strategy. Section 4 presents the main findings and robustness tests. Section 5 provides evidence ruling out alternative interpretations. Section 6 reports further analyses on the effect of housing speculation on entrepreneurial outcomes and on entrepreneurial activities around the world. Section 7 concludes.

2 Institutional background on China's HPR policy

China has been experiencing a vast and long-lasting housing boom since 2003 (Liu and Xiong, 2020). Average annual growth of house prices reaches 13.1% for the four metropolises (i.e., Beijing, Shanghai, Guangzhou, and Shenzhen), and 10.5% (7.9%) for the second tier (third-tier) cities (Fang et al., 2016). Likewise, many cities witness rapid growth of an over 20% annual compounding rate in real land prices during 2003-2011 (Wu et al., 2015).

Existing literature has found that China's house market booms induce strong and widespread speculating motives, which, in return, strengthen the housing bubble's self-fulfilling (see Liu and Xiong (2020) for a summary of typical facts and past literature). Since the real estate industry takes a major proportion in China's economic growth as well as government revenues, it leads to a strong perception that the housing market is "too important to fail" (Liu and Xiong, 2020), i.e., the government would do anything to avoid housing market crash. In addition, China's vast economic growth and quickly rising household incomes can be supportive enough to its surging housing market, and make the boom distinct from a pure bubble like in the ante-crisis U.S. market. As a result, people's solid confidence on implicit government guarantees on China's housing market, along with the cheerful economic fundamentals, reasonably encourage the contagion of speculating. In addition, China has never formally compelled any kind of taxation on real estate holdings, which further supports a surging housing speculation.

However, a "too important to fail" sector also implies an intensive and dangerous origin of systemic risks for the economy. The housing boom leads to substantial concerns, from not only academics but also China's policy makers, that the long-lasting real estate appreciation and widespread speculation could have developed into a real "gigantic housing bubble" (Song and Xiong, 2018), the burst of which could ultimately damage the whole financial systems and the economy (Wu et al., 2015; Glaeser et al., 2017; Chen and Wen, 2017).

To avoid coming to the break of an uncontainable real estate bubble, since 2010, China's central government has planned a nationwide intervention on the housing market by imposing the HPR policy. Basically, the policy is in the spirit of squeezing out housing speculation at the city-level through various tools, which is concluded by China's Central Economic Working Conference, the highest institute of making decisions on nationwide economic policies, with a highly influential slogan: houses should be for living in, but not for speculative investment.[5]

According to our hand-collected detailed information about the HPR policy, by the end of 2019, 68 out of 293 prefecture-level cities have ever adopted the restrictive policy on house purchases (we refer them as the HPR cities hereafter), geographically spreading across almost all the provinces of mainland China (see Internet Appendix A for detailed summary of the HPR policies across prefectures).[6,7] The HPR policies typically include four components. First, purchase restrictions, i.e., prohibiting families to buy a second or third house; second, loan restrictions, i.e., raising the requirement on down payment proportions; third, resale restrictions, i.e., requiring a lock-up period of 2-5 years for house resale; fourth, price restrictions, i.e., setting a maximum deal price that is fixed within a period (usually 2-12 months). The four components are not all indispensable for each HPR policy: usually the former two components, i.e., purchase restrictions and loan restrictions, are dominant, while the latter two are auxiliary. While there are variations on the supplementary measures of HPR policies (besides purchase

[5] Another well-known example is that Mr. Li Keqiang, the Premiere of China, once quoted an old saying in China, "a piece of land brings peace of mind", in a press conference, to describe the policy target, i.e., leading house prices back to the level reflecting real needs instead of speculative motives, which enhances the affordability of houses for majority people.
[6] From now on, to simplify, the word "city" in this paper denotes a prefecture-level city, i.e., we consider only prefectures in our paper, which is a standard execution in China-related studies (e.g., Huang et al., 2020).
[7] The only exception is Tibet where no city has ever been regulated by a HPR policy.

restrictions) across cities, the effect of those differences are merely limited, i.e., the effects of HPR policies mainly depend on the timing of policy adoptions and cancellations because all the policies contain purchase restrictions.

Although the HPR is prompted by the central government that sets general policy targets, decisions on concrete components and adoption timing are delegated to provincial governments. Moreover, after deciding the local restrictive policy, provincial governments further delegate the HPR implementation and supervision to municipal governments of those regulated cities. As a result, this top-down process leads to differences in the exact time of policy adoptions and cancellations across regulated cities all over the country. This is because local governments at each layer are different from each other in their general consideration as well as administrative efficiency to carry out the HPR policy.[8]

In general, the adoption of the HPR policy can be roughly divided into two major waves: the 2010-2014 wave (Round 1) and the 2016-17 wave (Round 2). The former wave contains 45 HPR cities, of which the starting point of adoptions ranges from 2010 to 2011. Since 2013, the HPR cities begin to repeal the policy in succession, except for five cities that never quit the policy. While the process of HPR cancellations continues until 2016, most of the first-wave HPR cities had exited the restrictions by the end of 2014.[9] After that, except for the five always-regulated cities, 9 HPR cities has never adopted the policy again by 2020, while the other 31 cities re-implement the HPR policy in 2016 or 2017, which makes the latter wave of restrictions. The 2016-2017 wave contains 54 cities in total, 23 of which are regulated for the first time. Thus, the HPR policy generates much variation in timing, as well as regulated cities, of the HPR adoptions, cancellations, and re-implementations.

Looking back on China's decade of HPR policies, it effectively controls local house price growth. It, however, also spills the squeezed-out speculation over to surrounding cities and ultimately triggers house market booms in these never-regulated cities. As shown by Deng et al. (2021), house price growth and housing transaction volume drop significantly in the HPR cities, which is simultaneous with the unexpected increase in the non-HPR cities with a comparable magnitude. In addition, they show a diminishing-with-distance effect, i.e., the real

[8] Since real estate and related industries take a sizable proportion in local gross domestic production, municipal governments would typically weigh controlling housing bubbles with supporting economic growth in policy making (Liu and Xiong, 2020). For provincial governments, city-level economic growth is of certainly the main focus, but more or less taken into their general consideration.

[9] The last city to have quit the policy in the first HPR wave is Zhuhai, cancelled the HPR policy in 2016 but soon re-picked the HPR policy.

estate appreciation in nearby non-HPR cities is greater than that in faraway ones, which points to an uninvited geographical spillover of the HPR policy.

Overall, China's countrywide policy interventions in 2010s, of which the concrete components are largely consistent with, and efficiently supportive to, the announced policy target (i.e., houses should be for living in, but not for speculative investment) offer a staggered quasi-experiment that generate plausibly exogenous variation in house prices, which could help identify the effect of housing speculation on entrepreneurship.

3 Identification strategy, data, and sample selection

3.1 Identification strategy

To address the endogeneity concerns discussed before and establish a causal link between housing speculation and entrepreneurship, we rely on the HPR policy in China, which provides a quasi-natural experiment for our study (Chen et al., 2017). First, the policy is aimed at squeezing speculators out of the local house market. Looking into the policy in details, most of policy components are targeted at purchase restrictions on a family's second or third house.[10] This is because, compared to individuals with no house or only one house to live in, multi-property owners are more likely housing speculators. The explicit policy target, as well as consistent enforcement, helps us identify the speculating part of housing price changes.

Second, while the restriction plan is prompt by the central government, the timing of enforcement is determined by provincial governments, and thus is different across HPR cities due to the fact that local governments vary in efficiency and consideration. Consequently, most of HPR cities have experienced a cycle of policy adoptions, cancellations, and re-implementations during 2010-2020, which offers staggered shocks on house prices. A key advantage of this identification strategy is that there are multiple shocks that affect different cities at exogenously different times, which avoids a common identification difficulty faced by studies with a single shock, namely, the existence of potential omitted variables coinciding with the shock that directly affect entrepreneurship.

Third, and perhaps more importantly, China's top-down policy making process assures that the policy is hardly resulted from lobbying by potential entrepreneurs (i.e., individuals who are interested in starting new businesses lobby local governments to contain the housing market

[10] Typically, for a local family, the policy restricts purchases of a third house, while non-local and unmarried persons are only allowed to own one house. The reason is that second houses can be out of real needs, instead of speculation, for families with a married couple, e.g., it is common for parents in China to buy an extra house for their child (especially males) who is getting married.

in order to reduce the entry cost or opportunity cost for starting new ventures), which is commonly observed in the U.S. and European countries. Hence, the variation in house price changes generated by the HPR policy is likely exogenous.

However, one major concern of using the HPR policy to tackle the endogeneity issue is that the identification would still be problematic if we directly compared individuals in HPR cities and those in non-HPR cities in a DiD framework, because HPR and non-HPR cities may not be completely randomly assigned. To elaborate, a HPR city is typically the capital city or an utmost important city within a province, and is therefore systematically different from non-HPR cities; individuals living in a core city within the province are likely better educated, wealthier, and frequently exposed to entrepreneurial ideas, and thus are not comparable to those in other cities. In addition, since real estate correspond to a high proportion of local economy development, which is crucial for local official promotions, policy makers in China's provincial governments always weight the benefit and cost of containing housing markets of core cities. In other words, while the timing of HPR implementations and cancellations is determined by provincial governments instead of municipal governments, local housing market and economic development could still affect HPR related decisions, especially the timing of HPR policies.[11] Therefore, given that entrepreneurship is a key engine of local economic growth, the comparison between HPR and non-HPR cities may not effectively address endogeneity concerns.

To alleviate this concern, we follow Deng et al. (2021) and use the variation in housing speculation generated by unintended geographical spillovers of HPR policies as our main identification strategy. The spirit of this strategy is that real estate speculators, crowded out of the housing market in the HPR cities, continue their speculating behavior but turn to those surrounding non-HPR cities, with the expectation that the housing booms would systematically continue in the cities that are not regulated by HPR policies. Consequently, as explicitly shown by Deng et al. (2021), for non-HPR cities, the arrival of those crowd-in speculators exogenously gives rise to house prices in the local housing market. Likewise, usually accompanied with restrictions on banks' mortgage loans, HPR policies force commercial banks redistribute their loan resources out of HPR cities to non-HPR cities, which also helps push up house prices and makes it easier for housing speculation. In summary, while the HPR policy negatively affects house price growth in the HPR cities, which may be subject to endogeneity concerns, HPR

[11] In fact, leaders of a core city are often in the standing committee of their provincial government with a relatively high rank, which further amplifies the voice from core cities in the consideration of whether, when, and how to enforce or cancel an HPR policy.

spillovers generate plausibly exogenous variation in house prices in non-HPR cities.[12] As Deng et al. (2021) has documented, HPR spillovers are descending with the distance to the HPR cities, i.e., those non-HPR cities closer to HPR cities have stronger HPR spillovers than those non-HPR cities that are farther away from HPR cities, which naturally constructs a DiD setting. Specifically, the treatment group consists of individuals in closer non-HPR cities and the control group consists of those in farther non-HPR cities.

While Deng et al. (2021) have documented the geographical spillover effects of HPR policies on non-HPR cities' house prices, they mainly focus on the second wave of HPR policy implications (the 2016-2017 wave). To reassure that the positive HPR spillovers on house prices (i.e., the first stage of our identification strategy) exist in both waves of house purchase restrictions, we plot the monthly time series of the house price differences between nearer non-HPR cities and farther ones (defined in details in Section 4.1). We obtain city-level house price data from WIND's 100-city average house price database, out of which 42 cities are non-HPR cities, covering June 2011 to December 2018.[13] House prices are taken logarithms and scaled by city means to make them comparable across cities, because we intend to capture the changes in house prices due to the policy spillovers rather than the time-invariant price gaps between the two groups of cities.

Figure 1 plots the monthly estimated logarithm differences on house prices between nearer (treated) cities and farther (control) cities defined in our main analysis, which clearly show two waves of house price divergence between the two groups of cities following the arrival of spillovers of the HPR policies. Specifically, after the entering phase (i.e., first-wave HPR cities successively adopt the policies for the first time and gradually finalize the concrete measures) during HPR wave 1, house prices in nearer non-HPR cities start to outgrow those in farther cities, especially since 2012 fall. The positive price gap becomes larger until May 2014. After that, HPR cities exit the restrictions in succession, and the house price gap between non-HPR cities begin to regress to zero. The diminished house price differences persist until June 2016, the starting point of the second wave of the house price restrictions. It seems that, while the first wave of house purchase restrictions witnesses a roughly one-year time lag between the adoption of policies and the divergence on house prices, the second wave of HPR policies has

[12] Another advantage of excluding HPR cities from our identification tests is that it helps rule out the possibility that mega-cities, such as Beijing, Shanghai, and Shenzhen, drive our results. This is because these cities are not in the sample.

[13] Deng et al. (2021)'s house price database has a better coverage (307 cities), but it ends in 2017 and is non-public. WIND's 100-city house price database is one of the best publicly accessible coverages of China's city-level house prices as far as we know. In addition, we are able to replicate Deng et al. (2021)'s first stage result using our data and get similar coefficient estimates as those in their Table 2 Column 1, suggesting that our house price data are comparable and representative.

larger and more immediate spillovers. The price gap become positive and significant again, and keep enlarging through the end of 2018 when the gap size is around tripled compared to the peak value in the first wave of restrictions.

To sum up, exploiting the geographical spillovers of China's HPR policies, our identification strategy captures uninvited and unexpected house booms (due to the arrival of out-of-town housing speculators), which possibly triggers local housing speculation and hence crowds out entrepreneurship.

3.2 Data and baseline sample construction

To study the effect of housing speculation on entrepreneurship, we use the China Family Panel Studies (CFPS).[14] The CFPS is a countrywide longitudinal survey that focuses on all aspects of each respondent's characteristics in a year, including the respondent's business ownership, home ownership, background, wealth and income, living environment, health, family and children, and even attitudes toward her own life and the society. This data set provides panel data at both the household level and the individual level. After a two-year pilot execution, the survey formally begins in 2010 on a biennial basis. In this paper, we use all five waves of surveys (i.e., in 2010, 2012, 2014, 2016, and 2018), with a representatively wide geographical coverage of 122 major prefectures of 25 provinces.[15]

Compared to other similar data sets used in previous entrepreneurship research (e.g., the U.S. Panel Study of Income Dynamic or the French Labor Force Statistics), the CFPS is either analogous or advantageous in three ways. First, the survey provides longitudinal data at the individual level, i.e., each respondent (the unit of observation) is assigned with a unique identifier and is repeatedly surveyed. The individual-year panel allows us to explore the transition between non-entrepreneurs and entrepreneurs by including the individual fixed effects in regressions, instead of only capturing a compositional effect. Second, the five waves of survey cover a 10-year panel, which supplies a rich set of variations to investigate the effects of staggered implementations and cancellations of the HPR policy on entrepreneurship. Third, the CFPS provides granular information on real estate properties, including not only the number of properties one owns, but also the ownership status and the estimated market value of each

[14] Supported by the National Population and Family Planning Commission of China, the Institute of Social Science Survey (ISSS) of Peking University is authorized to conduct the survey. The ISSS is referred to the Survey Research Center of the University of Michigan for instruction and cooperation in designing the questionnaires and formalizing the procedures of the survey, which strengthens the reliability of the project.

[15] We notice that the 2020 wave data have been released recently, but its quality is under further investigation, especially given the large impacts of the COVID-19 pandemic.

house. Rather than using average prices at the regional or the ZIP-code level as a proxy for house value (Schmalz et al., 2017; Gao et al., 2020), the high granularity of housing ownership information in our data allows us to observe the collateral value (i.e., the market value of one's pledgeable houses) precisely as well as multi-property ownership of each potential entrepreneur. We are aware that, despite these advantages of using individual-level longitudinal data, it is hard to compute the aggregate economic magnitude. Hence, in robustness checks, we estimate economic magnitude at the city-level based on a city's number of newly registered firms.

We merge the CFPS data with the timing information of HPR policies. We manually collect detailed information of HPR policies across cities from 2010 to 2020 by web search, which is more inclusive than previous studies (see Internet Appendix A for details). Specifically, Chen et al. (2017) use the HPR policy in 2010-2014, the ever-initial wave of HPR and its cancellation, as their key identification strategy, while Deng et al. (2021) investigate the effects of the HPR policy in 2016-2017. Our HPR policies include both Chen et al. (2017)'s and Deng et al. (2021)'s data, as well as those not incorporated in either of the previous studies. In other words, we consider all the cases of HPR adoptions, cancellations, and re-implementations from 2010 to 2020 in China.

We then restrict the sample to urban citizens because HPR policies are implemented at the city level and there is almost no housing market in rural China. We also exclude individuals who are under 18 or above 65 because they are unlikely to start new businesses.[16] The filters leave us with an initial sample of 45,771 observations, 18,681 of which are in the HPR cities and are then excluded, because our identification strategy requires the inclusion of those unregulated cities only as discussed in Section 3.1. Thus, the final sample for the baseline regressions consists of 27,090 observations.[17]

Table 1 presents summary statistics on individual characteristics.[18] The average entrepreneurship probability is 12%.[19] 90% of surveyed individuals own at least one house and 17% have multiple properties. A median respondent is 44 years old, living in a family with 4

[16] While these criteria are standard in the entrepreneurship literature, our results remain robust without these filters.
[17] We winsorize all the continuous variables at the 5th and 95th percentiles of their distributions to reduce the influence of extreme values.
[18] See Table IA1 for the summary statistics of the initial sample of 45,771 observations that include individuals in the HPR cities. The distribution of all the variables (except for house values and incomes) is very similar to that of the final sample reported in Table 1. Thus, the exclusion of observations in HPR cities for the sake of cleaner identification does not materially undermine the representativeness of the sample or the reliability of our findings.
[19] Entrepreneurs in our sample include those who run small businesses, such as a grocery, a restaurant, or a retail shop. While their entrepreneurship is not Silicon-Valley-style (i.e., commercializing technological innovation, new products, or novel business models), we still include these individuals in the group of entrepreneurs because they are also risk bearers that support economic growth (especially in China). Hence, they are exactly in accord with the original conception of entrepreneurship that can be dated back to Richard Cantillon's seminal "*Essai Sur La Nature Du Commerce En Général*" in the eighteenth century.

members. 48% of respondents are female, 86% of respondents are married, and 69% of respondents are employees working for a registered entity. With regard to education levels, 13% of respondents have a college degree or above, while 16% of them are non-literate. Finally, 25% of respondents are in good health, while 43% of respondents feel they are unhealthy. The distribution of our sample is comparable with previous studies (e.g., Djankov et al., 2006; Li et al., 2020).

