

Internet Appendix to “Housing Speculation and Entrepreneurship”

(Not to be published)

Internet Appendix A: Details for the conceptual framework model

To clarify our paper’s logic, we present a suggestive (“toy”) model formalizing our arguments. We follow Wang, Wang, and Yang (2012) to construct a binary career choice model and further incorporate housing wealth and housing market return into the model. Here are some key procedures to solve the model, which is omitted in the main body of the paper.

Based on the setting of the model in the main context, the agent compares the optimal CE wealth of being an entrepreneur against taking the outside option, and chooses to operate own business if and only if the former exceeds the latter, which constructs the participant incentive of entrepreneurship:

$$P(K_0^*, W_0 - K_0^* - \varphi) \geq W_0 + \Pi + R_H^*(g_H) \quad (\text{A.1})$$

Applying *Euler’s theorem* for $P(K, W)$, we have

$$\begin{aligned} P(K_0^*, W_0 - \varphi - K_0^*) &= P_K^* \times K_0^* + P_W^* \times (W_0 - \varphi - K_0^*) \\ &\stackrel{(\text{A.1})}{\implies} p'(w^*)(W_0 - \varphi). \end{aligned} \quad (\text{A.2})$$

Hence, substituting (A.4) into (A.3), the threshold wealth level satisfies

$$P(K_0^*, \widetilde{W}_0 - K_0^* - \varphi) = p'(w^*)(\widetilde{W}_0 - \varphi) = \widetilde{W}_0 + \Pi + R_H^*(g_H), \quad (\text{A.3})$$

Hence, let $p(w) = P(K, W)/K$, the participate incentive gives the endogenous threshold wealth level

$$\widetilde{W}_0 = \frac{p'(w^*)\varphi + \Pi + R_H^*(g_H)}{p'(w^*) - 1}, \quad (\text{A.4})$$

Following Wang, Wang, and Yang (2012), the entrepreneur, conditional on entry, chooses their initial size of business ownership K_0^* to maximize utility given by (A.1). In equilibrium, the maximization implies the equality between marginal value of capital investment and marginal value of remaining wealth, which is

$$P_K(K_0^*, W_0 - \varphi - K_0^*) = P_W(K_0^*, W_0 - \varphi - K_0^*). \quad (\text{A.5})$$

Let $p(w) = P(K, W)/K$, we have

$$\begin{cases} P_w(K, W) = p'(w) \\ P_K(K, W) = p(w) - wp'(w) \end{cases} \quad (\text{A.6})$$

Therefore, using (A.8) to simplify (A.7), we can get w^* which is given by

$$p'(w^*) = \frac{p(w^*)}{1 + w^*} \quad (\text{A.7})$$

that is independent of the fixed entry cost and the value of outside option, but merely determined by the marginal value of wealth $P_w(K, W) = p'(w^*)$. H^* is determined by the agent's level of risk aversion and the expected CE return via housing speculation. As stated above, the agent chooses to operate own business if and only if $W_0 \geq \widetilde{W}_0$. Correspondingly, the entrepreneurial firm's optimal initial size is given by

$$K_0^* = \frac{W_0 - \varphi}{1 + w^*}, \quad (\text{A.8})$$

which is independent of the outside option value.

Notably, in the main text we only present a now-or-never choice model (i.e., "European" style option). However, there is no qualitative difference if we allow for the agent's timing the transition between entering entrepreneurship and simply waiting. For example, following Wang, Wang, and Yang (2012), in a continuous-time model (i.e., "American" style option), the optimal threshold wealth level for entry is given as

$$\widetilde{W} = \lambda(p'(w^*)) \cdot [\varphi + \Pi + R_H^*(g_H)] + \varphi, \quad (\text{A.9})$$

where w^* is given by (A.9) and $\lambda(p'(w^*)) > 0$ depends on an agent's risk aversion (γ), EIS, subjective discount factor, and the Sharpe ratio of market portfolio.

Internet Appendix B: Institutional background on China's HPR policy

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 and administrative efficiency to carry out the HPR policy.¹

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 2014.² 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 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.

While Deng et al. (2021) focus on the second wave of HPR spillovers (2016–2017), we extend the analysis to both waves. 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

¹ 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.

² 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.

details in Section 5.1). 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 2018.³ Prices are log-transformed and normalized by city means to ensure comparability.

Figure IA2 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 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.

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.

Internet Appendix C: Identifying industry classifications with machine learning techniques

³ 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.

