

Supplementary Information: Online Appendix

The Effect of Urban Density on Stock Returns: Evidence from Satellite Imagery and Firms in the United States

Abstract. We study how the urban density of the areas where firms are located impacts their stock returns. Using high-resolution satellite images and image segmentation algorithms, we develop measures of building density for the most populated metropolitan statistical areas (MSAs) in the US. We find that firms in urban areas with a higher potential for density increase are perceived as less risky in the stock market because they can feasibly sustain their growth, leading to lower stock returns. This effect is stronger for smaller and younger firms.

Keywords. Urban density; density increase; stock returns.

JEL Codes. R30; R14; G10.

This Online Appendix includes the supplementary material for the paper.

Section A.1 provides a comprehensive explanation of the land classification approach. We apply an image segmentation algorithm on high-resolution satellite images sourced from Google Earth. Its purpose is to derive a categorization of urban land characteristics based on building density.

Section A.2 reports the summary statistics and correlation matrices of the variables at the MSA and firm levels.

Section A.3 presents supplementary information about the characteristics of the Metropolitan Statistical Areas (MSAs). This includes the list of MSAs that are part of the Compustat data and the subset of MSAs used for the empirical analysis. It also provides the values of urbanization measures obtained from the land classification process, as well as other important variables specific to each MSA.

Section A.4 describes the procedure to estimate the productivity of firms.

Section A.5 includes some additional figures that complement the main text.

Section A.6 reports additional results that complement the main analysis discussed in the text, including several robustness checks.

A.1 Land classification

Our research focuses on listed firms on the stock market. We aim to assess the role of the potential increase in building density of the city where the firm’s headquarters is located on its perceived risk, measured as stock excess returns. To measure urban density in terms of buildings, we rely on high-resolution satellite imagery of the major metropolitan statistical areas (MSAs) of the US from Google Earth. We focus on the MSAs with a population greater than 500,000 in the 2000 Census. The 86 MSAs considered for the analysis are listed in Tables 1 and A.7 and a map is reported in Figure A.6. The idea at the core of our approach is that areas with different levels of building development appear with different colors in satellite images, allowing the human eye to identify urban, semi-urban, and rural areas. Therefore, we develop a process to simplify the satellite images to categorize and measure the extension of different types of land in terms of building density within the territory of each MSA.

Polygon definition. First, we define a polygon around the main city of each MSA, starting from its city hall. Using Google Maps, we find the points that correspond to a maximum of approximately 60-minute drive in non-rush hour periods along the main highway routes. We pin those points on Google Earth and we define the borders of the polygon accordingly. If a city borders the seaside or a major lake, the waterbody around it is not considered as part of the polygon. We exclude from the Google Earth image the area outside the polygon (using the yellow color). This approach implies that we have a polygon with a different shape and surface for each MSA. Figure A.1 shows examples of polygons, highlighting their different shape and extension.

Manual exclusion of undevelopable land. Starting from the 60-minute drive polygon, we manually exclude from the image the *undevelopable* areas, where construction is totally restricted due to geographical constraints (water bodies such as rivers, lakes, reservoirs, water basins; lagoons and swamps; steeply sloped terrain) or due to protection initiatives (natural reserves, national/state parks, large city parks). We define polygons on the Google Earth image following the shape of the area to be excluded due to the impossibility of building (using the yellow color). We gather the area of each one of these polygons of undevelopable land from Google Earth, expressed in square kilometers. For each MSA, we sum up these areas to obtain the total undevelopable land within the 60-minute polygon. The input images reported in Figure A.1 are also examples of MSA with different levels of areas unavailable for construction within the 60-minute drive polygon.

Land categories according to building density. Next, we need to move from the complex satellite image to a simplified image where land is categorized according to the level of building density. In and around a city, it is possible to identify three types of land. First, *highly developed* (HD) areas are characterized by dense residential and commercial buildings, where the available observable space for new developments is negligible. Second, *developed* (D) areas correspond to semi-urban sectors, featuring residential or commercial buildings with some visible space for po-

tential new constructions. These areas of the city offer great potential for rapid growth in urban density due to existing infrastructure and the availability of land for further development.¹ Third, *low-developed* (LD) areas correspond to sparsely built or plain rural areas with high availability of unconstructed land, where development is costly due to the lack of urban infrastructure. Figure A.2 shows examples of each type of land, obtained by zooming in Google Earth images.

Image segmentation and area calculation. To categorize the types of land from the input satellite images, we adopt a process that mixes a computerized image segmentation algorithm with the manual classification by the authors. We use the Image Processing Toolbox in MATLAB to develop an algorithm that classifies the pixels of the input image into five categories, according to their similarity in color.² It is important to notice that satellite images of the US territory vary considerably depending on the region. For instance, Figure A.1 shows the polygon for Baton Rouge, in Louisiana, where the dominant color is dark green, and the polygon for Phoenix, in Arizona, where the dominant color is ocher. Therefore, our image segmentation approach is not based on a manually classified training set, which would be difficult to apply to images with a very different color palette. Instead, the algorithm takes the input image and clusters its pixels into the established number of categories, to maximize similarity in color within the category and maximize dissimilarity in color between the categories. This approach implies that the categories given as the outcome of the segmentation are “meaningless” regarding the type of land. In other words, the authors must compare the output image from the MATLAB process with the original image and manually classify each category as high-developed, developed, and low-developed land, based on the actual building density identified by zooming in the image on Google Earth. Figure A.3 provides an example of this process.