3.3 Sample for intensive-margin analysis and industry classifications

3.3.1 Measuring entrepreneurial outcomes

Besides examining the effects of housing speculation on individuals' choices on business ownership (i.e., the extensive margin), we explore the intensive margin as well, i.e., conditional on entry, how housing speculation affects the quality (i.e., the outcomes) of entrepreneurial activities. To this end, we use firm registration data, obtained from China's Administration of Market Regulation. This database provides detailed information on more than 60 million market entities (at their creation) from 1994 to 2020, including business registration code, organizational code, business name, incorporation location, business scope, registered capital, date of registration, as well as their post-entry survival.[20] Firm registration data make exploring entrepreneurial activities on the intensive margin plausible, because whoever starts a new startup must at first get legally qualified by registration in China.

We exclude entities that are not a business (e.g., associations, public universities, charity organizations, among others), not privately owned (e.g., state-owned enterprises, collective enterprises, among others), or is a spin-off of existing companies, to capture authentic entrepreneurial activities for the sake of business profits. We further restrict the sample period to 2004-2020, because city-level controls are available only after 2003 and we use entrepreneurial outcome measures one year ahead. Finally, we exclude observations in HPR cities for identification purposes, which leave us with 15,700,000 startups, corresponding to 4,711 city-year observations.

We construct two measures for entrepreneurial outcomes both at the startup level and at the city level: startup size at creation and post-entry survival. For a startup's size at creation, we use *Ln(Registered Capital)*, the logarithm of its registered capital. For post-entry survival, we use *No. of survival months*, i.e., the number of months, from a startup's establishment to the

[20] Unlike firms in the US, entrepreneurial entities in China typically register their real headquarter addresses as the locations of incorporation, because the Administration of Market Regulation requires verifiable addresses for firm registration.

date of its cancellation or last record of normal operating activities. We are aware that our duration measure could be subject to truncation problems, e.g., a startup starts two years ago would by no means achieve a 4-year survival. Thus, we construct two alternative dummy variables, *3-year survival dummy* and *5-year survival dummy*, denoting whether a firm survives a 3-year or a 5-year window. To conduct city-level estimations, we aggregate the four startup-level measures, adjusted for industry-by-organization-type fixed effects by scaling the industry-by-organization-type group mean in each year, to the city-year level. Table 2 tabulates the summary statistics of the above-mentioned quality measures for the intensive margin analysis. The pooling average of startups' size at creation is 4.91 million RMB yuan (roughly 0.77 million in US dollars), and the average survival duration is 23.7 months.

3.3.2 Industry classifications

Entrepreneurial activity outcomes are by nature varying across industries, e.g., starting a real estate firm typically requires far more registered capital than simply opening a grocery. Thus, it is necessary in our intensive margin analysis to take differences across industries into consideration. Unfortunately, we cannot directly observe an industrial classification code of each startup, because startups are not categorized into different industries at their registration.

To overcome this barrier, we develop a machine learning algorithm to consistently identify a startup's industrial classifications, using their business scope information that is an obliged disclosure item at registration delineating its legally permitted operating activities (see Internet Appendix B for details on the machine learning algorithm).

Figure 2 illustrates the identified industry distribution of the startups in our sample.[21] Out of 19 industries, newly registered startups cluster mostly in the wholesale/retail, manufacturing, and leasing commercial service industries, which take the value of 32.6%, 24.4%, and 12.7%, respectively.[22]

4 Main Results

4.1 Empirical model

We examine the effect of housing speculation on entrepreneurship by estimating the following equation using the staggered DiD framework:

[21] The industry distribution of the full sample with over 50 million startups is similar to that shown in Figure 2.

[22] Note that the ranks of industries in the distribution are subject to the classification of each specific industry. For example, the reason why manufacturing firms take a large proportion in our sample is that this industry covers a large variety of subcategories. However, this 19-category classification is the highest granularity we can achieve by far.

$$Entrepreneurship_{i,j,t} = \alpha + \beta \cdot HPR\ spillover_{j,t\text{-}1} + \gamma' \cdot X_{i,j,t} + Individual_i + Year_t + \epsilon_{i,j,t}, \qquad (1)$$

where $Entrepreneurship_{i,j,t}$ is a dummy variable that equals one if individual i living in city j is an entrepreneur at year t and zero otherwise. The key explanatory variable is $HPR\ spillover_{j,t\text{-}1}$, which equals one if city j is within 200 km to the closest regulated city and the regulated city is presently under the HPR policy in year *t-1*, and zero otherwise.[23] We lag *HPR spillover* for one year because the CFPS is typically finished at Q2 in a year, with the questions asking about the situation in the past year. Therefore, we use the lagged explanatory variable to alleviate reverse causality concern and also allow for a period for the policy spillovers to affect the real estate market and thereby entrepreneurship. X represents the array of control variables. At the city level, since the decision of starting a new business is associated with local investment opportunities as well as local business environment, we follow the existing studies (e.g., Huang et al., 2020) to include control variables: *GDP*, *GDP per capita*, *Fiscal revenue*, *Fiscal expenditure*, *Average salary*, and *Unemployment rate.*[24] At the individual level, we control for personal characteristics including *Housing collateral value*, *Family size*, *Salary*, *Ethnicity*, *Age*, *Marriage*, *Gender*, *Employed*, education dummies and health condition dummies.[25] We also include individual fixed effects and year fixed effects in the regressions to account for the effect of time-invariant personal characteristics and aggregate time trends, respectively.[26] We cluster robust standard errors by individual.[27] We run the linear probability model in the main specification because it is advantageous in interpretation and accounting for multiple fixed effects. The coefficient on the key variable of interest, β, captures the effect of house prices caused by HPR spillovers on entrepreneurship.

[23] We use a 200 km threshold to divide the sample into treatment group and control group for two reasons. First, compared to estimations on a continuous variable, a DiD setting facilitates interpretation of the results; second, unlike Deng et al. (2021) who only consider the second wave of HPR policies in 2016 and 2017, we take into account all the HPR policies (i.e., both the 2010-14 wave and the 2016-17 wave), which leads to the fact that our sample contains more HPR cities, and consequently more treated (nearer) non-HPR cities. With the full coverage of HPR policies in this paper, there are much more treated cities than control cities if we use a 250 km threshold. Therefore, we choose 200 km as the threshold to ensure that the treated group and the control group construct a balanced panel, with the same number of cities and comparable number of observations. However, the baseline results remain intact if we alternate the threshold or use a continuous measure, which we report in the first set of robustness checks.

[24] The city-level control variables are reported in the *China Yearbook of Statistics*, taken logarithms and measured in year *t*.

[25] Some variables are absorbed after controlling for individual fixed effects.

[26] Our results remain largely unchanged if we include city fixed effects as well. In addition, we could not include city-year fixed effects in the regressions because the key variable of interest, *HPR spillover*, is at the city-year level.

[27] Our choice of clustering the standard errors by individual follows the standard execution of the DiD regressions with two-way fixed effects. However, we are aware that individuals who are in the same city and sharing the same home ownership status (i.e., multi-property owners, one-house owners, or renters) are likely to be subject to the same shocks in the real estate market, and thus it is plausible to cluster the standard errors by city-ownership status following Han et al. (2020). While we cluster the standard errors by individual to account for the situation of migration and to avoid unnecessarily over-conservative standard errors (Abadie et al., 2022), our results would not materially alter if we cluster the standard errors by city-ownership status.

4.2 Baseline results

We undertake the baseline analyses by estimating Eq. (1) and report the results in Table 3. Column (1) of Table 3 Panel A presents the result of a parsimonious regression, i.e., we regress the entrepreneurship dummy on the key variable of interest: *HPR spillover*$_{i,j,t}$, and control only for city and year fixed effects. Thus, the regression captures within-city estimations. The coefficient estimate is negative and significant at the 1% level, suggesting that HPR spillovers lead to 2.3% fewer entrepreneurs in closer non-HPR cities compared to farther non-HPR cities.

Column (2) of Table 3 Panel A reports the result after controlling for individual characteristics and city-level controls. The coefficient estimate on *HPR spillover* continues to be negative and significant at the 1% level with an almost identical magnitude. Regarding controls, the result shows that women and aged people are less likely to start new businesses, while marriage is positively related to entrepreneurship. Compared to non-literates, individuals with primary/middle/high school levels of education are more likely to become entrepreneurs, while those with a college degree or above are less likely to start new businesses. This observation might be due to the fact that college degrees in China are a strong signal in the job market to get a decent, well-paid, and stable job, which significantly reduces one's incentives to start a new business.

We next control for individual fixed effects to get intensive-margin estimators and tabulate the result in Column (3) of Table 3 Panel A. While the personal characteristics may be largely absorbed by individual fixed effects, we still include marriage dummies, education dummies, and health dummies to account for factors that could change with time, to thus further alleviate omitted variable concerns. The coefficient estimate on the key explanatory variable continues to be negative and significant at the 1% level with a larger magnitude, suggesting that HPR spillovers lead to a 3.2% smaller probability of entrepreneurs in closer non-HPR cities compared to farther non-HPR cities. Given that the unconditional probability of entrepreneurship is 12% in our sample, it represents a sizable economic significance, i.e., a 26.7% drop in entrepreneurship.

Column (4) of Table 3 Panel A reports the baseline regression in which we further include *Housing collateral value*, *Family size*, and *Salary* in the estimation. The coefficient estimate on *Ln(Housing collateral value)* is positive and significant at the 5% level, suggesting that the collateral channel documented by previous studies (e.g., Corradin and Popov, 2015; Schmalz et al., 2017) is valid in our sample. Meanwhile, the coefficient estimate on *HPR spillover*, the key variable of interest, is negative and significant at the 1% level with the same magnitude as that in Column (3), suggesting that the collateral channel cannot explain away the speculation channel we have documented.

Overall, the findings in Table 3 Panel A suggest that the presence of HPR spillovers (i.e., a positive shock to house prices and housing speculation) has a negative effect on local entrepreneurship.

4.3 Tests on reverse causality

While the spillover effects of HPRs are unlikely to be endogenous, one may still argue that our results only reflect feedback in local house prices from entrepreneurship rather than a treatment effect, which is the typical reverse causality concern. Specifically, *ex ante* surging entrepreneurship could give "birth" to more successful entrepreneurs, and those "new money" might tend to buy houses in neighbor cities to get better educational resources for their children or better hygienic service. Consequently, the arrival of extrinsic housing demand would give rise to house prices of these cities, as well as the possibility that the HPR policy is implemented in these cities. If this argument is supported, our results could merely be a manifestation of reverse causality instead of the causal effect of housing speculation on entrepreneurship.

To address this concern, we follow Bertrand and Mullainathan (2003) to examine the dynamics of entrepreneurship by decomposing the key explanatory variable, *HPR spillover*, into five indicators: *Before* 2, *Before* 1, *Year* 0, *After* 1, and *After* 2+. *Before p* (*After p*) equals one for individuals in treated cities if the survey year is *p* period(s) before (after) the HPR implementation, and zero otherwise.[28] *After* 2+ is a dummy variable that equals one for individuals in treated cities if the year is two periods or above since the HPR implementation and the policy is not yet cancelled, and zero otherwise. We also construct a parsimonious version of the tests on dynamics, i.e., *Before* and *After*, which indicates the years before and after Year 0, respectively. If reverse causality drives our results, we expect to observe significant coefficient estimates on *Before* dummies above.

Table 3 Panel B reports the results. In Column (1), the coefficient estimate on *Before* is statistically insignificant while the coefficient estimate on *After* is positive and significant at the 1% level. In Column (2), the coefficient estimates on *Before 1*, *Before 2*, *Year 0* are all statistically insignificant and economically close to zero, suggesting that the dynamic trend of entrepreneurship before the HPR spillover is parallel between the treatment group and control group. The estimates on *After 1* and *After 2+* are both negative and significant at the 5% level.

[28] Remember the CFPS survey is conducted on a biennial basis. Hence, each period in this analysis is equivalent to two years.

These results suggest that individuals in closer non-HPR cities are less likely to start new businesses than those in farther cities from the HPR year onward and the effect becomes stronger in the subsequent years. Therefore, the above analyses suggest that the causal link seems to be from HPR spillovers (and thus from housing speculation) to entrepreneurship, not the other way around.

4.4 Robustness checks

We perform a battery of additional tests to ensure the robustness of our baseline results. For brevity, we report the results in Table 4 with only the coefficient estimates of the key variables tabulated.

First, we test the robustness of our baseline results with respect to alternative definitions of the key explanatory variable, *HPR spillover*. Specifically, we use five alternative definitions that can be divided into three sets: (a) using alternative thresholds of distance, 150 km, 175 km, or 250 km, when defining the treatment group and the control group; (b) requiring that the treatment group and the control group are in the same province; (c) using a continuous measure, i.e., *HPR* × *Nearness*, as the explanatory variable, where *HPR* denotes whether the closest HPR city (for the focal non-HPR city) is under a HPR policy.[29,30] Our main findings are not altered by these alternative definitions of the key variable. In particular, the economical magnitude of the coefficient estimates, in Columns (1)-(3) of Table 4 Panel A, is monotonically diminishing with the increase in threshold distance. This pattern is consistent with a geographical spillover argument claimed above, which further ensures the internal validity of our empirical method.

Second, we check whether our baseline results are robust to alternative sample selection criteria. Specifically, we repeat the baseline analyses in alternative samples that (a) include individuals that are non-urban citizen; (b) include aged people (older than 65); (c) exclude individuals who are not in the first wave (i.e., the 2010 wave) of the CFPS survey to construct a perfectly balanced panel; (d) use the full sample (i.e., all individuals in the non-HPR cities); (e) consider three extra filters that require the focal individual to be healthy, educated, and belong to the majority people (i.e., the Han people). Table 4 Panel B suggests that our results remain robust to all these alternative sample selection filters.

[29] *Nearness* is the additive inverse of the distance from the focal city to the nearest regulated city. Therefore, a larger value of *Nearness* indicates that the focal city is closer to the closest regulated city. We use *Nearness* instead of distance in order to make the result easier to interpret.
[30] Note that *HPR spillover* is actually *HPR* × *1[whether the focal non-HPR city is within 200km to the nearest HPR city]*.

Third, we include additional controls in our baseline regressions to check whether our main results are altered. These controls are: (a) whether the focal individual is employed or not;[31] (b) city-level controls, such as consumption, bank loans, citizen deposit, income Gini index, and the number of primary schools, high schools, and colleges.[32] Columns (1)-(2) of Table 4 Panel C show that the main results remain unaltered. Fourth, to address the concern that individuals who are in the same city and sharing the same home ownership status are likely to experience the same shocks on the real estate market, we follow Han et al. (2020) to cluster the standard errors at the city-ownership status level, where ownership status denotes whether an individual is a multi-property owner, a one-house owner, or a renter. We also follow the recommendation of Petersen (2009) to estimate Eq. (1) with standard errors clustered at both the individual and the city-year level to mitigate the concern regarding the presence of residual correlation in these dimensions. Columns (3)-(4) of Table 4 Panel C suggest that our baseline results are robust to city-ownership status-level standard error clustering and multi-way standard error clustering.

Fifth, in Column (5) of Table 4 Panel C, we use a panel Probit model instead of the linear probability model to check whether our main results are robust to alternative econometric models. The main results remain intact. Actually, the Probit estimation shows a larger (in absolute value) coefficient estimate on the key variable of interest, suggesting that our main results are likely an underestimation yet the real effect could be even stronger than what we document. Sixth, we exclude the last wave of survey to alleviate the concern that our results are driven by sample truncation. Column (1) of Table 4 Panel D shows that the results remain largely unchanged.

Seventh, Baker et al. (2022) point out that staggered DiD estimations could be misleading under some conditions, e.g., the effects are changing with time, because part of the estimates can be negatively weighted and driving the result which is actually a weighted-average estimate. To address this concern, we follow their recommendation to conduct two diagnostic tests and report the results in Columns (2) and (3) of Table 4 Panel D: (a) we perform a stacked regression and get similar result as the baseline result; (b) we conduct Goodman-Bacon (2021)'s decomposition, and the result shows that each part of the averaged DiD estimate has a positive

[31] We do not include employment status in the baseline regression because *Employed* could potentially be a "bad control" (Angrist and Pischke, 2008), i.e., being an entrepreneur and an employee are by nature mutually exclusive. But the inclusion of employment status does alter our results.

[32] The income Gini index is estimated based on the CFPS data, while other city-level variables are reported in the *Year Book of Statistics.*

weight and a negative value. Both tests suggest that our results are unlikely driven by a misleading staggered DiD estimation.

Eighth, while we use the CFPS's individual-level panel data to better account for inter-personal differences, one might concern that this execution could be subject to the double-counting problem. Thus, we re-run the baseline regressions at the household level to keep consistent with Schmalz et al. (2017)'s specification, i.e., we only consider entrepreneurial activities of household heads.[33] As shown in Column (4) of Table 4 Panel D, the baseline results remain robust.

Ninth, to address the concern that our baseline results are driven by the exclusion of HPR cities, we repeat the baseline regressions in the sample with the HPR cities included. Consistent with our main results, we find that individuals in HPR cities are more likely to start new businesses than those in non-HPR cities because of the restrictions on housing speculation in HPR cities. We provide detailed discussions in Internet Appendix C and present the results in Table IA2.

Finally, we use an alternative data set, i.e., firm registration, to re-estimate the baseline model at the city level, of which the dependent variable is the logarithm of the number of a city's newly registered businesses per capita in year $t+1$. While the use of individual-level data for baseline tests is aimed at distinguishing housing speculators with personal traits, the use of this city-level panel based on the number of firm registration records helps to account for the economic magnitude at the aggregate level. We tabulate the result in Column (5) of Table 4 Panel D. The city-level estimation is qualitatively consistent with our baseline results at the individual level, and quantitatively comparable with the Probit estimation, which further ensures the robustness of the individual-level results.

4.5 Placebo tests

To address the concern that our results could be driven by chance, we conduct two placebo tests. First, although Deng et al. (2021) has shown the effectiveness of HPR spillovers on home price growth, a reasonable concern is that our key variable, *HPR spillover*, may fail to capture house price changes and housing speculation, but rather reflect the effects of some other latent factors. To mitigate this concern, following Kerr et al. (2015) and Deng et al. (2021), we focus

[33] Schmalz et al. (2017) focus on household heads because their research question is on housing collateral, i.e., only one person per household could pledge the house and this person is likely to be the household head.

on renters, i.e., individuals who have no house, and re-run baseline regressions in Column (4) of Table 3 Panel A.

Intuitively speaking, renters' entrepreneurship activities in our setting should not be affected by house prices: on one hand, they do not have any house to pledge; on the other hand, they do not even have a first house to live in, not to mention speculating on extra houses. Hence, we expect that *HPR spillover* has no effect on renters' entrepreneurial activities. Table 5 Panel A presents the results of the placebo test with renters. Consistent with our conjecture, the coefficient estimates on *HPR spillover* are statistically insignificant and economically close to zero, suggesting that renters' entrepreneurial activities are not affected by housing speculation. Thus, the key explanatory variable is unlikely to capture variation other than house prices.