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 compulsory 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 based on supervised machine learning 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

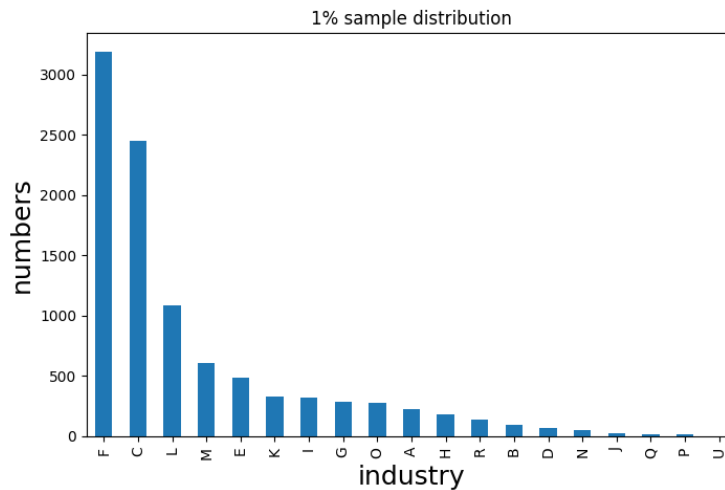
Our data contains more than 50 million records of firm registration obtained from State Administration for Industry and Commerce of the People’s Republic of China (SAIC). We begin with one-year registration data and manually identify an industry (out of the 19 major classifications) for each firm by 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
<u>Linear SVC</u>	<u>0.739483</u>
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

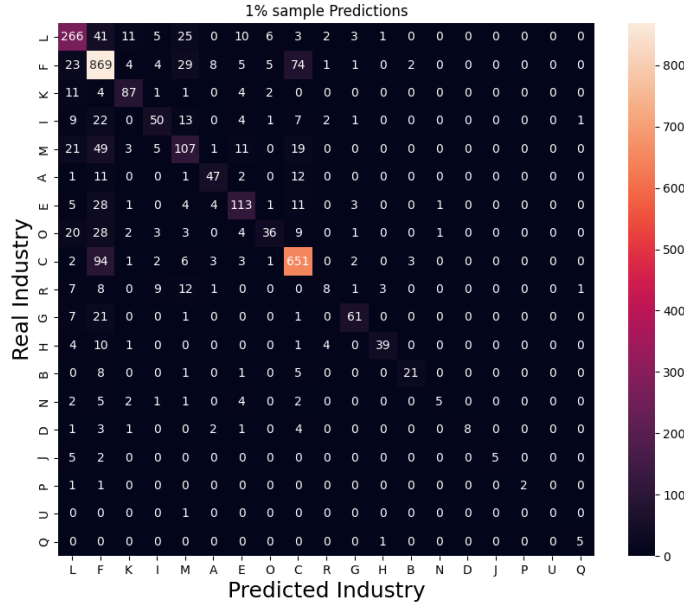
learning) approach, consisting of multiple sub-classifiers, and therefore not suitable to cope with high-dimension data with too many eigenvalues, such as textual data, by nature.

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 D: Robustness checks, placebo tests, and tests on reversal shocks

D.1 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.⁴

To address this concern, we follow the spirit of Bertrand and Mullainathan (2003) to decompose the key explanatory variable, *HPR spillover*, into two indicators: *Before* and *After*. *Before* (*After*) equals one for individuals in treated cities if the survey year is before (after) the HPR implementation (and the policy is not yet cancelled), and zero otherwise.⁵ If reverse causality drives our results, we should expect to observe significant coefficient estimate on the *Before* dummy above. In Column (1) of Table IA3, the coefficient estimate on *Before* is statistically insignificant while the coefficient estimate on *After* is positive and significant at the 1% level, which is consistent to our conjecture.

We acknowledge that there still could be pre-existing trends (i.e., changes in entrepreneurship before the arrival of HPR spillovers) that might drive our results. Hence, to further rule out the possible effect of pre-existing trends, if any, we make another attempt to include the effect of trends into the baseline model. Specifically, we construct a *Trend* variable that denotes the number of waves relative to the year of HPR implementation. In Column (2), the coefficient estimate on *HPR spillover* remains negative and significant as those in the baseline results even after controlling for the effect of pre-trends, suggesting that our baseline findings are not driven by the effect of pre-existing trends; moreover, the

⁴ Likewise, the HPR policy is mandated by the central government since 2010 (although delegated to local governments afterwards), and thus speculators could reasonably anticipate that the policy would be implemented in the near future. Consequently, they would likely find a surrounding city with a lower level of entrepreneurial activities (and therefore lower house prices) to forestall the real estate market with the expectation of a higher speculative return in the city’s housing market, which could also produce a negative relation between house prices and entrepreneurship even before the arrival of HPR spillovers. If these arguments are supported, our results could merely be a manifestation of reverse causality instead of the causal effect of housing speculation on entrepreneurship.

⁵ The omitted group (i.e., benchmark year) is the year of policy change. Note that the year of (or right after) a cancellation of HPR policy is not regarded as a *Before* year, because the year is more capturing the residual effect of the just-cancelled policy rather than a year before a city is treated.

coefficient estimate on *Trend* is largely insignificant and close to zero, standing against the possibility that pre-existing trends alter the baseline results in our sample.⁶

Together, 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.

D.2 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 IA4 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., $HPR \times \text{Nearness}$, as the explanatory variable, where *HPR* denotes whether the closest HPR city (for the focal non-HPR city) is under a HPR policy.^{7,8} 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 IA4 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

⁶ In unreported results, the coefficient estimates on *HPR spillover* remain negative and significant (at the 1% level) even after controlling for the effects of higher power of trends.

⁷ *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.