An additional feature of our MATLAB algorithm is that it includes a bar whose length can be modified by the user. After manually classifying the color categories, we adjust the bar to accurately match the scale reported in the bottom right corner of the Google Earth image. This map scale is the proportion between a distance on a map (the length of the scale rectangle) and a corresponding distance on the ground (reported in the middle of the scale rectangle). The length of the scale rectangle and the equivalent ground distance vary with each image. We superimpose the adjustable bar onto the scale rectangle and input the corresponding ground distance (in kilometers) into MATLAB to compute the total area for each color category, expressed in square kilometers. Since the area calculation by land category (high-developed, developed, and low-developed) depends

¹Recent studies focus on the density of buildings to define urban areas. For instance, [1] and [2] use unique databases that contain information on the universe of buildings on French and Spanish territory, respectively. The former defines as “urban” any area with statistically significant excess building density and distinguishes between “core” (highly urbanized) and “regular” urban pixels, whereas the latter defines urban areas as spaces with a minimum uninterrupted level of building density using a machine learning algorithm. Both works capitalize on exhaustive information on the location and characteristics of buildings; however, such rich datasets are a rarity. We propose an alternative method that is similar in spirit, as we also focus on building density and define different “types” of urban areas based on different levels of building concentration at the pixel level. However, our approach does not require specific databases and can be applied homogeneously across space and time.

²When we developed the image segmentation algorithm in MATLAB, we tried different numbers of categories. We decided to use five categories since it proved to be the minimum number that ensured the clear identification of high-developed, developed, and low-developed land across images taken from different regions of the United States. When we used lower numbers of categories, it was challenging to classify them into the types of land of interest, since areas with considerably different levels of building density were clustered into the same category. When we used larger numbers of categories, it was not easier to classify the types of land, as the distinction among types was approximately the same as the one with five categories, while the algorithm was distinguishing between slightly different land characteristics within the same type of interest.

on the automated image segmentation, we obtain slightly different numbers if we run the algorithm multiple times on the same image. This is due to the propensity of pixels that exhibit borderline coloration between categories to be variably classified within different groups throughout successive algorithm iterations.³ Even though the differences between iterations are slight, we execute the image segmentation algorithm five times for each input image, record the area corresponding to each type of land during each iteration, and subsequently calculate the average. To ensure that our main results are not affected by this margin of error, in Table A.24 we replicate our analysis excluding the MSAs with the highest variation in our PDI measure, and our conclusions remain consistent.

One drawback of our image segmentation approach is that the developed land around the urban center and the cultivated agricultural land often have similar colors in the satellite images (see Figure A.4)⁴. This means that it is impossible to precisely classify some of the color categories in the output image obtained from the MATLAB segmentation process. To overcome this inconvenience, in most cases, we implement a two-step approach. First, we use the original input image, apply the image segmentation algorithm, merge all automated color categories into one, and calculate the total area of the input polygon (excluding the undeveloped areas). Second, by examining and zooming in the image on Google Earth, we manually exclude from the input image the portions of land that are clearly rural. Then, we run the segmentation algorithm on this reduced input image and follow the usual procedure as described above. In this way, we obtain the total area of high-developed and developed land, in addition to the area of low-developed land only included in the reduced input image. We then subtract the surface of the high-developed and developed land from the total area calculated in the first step to obtain the total area of the low-developed land in the full initial input image. Figure A.5 shows an example of this two-step procedure.

³Each image segmentation outcome is independent of the outcomes of the previous iterations.

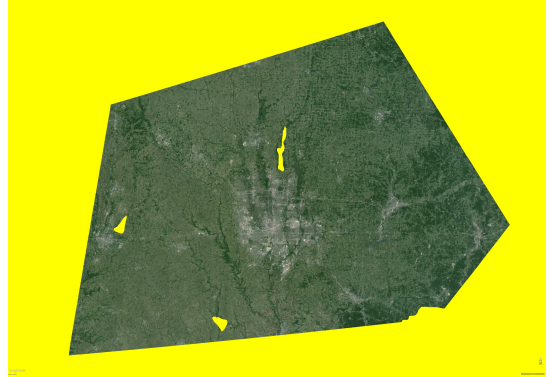
⁴This inconvenience is not solvable increasing the number of categories of the image segmentation algorithm, since in this case the color nuances are truly similar in the satellite image even though the land type is different.