Second, while our identification tests that exclude the HPR cities from the analyses enhance the comparability between the treatment and the control group, we are aware that it is still possible for some time-varying and unobservable differences between the two groups to drive our main results. To address this concern, we conduct a Monte Carlo analysis following Bekaert et al. (2005). Specifically, we first randomly assign falsified treatment and control non-HPR cities to the individuals but preserve the distribution of the actual time of shocks, and then re-estimate Eq. (1). We repeat the above procedures for 1000 times. This approach maintains the distribution of HPR shock years from our baseline specification, but it disrupts the proper assignment of HPT shock years to cities. Therefore, it an unobservable shock occurs at approximately the same time as the HPR policy, it should still reside in the testing framework, and thus have an opportunity to drive the results. If, however, no such shock exists, then our incorrect assignments of HPR years to cities should weaken our results when we re-estimate the baseline tests.

Table 5 Panel B reports the distribution of the coefficient estimates and corresponding *t*-statistics of the randomized falsified *HPR spillover* with the 1000 replications. The distribution exhibits a normal distribution with the mean of 0.0003, as well as the mean *t*-stat of 0.055, suggesting that falsified HPR spillover shocks do not have any statistically or economically significant effect on entrepreneurial activities. More importantly, the coefficient estimate reported in Table 3 (i.e., -0.032) are far out in the left tail of the distribution in the placebo test (i.e., more than four times larger in the absolute value than the corresponding 1st percentile, -0.0124, in Table 5 Panel B), suggesting that our main results are unlikely driven by unobservable shocks occurring at approximately the same time as the HPR policy.

4.6 Reversal shocks: the effect of HPR cancellations

The spirit of our identification strategy is that the HPR policies bring forth uninvited and unexpected foreign housing speculation in surrounding cities, which triggers real estate boom (and consequent local housing speculation) in these non-HPR cities. Based on the same idea, one would expect the negative spillover effect on entrepreneurship in non-HPR cities should disappear after the cancellation of HPR policies in the nearer HPR cities.

To test this conjecture, we further examine the effects of HPR policies in an event-study framework, that is, we investigate the year-by-year spillover differences in entrepreneurship before and after the cancellation of HPR policies. Specifically, we rerun the baseline regression using a set of dummies that indicate the normalized period relative to the cancellation of HPR policies and equal one for treated cities in each period and zero otherwise. Figure 3 shows the coefficient estimates and their confidence intervals, where *t-1* (*t-2*) denotes one (two) period(s) before, *t* denotes the period of, and *t+1* denotes one period after, the cancellation. A predicted symmetric pattern emerges from the estimates that the negative spillover effect of HPR policies diminishes after the HPR cities exit from the restrictions. Specifically, while *t-1* and *t-2* capture a significant negative effect (p-values < 0.05), the differences between individuals in nearer and farther non-HPR cities after the HPR policy cancellation are largely insignificant and close to zero. This symmetric pattern of reversal shocks is consistent with our conjecture and further strengthens our identification.

4.7 Heterogeneous tests

Thus far, we have shown that housing speculation crowds out entrepreneurship in a booming real estate market. To further strengthen the speculation channel argument, we examine how our baseline results are altered by individual-level and city-level characteristics that capture housing market speculation propensity, which allows us to further explore whether or not it is the speculator that drives our main results.

First, we use multi-property ownership to identify speculators. Chen and Wen (2017) show that China's housing boom is accompanied by high vacancy rates, which denotes widespread speculation in housing market. We use multi-property ownership as a proxy for an individual's housing speculation propensity. Intuitively, if an individual owns multiple properties, it is likely that the individual is wealthy and the extra house(s) is for speculative investment because one cannot physically live in many houses concurrently (Gao et al., 2020). In this subsection, we begin with providing evidence using the 2014 CFPS survey to gauge multi-property owners' housing speculation from three different aspects: expectation on house price changes, access to external finance, and risk attitudes.[34]

First, we consider price change expectation, because housing speculation typically anchors on an optimistic (extrapolative) belief of future price changes (e.g., Glaeser and Nathanson, 2017; Gao et al., 2020). Consistent with our conjecture, multi-property owners are more likely to expect future house price appreciation than single-property owners and renters, with the inter-group *t*-stat equal to 3.05. Moreover, as shown in Figure 4, 36.2% of multi-property owners expect local house prices to increase in near future, while the proportion for single-property-owners are less than 30%.

Second, we compare access to external finance of multi-property owners with that of single-property owners and renters. Since housing speculation by nature requires a sizable amount of cash, speculators should have easy access to external finance. Figure 5 illustrates the answers to the survey question on access to external finance by multi-property owners and others. Over 45% of multi-property owners response that their access to external finance is somewhat easy or very easy, while the proportion for single-property owners and renters is merely 25%. On the contrary, the percentage of multi-property owners who respond very difficult (somewhat difficult) in emergent borrowing is 7.5% (11.8%), while that for single-property owners and renters is almost doubled, reaching to 19.5% (20.6%). The inter-group *t*-stat is 19.23, suggesting that multi-property owners have significantly easier access to external finance.

Third, we examine one's risk attitude, because speculation, by definition, implies risk-taking. As illustrated in Figure 6, multi-property owners are over 15% less (33.3% compared to 49.5%) than those who are perfectly risk-averse than single-property owners and renters, yet almost 15% more (47.1% compared to 32.3%) than those who are weakly or strongly risk-loving than single-property owners and renters. The difference is statistically significant (*t*-stat = 11.98), suggesting that multi-property owners are more likely to be risk loving. Overall, our findings suggest that multi-property owners are more likely housing speculators.

Spirited by the above findings, we construct three proxies to capture an individual's housing speculator propensity. The first measure is *Multi-property owner*, a dummy variable that equals one for individuals who are multi-property owners and zero otherwise. Our second

[34] The CFPS conducted an extra survey uniquely in the 2014 wave, which consists of a variety of questions on a respondent's subjective belief and attitude. Out of the options for every question, each interviewee is required to choose one that is most consistent to their own belief or attitudes. Although this survey appears only once out of all the CFPS waves, we still benefit from comparing the cross-sectional distributions of the answers by different groups of respondents.

measure is *No. of extra houses*, the number of house(s) other than the house that an individual is currently living in. Note that *No. of extra houses* equals zero if an individual is a renter or owns one house. Our third measure is *Has property income*, which equals one if an individual has property income and zero otherwise. We then interact these three speculator variables with *HPR spillover* and expect that the baseline results are more pronounced for speculators. We also include *Single-property owner* (a dummy variable that equals one if an individual has only one house and zero otherwise) to account for differences between single-house owners and renters.

Table 6 Columns (1)-(3) tabulates the results. The coefficient estimates on the interaction terms between *HPR spillover* and speculator indicators are all negative and significant at the 1% or the 5% level, consistent with our conjecture. Moreover, the coefficient estimates on *HPR spillover* are all statistically insignificant and economically close to zero, suggesting that renters are not the drivers of the baseline results.

Beyond the above three proxies, we conduct a similar test based on a supplementary measure of one's intensity to be housing speculator: *High (house price) expectation*. As mentioned above, the CFPS includes a questionnaire about one's expectation on local house price changes in the near future, which provides an opportunity for us to directly observe a respondent's objective expectation on house price changes. The binary variable *High expectation* equals one if a respondent in the 2014 wave expects that the local house prices is going to experience a sizable increase in the near future, and zero otherwise. It is reasonable to assume that people's choice of being housing speculators is based on her expectation on high house price increases. With this supplementary proxy, we rerun a similar regression as those in Table 6 Columns (1)-(3), but only with the sample of the 2014 and the 2016 waves, because we cannot observe how the price expectation changes before and after the 2014 wave. Again, in Table 6 Column (4), the coefficient estimate on *HPR spillover × High expectation* is negative and significant at the 5% level, consistent with other pieces of evidence reported in Table 6.

Put together, the results in Table 6 suggest that the crowding-out effect of housing speculation on entrepreneurship is more pronounced for multi-property owners who are more likely to be housing speculators, consistent with the conjecture that speculation is the underlying channel through which house prices negative affect entrepreneurship.

5 Addressing alternative interpretations

Thus far, our test results suggest the speculation channel could explain the negative effect of house prices on entrepreneurship. However, there are three alternative interpretations that could potentially explain the results as well: reduced labor supply caused by changes on wealth, reduced capital supply, and heightened entry cost.[35] In this section, we undertake a few tests and provide evidence trying to rule out these alternative interpretations of our main findings.

5.1 Reduced labor supply

One alternative interpretation of our main results is that households that enjoy wealth gains from house price appreciation have less incentive to work hard and start new businesses, i.e., the labor supply argument (e.g., Li et al., 2020). This argument is also consistent with our earlier findings on multi-property owners because they are likely to experience a larger wealth appreciation in a house boom, which consequently crowds out their labor (including entrepreneurial labor) supply.

To rule out this alternative interpretation, we conduct two sets of tests. We first examine whether our main results are heterogeneous with respect to different levels of household wealth. We include *Ln(Net assets)*, the value of a household's total assets net of debt, and its interaction term with *HPR spillover*, as well as all the control variables, into the tests in Table 6 to account for the wealth effect caused by house price appreciation on labor supply.[36] Table 7 Panel A tabulates the results. In Column (1), the coefficient estimate on *HPR spillover* × *Ln(Net assets)* is largely statistically insignificant and close to zero, disparate to the significant heterogeneous effects shown in Table 6, suggesting that the baseline results are not altered by household wealth levels. Moreover, in Columns (2)-(5), the coefficient estimates on *Ln(Net assets)* are positive and statistically significant in two out of four specifications (and all the four coefficients are positive), consistent with the wealth effect in the entrepreneurship literature. However, the estimated heterogeneous effects based on one's multi-property ownership dummy, the number of extra houses, one's property income dummy, and one's house price expectation, are all largely intact, both statistically and economically. Considering that we have controlled for one's house value (the collateral value) in the baseline model, this test further accounts for the effect of a household's total wealth and suggests that the main findings we have documented is unlikely driven by the wealth effect on labor supply.

---

[35] We do not discuss in details some of other alternative interpretations that obviously cannot reconcile with the evidence above. For example, one might concern that there is possibility that the arrival of house booms may force people to save more money to buy their first houses, instead of investing in entrepreneurship, because in most cases they must have a house before starting new ventures, especially in China. However, this conjecture contradicts to the results in our first placebo test (Table 5 Panel A) as well as the heterogeneity tests above (Table 6) that the coefficient estimates on *HPR spillover*, capturing the effect on renters in these tests, are largely statistically insignificant and economically close to zero.

[36] We get similar results if we use total assets rather than net assets.

The second set of tests directly examines the labor supply argument by exploring whether labor supply is affected by house market prices.[37] To this end, we re-run our baseline regression, as well as the tests in Table 6 and Table 7 Panel A, but replace the dependent variable with individuals' labor supply. To measure one's labor supply (which is by no means directly observable), we use two proxies. One is the *Work-or-quit* dummy that probes an individual's choice on labor market participation, because people who work (either being employed or working as business owners) by definition supply more labor than those who do not; the other is *Ln(Job income)* that captures an individual's wage from working as an employee, because (theoretically) one's labor supply is priced by wage in an efficient labor market.[38] If our tests actually served in favor of the labor supply argument rather than the housing speculation channel, we should expect to get similar results with the two proxies of labor supply as the dependent variables.

We tabulate the results in Table 7 Panel B. For brevity, we only report the heterogeneity results based on *Multi-property owner*, while the results are largely similar if we alternatively use *No. of extra houses*, *Has property income*, or *High expectation*. In Columns (1) and (4), we re-estimate the baseline model with the alternative dependent variables. The coefficient estimates on our key explanatory variable, *HPR spillover*, are largely insignificant, suggesting that house market prices do not affect individuals' labor supply. We then further reconduct the heterogeneous tests with the alternative dependent variables. Again, the largely insignificant coefficient estimates in Columns (2) and (5) suggest that the heterogeneous effects on entrepreneurship shown in Table 6 are not viable on individuals' labor supply choices. Additionally, after including *Ln(Net assets)* in Columns (3) and (6), the heterogeneous effect based on *Multi-property owner* remains statistically insignificant.

Overall, the results in Table 7 consistently stand against the labor supply argument. The evidence also shows that our findings do not capture one's labor choice, but one's entrepreneurial choice.

---

[37] We are aware that in entrepreneurship literature (especially theoretical context), entrepreneurship, professionals and labor are mutually exclusive type of work one can choose. In this paper, we use the word "labor" in a more general way in the sense that it denotes the efforts an individual pay to get corresponding compensation, which can be either in the form of getting job income from the employers or in the form of operating self-owned businesses.

[38] When using *Ln(Job income)* as the dependent variable, the sample is naturally restricted to individuals who always work as an employee in our sample, because entrepreneurs and individuals who do not work are not likely to have a job income that is comparable to employees. Thus, this execution helps avoiding truncation biases, i.e., an individual's job income is missing after exiting the labor market. We acknowledge that unobservable job income for entrepreneurs makes the results not relevant to entrepreneurship. Thus, we pay less value on the results presented in Columns (4)-(6) of Table 7 Panel B, but merely rely this piece of evidence to further assure that housing speculation is not likely to affect labor supply even within the group of employees, and therefore it is less likely that the labor supply argument can explain away the housing speculation channel in this paper.

5.2 Reduced capital supply

Another alternative interpretation of our main results is that, as house prices surge, promising returns on real estate investment could attract away investors, and thereby capital, from entrepreneurial finance to the real estate market. In other words, the crowding-out effect we have documented might not reflect the choices of potential entrepreneurs *per se*, but rather the choices of investors as well as reduced capital supply of entrepreneurial finance. This alternative argument is reasonable because, holding the total level of capital fixed, more real estate investment would imply less funding for non-housing investment vehicles (including startups), and thus possibly lower entrepreneurship. This concern also has similar prediction on entrepreneurial quality as our speculation interpretation — the reduction of capital supply motivates investors to screen off low-quality startups, which would lead to the same results as our speculation argument that heightened opportunity cost crowd out less-promising entrepreneurs.

To rule out this alternative interpretation of our main findings, we conduct two sets of tests. Both of them use the standardized VC investment score (*VC inv. score*), a component of Dai et al. (2021)'s Index of Regional Innovation and Entrepreneurship in China (IRIEC), that reflects a city's level of VC investment activities in a year. The IRIEC is one of the best public indices capturing the dynamics of entrepreneurial finance. For the first set of tests, we examine whether HPR policies have negative spillover effects on VC dynamics, as predicted by the capital supply argument. Specifically, we conduct the same regression as that in Panel D Column (5) of Table 4, but replace the dependent variable with *VC inv. score* measured one year ahead. We also use the logarithm of *VC inv. score* for robustness check to avoid the influence of extreme values. In Table 8 Columns (1)-(2), the coefficient estimates on *HPR spillover* based on city-level regressions are largely insignificant. Compared to the result in Panel D Column (5) of Table 4 that captures significant and sizable negative effects of house prices on entrepreneurship measured by the number of newly registered firms, this insignificant finding suggests that our baseline regression is not likely driven by reduced capital supply.

The second set of tests explore heterogeneous effects based on *VC inv. score* at the individual level, which are analogous to those in Table 6. In Table 8 Columns (3)-(4), the coefficient estimates on *HPR spillover* × *VC inv. score* and *HPR spillover* × *Ln(VC inv. score)* are economically nominal and statistically insignificant, suggesting that reduced capital supply on entrepreneurial finance unlikely explains our baseline findings.

Overall, while the housing speculation channel does not exclude the possibility that real estate cycles can drive the participant incentives for entrepreneurial investors, analogous to our argument on entrepreneurship, the results in Table 8 show that the capital supply argument is not the main driving force on the results we have documented in this paper.

5.3 Heightened entry costs

The third alternative interpretation is that the uninvited house booms resulted from HPR spillovers can synchronize with increases in workplace rent as well as various sources of entry costs against entrepreneurial activities.[39] While this heightened entry cost interpretation could also explain a negative relation between house prices and entrepreneurship, it is distinct from the speculation channel in two aspects. First, the heightened entry cost explanation predicts that potential entrepreneurs with lower income and those faced with severer financial constraints are more likely to be crowded out because rents are likely to take a larger proportion of their entrepreneurial capital, while the speculation channel predicts that wealthier individuals and those financially privileged are more likely to be crowded out because usually they have more disposable finance to do housing speculation. Second, the heightened entry cost explanation predicts that smaller new businesses are more likely to be crowded out because workplace rents are typically not a major hurdle for large ventures, while the speculation channel predicts no differential effect on the size of business creation.

To test the first distinction between the speculation channel and the heightened entry cost explanation, we use individuals' personal traits and city-level supply of disposable finance to capture one's financial constraint level. First, since housing speculation typically requires sizable amount of disposable liquidity and anchors on optimism, individuals who have a high net worth (i.e., much redundant "cash in hand"), a wealthy life, a positive sense of worth, and optimistic attitudes toward the future are more likely to be housing speculators. Hence, we expect that our baseline results are more pronounced for individuals with these traits. On country, the heightened entry cost explanation predicts that individuals with severer financial constraints are more likely to be crowded out from business ownership.

To examine this conjecture, we explore the heterogeneous effect based on an individual's answers to the 2010 CFPS survey questions about their own attitudes toward life. Questions on an individual's attitudes toward their lives only appear in the first wave (2010) of the CFPS

[39] Actually, using their non-public data set, Deng et al. (2021) show that rent does not respond differently between treated and control cities. In this paper, we discuss the entry cost argument in a more general way.

survey. We consider four questions related to one's wealth and life: (a) are you ranking high in income locally? (b) do you agree that wealth reflects one's success? (c) do you feel confident to your future? (d) have you gotten any bad experience due to inequality? If a respondent gives positive answers to the first three questions and a negative answer to the last question, then they are likely to be speculators. We then construct four dummies based on the answers, which equal one if the respondent gives a positive answer (i.e., agree or highly agree) to the first three questions or a negative answer (i.e., disagree or highly disagree) to the last question, and zero otherwise. Again, we interact these dummies with the key explanatory variable, *HPR spillover*, respectively.[40] Columns (1) - (4) of Table 9 present the results. All of the coefficient estimates on the interaction terms are negative and significant at the 1% or 5% level. The findings suggest that our baseline result is more pronounced for individuals with personal traits that are more consistent with speculators, i.e., individuals who have higher income, are more positive to wealth, more confident to their future, and have no bad experience of inequality.[41] All these findings point to the speculation channel yet stand opposite to the heightened entry cost explanation.

A reasonable concern to this set of tests is that, since education help people enjoy decent social statuses and obtain high income, the personal traits examined before may not depict a wealthy and optimistic individual but rather a well-educated one. To address this concern, we interact education dummies with *HPR spillover* and report the results in Column (5) of Table 9. The result shows that, as educational levels increase, the effect of *HPR spillover* is monotonically diminishing, statistically and economically, ranging from -0.059 for non-literate to 0.001 for college graduates, which helps rule out the alternative explanation.[42] Overall, instead of heightened entry costs preventing individuals with severer financial constraints from being entrepreneurs, the heterogeneity test results in Table 9 suggest that individuals who are less educated yet wealthier and more optimistic (and thus are likely housing speculators) are more likely to drive the results, which is consistent with the speculation channel.