⁸ Note that *HPR spillover* is actually $HPR \times 1[\text{whether the focal non-HPR city is within 200km to the nearest HPR city}]$.

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 IA4 Panel B suggests that our results remain robust to all these alternative sample selection filters.

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;⁹ (b) city-level controls, such as consumption, bank loans, citizen deposit, income Gini index, and the number of primary schools, high schools, and colleges.¹⁰ Columns (1) - (2) of Table IA4 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, Han, and Zhou (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 IA4 Panel C suggest that our baseline results are robust to city-level standard error clustering and multi-way standard error clustering.¹¹

Fifth, in Column (5) of Table IA4 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

⁹ 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. In addition, the coefficient estimate on *Employed* is positive and significant at the 1% level, suggesting that the effect we capture is unlikely forced entrepreneurship (Hacamo and Kleiner 2022) resulted from unemployment.

¹⁰ The income Gini index is estimated based on the CFPS data, while other city-level variables are reported in the *Year Book of Statistics*.

¹¹ In addition, we present the results for the tests on the speculation channel with standard error clustered by city in Table IA4 Panel E. The significance does not diminish (except for the test with 2014 CFPS sample only due to the small sample size). In fact, most of our main results remain largely intact with the alternative standard error clustering.

survey to alleviate the concern that our results are driven by sample truncation. Column (1) of Table IA4 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 IA4 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 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, Sraer, and Thesmar (2017)'s specification, i.e., we only consider entrepreneurial activities of household heads.¹² As shown in Column (4) of Table IA4 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 E and present the results in Table IA5.

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.

¹² Schmalz, Sraer, and Thesmar (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.

We tabulate the result in Column (5) of Table IA4 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.

D.3 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, Kerr, and Nanda (2022) and Deng et al. (2021), we focus on renters, i.e., individuals who have no house, and re-run baseline regressions in Column (4) of Table 3.

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 IA6 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, although the insignificant coefficient estimates on renters could be due to the small number of renters in our sample, at least this piece of evidence supports that 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. (9). 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 IA6 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 IA6 Panel B), suggesting that our main results are unlikely driven by unobservable shocks occurring at approximately the same time as the HPR policy.

D.4 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 IA4 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.

Internet Appendix E: 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:

$$\begin{aligned} Entrepreneurship_{i,j,t} = & \alpha + \beta \cdot HPR_{j,t-1} + \gamma' \cdot X_{i,j,t} \\ & + Individual_i + Year_t + \epsilon_{i,j,t}, \end{aligned} \quad (A1)$$

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. (9). 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 IA2 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 IA5 tabulates the results. We begin with a naïve regression that compares entrepreneurship before and after HPR adoption within HPR cities only. The coefficient estimate on *HPR* is positive (though not significant, possibly due to various endogeneity issues), which is arguably consistent to our conjecture that, as a negative “shock” to house market appreciation, HPR is likely to result in positive changes in entrepreneurship (or at least is not likely to also capture a negative change as we have from the baseline regressions).

Then, we take a step further to formally investigate the effect of HPR using a DiD-style empirical setting. With almost the same empirical model as Eq. (9) but altering the key explanatory variable as the HPR policy, the estimator of *HPR* in Column (1) of Table IA5 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.

References

- Angrist, J. D., Pischke, J.-S., 2008. Mostly harmless econometrics. Princeton university press.
- 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), 370-395.
- Bekaert, G., Harvey, C. R., Lundblad, C., 2005. Does financial liberalization spur growth? *Journal of Financial Economics* 77, 3–55.
- Fang, H., Gu, Q., Xiong, W., Zhou, L.-A., 2016. Demystifying the Chinese housing boom. NBER macroeconomics annual 30, 105–166.
- Glaeser, E., Huang, W., Ma, Y., Shleifer, A., 2017. A real estate boom with Chinese Characteristics. *Journal of Economic Perspectives* 31, 93–116.
- Goodman-Bacon, A., 2021. Difference-in-differences with variation in treatment timing. *Journal of Econometrics* 225, 254–277.
- Hacamo, I., Kleiner, K., 2022. Forced entrepreneurs. *The Journal of Finance*, 77(1), 49-83.
- Petersen, M. A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22, 435–480.
- Wu, J., Gyourko, J., Deng, Y., 2015. Real estate collateral value and investment: The case of China. *Journal of Urban Economics* 86, 43–53.