Figure A.1: Examples of 60-minute drive polygon around the main city center of the MSA

(a) Chicago, IL



(b) Columbus, OH



(c) Baton Rouge, LA



(d) Phoenix, AZ



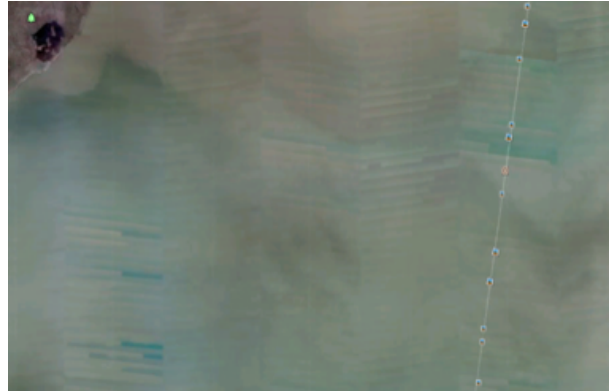
Note: This figure shows four examples of the polygon defined around the main city of the MSA. Figure A.1a displays the polygon for the Chicago-Naperville-Joliet, IL-IN-WI metropolitan area (whose main center is Chicago, IL); Figure A.1b displays the polygon for the Columbus, OH metropolitan area; Figure A.1c displays the polygon for the Baton Rouge, LA metropolitan area; and Figure A.1d displays the polygon for the Phoenix-Mesa-Scottsdale, AZ metropolitan area (whose main center is Phoenix, AZ). Each polygon is defined starting from its city hall, based on the Google Maps estimated driving distance of one hour by car in non-rush hour periods along the main highways. Therefore, each MSA has a polygon with a different shape and surface. The figures also include yellow polygons within the 60-minute drive areas that represent *undevelopable* areas (water bodies such as rivers, lakes, reservoirs, water basins; lagoons and swamps; steeply sloped terrain; protected areas such as natural reserves, national parks, large city parks). Figures A.1a and A.1b are examples of MSAs with limited undevelopable land, while Figures A.1c and A.1d are examples of MSAs with considerably large portions of undevelopable land. Finally, Figures A.1c and A.1d are also useful examples to illustrate that different color palettes are prevalent in different areas of the United States, with the color green dominating the first image and the color ochre dominating the second one. Source: Google Earth.

Figure A.2: Examples of types of areas according to their potential for urban development

(a) Undevelopable (steep terrain)



(b) Undevelopable (water body)



(c) High-developed



(d) Developed



(e) Low-developed (sparsely built land)

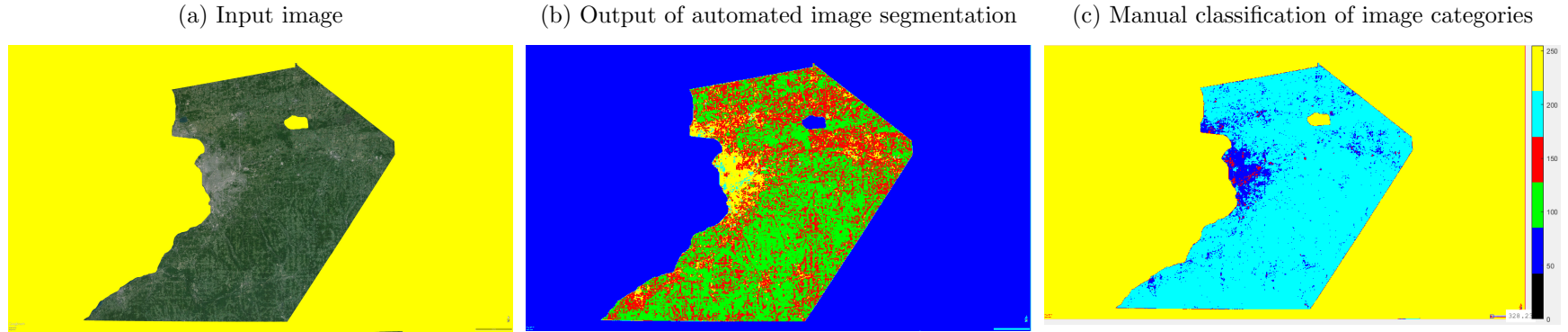


(f) Low-developed (agricultural land)



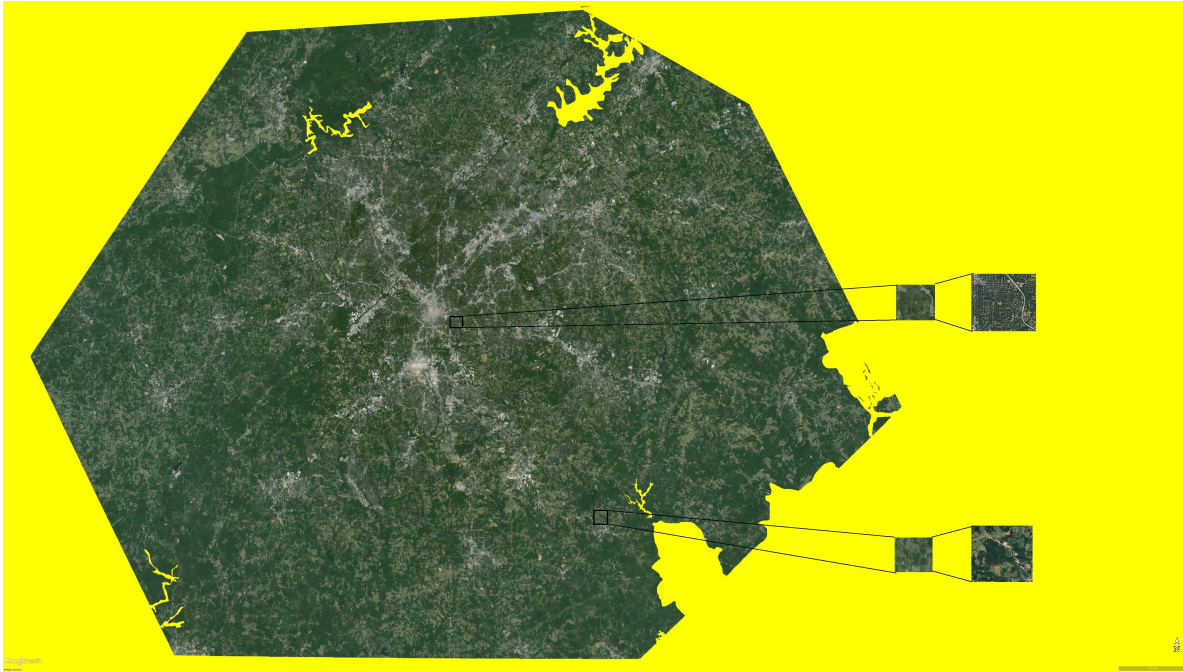
Note: This figure shows examples of aerial views of undevelopable (U) areas (mountainous terrain in Panel A.2a and waterbody in Panel A.2b), high-developed (HD) areas (Panel A.2c), developed (D) areas (Panel A.2d), and low-developed (LD) areas (Panels A.2e and A.2f). Source: Google Earth.