Second, we move on to examine city-level characteristics in terms of disposable finance availability for heterogeneity tests. Since China's financial market is geographically segmented

[40] We are aware that subjective survey data based on questionnaire answers might be vulnerable to measurement errors. Thus, we follow the recommendation by Bertrand and Mullainathan (2001) to only use these answers in constructing explanatory variables instead of dependent variables.
[41] Note that one who has "no bad experience of inequality" does not denote equalized environment, but rather the probability that the respondent, as well as their family, enjoys a more decent social status.
[42] The result regarding education levels makes sense, because the magnificent transition of China's econ-system since 1981 creates thousands of millionaires whose success is originated barely from their ambition and endeavor, instead of their diploma that has long been scarce within the nation. Moreover, compared to a college graduate, a non-literate individual seems more likely to be appealed by speculative impulse.

(Boyreau-Debray and Wei, 2005; Huang et al., 2020), the availability of financial resources varies significantly across cities. Thus, we use city-level characteristics as proxies for local individuals' probabilities of acquiring enough financing support for their housing speculation. While the heightened entry cost argument predicts that individuals who live in the cities lacking resources of disposable finance are likely to be the main force driving the results, we expect our baseline results to be more pronounced in cities where people are more likely be able to afford speculating on house markets. Specifically, we construct three proxies that capture a city's disposable liquidity availability. The first proxy is *Income inequality*, i.e., a dummy variable that equals one if a city's Gini index is larger than 0.4 and zero otherwise, because Zhang et al. (2016) find that a higher inequality drives up price-income ratio and housing vacancy rate.[43] Intuitively speaking, conditional on the level of local economic development, higher inequality means that the minority owns a major part of liquidity. Therefore, there would be a larger amount of "redundant money" for housing speculation since wealthy people have lower marginal consumption propensity. In Column (1) of Table 10 Panel A, the coefficient estimate on *HPR spillover* × *Income inequality* is negative and significant at the 1% level, suggesting that the crowding-out effect of housing speculation on entrepreneurship is more pronounced in cities with severer income inequality.[44]

We next consider citizen deposit and bank loans. Mian and Sufi (2022) show that the expansion of banking credit triggers housing speculation in a booming real estate market. Likewise, Chakraborty et al. (2018) suggest that bank loans play an important role in the crowding-out effect of housing appreciation. Also, Liu and Xiong (2020) find that mortgage loans take a dominant proportion in financing households' housing purchases. Following their logic, we measure the availability of financing resources with *Citizen deposit* (the logarithm of city-level volume of citizen deposits) and *Bank loan* (the logarithm of city-level volume of bank loans), because they directly capture bank credit supply that is the dominating financing tool in China. The coefficient estimates on the interaction terms, *HPR spillover* × *Citizen deposit* and *HPR spillover* × *Bank loan* in Columns (2)-(3) of Table 10 Panel A are negative and significant at the 5% level, suggesting that cities with better access and higher availability of liquidity are more likely to exhibit a stronger negative effect of housing speculation on entrepreneurship.

[43] The Gini index is estimated from household-level income (following Zhang et al., 2016) reported in the CFPS. We use a 0.4 threshold because the number is an *ad hoc* boundary to identify cities as equal or unequal.

[44] Moreover, the coefficient estimate on *Income inequality* is negative and marginal significant, consistent with the existing literature on inequality and entrepreneurship (e.g., Aghion and Bolton, 1997; Easterly, 2007).

One may concern that our city-level proxies of disposable finance availability are merely a reflection of local economic fundamentals, i.e., the effect of HPR spillovers could be more pronounced for wealthier cities. To rule out this alternative interpretation, we re-run the tests but interact *HPR spillover* with proxies for a city's economic fundamentals, i.e., *GDP*, *fiscal expenditure*, and *unemployment rate*, and report the results in Table 10 Panel B. We observe that the coefficient estimates on the interaction terms are all statistically insignificant and economically close to zero (i.e., -0.002, -0.004, and -0.004, respectively), suggesting that economic fundamentals cannot explain the city-level heterogeneity of our main results.

Overall, while each test in this subsection could be subject to various concerns, our findings from different tests collectively suggest that the crowding-out effect of HPR spillovers on entrepreneurship is more pronounced for individuals who are more likely housing speculators and who live in cities with easier access to external finance, supporting the speculation channel and standing against the heightened entry cost argument.

6 Further Analyses

6.1 Housing speculation and entrepreneurial outcomes

In this section, we examine the speculation channel on the intensive margin by exploring how house prices affect entrepreneurial outcomes conditional on entry. The speculation channel argues that individuals start a new business only if they anticipate the return from entrepreneurial activities exceeds the opportunity cost of capital (i.e., the return from speculate in the house market). This argument has two important implications regarding entrepreneurial outcomes. First, the speculation channel is only affected by the opportunity cost of capital but not by other frictions, such as information asymmetry or entry costs. Second, the increased opportunity cost of capital resulted from house market booms only crowd out "low quality" entrepreneurial projects but not "grand-slam" startups.

To test these conjectures, we examine two aspects of entrepreneurial outcomes: startup size at creation and post-entry survival. Specifically, we use firm registration data and run the same regressions as Eq. (1), at both the firm level and the city level, with the dependent variable replaced with entrepreneurial outcome measures (see Section 3.3 for the construction of outcome measures). For firm-level regressions, we include city-industry-organization type fixed effects and year fixed effects to account for differences in entrepreneurial outcomes across different comparable groups of firms; for city-level regressions, we follow standard execution to control for city and year fixed effects, and adjust the dependent variables to partial out the differences across industries, organization types, and years (see Section 3.3.1). If our conjecture is supported, we expect to observe no effect on startup size at creation and a positive effect on startup survivals.

Table 11 tabulates the firm-level results with more than 13 million startup observations (Panel A) and the city-level results with 3,683 city-year observations (Panel B). Consistent with our conjecture, the coefficient estimate on *HPR spillover* is statistically insignificant when startup size is the dependent variable at both the firm level analysis in Panel A (t-stat < 0.2) and the city level analysis in Panel B (t-stat < 0.02). However, the coefficient estimates on *HPR spillover* in Columns (2)-(4) in both panels when startup survival variables are used as the dependent variable are all positive and significant at the 1% or 5% level, suggesting that speculative house market booms induced by HPR spillovers are associated with better survival of new businesses conditional on entry. These two findings collectively suggest that the increase in the opportunity cost of capital resulted from house market booms crowds out low-quality entrepreneurial projects and thereby are associated with better entrepreneurial outcomes. The increase in the opportunity cost of capital resulted from house market booms, however, does not affect startup size at creation, which further helps rule out the heightened entry cost argument.

Overall, the entrepreneurial outcome results further validate the speculation channel underlying the negative effects of house prices on entrepreneurship along the intensive margin. Moreover, the findings imply that, unlike the collateral channel that emphasizes on the role of houses in reducing financial frictions, the speculation channel focuses on the opportunity cost of capital resulted from house price appreciation.

6.2 External validity: Cross-country evidence

One concern of our analyses thus far is external validity: the generalization of our results, i.e., the documented negative effect of housing speculation induced by house market booms may exhibit only in China. To address this concern and generalize our findings, we extend our sample to other countries and undertake a cross-country analysis. Specifically, we focus on the OECD countries, one of the most representative groups of major economies (especially developed countries) throughout the world, and explore the relation between house prices and entrepreneurship.

The OECD countries have been experiencing a booming real estate market since 2015, with low (or even negative) interest rate policies in Europe. The OECD's house price index shows that the average property price witnesses a 30% growth during 2003-2015, yet it takes only 5 years since 2015 to increase by another 30%. By 2020, the price index has even surpassed the historic peak before the 2007-2009 global financial crisis, which is even reported as "flashing the kind of 2008 style (bubble) warnings".[45] Given this similar yet more widespread setting of continuous house price appreciation in the OECD countries, we explore whether a negative effect of house prices on entrepreneurship could still be observed in another group of booming housing markets.

Panel (a) of Figure 7 plots changes in entrepreneurship against house price growth in the OECD countries.[46] For each OECD country, we capture entrepreneurship levels with the Global Entrepreneurship Index (GEI) published by the Global Entrepreneurship and Development Institute (GEDI) in U.S.[47] A country's house price growth is calculated from the house price index (HPI) on the OECD's website.[48] Consistent with our findings using China data, the dashed line in Panel (a), plotting the fitted values from ordinary least squares regressions, illustrates a negative relation between house price growth and entrepreneurship in OECD countries during the booming period.

Furthermore, we split the sample by the sample median of house price growth in Panel (a) and re-plot the relation, based on the rationale that countries with larger (smaller) increases in real estate price are more (less) likely to be booming markets. We conduct the same plotting with each of the two subsamples in Panels (b) and (c), respectively, of Figure 7. For economies with relatively stable real estate markets (i.e., house price growth below the sample median) in Panel (b), most of the observations (13 out of 17) are around or above zero, and the changes in entrepreneurship scores are ranging from -6.6 (Greece) to 13.6 (Switzerland). Meanwhile, for economies with booming real estate markets (i.e., house price growth above the sample median) in Panel (c), most of the observations (16 out of 18) are negative, and the changes in entrepreneurship are ranging from -15.2 (Latvia) to 6.0 (Ireland). Moreover, the dashed lines

---

[45] See https://www.bloomberg.com/news/articles/2021-06-15/world-s-most-bubbly-housing-markets-flash2008-style-warnings for more information.

[46] We exclude Iceland in this part of analysis, because surge in tourism during this period is believed to (partly) dramatically lift the economy, i.e., both housing price and aggregate demand (see a report by Reuters at https://www.reuters.com/article/us-iceland-tourism-idUSKCN12A1MT). Since the country has been witnessing an extremely boom (and thus is an outlier compared to other countries), the latent factor, i.e., tourism boom, would likely drift the results from our main focus. We, however, find similar patterns if Iceland is included in the sample.

[47] The GEDI is founded by world-leading entrepreneurship scholars from the LSE, George Mason University, University of Pecs and Imperial College London. The Global Entrepreneurship Index (GEI) methodology, has been validated in rigorous academic peer reviews (see Bonyadi and Sarreshtehdari, 2021) and has been widely reported in media, including in The Economist, The Wall Street Journal, Financial Times and Forbes (see media tab). The methodology has also been endorsed by the European Commission and has been used to inform the allocation of EU Structural and Cohesion Funds.

[48] The sample period of the raw data is 2015-2019 because the GEI is merely available for these five years. With 2015 as the benchmark, we calculate a country's change in entrepreneurship as the 2019 GEI minus the 2015 GEI, and the growth in house price as the 2017 OECD HPI (2015 OECD HPI = 100). We use the 2017 HPI to allow for a gestation period of the effect from housing market, but Figure 7's patterns remain qualitatively intact if we use HPI in some year later to account for a longer period, because the housing appreciation in our sample are quite standing that no country exhibit reversal of its increase or decrease during 2015-2019.

in Panels (b) and (c) exhibit different patterns: in stable housing markets, the relation between housing appreciation and entrepreneurship seems positive (though weakly), consistent with the collateral channel, while in booming housing markets, the relation is negative, consistent with the speculation channel. In addition, the pattern shown in Figure 7 does not depend on which entrepreneurial index (i.e., GEI) we use, since we observe highly similar plots (unreported) if we measure entrepreneurial activities with the nationwide number of new registered firms obtained from the Bureau van Dijk database.

In summary, the results from cross-country comparisons suggest that the speculation channel, i.e., the negative effect of house prices on entrepreneurship, can be observed not only in China, but also in other economies with booming real estate markets, such as the OECD countries. These results help generalize our baseline findings and ensure the external validity of our main results.[49] On top of that, these cross-country results also suggests that the market-contingent speculating behavior in the real estate market could "mask" the positive relation predicted by the collateral channel as well, especially during house market booms, which provides a possible explanation that helps reconcile the seemingly mixed findings in the existing literature summarized by Kerr et al. (2015). Put differently, the seemingly mixed observations on the relation between house prices and entrepreneurship depend on a country's housing market booming level

7 Conclusion

In this paper, we have documented a speculation channel through which house market booms negatively affect entrepreneurship, using a unique longitudinal individual-level data. To address endogeneity concerns, we use an identification strategy that exploits staggered and unintended policy spillovers in China. We find housing speculation caused by house market booms crowds out entrepreneurship, and the effect is more pronounced for individuals with personal traits that make them more likely speculators. Alternative interpretations, such as reduced labor supply, reduced capital supply, and heightened entry cost, are unlikely to be the main driving force of our main findings and cannot explain away the housing speculation channel. Along the intensive margin, house market booms have no effect on startup size at creation but are positively associated with startup survivals. A similar negative relation between

[49] We are aware of the fact that by no means the naïve cross-country comparison could provide causal interpretation as what we do with our main results. Thus, we are not intended to claim causality, but rather to reveal the possibility that the relation between house prices and entrepreneurship could depend on the condition of the real estate market (e.g., house price growth), which points to the potential of generalizing the housing speculation channel and reconciling the mixed evidence in the existing literature.

house market booms and entrepreneurship exhibits in the OECD countries as well. Our paper offers novel evidence on a previously under-explored adverse consequence of house market booms – their hindrance to entrepreneurship.

Note that the speculation channel by no means undermines, but rather complements, the collateral channel that is well documented in the existing literature. This is because the speculation channel considers house price appreciation as an alternative investment opportunity instead of a tool of alleviating financial constraints for a given level of capital cost. In addition, our paper highlights the point that the net effect of house prices on entrepreneurship is contingent on the dynamics of housing markets, which ultimately determines the trade-off between housing speculation in booming markets and the collateral value in stable markets. Finally, our paper has important policy implications: while real estate booms could have an expansionary effect on local economy, lasting booms could induce surging speculation and dynamically transit to a more pro-cyclical and unsustainable economy.

References

Abadie, A., Athey, S., Imbens, G., Wooldridge, J., 2022. When should you adjust standard errors for clustering? The Quarterly Journal of Economics, forthcoming.

Adelino, M., Schoar, A., Severino, F., 2015. House prices, collateral, and self-employment. Journal of Financial Economics 117, 288–306.

Aghion, P., Bolton, P., 1997. A theory of trickle-down growth and development. The Review of Economic Studies 64, 151–172.

Angrist, J. D., Pischke, J.-S., 2008. Mostly harmless econometrics. Princeton university press.

Bahaj, S., Foulis, A., Pinter, G., 2020. Home values and firm behavior. American Economic Review 110, 2225–70.

Baker, A.C., Larcker, D.F. and Wang, C.C., 2022. How much should we trust staggered difference-in-differences estimates? Journal of Financial Economics, 144(2), pp.370-395.

Bekaert, G., Harvey, C. R., Lundblad, C., 2005. Does financial liberalization spur growth? Journal of Financial Economics 77, 3–55.

Bertrand, M., Mullainathan, S., 2001. Do people mean what they say? Implications for subjective survey data. American Economic Review 91, 67–72.

Bertrand, M., Mullainathan, S., 2003. Enjoying the quiet life? Corporate governance and managerial preferences. Journal of Political Economy 111, 1043–1075.

Bonyadi, E., Sarreshtehdari, L., 2021. The global entrepreneurship index (GEI): a critical review. Journal of Global Entrepreneurship Research, 1–20.

Boyreau-Debray, G., Wei, S.-J., 2005. Pitfalls of a state-dominated financial system: The case of China. Working Paper.

Chakraborty, I., Goldstein, I., MacKinlay, A., 2018. Housing price booms and crowding-out effects in bank lending. Review of Financial Studies 31, 2806–2853.

Chaney, T., Sraer, D., Thesmar, D., 2012. The collateral channel: How real estate shocks affect corporate investment. American Economic Review 102, 2381–2409.

Charles, K. K., Hurst, E., Notowidigdo, M. J., 2015. House market booms and busts, labor market opportunities, and college attendance. Working Paper.

Charles, K. K., Hurst, E., Notowidigdo, M. J., 2019. House market booms, manufacturing decline and labor market outcomes. The Economic Journal 129, 209–248.

Chen, K., Wen, Y., 2017. The great housing boom of China. American Economic Journal: Macroeconomics 9, 73–114.

Chen, T., Liu, L., Xiong, W., Zhou, L.-A., 2017. Real estate boom and misallocation of capital in China. Working paper, Princeton University.

Corradin, S., Popov, A., 2015. House prices, home equity borrowing, and entrepreneurship. Review of Financial Studies 28, 2399–2428.

Dai, R., Zhu, Z., Zhang, X., 2021. Construction and spatial pattern of China's regional innovation and entrepreneurship index: 1990-2020. Working paper, Enterprise Research Center of Peking University.

Deng, Y., Liao, L., Yu, J., Zhang, Y., 2021. Capital spillover, house prices, and consumer spending: Quasi-experimental evidence from house purchase restrictions. Review of Financial Studies 35, 3060–3099.

Djankov, S., Qian, Y., Roland, G. and Zhuravskaya, E., 2006. Who are China's entrepreneurs? American Economic Review, 96, 348-352.

Easterly, W., 2007. Inequality does cause underdevelopment: Insights from a new instrument. Journal of Development Economics 84, 755–776.

Fang, H., Gu, Q., Xiong, W., Zhou, L.-A., 2016. Demystifying the Chinese housing boom. NBER macroeconomics annual 30, 105–166.

Favilukis, J. and Van Nieuwerburgh, S., 2021. Out-of-Town Home Buyers and City Welfare. Journal of Finance 76, 2577-2638.