Table IA1: 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

Table IA2 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
<i>Entrepreneurship</i>	0.11	0.31	0.00	0.00	0.00	1.00	1.00
<i>Homeowner</i>	0.88	0.32	0.00	0.00	1.00	1.00	1.00
<i>Multi-property owner</i>	0.19	0.39	0.00	0.00	0.00	1.00	1.00
<i>Housing collateral value</i>	437,237.7	478,182.2	0.00	10,000	25,000	1,400,000	1,600,000
<i>Ln(Housing collateral value)</i>	11.37	3.78	0.00	9.21	12.43	14.15	14.29
<i>Family size</i>	4.01	1.55	2.00	2.00	4.00	6.00	8.00
<i>Salary</i>	15,617.1	11,695.2	726	2,857	12,500	36,800	40,000
<i>Ln(Salary)</i>	9.28	0.99	6.59	7.96	9.43	10.51	10.60
<i>Female</i>	0.48	0.50	0.00	0.00	0.00	1.00	1.00
<i>Age</i>	43.56	12.63	18.00	26.00	44.00	61.00	65.00
<i>Marriage</i>	0.85	0.36	0.00	0.00	1.00	1.00	1.00
<i>Ethnicity</i>	0.97	0.18	0.00	1.00	1.00	1.00	1.00
<u><i>Education</i></u>							
<i>Non-literate</i>	0.14	0.35	0.00	0.00	0.00	1.00	1.00
<i>Primary school</i>	0.17	0.37	0.00	0.00	0.00	1.00	1.00
<i>Middle school</i>	0.32	0.47	0.00	0.00	0.00	1.00	1.00
<i>High school</i>	0.21	0.41	0.00	0.00	0.00	1.00	1.00
<i>College or above</i>	0.16	0.37	0.00	0.00	0.00	1.00	1.00
<u><i>Health</i></u>							
<i>Very bad</i>	0.21	0.40	0.00	0.00	0.00	1.00	1.00
<i>Bad</i>	0.23	0.42	0.00	0.00	0.00	1.00	1.00
<i>Neutral</i>	0.32	0.47	0.00	0.00	0.00	1.00	1.00
<i>Good</i>	0.14	0.35	0.00	0.00	0.00	1.00	1.00
<i>Very good</i>	0.10	0.30	0.00	0.00	0.00	0.00	1.00

Table IA3 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. *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. *Trend* is the number of wave(s) relative to the policy implementation year. 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.

Variables	Dependent variable: <i>Entrepreneurship</i> dummy		
	(1)	(2)	(3)
<i>Before</i>	0.006 (0.02)		
<i>After</i>	-0.021*** (0.01)		
<i>Before 2</i>		-0.001 (0.02)	
<i>Before 1</i>		-0.006 (0.01)	
<i>Year 0</i>		-0.023*** (0.01)	
<i>After 1</i>		-0.022** (0.01)	
<i>After 2+</i>		-0.028 (0.02)	
<i>HPR spillover</i>			-0.034*** (0.01)
<i>Trend</i>			0.001 0.005
All controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
No. of observations	23,643	23,643	23,643
R-squared	0.535	0.535	0.535

Table IA4 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					
Variable	Dependent variable: <i>Entrepreneurship</i> dummy				
	(1) 150km	(2) 175km	(3) 250km	(4) Same province	(5) Continuous measure
<i>HPR spillover</i>	-0.046*** (0.01)	-0.038*** (0.01)	-0.021*** (0.01)	-0.037*** (0.01)	
HPR \times 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					
Variable	(1) Non-citizen included	(2) Age > 65 included	(3) Excluding individuals not in the first wave	(4) No filter	(5) Extra filters
<i>HPR spillover</i>	-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					
	(1)	(2)	(3)	(4)	(5)
Variable	Including employment status	Including extra city-level controls	Clustering by city	Two-way clustering	Probit regression
<i>HPR spillover</i>	-0.030*** (0.01)	-0.022*** (0.01)	-0.032* (0.02)	-0.032** (0.01)	-0.188*** (0.02)
<i>Employed</i>	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					
	(1)	(2)	(3)	(4)	(5)
Variable	Excluding the last wave	Stacked regression	Goodman-Bacon (2021)'s decomposition	Household-level estimation	City-level estimation with firm registration data
<i>HPR spillover</i>	-0.037*** (0.01)	-0.042*** (0.00)		-0.030*** (0.01)	-0.155** (0.08)
<i>Treated vs. Never T.</i>			-0.022 [weight=0.882]		
<i>Earlier T. vs. Later C.</i>			-0.142 [weight=0.081]		
<i>Later T. vs Earlier C.</i>			-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

Panel E: Heterogeneous tests for speculation channel with city-level clustering

Variables	Dependent variable: <i>Entrepreneurship</i> dummy				
	(1)	(2)	(3)	(4)	(5)
<i>HPR spillover</i> × <i>Multi-property owner</i>	-0.054** (0.03)				
<i>Multi-property owner</i>	0.132 (0.56)				
<i>HPR spillover</i> × <i>No. of extra houses</i>		-0.037** (0.02)			
<i>No. of extra houses</i>		-0.298 (0.44)			
<i>HPR spillover</i> × <i>Has property income</i>			-0.030* (0.02)		
<i>Has property income</i>			-0.216 (0.46)		
<i>HPR spillover</i> × <i>Single-property owner</i>	-0.037 (0.02)	-0.030 (0.02)	-0.006 (0.02)		
<i>Single-property owner</i>	-0.073 (0.40)	-0.360 (0.39)	-0.194 (0.30)		
<i>HPR spillover</i> × <i>High expectation</i>				-0.199 (0.13)	
<i>HPR spillover</i> × <i>Trading activeness</i>					-0.034* (0.02)
<i>HPR spillover</i>	0.007 (0.03)	-0.000 (0.02)	-0.021 (0.02)	0.011 (0.03)	-0.021 (0.02)
Controls & interactions	Yes	Yes	Yes	Yes	Yes
Year & Individual FEs	Yes	Yes	Yes	Yes	Yes
No. of observations	23,643	23,641	23,527	3,522	23,052
R-squared	0.537	0.537	0.532	0.781	0.534

Table IA5 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.