Figure A.3: Example of the image segmentation and land classification process (Buffalo, NY)



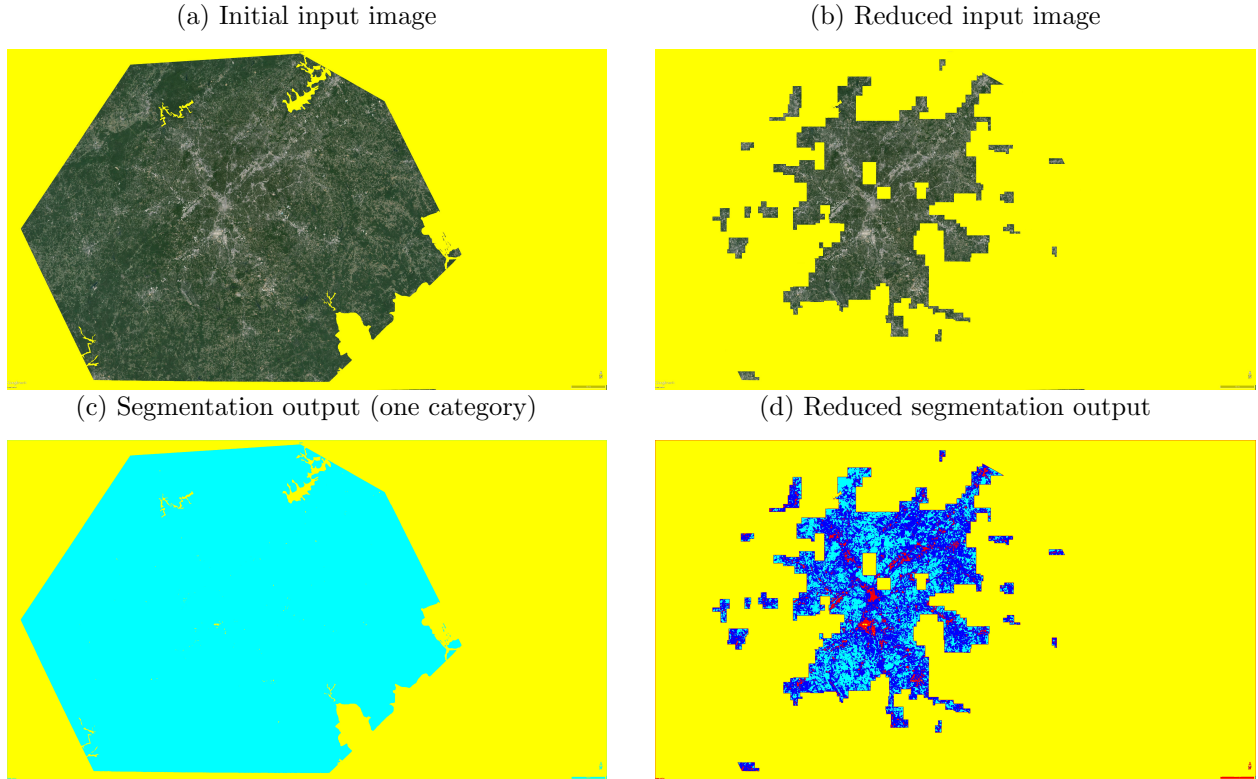
Note: This figure shows the example of the images in each step of the image segmentation process for the Buffalo, NY metropolitan area. Figure A.3a is the input image, with the 60-minute drive polygon around the city center and the undevelopable areas defined in Google Earth using the yellow color. This is the image that we upload into MATLAB for the segmentation. Figure A.3b is the outcome of the MATLAB image segmentation. The algorithm classifies the image pixels according to their color similarity into five categories. Colors are meaningless in terms of land type and are assigned randomly. Figure A.3c is the output image from the manual classification of the categories into land types. This step is manually done by the authors, by comparing the automated output (A.3b) with the input image (A.3a) in Google Earth, which allows to zoom in the image to clarify which type of areas (in terms of building density) is represented by each color category (see Figure A.2). In this case, we change to color blue in Figure A.3b to yellow Figure A.3c (area outside the polygon and undevelopable land within the polygon). Areas in green and red in Figure A.3b correspond to sparsely built and rural land, so we change it to cyan Figure A.3c (low-developed land). Yellow and Cyan in Figure A.3b correspond to urban land. By examining the zoomed-in image on Google Earth, we identify that the areas reported in cyan in Figure A.3b represent fully built areas of the city, so we change it to red in Figure A.3c (high-developed land). Finally, we identify that the yellow areas in Figure A.3b correspond to semi-urban land with potential for building density increase, so we change it to blue in Figure A.3c (developed land). In all output images, we classify land types according to these colors (yellow for undevelopable, cyan for low-developed, blue for developed, and red for high-developed). Source: Google Earth (input) and MATLAB (outputs).