Gao, Z., Sockin, M., Xiong, W., 2020. Economic consequences of housing speculation. Review of Financial Studies 33, 5248–5287.
Glaeser, E., Huang, W., Ma, Y., Shleifer, A., 2017. A real estate boom with Chinese Characteristics. Journal of Economic Perspectives 31, 93–116.
Glaeser, E. L., Nathanson, C. G., 2017. An extrapolative model of house price dynamics. Journal of Financial Economics 126, 147–170.
Goodman-Bacon, A., 2021. Difference-in-differences with variation in treatment timing. Journal of Econometrics 225, 254–277.
Han, B., Han, L., Zhou, Z., 2020. Housing market and entrepreneurship: Micro evidence from China. Working paper, available at SSRN 3676832.
Huang, Y., Pagano, M., Panizza, U., 2020. Local crowding-out in China. Journal of Finance 75, 2855–2898.
Hurst, E., Lusardi, A., 2004. Liquidity constraints, household wealth, and entrepreneurship. Journal of Political Economy 112, 319–347.
Hvide, H. K., Panos, G. A., 2014. Risk tolerance and entrepreneurship. Journal of Financial Economics 111, 200–223.
Iyigun, M. F., Owen, A. L., 1998. Risk, entrepreneurship, and human-capital accumulation. American Economic Review 88, 454–457.
Jensen, T. L., Leth-Petersen, S., Nanda, R., 2021. Financing constraints, home equity and selection into entrepreneurship. Journal of Financial Economics .
Kerr, S. P., Kerr, W. R., Nanda, R., et al., 2015. House money and entrepreneurship. No. w21458, National Bureau of Economic Research.
Knight, F. H., 1921. Risk, uncertainty and profit, vol. 31. Houghton Mifflin.
Li, H., Li, J., Lu, Y. and Xie, H., 2020. Housing wealth and labor supply: Evidence from a regression discontinuity design. Journal of Public Economics, 183, p.104139.
Li, L., Wu, X., 2014. Housing price and entrepreneurship in China. Journal of Comparative Economics 42, 436–449.
Liu, C., Xiong, W., 2020. China's real estate market. The Handbook of China's Financial System; Walter de Gruyter, Inc.: Berlin, Germany pp. 183–207.
Mao, Y., 2021. Managing innovation: The role of collateral. Journal of Accounting and Economics 72, 101419
Mian, A., Sufi, A., 2011. House prices, home equity-based borrowing, and the US household leverage crisis. American Economic Review 101, 2132–56.
Mian, A., Sufi, A., 2022. Credit supply and housing speculation. Review of Financial Studies 35, 680–719.
Mian, A., Sufi, A., Trebbi, F., 2015. Foreclosures, house prices, and the real economy. Journal of Finance 70, 2587–2634.
Petersen, M. A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. Review of financial studies 22, 435–480.
Schmalz, M. C., Sraer, D. A., Thesmar, D., 2017. Housing collateral and entrepreneurship. Journal of Finance 72, 99–132.
Shiller, R. J., 2009. Unlearned lessons from the housing bubble. The Economists' Voice 6.
Song, Z., Xiong, W., 2018. Risks in China's financial system. Annual Review of Financial Economics 10, 261–286.
Stiglitz, J. E., Weiss, A., 1981. Credit rationing in markets with imperfect information. American Economic Review 71, 393–410.
Wu, J., Gyourko, J., Deng, Y., 2015. Real estate collateral value and investment: The case of China. Journal of Urban Economics 86, 43–53.
Zhang, C., Jia, S., Yang, R., 2016. Housing affordability and housing vacancy in China: The role of income inequality. Journal of Housing Economics 33, 4–14.

Table 1 Summary statistics

The sample contains all the urban-citizen respondents living in the unregulated cities, who are surveyed by CFPS and between 18 and 65, with 27,090 individual-year observations of each variable. *Entrepreneurship* equals one if the respondent is running their own business, and zero otherwise. *Homeowner* equals one if the respondent owns at least one house, and zero otherwise. *Multi-property owner* equals one if the respondent own more than one house, and zero otherwise. *Housing collateral value* is the total market value of the respondent's house(s) net of the total value of mortgage loan(s) on the house(s). *Family size* is the number of a respondent's family member(s). *Salary* is the average yearly income of a respondent's family. *Ethnicity* equals one if the respondent belongs to Han people, the majority people of China, and zero otherwise. *Female* equals one for female, and zero for male. *Age* is the respondent's age. *Marriage* equals one if the respondent is in a marriage, and zero otherwise. *Education* and *Health* are defined as a bundle of dummy variables base on the respondent's answers (checked by the interviewer) in the survey.

	Mean	SD	Min	p(10)	p(50)	p(90)	Max
Entrepreneurship	0.12	0.32	0.00	0.00	0.00	1.00	1.00
Homeowner	0.90	0.30	0.00	0.00	1.00	1.00	1.00
Multi-property owner	0.17	0.38	0.00	0.00	0.00	1.00	1.00
Housing collateral value	297,529	339,264.7	0	15,768	200,000	710,000	1,600,000
Ln(Housing collateral value)	11.26	3.34	0.00	9.67	12.21	13.47	14.29
Family size	4.15	1.62	2.00	2.00	4.00	6.00	8.00
Salary	12,526	10,147	726	2,067	10,000	27408	40,000
Ln(Salary)	9.04	1.00	6.59	7.63	9.21	10.22	10.60
Female	0.48	0.50	0.00	0.00	0.00	1.00	1.00
Age	43.47	12.49	18.00	26.00	44.00	60.00	65.00
Marriage	0.86	0.35	0.00	0.00	1.00	1.00	1.00
Ethnicity	0.96	0.19	0.00	1.00	1.00	1.00	1.00
Education							
Non-literate	0.16	0.37	0.00	0.00	0.00	1.00	1.00
Primary school	0.19	0.39	0.00	0.00	0.00	1.00	1.00
Middle school	0.33	0.47	0.00	0.00	0.00	1.00	1.00
High school	0.20	0.40	0.00	0.00	0.00	1.00	1.00
College or above	0.13	0.33	0.00	0.00	0.00	1.00	1.00

<u>*Health*</u>							
Very bad	0.21	0.41	0.00	0.00	0.00	1.00	1.00
Bad	0.22	0.42	0.00	0.00	0.00	1.00	1.00
Neutral	0.32	0.47	0.00	0.00	0.00	1.00	1.00
Good	0.14	0.35	0.00	0.00	0.00	1.00	1.00
Very good	0.11	0.31	0.00	0.00	0.00	1.00	1.00

Table 2 Summary statistics for the intensive margin analysis

The sample contains registration records of all non-foreign firms during 2004-2020 in all the non-HPR-regulated cities. At the firm level, *Ln(Registered capital)* is the logarithm of a firm's registered capital. *No. of survival months* is the monthly duration from a firm's established date to the date of its write-off or last record of normal operating activities. *3-year (5-year) survival dummy* is a binary variable that equals one if a firm is still not written off after 3 years (5 years) from its registration, and zero if a firm fails to survive 3 years (5 years). At the city level, *Average size at creation* is the within-city average of *Registered capital* in a specific year. *Average number of survival months*, *Rate of 3-year survival*, and *Rate of 5-year survival* are city-level averages of corresponding firm-level variations, respectively. We adjust all the city-level measures by aggregating corresponding firm-level measures after scaling the sample means within each industry-by-type category, where type denotes a firm's registered type of organization, including limited liability, limited shares, partnership, and others.

	Mean	SD	Min	p(10)	p(50)	p(90)	Max	N
Firm level:								
Registered capital (10 thousand RMB)	491.21	1258.32	1.00	10.00	100.00	1,000.00	1,0000.00	15,700,000
Ln(Registered capital)	4.67	1.90	0.00	2.30	4.61	6.91	9.21	15,700,000
No. of survival months	23.70	34.70	0.00	0.00	7.00	73.00	190.00	15,700,000
3-year survival dummy	0.49	0.50	0.00	0.00	0.00	1.00	1.00	9,078,185
5-year survival dummy	0.33	0.47	0.00	0.00	0.00	1.00	1.00	7,410,797
City level:								
Average size at creation	929.34	2,200.29	13.97	397.09	675.35	1,234.61	53,790.12	4,711
Average size at creation (adjusted)	1.10	0.96	0.04	0.56	0.88	1.64	12.57	4,711
Average number of survival months	41.38	33.32	0.00	2.48	37.72	87.17	142.00	4,711
Average number of survival months (adjusted)	0.99	0.32	0.00	0.69	0.96	1.32	2.71	4,711
Rate of 3-year survival	0.56	0.28	0.00	0.12	0.63	0.89	1.00	3,875
Rate of 3-year survival (adjusted)	1.04	0.37	0.00	0.67	1.00	1.39	2.97	3,324
Rate of 5-year survival	0.46	0.27	0.00	0.08	0.48	0.79	1.00	3,324
Rate of 5-year survival (adjusted)	1.06	0.56	0.00	0.57	0.96	1.63	3.86	3,046

Table 3 Spillover of house purchase restrictions and entrepreneurship

The sample contains CFPS-surveyed individuals in non-HPR cities during 2010-2018. *Entrepreneurship* equals one if the respondent is running their own business, and zero otherwise. In Panel A, *HPR spillover* equals one if individual *i* lives in a city within 200km to the closest regulated city and the regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. In Panel B, *Before* equals one for persons in treated cities if the survey year is before HPR implementations, and zero otherwise. *After* equals one for persons in treated cities if the survey year is after HPR implementations and the policy is not yet cancelled in year *t*, and zero otherwise. *Before* 1 (*Before* 2) is a binary variable that equals one for persons in treated cities if the survey year is one (two) wave(s) before HPR implementations, and zero otherwise. $Year^0$ is a binary variable that equals one for persons in treated cities at the year of HPR implementations, and zero otherwise. *After* 1 is a binary variable that equals one for persons in treated cities if the survey year is one wave after HPR implementations and the policy is not yet cancelled, and zero otherwise. *After* 2+ is a binary variable that equals one for persons in treated cities if the year is two waves or above after HPR implementations and the policy is not yet cancelled, and zero otherwise. Other variables are defined in Table 1. Robust standard errors in parentheses are clustered by individual. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Panel A: Baseline results				
	Dependent variable: Entrepreneurship dummy			
Variables	(1)	(2)	(3)	(4)
HPR spillover	-0.023***	-0.022***	-0.032***	-0.032***
	(0.01)	(0.01)	(0.01)	(0.01)
Ln(Housing collateral value)				0.002**
				(0.00)
Family size				0.004
				(0.00)
Ln(Salary)				-0.018***
				(0.00)
Female		-0.035***		
		(0.01)		
Age		-0.002***		
		(0.00)		
Married		0.061***		
		(0.01)		
<u>*Education*</u>				
Primary/middle school		0.035***		
		(0.01)		
High school		0.022**		
		(0.01)		
College or above		-0.042***		
		(0.01)		
Ethnicity (Han =1)		-0.035**		
		(0.02)		
<u>*Health*</u>				
Very good		-0.004		
		(0.01)		
Good		0.001		
		(0.01)		
Not good		-0.017***		
		(0.01)		
Bad		-0.033***		
		(0.01)		
City-level controls		Yes	Yes	Yes

Age/Marriage/Education/Health dummies			Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Individual FE			Yes	Yes
No. of observations	27,090	27,090	23,643	23,643
R-squared	0.040	0.058	0.533	0.535

Panel B: Tests on reverse causality

	Dependent variable: Entrepreneurship dummy	
Variables	(1)	(2)
Before	0.006	
	(0.02)	
After	-0.021***	
	(0.01)	
Before 2		-0.007
		(0.03)
Before 1		0.0001
		(0.02)
Year 0		-0.005
		(0.01)
After 1		-0.022**
		(0.01)
After 2+		-0.023**
		(0.01)
All controls	Yes	Yes
Year FE	Yes	Yes
Individual FE	Yes	Yes
No. of observations	23,643	23,643
R-squared	0.535	0.535

Table 4 Robustness checks

The sample contains CFPS-surveyed individuals in non-HPR cities during 2010-2018. *Entrepreneurship* dummy equals one if the respondent is running their own business, and zero otherwise. *HPR spillover* equals one if a city is within 200km to the closest regulated city and the regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. In Panel A, the 200km threshold is altered as other specific distances. *HPR (in regulated city)* equals one if a city's closest regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. *Nearness* is the additive inverse of a city's distance to the closest regulated city. In Panel C, *Employed* equals one if an individual is working for a registered entity. Extra city-level controls are a city's consumption, bank loans, citizen deposit, income Gini index, and the number of primary schools, high schools, and colleges in the reference year. All controls (as well as city-level controls) and FEs denote those in the baseline model, i.e., Column (4) of Table 3. Robust standard errors in parentheses are clustered by individual except for Column (4) of Panel C. In Panel D Column (5), the sample contains all the non-HPR cities during 2003-2019, and the dependent variable is the logarithm of the number of a city's newly registered businesses per capita in year *t+1*, which is obtained from firm registration data. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Panel A: Alternative definitions of the explanatory variable					
	Dependent variable: Entrepreneurship dummy				
	(1)	(2)	(3)	(4)	(5)
Variable	150km	175km	250km	Same province	Continuous measure
HPR spillover	-0.046***	-0.038***	-0.021***	-0.037***	
	(0.01)	(0.01)	(0.01)	(0.01)	
HPR × Nearness					-1.922***
					(0.51)
HPR in regulated city					-0.033***
					(0.01)
All Controls & FEs	Yes	Yes	Yes	Yes	Yes
No. of observations	23,643	23,643	23,643	23,643	23,643
R-squared	0.536	0.536	0.535	0.535	0.535

Panel B: Alternative sample selection					
	(1)	(2)	(3)	(4)	(5)
Variable	Non-citizen included	Age > 65 included	Excluding individuals not in the first wave	No filter	Extra filters
HPR spillover	-0.013***	-0.028***	-0.035***	-0.012***	-0.044***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)
All Controls & FEs	Yes	Yes	Yes	Yes	Yes
No. of observations	67,466	29,400	22,590	81,540	6,174
R-squared	0.505	0.538	0.526	0.505	0.538

Panel C: Alternative model specification

Variable	(1) Including employment status	(2) Including extra city-level controls	(3) City-ownership status clustering	(4) Two-way clustering	(5) Probit regression
HPR spillover	-0.030***	-0.022***	-0.032**	-0.032**	-0.188***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Employed	0.149***				
	(0.01)				
Extra city-level controls		Yes			
All Controls & FEs	Yes	Yes	Yes	Yes	Yes
No. of observations	23,643	23,313	23,643	23,643	23,643
R-squared	0.551	0.540	0.535	0.535	

Panel D: Diagnostic tests for staggered DiD & Alternative data source/estimations

Variable	(1) Excluding the last wave	(2) Stacked regression	(3) Goodman-Bacon (2021)'s decomposition	(4) Household-level estimation	(5) City-level estimation with firm registration data
HPR spillover	-0.037***	-0.042***		-0.030***	-0.155**
	(0.01)	(0.00)		(0.01)	(0.08)
Treated vs. Never T.			-0.022		
			[weight=0.882]		
Earlier T. vs. Later C.			-0.142		
			[weight=0.081]		
Later T. vs Earlier C.			-0.026		
			[weight=0.037]		
City-level controls					Yes
City FEs and Year FEs					Yes
All Controls & FEs	Yes	Yes		Yes	
No. of observations	19,108	23,643		8,540	3,664
R-squared	0.530			0.600	0.790

Table 5 Placebo tests

In Panel A, the sample contains all the CFPS-surveyed renters (i.e., individuals have no house) in non-HPR cities during 2010-2018. In Panel B, the sample for all the regressions contains CFPS-surveyed individuals in non-HPR cities during 2010-2018, and the panel reports the distribution of the coefficient estimates (and *t*-stats) of randomly falsified *HPR spillover* with 1000 replications. The dependent variable *Entrepreneurship* equals one if the respondent is running their own business in year $t + 1$, and zero otherwise. *HPR spillover* equals one if individual *i* lives in a city within 200km to the closest regulated city and the regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. All the regressions include the same control variables and fixed effects as those in Table 3, but they are not tabulated. Robust standard errors in parentheses are clustered by individual. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Panel A: Subsample of renters			
	Dependent variable: Entrepreneurship dummy		
Variables	(1)	(2)	(3)
HPR spillover	0.002	-0.0004	0.018
	(0.02)	(0.03)	(0.03)
Family size			0.011
			(0.02)
Ln(Salary)			-0.039**
			(0.02)
City controls		Yes	Yes
Year FEs	Yes	Yes	Yes
Individual FEs		Yes	Yes
No. of observations	3,550	1,462	1,395
R-squared	0.001	0.641	0.606

Panel B: Monte Carlo tests with randomly falsified shocks						
	Mean	p(5)	p(25)	p(50)	p(75)	P(95)
Coefficients of falsified *HPR spillover*	0.000285	-0.00763	-0.00210	0.00062	0.00310	0.00717
t-stats	[0.05493]	[-1.549]	[-0.411]	[0.124]	[0.607]	[1.434]

Table 6 Heterogeneity tests based on individual-level multi-property ownership

The sample contains CFPS-surveyed individuals in non-HPR cities during 2010-2018. *Entrepreneurship* equals one if the respondent is running their own business, and zero otherwise. *HPR spillover* equals one if individual *i* lives in a city within 200km to the closest regulated city and the regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. The binary variable *Multi-property ownership* equals one if a respondent has more than two houses, and zero otherwise. *No. of extra houses* equals one's number of house(s) other than the house they are living in if they have multiple properties, and zero if they have no house or only one house. *Single-property owner* equals one if an individual has only one house (full ownership), and zero otherwise. *Has property income* equals one if a respondent's family earn money from their property, and zero otherwise. In Column (4), the sample contains the 2014 and 2016 waves of the CFPS survey that contains the individuals who answered the question about house price expectation in the 2014 wave; *High expectation* equals one if a respondent believes that the local house prices is going to increase sizably, and zero otherwise. All the regressions include the same control variables as those in Table 3, as well as their interactions with the newly added variables, but they are not tabulated. Robust standard errors in parentheses are clustered by individual. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

	Dependent variable: Entrepreneurship dummy			
Variables	(1)	(2)	(3)	(4)
HPR spillover × Multi-property owner	-0.054***			
	(0.02)			
Multi-property owner	0.132			
	(0.39)			
HPR spillover × No. of extra houses		-0.037***		
		(0.01)		
No. of extra houses		-0.298		
		(0.28)		
HPR spillover × Has property income			-0.030**	
			(0.01)	
Has property income			-0.216	
			(0.31)	
HPR spillover × Single-property owner	-0.037**	-0.030**	-0.006	
	(0.02)	(0.01)	(0.01)	
Single-property owner	-0.073	-0.360	-0.194	
	(0.33)	(0.29)	(0.23)	
HPR spillover × High expectation				-0.199**
				(0.10)
HPR spillover	0.007	-0.000	-0.021*	0.011
	(0.02)	(0.01)	(0.01)	(0.02)
Controls & interactions	Yes	Yes	Yes	Yes
Year/Individual FEs	Yes	Yes	Yes	Yes
No. of observations	23,643	23,641	23,527	3,522
R-squared	0.537	0.537	0.532	0.781