Panel A: Naïve regression – entrepreneurship before and after HPR implementation in HPR cities					
	Dependent variable: <i>Entrepreneurship</i> dummy				
	Sample: Individuals in HPR cities only				
Variables	(1)				
<i>HPR</i>	0.004 (0.007)				
Year FEs	Yes				
Individual FEs	Yes				
No. of observations	19,994				
R-squared	0.542				
Panel B: DiD-style regression					
	Dependent variable: <i>Entrepreneurship</i> dummy				
	Sample: Individuals in HPR cities or non-HPR cities				
Variables	(1)	(2)	(3)	(4)	(5)
<i>HPR</i>	0.017*** (0.01)				
<i>HPR (First round only)</i>		0.014** (0.01)			
<i>HPR (Loan restriction excluded)</i>			0.017*** (0.01)		
<i>HPR (Second house and above)</i>				0.017*** (0.01)	
<i>HPR (Third house and above)</i>					0.011** (0.01)
<i>Ln(Housing collateral value)</i>	0.002*** (0.00)	0.002*** (0.00)	0.002*** (0.00)	0.002*** (0.00)	0.002*** (0.00)
<i>Family size</i>	0.003 (0.00)	0.003 (0.00)	0.003 (0.00)	0.003 (0.00)	0.003 (0.00)
<i>Ln(Salary)</i>	-0.019*** (0.00)	-0.019*** (0.00)	-0.019*** (0.00)	-0.019*** (0.00)	-0.019*** (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

Table IA6 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: <i>Entrepreneurship</i> dummy					
Variable	(1)		(2)		(3)	
<i>HPR spillover</i>	0.002		-0.0004		0.018	
	(0.02)		(0.03)		(0.03)	
<i>Family size</i>					0.011	
					(0.02)	
<i>Ln(Salary)</i>					-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 <i>HPR spillover</i>	0.000285	-0.00763	-0.00210	0.00062	0.00310	0.00717
<i>t</i> -stats	[0.05493]	[-1.549]	[-0.411]	[0.124]	[0.607]	[1.434]

Table IA7 Correlations between and the role of wealth variables

The sample contains all the CFPS-surveyed renters (i.e., individuals have no house) in non-HPR cities during 2010-2018. *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 A, the symbol * denotes significance at or above the 5% level for the Pearson correlation matrix. In Panel B, 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: Pearson correlation matrix of wealth variables					
	<i>HPR spillover</i>	<i>Ln(Housing collateral value)</i>	<i>Multi-property owner</i>	<i>Ln(Asset)</i>	
<i>HPR spillover</i>	1				
<i>Ln(Housing collateral value)</i>	0.0722*	1			
<i>Multi-property owner</i>	0.0037	0.2419*	1		
<i>Ln(Asset)</i>	0.1121*	0.6884*	0.3236*	1	

Panel B: Tests on the effect of each wealth variable						
Variable	Dependent variable: <i>Entrepreneurship</i> dummy					
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Ln(Housing collateral value)</i>	0.002*** (0.00)	0.002*** (0.00)				
<i>Multi-property owner</i>			0.009* (0.01)	0.007 (0.01)		
<i>Ln(Asset)</i>					0.014*** (0.00)	0.018*** (0.00)
<i>HPR spillover</i>		-0.032*** (0.01)		-0.032*** (0.01)		-0.032*** (0.01)
Controls & FEs	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	23,643	23,643	23,643	23,643	23,075	23,075
R-squared	0.541	0.535	0.541	0.535	0.541	0.536