Figure A.4: Examples of problematic colors for image segmentation (Atlanta, GA)



Note: This figure shows the input image of the Atlanta-Sandy Springs-Marietta, GA metropolitan area (whose main center is Atlanta, GA). This is an example of input image where the colors of developed area around the urban center are similar to the colors of some parts of cultivated agricultural land, leading to imprecise image segmentation. To solve this limitation, we implement the two-step procedure described in the text (see also Figure A.5). Source: Google Earth.

Figure A.5: Example of the two-step image segmentation process (Atlanta, GA)



Note: This figure shows the example of the two-step image segmentation process for the Atlanta-Sandy Springs-Marietta, GA metropolitan area (whose main center is Atlanta, GA). Figure A.5a is the initial input image, with the 60-minute drive polygon around the city center and the undevelopable areas defined in Google Earth using the yellow color. This is the image that we upload into MATLAB for the first-step image segmentation, where we merge all color categories into one to calculate the total area of the polygon, in square kilometers. (Figure A.5c). Figure A.5b is the reduced input image we obtain by manually eliminating portions of low-developed area using the zoomed-in image on Google Earth. The reduced input image focuses on the urban areas within the polygon and their surrounding space. We then use this reduced input image for the second-step image, following the same steps described under Figure A.3. Figure A.5d is the reduced output image, where we classify land into high-developed (red), developed (blue), and low-developed (cyan). From the reduced output we obtain the total area of the high-developed and developed land, in square kilometers. We obtain the total area of the low-developed land, in square kilometers, by subtracting the high-developed and developed areas from the total calculated in the first step. Source: Google Earth (inputs) and MATLAB (outputs).

A.2 Summary statistics

This section includes the summary statistics of the MSA-level and firm-level variables used in the empirical analysis.

Table A.1 describes the variables at the MSA level and their sources. Table A.2 reports a detailed description of the variables at the firm level, including the formula and the Compustat and CRSP variables used to calculate the measures.

Table A.3 reports the summary statistics for the variables used in the empirical analysis. The top panel lists the variables at the MSA level, while the bottom panel lists the variables at the firm level.

Table A.4 reports the correlation matrix among the MSA-level variables.

Table A.5 reports the correlation matrix among the firm-level variables.

Table A.6 reports the correlation matrix between the MSA-level housing supply elasticity, its main components (the geographic constraints/share of undevelopable area and the regulatory constraints/Wharton regulation index), and population in 2010.

Table A.1: Detailed description of MSA-level variables

Variable	Description	Source
Area of one-hour polygon	Thousands of km ²	Defined by the authors on Google Earth images; driving distance calculated using Google Maps on main highways around the city center in non-rush hours
High-developed share	$\frac{area_{HD}}{area_{HD}+area_D+area_{LD}}$	Calculated by the authors using the output of the land classification algorithm
Developed share	$\frac{area_D}{area_{HD}+area_D+area_{LD}}$	Calculated by the authors using the output of the land classification algorithm
PDI	$\frac{area_D}{area_D+area_{LD}}$	Calculated by the authors using the output of the land classification algorithm
Undevelopable area	Percentages of undevelopable area (geographically constrained) within a MSA with population over 500,000 in the 2000 Census	[3]
Wharton regulation index	Index of the level of regulations on zoning and real estate development	[4]
Housing supply elasticity	Elasticity of the housing supply, calculated considering geographical and regulatory constraints, and population levels in 2000	[3]
House price index rate	Measure of the local movement of single-family house prices, seasonally adjusted; Annual percentage change in HPI	<i>“House Price Index, Quarterly Data - Purchase-Only Indexes (Estimated using Sales Price Data) - 100 Largest Metropolitan Statistical Areas (Seasonally Adjusted)”</i> . Retrieved in July 2023 from the US Federal Housing Finance Agency.
Population density	Log of $\frac{Population}{Area}$	Area in km ² : calculated by the authors using ArcGIS, from the shapefiles for MSAs with more than 500,000 people. Retrieved in May 2023 from the Census Bureau website. Population estimates: <i>“Metropolitan and Micropolitan Statistical Area Population Estimates and Estimated Components of Change: April 1, 2010 to July 1, 2019”</i> . Retrieved in June 2023 from the Census Bureau website.
Seacoast	Dummy= 1 if there is a seacoast within 80 km from the city center	[5]. Data are available from the paper’s supplementary material. Retrieved in May 2023 from Rappaport’s website.
Elevation	Log of average elevation in meters above sea level	Calculated by the authors using ArcGIS, using the worldwide elevation raster data. Retrieved in December 2022 from Global Solar Atlas.
Slope	Log of average slope, defined as the gradient (steepness) of each cell of the elevation raster	Calculated by the authors using ArcGIS, using the worldwide elevation raster data. Retrieved in December 2022 from Global Solar Atlas.
GDP	Gross domestic product of the MSA (Log of thousands of USD)	<i>“Real GDP by county and metropolitan area; all industry total (thousands of chained 2012 dollars)”</i> . Retrieved in July 2023 from the Bureau of Economic Analysis (BEA) website.
Population in 2000	Population in year 2000, in millions	<i>“Patterns of Metropolitan and Micropolitan Population Change: 2000 to 2010 Tables. Chapter 1 Data”</i> . Retrieved in August 2024 from the Census Bureau website.