Table 7 Ruling out alternative interpretation: reduced labor supply

The sample contains CFPS-surveyed individuals in non-HPR cities during 2010-2018. In Columns (4)-(6) of Panel B, the sample is restricted to the individuals that always work as employees, because the job income of those who do not work for others is unobservable. *Entrepreneurship* equals one if the respondent is running their own business, and zero otherwise. *HPR spillover* equals one if individual *i* lives in a city within 200km to the closest regulated city and the regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. *Ln(Net assets)* is the log value of individual's family assets net of debts. In Panel B, *Work-or-quit* dummy denotes whether an individual work (as a business owner or an employee) or not, and *Ln(Job income)* denotes an individual's wage from working as an employee. The other variables are defined the same as those in Table 6. All the regressions include the same control variables as those in Table 3, as well as their interactions with the newly added variables, but they are not tabulated. Robust standard errors in parentheses are clustered by individual. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Panel A: Ruling out the wealth effect on labor supply					
	Dependent variable: Entrepreneurship dummy				
Variables	(1)	(2)	(3)	(4)	(5)
HPR spillover × Multi-property owner		-0.058**			
		(0.02)			
Multi-property owner		-0.286			
		(0.46)			
HPR spillover × No. of extra houses			-0.031**		
			(0.01)		
No. of extra houses			-0.514*		
			(0.31)		
HPR spillover × Has property income				-0.030**	
				(0.01)	
Has property income				-0.328	
				(0.32)	
HPR spillover × Single-property owner		-0.046**	-0.031**	-0.011	
		(0.02)	(0.02)	(0.01)	
Single-property owner		-0.322	-0.476	-0.186	
		(0.36)	(0.30)	(0.23)	
HPR spillover × High expectation					-0.211**
					(0.10)
HPR spillover × Ln(Net assets)	-0.004	0.003	0.001	-0.002	-0.295
	(0.00)	(0.00)	(0.00)	(0.00)	(0.38)
Ln(Net assets)	0.123	0.144	0.181**	0.144*	0.000
	(0.08)	(0.09)	(0.09)	(0.08)	(0.01)
HPR spillover	0.013	-0.024	-0.008	0.006	0.009
	(0.05)	(0.05)	(0.05)	(0.05)	(0.13)
Controls & interactions	Yes	Yes	Yes	Yes	Yes
Year/Individual FEs	Yes	Yes	Yes	Yes	Yes
No. of observations	23,075	23,057	23,057	22,959	3,388
R-squared	0.537	0.539	0.538	0.533	0.785

Panel B: Ruling out the labor supply argument

Variables	Dependent variable: *Work-or-quit* dummy			Dependent variable: *Ln(Job income)*		
	(1)	(2)	(3)	(4)	(5)	(6)
HPR spillover × Multi-property owner		-0.020	0.024		0.046	-0.156
		(0.02)	(0.03)		(0.23)	(0.28)
Multi-property owner		0.180	-0.542		-1.464	-4.634
		(0.54)	(0.61)		(4.67)	(5.33)
HPR spillover × Ln(Net assets)			-0.017***			0.046
			(0.01)			(0.06)
Ln(Net assets)			0.192			0.895
			(0.13)			(1.16)
HPR spillover × Single-property owner		-0.042**	-0.020		-0.021	-0.182
		(0.02)	(0.02)		(0.21)	(0.24)
Single-property owner		-0.697	-1.233**		-4.894	-7.232
		(0.47)	(0.49)		(4.15)	(4.56)
HPR spillover	-0.012	0.023	0.214***	-0.099	-0.115	-0.519
	(0.01)	(0.02)	(0.07)	(0.08)	(0.21)	(0.63)
Controls & interactions	Yes	Yes	Yes	Yes	Yes	Yes
Year/Individual FEs	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	23,643	23,643	23,075	10,244	10,244	10,000
R-squared	0.622	0.624	0.628	0.748	0.750	0.754

Table 8 Ruling out alternative interpretation: reduced capital supply

In Columns (1)-(2), the sample contains IRIEC-surveyed non-HPR cities during 2009-2019. In Columns (3)-(4), the sample contains CFPS-surveyed individuals in non-HPR cities during 2010-2018. *VC inv. score* is a standardized score reflecting a city's level of VC investment activities in a specific year. *Entrepreneurship* equals one if the respondent is running their own business, and zero otherwise. *HPR spillover* equals one if individual *i* lives in a city within 200km to the closest regulated city and the regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. In Columns (1)-(2), the regressions include the same city-level control variables as those in Table 3, but they are not tabulated; the dependent variable is measured one year ahead; and robust standard errors in parentheses are clustered by city. In Columns (3)-(4), the regressions include the same control variables as those in Table 3, as well as their interactions with the newly added variables, but they are not tabulated; and robust standard errors in parentheses are clustered by individual. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Variables	(1) *VC inv. score*	(2) *Ln(VC inv. score)*	(3) *Entrepreneurship*	(4) *Entrepreneurship*
HPR spillover ×VC inv. score			0.0001	
			(0.0002)	
VC inv. score			0.006	
			(0.005)	
HPR spillover × Ln(VC inv. score)				0.003
				(0.01)
Ln(VC inv. score)				0.187
				(0.13)
HPR spillover	-0.793	-0.018	-0.041***	-0.048*
	(1.39)	(0.06)	(0.01)	(0.03)
All controls & interactions			Yes	Yes
Year/Individual FEs			Yes	Yes
City-level controls	Yes	Yes		
Year/City FEs	Yes	Yes		
No. of observations	2,126	2,126	23,415	23,415
R-squared	0.607	0.432	0.533	0.533

Table 9 Heterogeneity tests based on one's personality and attitude

The sample contains CFPS-surveyed individuals in non-HPR cities during 2010-2018. *Entrepreneurship* equals one if the respondent is running their own business, and zero otherwise. *HPR spillover* equals one if individual *i* lives in a city within 200km to the closest regulated city and the regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. The newly added dummy variables indicate the answers of corresponding questions in CFPS survey. All the regressions include the same control variables as those in Table 3, as well as their interactions with the newly added variables, but they are not tabulated. Robust standard errors in parentheses are clustered by individual. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

	Dependent variable: Entrepreneurship dummy				
Variables	(1)	(2)	(3)	(4)	(5)
HPR spillover × I'm ranking high in income	-0.032***				
	(0.01)				
HPR spillover × I agree with "Wealth is success"		-0.037***			
		(0.01)			
HPR spillover × I'm confident to my future			-0.034**		
			(0.01)		
HPR spillover × I have no experience of inequality				-0.073***	
				(0.02)	
HPR spillover	-0.024***	-0.008	-0.009	-0.097***	-0.059***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
HPR spillover × Primary school					-0.043***
					(0.01)
HPR spillover × Middle school					-0.031***
					(0.01)
HPR spillover × High school					-0.021*
					(0.01)
HPR spillover × College and above					0.001
					(0.01)
Controls & interactions	Yes	Yes	Yes	Yes	Yes
Year/Individual FEs	Yes	Yes	Yes	Yes	Yes
No. of observations	19,443	19,523	19,987	20,708	23,643
R-squared	0.515	0.511	0.515	0.516	0.537

Table 10 Heterogeneity tests based on city-level supply of disposable liquidity

The sample contains CFPS-surveyed individuals in non-HPR cities during 2010-2018. *Entrepreneurship* equals one if the respondent is running their own business, and zero otherwise. *HPR spillover* equals one if individual *i* lives in a city within 200km to the closest regulated city and the regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. The binary variable *Income inequality* equals one if the income Gini index for the city is bigger than 0.4, and zero otherwise. *Citizen deposit*, *Bank loan*, *GDP*, *Fiscal expenditure*, and *Unemployment rate* are reported in the *Yearbook of Statistics* and taken logarithms. All the regressions include the same control variables as those in Table 3, as well as their interactions with the newly added variables, but they are not tabulated. Robust standard errors in parentheses are clustered by individual. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Panel A: Supply of disposable liquidity			
	Dependent variable: Entrepreneurship dummy		
Variables	(1)	(2)	(3)
HPR spillover × Income inequality	-0.035***		
	(0.01)		
Income inequality (Gini index >0.4)	-0.394*		
	(0.21)		
HPR spillover × Ln(Citizen deposit)		-0.021**	
		(0.01)	
Ln(Citizen deposit)		0.475***	
		(0.18)	
HPR spillover × Ln(Bank loan)			-0.019**
			(0.01)
Ln(Bank loan)			0.503***
			(0.18)
HPR spillover	-0.008	0.312**	0.275**
	(0.01)	(0.13)	(0.13)
Controls & interactions	Yes	Yes	Yes
Year/Individual FEs	Yes	Yes	Yes
No. of observations	23,643	23,643	23,643
R-squared	0.537	0.538	0.537

Panel B: Ruling out alternative explanations

Variables	(1)	(2)	(3)
HPR spillover × Ln(GDP)	-0.002		
	(0.01)		
Ln(GDP)	0.519***		
	(0.18)		
HPR spillover × Ln(Fiscal expenditure)		-0.004	
		(0.01)	
Ln(Fiscal expenditure)		0.606**	
		(0.25)	
HPR spillover × Unemployment rate			-0.004
			(0.01)
Unemployment rate			0.789***
			(0.28)
HPR spillover	-0.062	0.024	0.010
	(0.14)	(0.18)	(0.08)
Controls & interactions	Yes	Yes	Yes
Year/Individual FEs	Yes	Yes	Yes
No. of observations	23,643	23,643	23,643
R-squared	0.536	0.537	0.536

Table 11 Intensive margin: housing speculation and entrepreneurial outcomes

In Panel A, the sample contains registration records of all non-foreign firms during 2004-2020 in all the non-HPR-regulated cities. In Panel B, the sample contains all the non-HPR cities. *HPR spillover* equals one if a city is within 200km to the closest regulated city and the regulated city is presently under the house purchase restrictions in year *t*, and zero otherwise. All the dependent variables are measured one year ahead. City-level control variables are as those in Table 3 and obtained from China's city-level *Yearbook of Statistics*. All the regressions include city-level controls, city-by-industry-by-type fixed effects, and year fixed effects, but they are not tabulated. Robust standard errors in parentheses are clustered by city-by-industry in Panel A, and by city in Panel B. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

Panel A: Firm-level estimation				
	Size at creation	Survival		
	(1)	(2)	(3)	(4)
Variables	*Ln(Registered capital)*	*Ln(No. of months)*	*3-year survival dummy*	*5-year survival dummy*
HPR spillover	-0.002	0.017**	0.024***	0.016**
	(0.01)	(0.01)	(0.01)	(0.01)
City-level controls	Yes	Yes	Yes	Yes
City × Industry × Org. type FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
No. of observations	13,613,497	7,345,324	6,266,182	5,455,288
R-squared	0.314	0.467	0.224	0.352

Panel B: City-level estimation				
	Size & Survival at the city level (adjusted for industry and registered organizational type fixed effects)			
	(1)	(2)	(3)	(4)
Variables	*Average size at creation*	*Average survival duration*	*Rate of 3-year survival*	*Rate of 5-year survival*
HPR spillover	0.001	0.050**	0.078***	0.156**
	(0.06)	(0.02)	(0.01)	(0.07)
City-level controls	Yes	Yes	Yes	Yes
City FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
No. of observations	3,683	3,683	2,611	2,410
R-squared	0.365	0.543	0.556	0.450

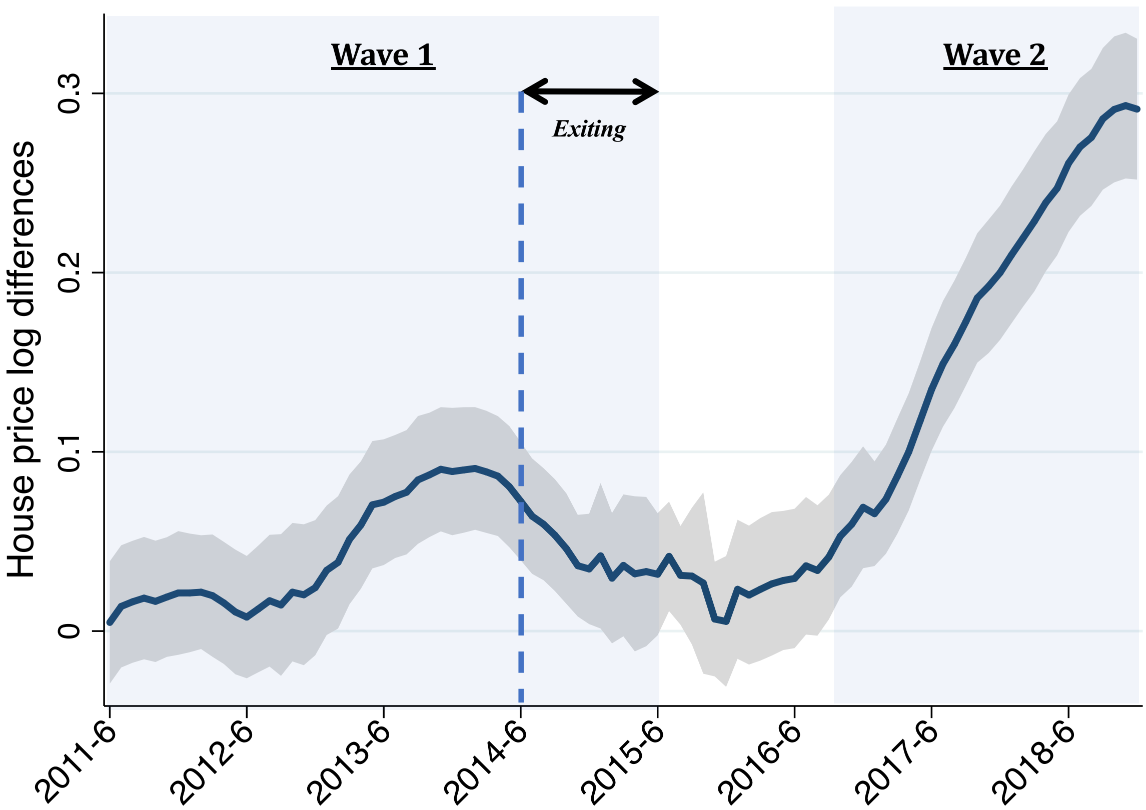


Figure 1. First stage – The geographical spillover effects of the two-round HPR policies on non-HPR cities' house prices. The sample contains monthly average house prices of 42 non-HPR cities (out of 100 cities in China) from June 2011 to December 2018. A city's house prices are scaled by its city-level mean to make the prices comparable between cities. The solid line captures the monthly differences in log house prices between nearer cities (treatment group) and farther cities (control group) defined in our baseline model. The shade around the solid line denotes 95% confidence intervals (robust standard errors are clustered by city and year-month). In the first wave restrictions, *Exiting* denotes the period during which most of the HPR cities gradually exit the first wave policies.

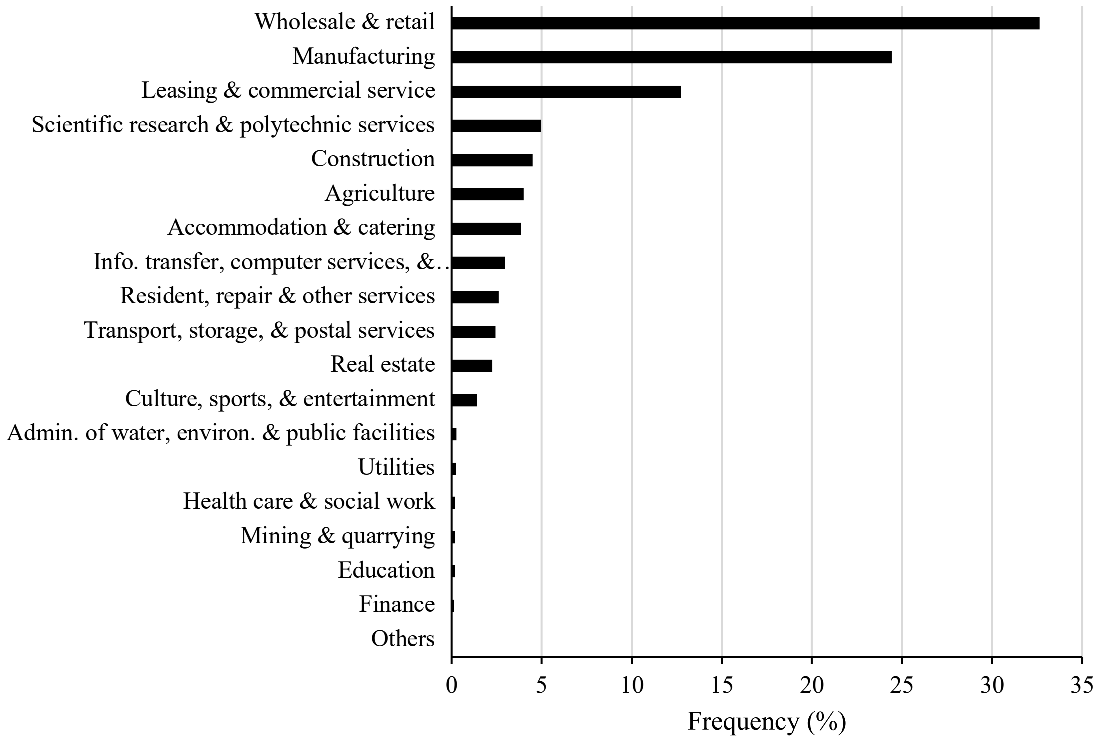


Figure 2. Industry distributions of newly created businesses based on industry classifications defined by a machine learning technique. The graph shows the industry distribution of the businesses created during 2004-2020 that are in our sample. The industries are classified by our machine learning approach.

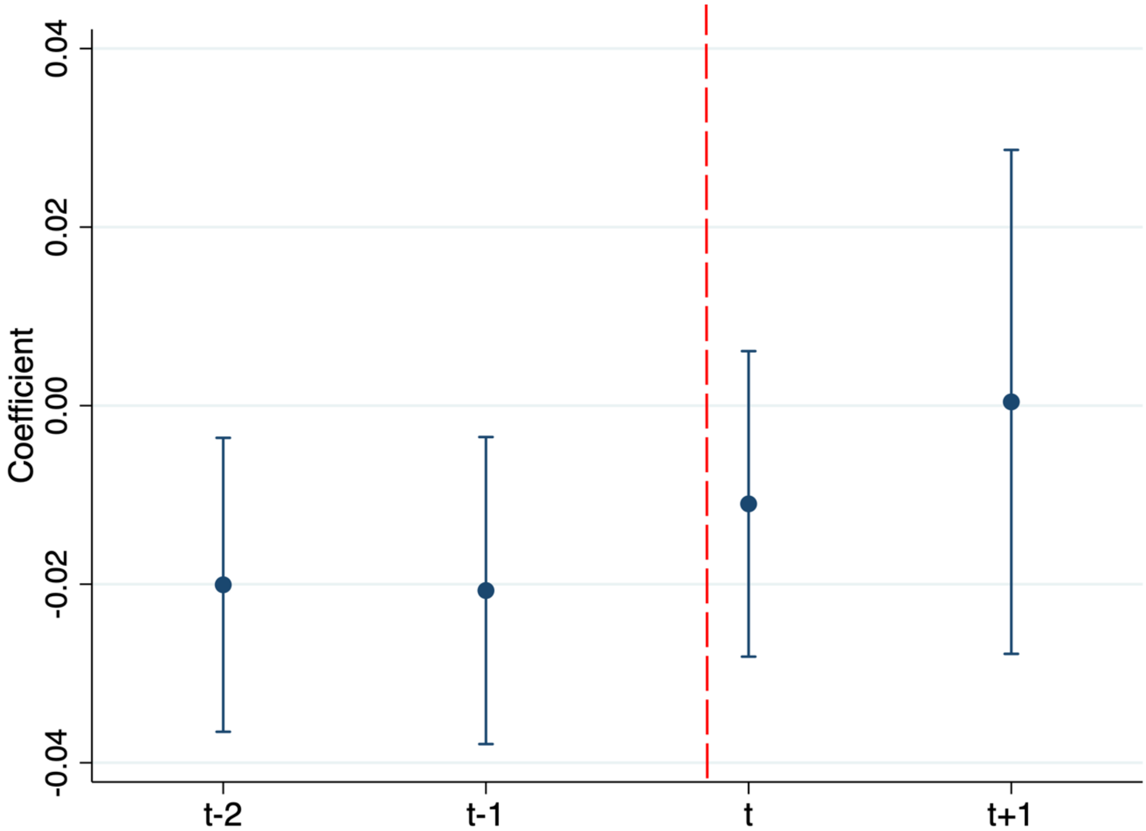


Figure 3. The effect of HPR policy cancellations in an event-study framework. The figure shows the coefficient estimates and their 90% confidence intervals of a set of dummies that equal one if an individual is in a nearer non-HPR (treated) city in a specific year around HPR cancellations, and zero otherwise. For the time indicators, *t-2*, *t-1*, *t*, and *t+1* denotes the year that is two periods before, one period before, the period of, and one period after, the cancellation, respectively. The vertical dashed line represents the time of HPR cancellations in a focal city's nearest HPR city.