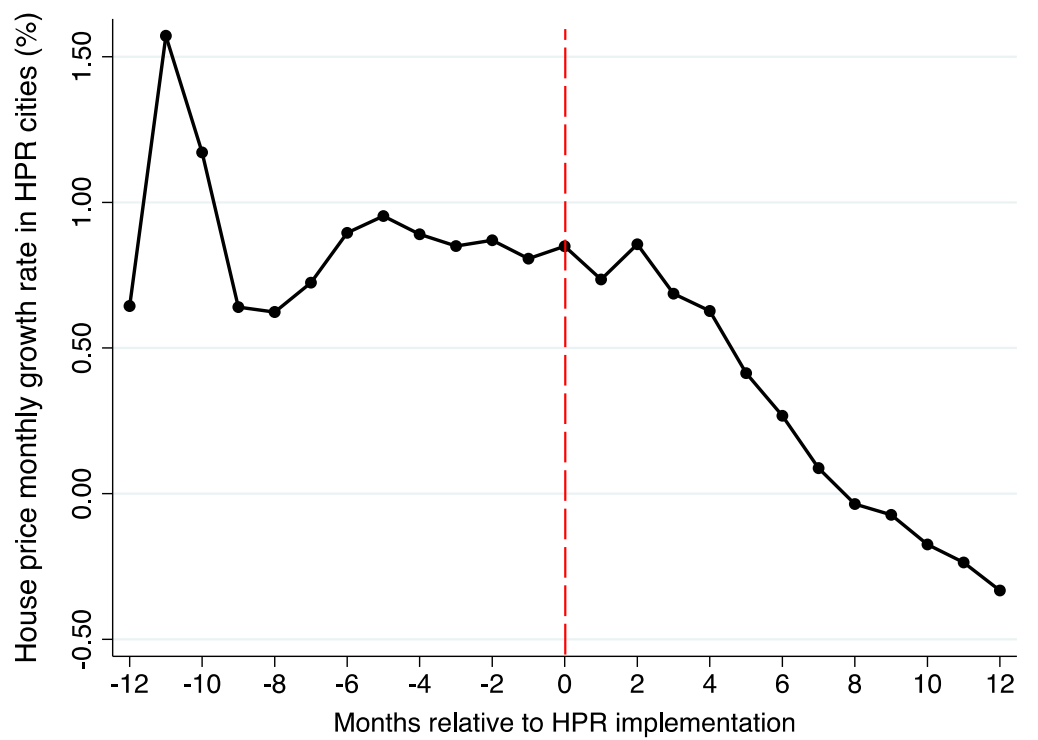


Figure IA1. Event study of the effect of HPR on house price growth in HPR cities. The sample contains monthly house price growth of 68 HPR cities (out of 100 cities in China) from June 2011 to December 2018. The numbers on X-axis denote the number of months relative to the implementation of HPR policy in a specific city.

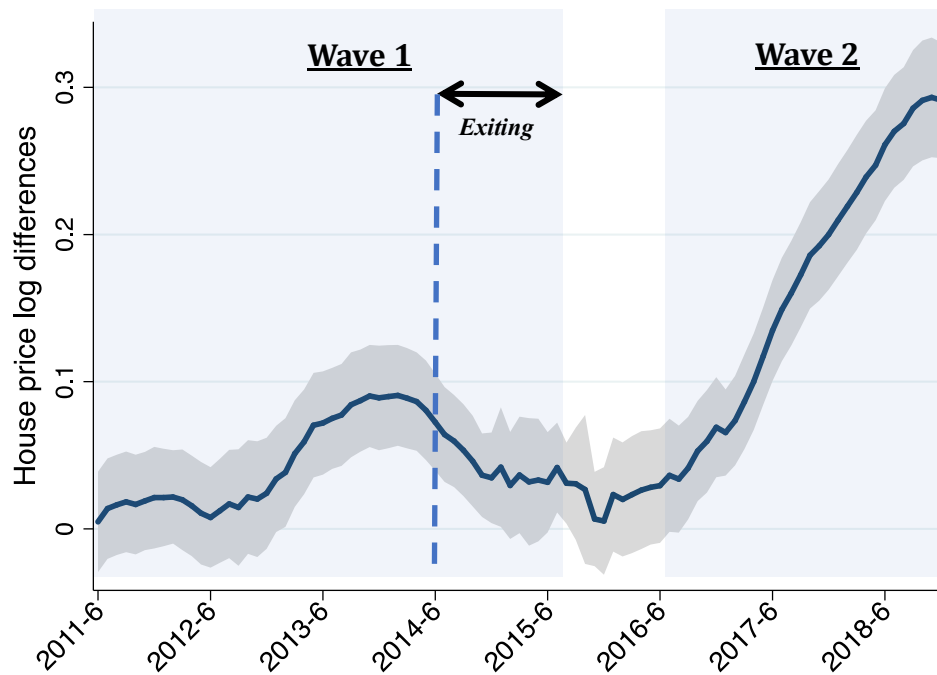


Figure IA2. The geographical spillover effects of the two-round HPR policies on non-HPR cities' house prices. The sample contains monthly average house prices of 42 non-HPR cities (out of 100 cities in China) from June 2011 to December 2018. A city's house prices are scaled by its city-level mean to 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.