Note: This table reports the definition and sources of MSA-level variables in our database.

Table A.2: Detailed description of firm-level variables

Variable	Formula	Formula (Compustat/CRSP variables)
Monthly stock return	Monthly stock return	MTHRET
Risk-free rate	10-year bond rate	B10RET
Stock excess return	See Section 2.2 of the main text	
Stock return	$\frac{\text{Average stock price}_t - \text{Average stock price}_{t-1}}{\text{Average stock price}_{t-1}}$	$\frac{\left(\frac{\sum_{m=1}^{12} \text{MTHPRC}_m}{12}\right)_t - \left(\frac{\sum_{m=1}^{12} \text{MTHPRC}_m}{12}\right)_{t-1}}{\left(\frac{\sum_{m=1}^{12} \text{MTHPRC}_m}{12}\right)_{t-1}}$
Average stock price	Average stock price _t	$\frac{\sum_{m=1}^{12} \text{MTHPRC}_m}{12}$
Stock price volatility	Stock price standard deviation _t * $\sqrt{\# \text{ months}_t}$	$\sqrt{\sum_{m=1}^{12} (\text{MTHPRC}_m - \text{mean}(\text{MTHPRC})_t)^2}$
Real estate ratio	$\frac{\text{Buildings at cost} + \text{Land at cost} + \text{Leases at cost}}{\text{Property, plant and equipment}}$	$\frac{\text{FATB} + \text{FATL} + \text{FATP}}{\text{PPENT}}$
Leverage	$\frac{\text{Long-term debt}}{\text{Total assets}}$	$\frac{\text{DLTT}}{\text{AT}}$
Market capitalization	Annual closing price × Common shares outstanding	PRCC_C × CSHO
Years since first listed	Current year – Year of IPO	Year – Year(IPODATE)
Years since first linked	Current year – Year of first link	Year – Year(LINKDT)
Seniority in stock market	min{Years since first listed; Years since first linked}	min{Year – Year(IPODATE); Year – Year(LINKDT)}
R&D expenditure	$\frac{\text{R\&D expenses}}{\text{Property, plant and equipment}}$	$\frac{\text{XRD}}{\text{PPEGT}}$
Return on assets (ROA)	$\frac{\text{Income before extraordinary items} + \text{Income taxes} - \text{Preferred dividends}}{\text{Total assets}}$	$\frac{\text{IB} + \text{TXDI} - \text{DVP}}{\text{TA}}$
Return on equity (ROE)	$\frac{\text{Income before extraordinary items}}{\text{Shareholders' equity}}$	$\frac{\text{IBCOM}}{\text{TEQ}}$
Market-to-book value ratio	$\frac{\text{Market value}}{\text{Book value per share} \times \text{Shares outstanding}}$	$\frac{\text{MKVALT}}{\text{BKVLPS} \times \text{CSHO}}$
Asset growth	$\frac{\text{Assets}_t - \text{Assets}_{t-1}}{\text{Assets}_{t-1}}$	$\frac{\text{AT}_t - \text{AT}_{t-1}}{\text{AT}_{t-1}}$
Fixed investment ratio	$\frac{\text{Capital expenditure}}{\text{Common stock}}$	$\frac{\text{CAPX}}{\text{CSTK}}$
Value added	$\frac{\text{Operating income} + \text{staff expenses}}{\text{GDP price deflator}}$	$\frac{\text{OIBDP} + \text{XLR}}{\text{GDP price deflator}}$
Capital stock	$\frac{\text{Property, plant and equipment}}{\text{Private investment price deflator}}$	$\frac{\text{PPEGT}}{\text{Private investment price deflator}}$
Labor stock	Number of employees	EMP

Note: This table reports the definition of firm-level variables in our database.