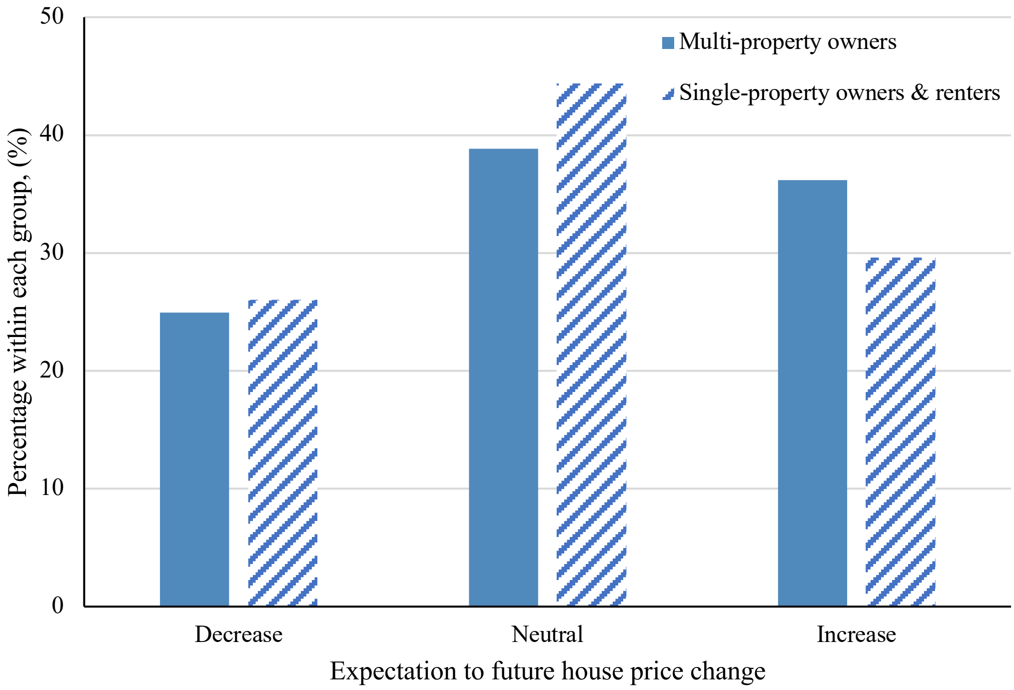


Figure 4. Distribution of the expectation on future house price change (Multi-property owners vs. Others). The sample contains the cross-section of all the CFPS-surveyed individuals, in treated cities and in control cities, responding to the extra survey of the 2014 wave, in which an individual is asked ”In your opinion, how will the housing price change in your living area?” *Multi-property owners* are those who own multiple houses, while *Others* are single-property owners and renters.

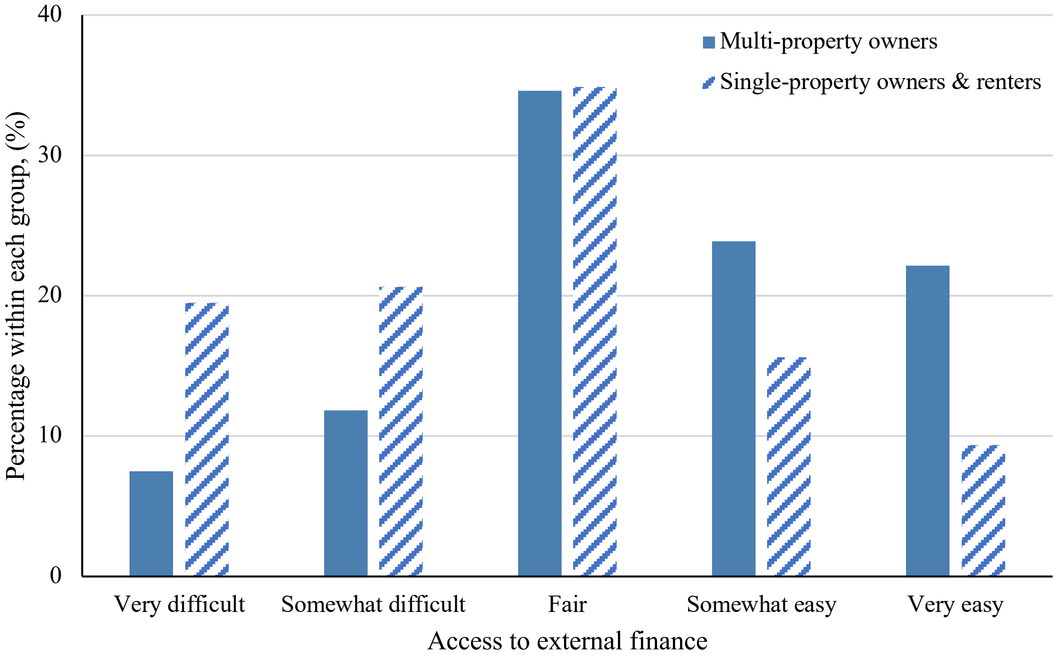


Figure 5. Distribution of the access to external finance. The sample contains the cross-section of all the CFPS-surveyed individuals responding to the extra survey of the 2014 wave, in which an individual is asked ”If you have to borrow a total amount of 20,000 yuan in case of some emergency. How difficult will it be to raise the money?” *Multi-property owners* are those who own multiple houses, while *Others* are single-property owners and renters.

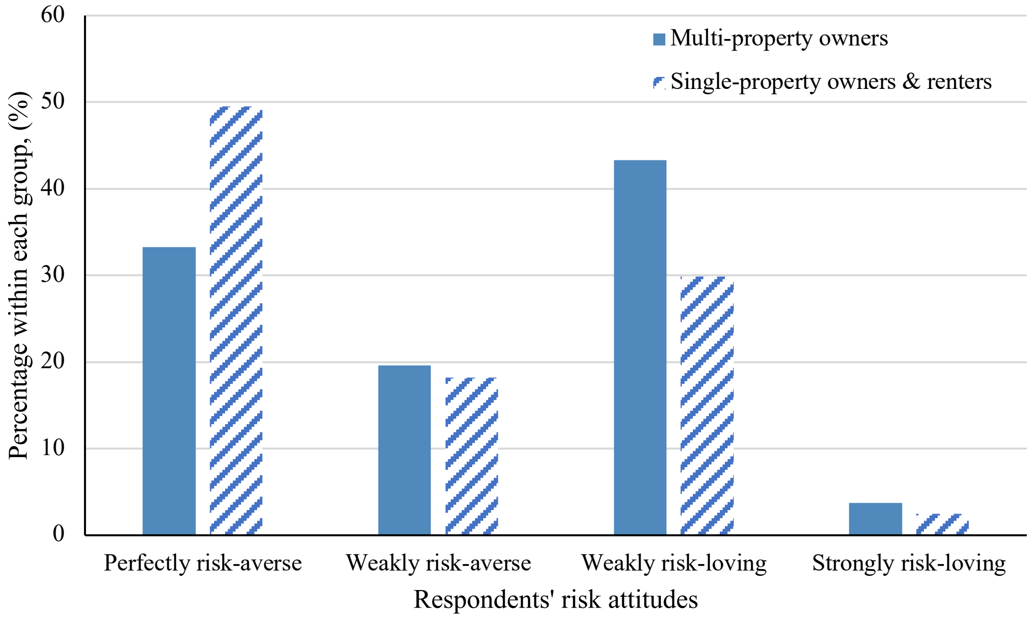
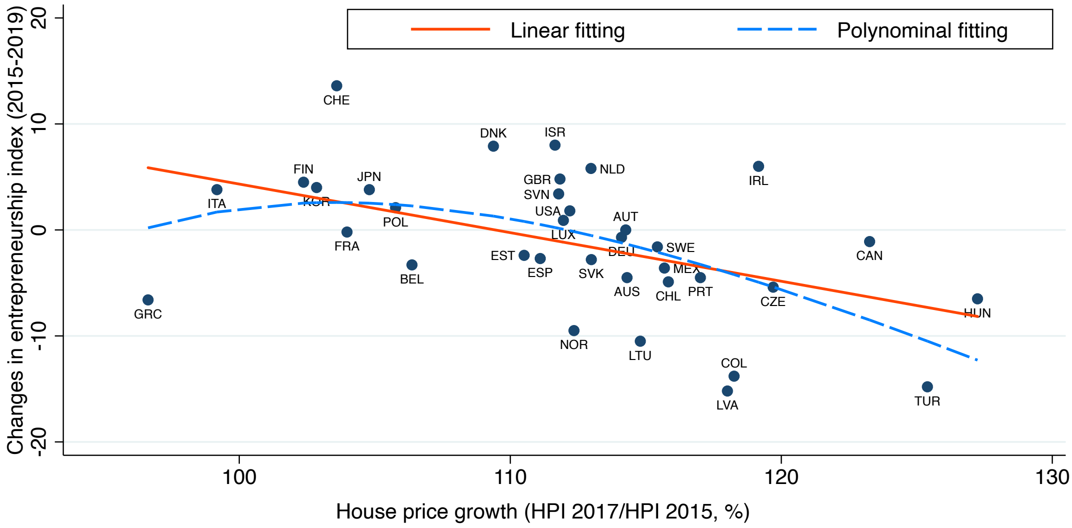
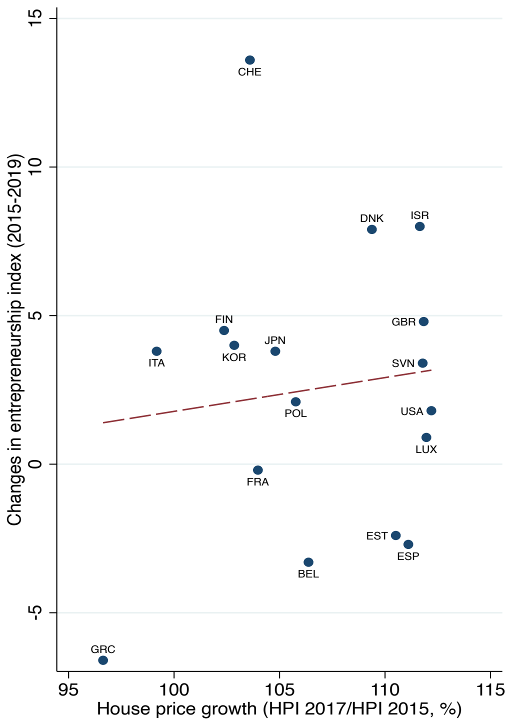


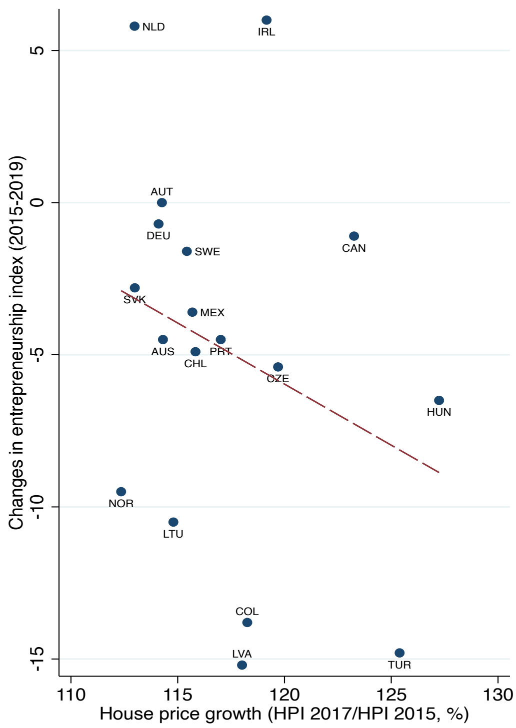
Figure 6. Distribution of the risk attitudes. The sample contains the cross-section of all the CFPS surveyed individuals responding to the extra survey of the 2014 wave, in which an individual is asked "If your family invest/In investment, what kind of risk are you willing to take?" *Perfectly risk-averse*, *Weakly risk-averse*, *Weakly risk-loving*, and *Strongly risk-loving* correspond to the answers of *Unwilling to take any investment risk*, *Low risk and low return*, *Moderate risk and steady return*, and *High risk and high return*, respectively. *Multi-property owners* are those who own multiple houses, while *Others* are single-property owners and renters.



(a) House price growth and entrepreneurship in OECD countries (2015-2019)



(b) Sample split: stable markets



(c) Sample spilt: booming markets

Figure 7. Housing price and entrepreneurship: global perspective.
The figure plots changes in Global Entrepreneurship Index (from 2015 to 2019) against the house price growth (from 2015 to 2017), of OECD countries. The Global Entrepreneurship Index is published by the Global Entrepreneurship and Development Institute (GEDI) in the US. The house price growth is calculated from the house price index on OECD's website. The dashed line plots the fitted values from ordinary least squares (OLS) regressions. Panel (b) and (c) are subsamples of those in Panel (a), based42 on the sample median of the house price index.

Internet Appendix (Not to be published)

Internet Appendix A: House purchase restrictions in China (2010-2019)

This table summarizes all the house purchase restriction policies at the city (prefecture) level from 2010 to 2019, in terms of the name of the regulated city, date of policy adoptions, cancellations, and re-implementations, and whether the policy measures contain purchase restrictions, loan restrictions, resale restrictions, and price restrictions. “Y” denotes that the corresponding measure is always in use during the period of regulation. “R” denotes that the corresponding measure is effective only within some sub-period(s) of regulation. Blank in the last four columns means a city has never include the corresponding measure in their HPR policy.

Province	City	Adoption	Cancellation	Re-implementation	Purchase	Loan	Resale	Price
Beijing	Beijing	2010/4/30			Y	Y	R	R
Guangdong	Shenzhen	2010/9/30			Y	Y	R	R
Hainan	Sanya	2010/10/12			Y	Y	R	R
Guangdong	Guangzhou	2010/10/15			Y	Y	R	R
Shanghai	Shanghai	2011/1/31			Y	Y	R	R
Zhejiang	Shaoxing	2011/4/8	2014/8/4		Y	Y	R	R
Inner Mongolia	Hohhot	2011/4/14	2014/6/26		Y	Y		
Zhejiang	Quzhou	2011/9/9	2014/8/1		Y	Y	R	R
Zhejiang	Wenzhou	2010/10/12	2014/7/28		Y	Y		
Zhejiang	Zhoushan	2010/10/12	2014/8/2		Y	Y		
Zhejiang	Taizhou	2011/8/25	2014/8/19		Y	Y		
Ningxia	Yinchuan	2011/2/22	2014/8/22		Y	Y		
Qinghai	Xining	2011/2/25	2014/9/19		Y	Y		
Xinjiang	Urumchi	2011/2/28	2014/10/23		Y	Y		
Jilin	Changchun	2011/1/28	2015/6/4	2016/6/13	Y	Y	R	R
Fujian	Xiamen	2010/9/29	2015/1/16	2016/8/31	Y	Y	R	R
Zhejiang	Hangzhou	2010/10/11	2014/8/29	2016/9/19	Y	Y	R	
Jiangsu	Nanjing	2011/2/19	2014/9/21	2016/9/26	Y	Y	R	Y
Sichuan	Chengdu	2011/2/15	2014/7/24	2016/10/1	Y	Y	R	R
Henan	Zhengzhou	2010/6/7	2014/8/9	2016/10/1	R	Y	R	R
Tianjin	Tianjin	2010/10/13	2014/10/17	2016/10/1	Y	Y		
Shandong	Jinan	2011/2/21	2014/7/10	2016/10/2	Y	Y	R	
Anhui	Hefei	2011/1/25	2014/8/1	2016/10/2	Y	Y		R

Jiangsu	Wuxi	2011/2/20	2014/8/29	2016/10/2	Y	Y	R	R
Hubei	Wuhan	2011/1/14	2014/9/23	2016/10/3	Y	Y		
Jiangsu	Suzhou	2010/11/3	2014/9/26	2016/10/4	Y	Y	R	R
Fujian	Fuzhou	2010/10/11	2014/8/1	2016/10/6	Y	Y	R	R
Guangdong	Zhuhai	2011/10/31	2016/5/3	2016/10/6	Y	Y	R	R
Jiangxi	Nanchang	2011/2/23	2014/8/12	2016/10/8	Y	Y	R	R
Guangdong	Foshan	2011/3/18	2015/4/30	2016/10/8	Y	Y		
Shaanxi	Xi'an	2011/2/25	2014/9/1	2017/1/1	Y	Y	R	R
Shandong	Qingdao	2011/1/30	2014/9/1	2017/3/16	Y	Y	R	
Hebei	Shijiazhuang	2011/2/19	2014/9/25	2017/3/17	Y	Y	R	R
Hunan	Changsha	2011/3/4	2014/8/6	2017/3/18	Y	Y	R	R
Gansu	Lanzhou	2010/7/7	2014/9/3	2017/4/7	Y	Y	R	R
Hainan	Haikou	2010/12/30	2014/7/22	2017/4/14	Y	Y	R	R
Zhejiang	Ningbo	2010/10/9	2014/8/30	2017/4/24	Y	Y	R	R
Guangxi	Nanning	2011/2/18	2014/10/1	2017/5/26	Y	Y	R	R
Jiangsu	Xuzhou	2011/4/16	2014/8/1	2017/6/1	Y	Y	R	R
Guizhou	Guiyang	2011/2/11	2014/9/1	2017/9/23	Y	Y	R	
Yunnan	Kunming	2011/1/14	2014/8/11	2018/3/1	Y	Y	R	R
Liaoning	Dalian	2010/10/11	2014/9/9	2018/3/21	Y	Y	R	R
Liaoning	Shenyang	2011/2/25	2015/10/19	2018/4/15	Y	Y	Y	Y
Heilongjiang	Harbin	2011/2/18	2014/8/16	2018/5/7	Y	Y	R	R
Shanxi	Taiyuan	2011/2/19	2014/8/4	2018/5/18	Y	Y	R	R
Hebei	Langfang	2016/4/2			Y	Y		
Guangdong	Dongguan	2016/10/7			Y	Y	R	R
Zhejiang	Jiaxing	2016/12/3			Y	Y	Y	
Jiangxi	Ganzhou	2017/3/14			Y	Y		Y
Hebei	Baoding	2017/3/19			Y	Y	Y	Y
Hebei	Cangzhou	2017/3/23			Y	Y		
Guangdong	Zhongshan	2017/3/26			Y	Y		
Hebei	Tangshan	2017/4/14			Y	Y	R	R
Hebei	Qinhuangdao	2017/4/15			Y	Y		
Fujian	Quanzhou	2017/4/17			Y	Y	Y	
Guangdong	Qingyuan	2017/4/21			Y	Y		
Guangdong	Jiangmen	2017/4/22			Y	Y		R
Hebei	Chengde	2017/5/14			Y	Y	Y	Y
Hebei	Zhangjiakou	2017/5/26			Y	Y	Y	Y
Jiangsu	Huai'an	2017/5/30			Y	Y		

Jiangxi	Jiujiang	2017/7/13	Y	Y		
Hubei	Xiaogan	2017/9/13	Y	Y		
Guangxi	Beihai	2017/9/30	Y		Y	
Jiangsu	Yangzhou	2017/11/30	Y	Y	Y	Y
Hubei	E’zhou	2017/12/27	Y	Y	Y	Y
Fujian	Ningde	2018/7/11	Y	Y	Y	Y
Yunnan	Pu'er	2018/7/19	Y	Y	Y	Y
Liaoning	Dandong	2019/4/30	Y	Y	Y	Y

Internet Appendix B: Identifying industry classifications with machine learning techniques

In empirical studies on entrepreneurship, researchers are often faced with a common challenge of identifying systematically unified industry classifications for startups, because researchers need the inclusion of industry-related fixed effects to account for various differences across industries. Neglect of these differences would likely induce misleading results. However, entrepreneurial companies are by nature privately owned at creation, which means that there is no compulsorily industry classification such as SIC codes for listed firms.