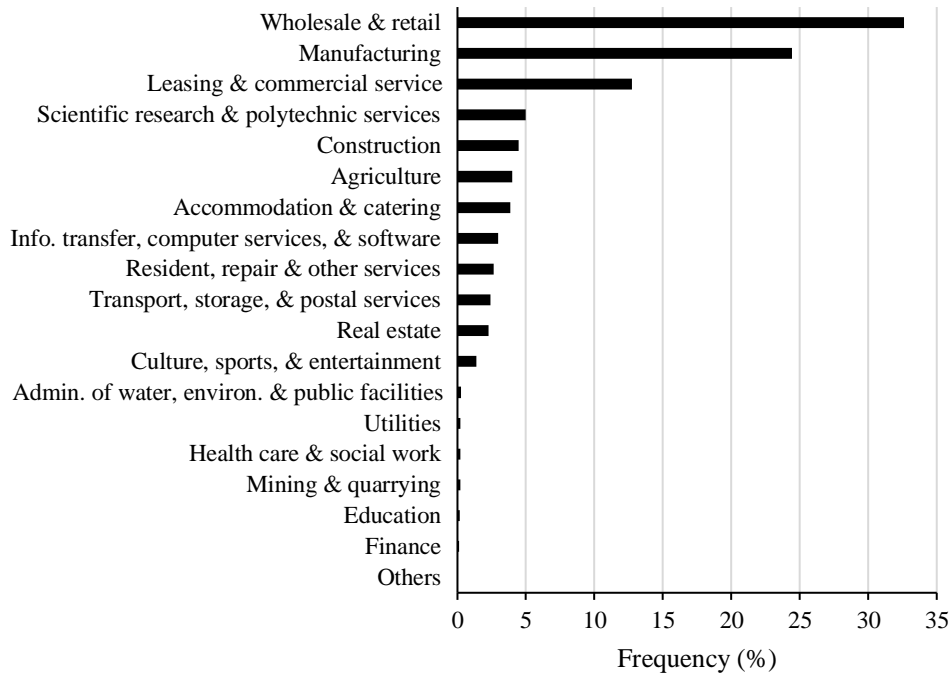


Figure IA3. Industry distributions of newly created businesses based on industry classifications defined by a machine learning technique. The graph shows the industry distribution of the businesses created during 2004-2020. The industries are classified by our machine learning approach.

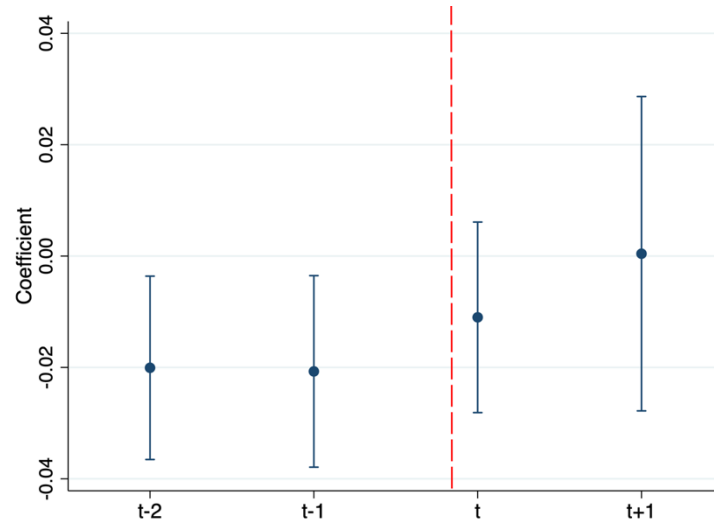


Figure IA4. The effect of HPR policy cancellations in an event-study framework. The figure shows the coefficient estimates and their 90% confidence intervals of a set of dummies that equal one if an individual is in a nearer non-HPR (treated) city in a specific year around HPR cancellations, and zero 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.

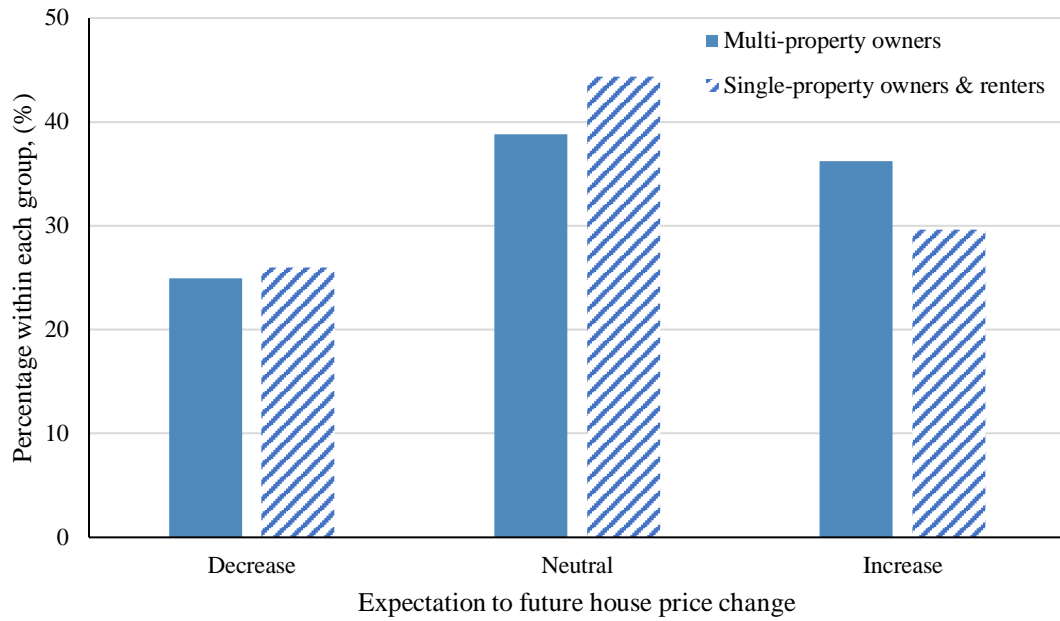


Figure IA5. Distribution of the expectation on future house price change (Multi-property owners vs. Others). The sample contains the cross-section of all the CFPS-surveyed individuals, in treated cities and in control cities, responding to the extra survey of the 2014 wave, in which an individual is 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.