Table A.3: Summary statistics

	Num.	Mean	St. Dev.	Min.	Max.
MSA-level variables					
Number of observations by city	82	1,192	1,779.37	21	10,255
Area of one-hour polygon (<i>Log</i>)	82	14.444	4.676	5.137	26.008
High-developed share	82	0.058	0.086	0	0.463
Developed share	82	0.169	0.126	0.016	0.721
Potential for density increase (PDI)	82	0.190	0.169	0.016	0.821
Undevelopable area	82	0.266	0.225	0.010	0.861
Wharton regulation index	82	0.089	0.691	-1.24	1.89
Housing supply elasticity	82	1.911	0.998	0.60	5.45
House price index (<i>Y-t-Y % change</i>)	75	0.044	0.223	0.003	0.097
Population density (<i>Log</i>)	82	5.184	0.717	3.629	7.006
Seacoast within 80 km	82	0.354	0.481	0	1
Elevation (<i>Log</i>)	82	5.254	1.240	1.522	7.700
Slope (<i>Log</i>)	82	-0.777	1.093	-3.659	1.487
GDP (<i>Log</i>)	81	18.265	0.960	16.744	21.093
Popul. year 2000 (<i>Millions</i>)	82	2.073	2.657	0.323	18.323
Firm-level variables					
Stock excess return (<i>4 Fama & French factors</i>)	17,756	0.056	0.494	-0.967	19.018
Stock return – Risk-free rate	17,756	0.106	0.539	-0.975	15.113
Stock excess return (<i>5 Fama & French factors – Def. 1</i>)	17,756	0.056	0.490	-0.968	18.602
Stock excess return (<i>5 Fama & French factors – Def. 2</i>)	17,756	0.057	0.494	-0.967	19.075
Stock excess return (<i>6 Fama & French factors</i>)	17,756	0.056	0.490	-0.968	18.637
Stock return (<i>Annual perc. change in price</i>)	14,720	0.097	0.387	-0.848	10.012
Average stock price (<i>Annual; Log</i>)	17,750	3.007	1.178	-1.894	8.338
Stock price volatility (<i>Annual; Log</i>)	17,669	2.151	1.257	-3.912	7.442
Real estate ratio (<i>1-year lagged</i>)	18,084	0.668	0.939	0	57.25
Leverage (<i>1-year lagged</i>)	17,986	0.203	0.235	0	4.059
Market capitalization (<i>Log; 1-year lagged</i>)	18,075	6.727	2.132	-2.621	13.673
Listed in the stock market (<i>Years</i>)	21,510	18.76	15.32	-1	73
R&D expenditure (<i>1-year lagged</i>)	18,084	1.019	8.737	-0.177	843.524
ROE (<i>1-year lagged</i>)	18,084	-0.143	10.109	-1,096.491	143.588
ROA (<i>1-year lagged</i>)	17,704	-0.050	0.324	-12.459	2.028
Market-to-book value (<i>1-year lagged</i>)	18,008	4.147	116.798	-4,062.5	14,450.66
Asset growth (<i>1-year lagged</i>)	14,896	0.114	0.429	-0.839	11.641
Fixed investment ratio (<i>1-year lagged</i>)	17,425	1,078.851	8,632.276	0.000	406,422
Value added (<i>Log</i>)	16,787	0.392	2.181	-10.386	6.982
Capital stock (<i>Log</i>)	21,510	0.672	2.580	-9.576	8.419
Labor stock (<i>Log</i>)	21,510	0.564	2.183	-6.215	6.682
Productivity	16,787	-0.178	1.652	-9.683	4.786

Note: This table reports the summary statistics (number of observations, mean, standard deviation, minimum and maximum values) of MSA-level variables (upper panel) and of firm-level variables (bottom panel).

Table A.4: Correlation matrix – MSA-level variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) Polygon area	1													
(2) High-developed	-0.522	1												
(3) Developed	-0.417	0.582	1											
(4) PDI	-0.464	0.755	0.961	1										
(5) Undevelop. area	-0.548	0.400	0.421	0.421	1									
(6) Regulation ind.	-0.272	0.300	0.300	0.295	0.248	1								
(7) Housing supply elast.	0.636	-0.451	-0.509	-0.501	-0.612	-0.453	1							
(8) House price rate	-0.148	0.287	0.139	0.163	0.249	0.110	-0.203	1						
(9) Popul. density	-0.369	0.354	0.438	0.409	0.243	0.340	-0.409	0.013	1					
(10) Seacoast $\leq$ 80 km	-0.413	0.314	0.356	0.330	0.535	0.285	-0.506	0.079	0.475	1				
(11) Elevation	0.061	0.166	0.004	0.068	-0.279	0.113	0.168	0.061	-0.302	-0.572	1			
(12) Slope	-0.229	0.380	0.144	0.222	0.001	0.282	-0.250	0.089	-0.122	-0.109	0.770	1		
(13) GDP	-0.201	0.546	0.508	0.527	0.135	0.221	-0.391	0.228	0.601	0.298	-0.046	0.079	1	
(14) Popul. in 2000	-0.241	0.583	0.501	0.555	0.157	0.217	-0.339	0.035	0.605	0.274	-0.054	0.040	0.807	1

Note: This table reports the correlation among city-level variables.