To address this challenge and construct an internally consistent industry classification, we develop a simple machine learning approach to identify (or predict) a firm's industry, using the information in mandatory disclosure of firm registration. Our approach is of an "AI + Human" style and mainly relies on the "sklearn" package on Python. Specifically, we first leverage textual analysis to identify a distribution of key words in a firm's business scope, and then use multiple textual-polychotomous machine learning models to learn and predict based on the assigned TF-IDF (Term frequency – Inverse document frequency) eigenvalues, in order to ultimately pick out a winner model with the highest precision. In addition, we also manually audit the results in the process of optimization to identify, which helps in enhancing the effectiveness due to the inclusion of more prior knowledge. The industry codes in our training data and testing data are manually assigned, but this would not materially undermine the potential of generalization of our approach because one can easily figure out the industry at reading the business scope. The highest prediction accuracy reaches 73.95%. In this paper, we use this approach to deal with firm registration data in Chinese, but the methodology is potentially not subject to specific language.

A brief introduction of the four major steps of our machine learning approach in this paper is as follows.

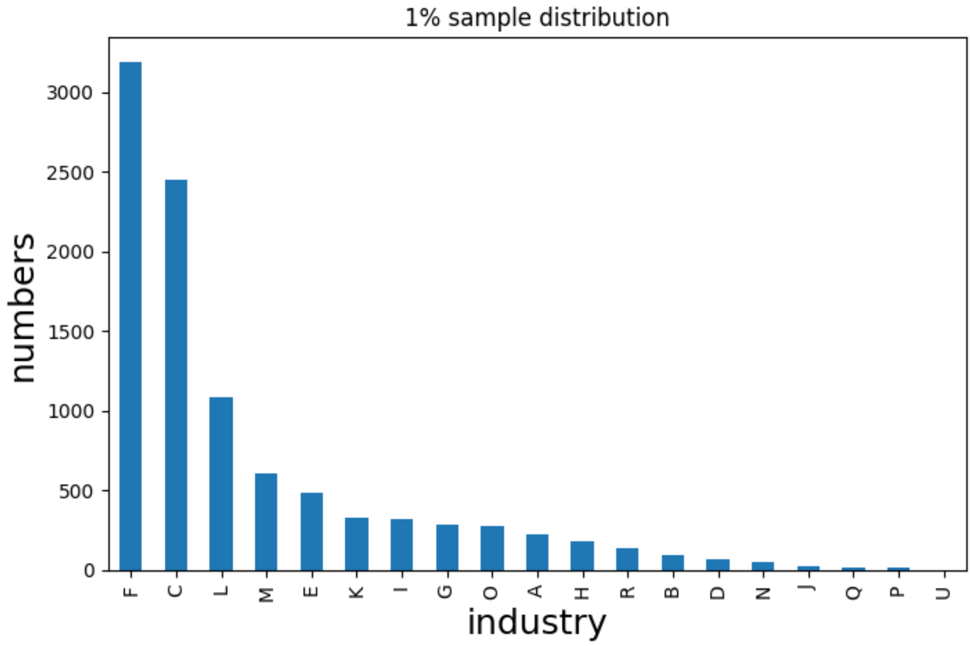
Step 1: Data

Our data contains more than 50 million records of firm registration obtained from State Administration for Industry and Commerce of the People's Republic of China (SAIC). We begin with one-year registration data and manually identify an industry (out of the 19 major classifications) for each firm by reading their description of main business. To enhance the coverage of the training data, we also include all the publicly traded firms (with their business scope), to each of which a unique industry code is officially assigned at listing. We presently choose a 19-industry classification, instead of those subcategories, because the 19 industries are more likely to significantly differ from each other, and thus can help prevent arbitrary biases in our manually assigning industries.

For brevity, hereafter, we use one-digit letter to represent the 19 industries as follows:

A	Agriculture	K	Real estate
B	Mining & quarrying	L	Leasing & commercial service
C	Manufacturing	M	Scientific research & polytechnic services
D	Utilities	N	Administration of water, environment & public facilities
E	Construction	O	Resident, repair & other services
F	Wholesale & retail	P	Education
G	Transport, storage, & postal services	Q	Health care & social work
H	Accommodation & catering	R	Culture, sports, & entertainment
I	Information transfer, computer services, & software	U	Others
J	Finance		

Our full sample for machine learning is a mix of private firms (dominant majority) and public firms that contains 987,387 observations. We randomly draw 70% of the sample as the training data and the left 30% as the out-of-sample testing data. Here is a representative distribution (1% of the sample) of the 19 industries.



Step 2: Pre-processing

We then step forward to identify separate words. Since Chinese sentences are sequential characters, unlike English that puts blank space between words, word identifying is even more challenging. We first use “stopwords” to delete meaningless characters and identify words, and then put blanks between these words.

Next, we leverage TF-IDF (Term frequency – Inverse document frequency) to compute the eigenvalues of each word, which denote the importance of each word within each observation of business scope. The TF-IDF eigenvalues increase with the total counts of the words yet decrease with the frequencies of the words.

Then we use sklearn.feature_extraction.text.TfidfVectorizer to get eigenvalues. Here we set ngram_range = (1, 2), which means we not only consider each identified word per se, but also take the combination of a word and its neighboring words, i.e., word pairs, into account. This execution helps the machine better understand the context meaning by expanding the variety of features, and ultimately enhance precision. We additionally set norm='l2' and normalize the eigenvalues between -1 and 1.

Step 3: Model selection and manual auditing

We try four alternative machine learning model for horseracing: Logistic Regression, (Multinomial) Naïve Bayes, Linear Support Vector Machine, and Random Forest Classifier. We then using testing data to evaluate the predicting accuracy of each model and list the results as follows:

Model name	Predicting accuracy
Logistic Regression	0.697415
Multinomial NB	0.536442
Linear SVC	**0.739483**
Random Forest Classifier	0.323163

Linear SVC exhibits the highest predicting accuracy, which we choose as the final model. In addition, notice that Random Forest Classifier merely makes 32.32% correct prediction. This is consistent to our expectation because random forest is a committee-based learning (or ensemble learning) approach, consisting of multiple sub-classifiers, and therefore not suitable to cope with high-dimension data with too many eigenvalues, such as textual data, by nature.

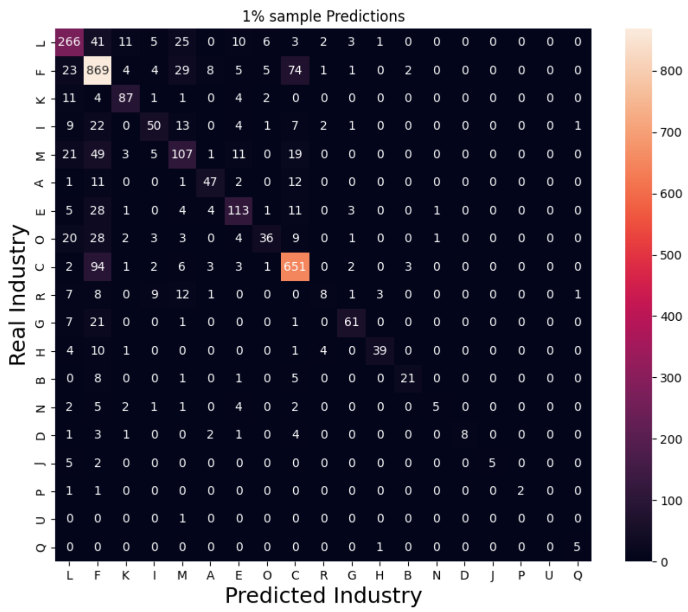
Step 4: Model evaluation and optimization with auditing

Typically, researchers use F1 scores, instead of accuracy, to evaluate the precision of polychotomous models, because accuracy, a de facto averaged precision, cannot reveal the precision in each category. Thus, we further check F1 scores of each industry for model evaluation and list the results (with 1% sample) as follows:

	Precision	Recall	F1 score	support
L	0.69	0.71	0.70	373
F	0.72	0.85	0.78	1025
K	0.77	0.79	0.78	110
I	0.62	0.45	0.53	110
M	0.52	0.5	0.51	216
A	0.71	0.64	0.67	74

E	0.7	0.66	0.68	171
O	0.69	0.34	0.45	107
C	0.81	0.85	0.83	768
R	0.47	0.16	0.24	50
G	0.84	0.67	0.74	91
H	0.89	0.66	0.76	59
B	0.81	0.58	0.68	36
N	0.71	0.23	0.34	22
D	1	0.4	0.57	20
J	1	0.42	0.59	12
P	1	0.5	0.67	4
U	NA	NA	NA	NA
Q	0.71	0.83	0.77	6
Weighted avg.	0.73	0.73	0.72	3255

Here is a confusion matrix depicting the relation between the LSVC-predicted industries and the real ones (with 1% sample).



We can find that the precision (F1 score) is largely good for most of the industries except O, N, and R (F1 scores below 0.5). Also, "F" is the industry that firms are more likely to be incorrectly assigned. Thus, we further audit these cases of misclassification, and set some extra artificial standards to the model according to government's official description of each industry.

For example, firms majoring in selling construction materials are easily to be misclassified into Wholesale & retail industry since the high frequency of the word "sell", but should have been labeled as manufacturing firms because they are actually suppliers of the construction materials produced by their own. With prior knowledge added, we expect to enhance the precision for these categories.

We are now in exploration of using more complicated textual-analysis-based machine learning approach in industry identifying, such as deep learning with BERT model. And we hope that this trial can help exploit various firm-registration data, or more broadly, any data of private firms without an officially assigned industry classification.

Internet Appendix C: The local effect of HPR: additional evidence

In this paper, we exploit the geographic spillovers of China's HPR policy, instead of the HPR policy itself, as our main identification strategy, because directly comparing individuals in HPR cities to those in non-HPR cities can be subject to various endogeneity issues (see Section 3.1). Following Deng et al. (2021), our identification strategy requires the exclusion of HPR cities. However, one might concern that our baseline results are driven by this execution rather than the effects of housing speculation.

To address this concern, we provide the results that document the effects of HPR policy itself on local entrepreneurship as a piece of auxiliary evidence. Specifically, we use DiD estimation to compare individuals in HPR cities and those in non-HPR cities with the full sample (i.e., the baseline sample without excluding individuals in HPR cities). The empirical model is as follows:

$$Entrepreneurship_{i,j,t} = \alpha + \beta \cdot HPR_{j,t\text{-}1} + \gamma' \cdot X_{i,j,t} + Individual_i + Year_t + \epsilon_{i,j,t}, \qquad \text{(A1)}$$

where *HPR* is a binary variable that equals one if the city is an HPR city and the restriction policy is effective in year *t-1*. Other components of Eq. (A1) are the same as those in Eq. (1). Since HPR policy is a negative shock to housing speculation, we expect HPR policy to have a significantly positive effect on local entrepreneurship, based on the logic that the HPR policy crowds the speculators out, which promotes entrepreneurship.

Table IA1 presents the summary statistics of the initial sample of 45,771 observations that include individuals in the HPR cities. The distribution of all the variables (except for house values and incomes which are significantly higher for individuals in HPR cities) is similar to that of the final sample reported in Table 1. Thus, the exclusion of observations in HPR cities for the sake of cleaner identification does not materially undermine the representativeness of the sample or the reliability of our findings.

Table IA2 tabulates the results. With almost the same empirical model as Eq. (1) but altering the key explanatory variable as the HPR policy, the estimator of *HPR* in Column (1) of Table IA2 is positive and significant at the 1% level, consistent to our expectation. Columns (2)-(5) replace *HPR* with four alternative definitions of the policy. The four alternative definitions are as follow: (a) considering the initial round of implementation as the shock only; (b) excluding policies with restrictions on bank loans only, i.e., considering direct HPR on purchase only; (c) only regarding policies with purchase restriction on a family's second house (and above) as the shocks; (d) only regarding policies with purchase restriction on a family's third house (and above) as the shocks. Similar to that in Column (1), the coefficient estimates are positive and significant at the 1% or 5% level, suggesting that individuals in HPR cities are more likely to start new businesses than those in non-HPR cities because of the restrictions on housing speculation.

As an auxiliary test, while the HPR policy is subject to various endogeneity concerns, for which we choose not to use it as our main identification strategy, the positive coefficient (an opposite sign to the baseline results) estimates of the key explanatory variable could help enhance the credence of our baseline results in the sense that our baseline findings are unlikely driven by coincidence or measurement errors.

Table IA1 Summary statistics of the full sample (HPR cities included)

The sample contains all the urban-citizen respondents, who are surveyed by CFPS and between 18 and 65, with 45,771 individual-year observations of each variable. *Entrepreneurship* equals one if the respondent is running their own business, and zero otherwise. *Homeowner* equals one if the respondent owns at least one house, and zero otherwise. *Multi-property owner* equals one if the respondent own more than one house, and zero otherwise. *Housing collateral value* is the total market value of the respondent's house(s) net of the total value of mortgage loan(s) on the house(s). *Family size* is the number of a respondent's family member(s). *Salary* is the average yearly income of a respondent's family. *Ethnicity* equals one if the respondent belongs to Han people, the majority people of China, and zero otherwise. *Female* equals one for female, and zero for male. *Age* is the respondent's age. *Marriage* equals one if the respondent is in a marriage, and zero otherwise. *Education* and *Health* are defined as a bundle of dummy variables base on the respondent's answers (checked by the interviewer) in the survey.

	Mean	SD	Min	p(10)	p(50)	p(90)	Max
Entrepreneurship	0.11	0.31	0.00	0.00	0.00	1.00	1.00
Homeowner	0.88	0.32	0.00	0.00	1.00	1.00	1.00
Multi-property owner	0.19	0.39	0.00	0.00	0.00	1.00	1.00
Housing collateral value	437,237.7	478,182.2	0.00	10,000	25,000	1,400,000	1,600,000
Ln(Housing collateral value)	11.37	3.78	0.00	9.21	12.43	14.15	14.29
Family size	4.01	1.55	2.00	2.00	4.00	6.00	8.00
Salary	15,617.1	11,695.2	726	2,857	12,500	36,800	40,000
Ln(Salary)	9.28	0.99	6.59	7.96	9.43	10.51	10.60
Female	0.48	0.50	0.00	0.00	0.00	1.00	1.00
Age	43.56	12.63	18.00	26.00	44.00	61.00	65.00
Marriage	0.85	0.36	0.00	0.00	1.00	1.00	1.00
Ethnicity	0.97	0.18	0.00	1.00	1.00	1.00	1.00
Education							
Non-literate	0.14	0.35	0.00	0.00	0.00	1.00	1.00
Primary school	0.17	0.37	0.00	0.00	0.00	1.00	1.00
Middle school	0.32	0.47	0.00	0.00	0.00	1.00	1.00
High school	0.21	0.41	0.00	0.00	0.00	1.00	1.00
College or above	0.16	0.37	0.00	0.00	0.00	1.00	1.00
Health							
Very bad	0.21	0.40	0.00	0.00	0.00	1.00	1.00

Bad	0.23	0.42	0.00	0.00	0.00	1.00	1.00
Neutral	0.32	0.47	0.00	0.00	0.00	1.00	1.00
Good	0.14	0.35	0.00	0.00	0.00	1.00	1.00
Very good	0.10	0.30	0.00	0.00	0.00	0.00	1.00

Table IA2 Housing purchase restrictions and local entrepreneurial activities

The sample contains CFPS-surveyed individuals in HPR cities and non-HPR cities during 2010-2018. *Entrepreneurship* equals one if the respondent is running their own business, and zero otherwise. *HPR* equals one if an individual lives in a city within 200km to the closest regulated city that is presently under the house purchase restrictions in year *t*, and zero otherwise. Robust standard errors in parentheses are clustered by individual. The symbols ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

	Dependent variable: Entrepreneurship dummy				
Variables	(1)	(2)	(3)	(4)	(5)
HPR	0.017***				
	(0.01)				
HPR (First round only)		0.014**			
		(0.01)			
HPR (Loan restriction excluded)			0.017***		
			(0.01)		
HPR (Second house and above)				0.017***	
				(0.01)	
HPR (Third house and above)					0.011**
					(0.01)
Ln(Housing collateral value)	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Family size	0.003	0.003	0.003	0.003	0.003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Ln(Salary)	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
City controls	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes
Individual FEs	Yes	Yes	Yes	Yes	Yes
No. of observations	42,311	42,311	42,311	42,311	42,311
R-squared	0.541	0.541	0.541	0.541	0.541