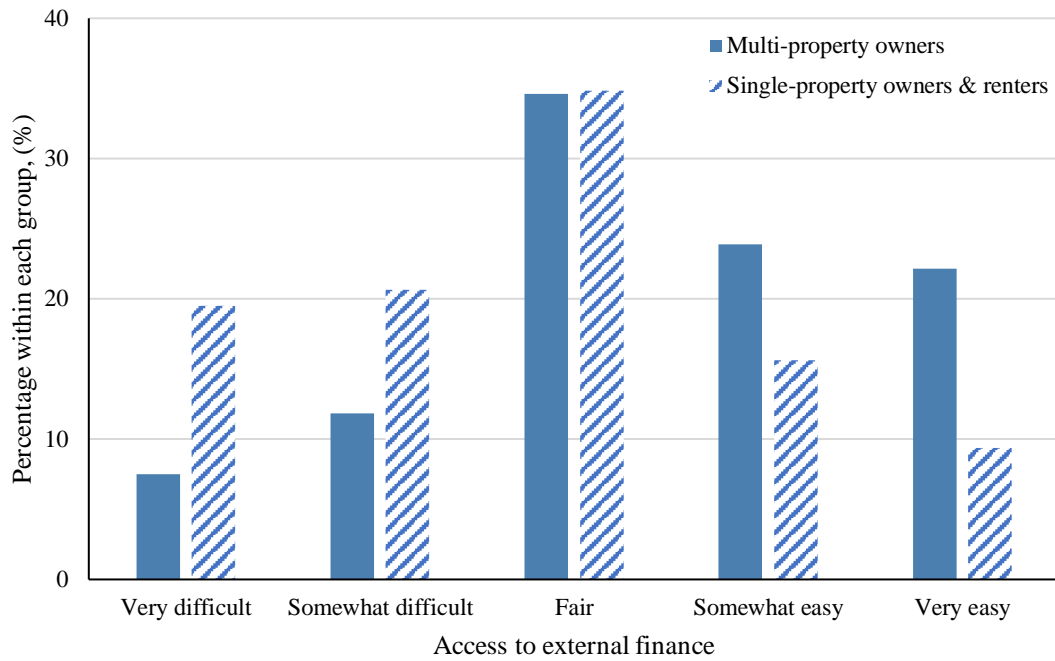


Figure IA6. Distribution of the access to external finance. The sample contains the cross-section of all the CFPS-surveyed individuals responding to the extra survey of the 2014 wave, in which an individual is asked “If you have to borrow a total amount of 20,000 yuan in case of some emergency. How difficult will it be to raise the money?” *Multi-property owners* are those who own multiple houses, while *Others* are single-property owners and renters. 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 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.

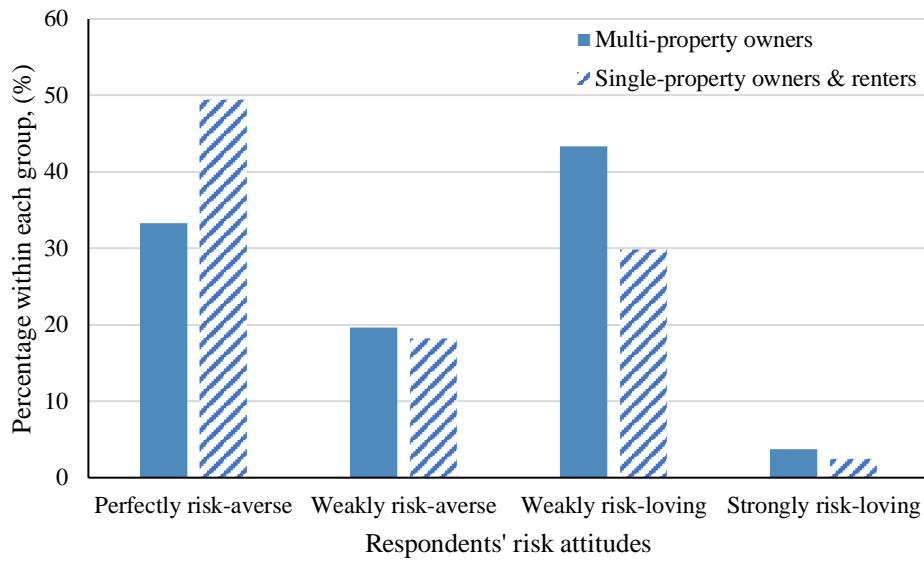


Figure IA7. Distribution of the risk attitudes. The sample contains the cross-section of all the CFPS surveyed individuals responding to the extra survey of the 2014 wave, in which an individual is asked “If your family invest/In investment, what kind of risk are you willing to take?” *Perfectly risk-averse*, *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. 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 intergroup difference is statistically significant (t-stat = 11.98).