Table A.5: Correlation matrix – Firm-level variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1) Excess ret. (4FF)	1																	
(2) Stock ret. - Risk-free	0.94	1																
(3) Excess ret. (5FF-def.1)	1	0.94	1															
(4) Excess ret. (5FF-def.2)	1	0.94	1	1														
(5) Excess ret. (6FF)	1	0.94	1	1	1													
(6) Stock ret. (% Δ Pr)	0.62	0.58	0.62	0.62	0.62	1												
(7) Aver. stock price	0.12	0.10	0.12	0.12	0.12	0.23	1											
(8) Stock price volat.	0.18	0.17	0.18	0.18	0.18	0.25	0.85	1										
(9) Real estate ratio	0.03	0.03	0.03	0.03	0.03	0.04	-0.12	-0.09	1									
(10) Leverage	0.04	0.03	0.03	0.03	0.03	-0.004	0.09	0.07	-0.06	1								
(11) Market capitaliz.	-0.01	-0.03	-0.01	-0.01	-0.01	0.06	0.77	0.62	-0.13	0.121	1							
(12) Market seniority	-0.01	-0.002	-0.01	-0.01	-0.01	-0.01	0.28	0.13	0.01	0.03	0.30	1						
(13) R&D expendit.	0.01	0.00	0.01	0.01	0.01	0.02	-0.07	-0.01	-0.02	-0.05	-0.06	-0.09	1					
(14) ROE	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.02	-0.01	-0.00	0.04	0.02	-0.00	1				
(15) ROA	-0.02	-0.02	-0.02	-0.02	-0.02	0.03	0.37	0.24	-0.08	0.03	0.35	0.23	-0.22	0.05	1			
(16) Mkt-to-book value	0.03	0.02	0.03	0.03	0.03	0.03	0.01	0.02	-0.00	-0.01	0.00	0.00	0.00	-0.33	-0.00	1		
(17) Asset growth	-0.03	-0.04	-0.03	-0.03	-0.03	0.05	0.10	0.14	-0.06	0.00	0.08	-0.08	0.05	0.00	0.09	0.00	1	
(18) Fixed invest.	0.01	0.01	0.00	0.01	0.00	-0.00	0.05	0.06	-0.04	0.05	0.07	-0.05	-0.01	0.00	0.03	-0.00	0.02	1

Note: This table reports the correlation among firm-level variables.

Table A.6: Correlation matrix of housing supply elasticity and its components

	(1)	(2)	(3)	(4)
(1) Geographic constraints	1			
(2) Regulatory constraints	0.248	1		
(3) Population in 2010	0.142	0.215	1	
(4) Housing supply elasticity	-0.612	-0.453	-0.331	1

Note: This table reports the pairwise correlations among housing supply elasticity and its main components, i.e., physical/geographical constraints, regulatory constraints, and population, as described in [3].

A.3 Additional MSA-level measures

In this section, we report additional information regarding the metropolitan statistical areas (MSAs) of the US considered in the paper.

Table A.7 reports the list of MSAs included in the Compustat database (2010-2019) after excluding firms listed in the United States but located abroad, firms located in the United States but outside the metropolitan statistical areas (e.g., those located in micropolitan statistical areas), as well as firms in the finance, insurance, real estate, nonprofit, government, construction, and mining industries. The table includes a column that reports whether the MSA is included in the final sample used for the empirical analysis (i.e., whether we calculate the corresponding urbanization measures), and three columns to identify the largest MSAs in the sample (top 3 and top 10 most populous; top 3 most densely populated).

Table A.8 reports the list of MSAs we include in the empirical analysis. The table shows the extension of the area of the one-hour drive polygon we build around the main city center of the MSA, the urbanization measures we calculate from the output of the land classification algorithm,⁵ as well as the average population density in the period 2010-2019 and the number of firm-year observations included in the database.

Figure A.6 shows a map of US MSAs and the corresponding PDI measure. The state borders are depicted in blue, while the MSA borders are depicted in black. Colored MSA areas indicate that the MSA is included in our sample, while blank MSA areas indicate that the MSA is not included. Darker colors indicate higher PDI. For each MSA included in the sample, the small green dot indicates the location of the main city center within the MSA territory.

Table A.9 shows the relationships among the real estate supply elasticity, geographic and regulatory constraints, and the MSA-level geographic controls used in the main analysis. These results are obtained from the following regression at the level of MSA c located in state s :

$$Supply\ Elasticity_c = \alpha + \beta_1 Geographic\ Constraints_c + \beta_2 Regulatory\ Constraints_c + \mathbf{X}'_c \Gamma + \zeta_{s(c)} + \varepsilon_c$$

where $Supply\ Elasticity_c$ is obtained from [3]; $Geographic\ Constraints_c$ (i.e., share undevelopable land) and $Regulatory\ Constraints_c$ (i.e., Wharton Regulatory Index) are the main components used to calculate the supply elasticity in [3]; $\mathbf{X}_c$ is a vector of MSA-level controls used in the main analysis; $\zeta_{s(c)}$ indicates state fixed-effects; and ε_c is the error component.

Table A.10 reports the residual real estate supply elasticity by MSA:

$$Residual\ Supply\ Elasticity_c = Supply\ Elasticity_c - \widehat{Supply\ Elasticity}_c$$

⁵Except for the potential density increase index, which is reported in Table 1 in the main text.

where $\widehat{Supply\ Elasticity}_c$ is estimated from the following regression:

$$\widehat{Supply\ Elasticity}_c = \alpha + \beta_1 Geographic\ Constraints_c + \beta_2 Regulatory\ Constraints_c + \varepsilon_c$$

This residual real estate supply elasticity is the source of variation that we exploit for our identification strategy.

Figure A.7 shows a map of US MSAs and the corresponding residual real estate supply elasticity measure. The state borders are depicted in blue, while the MSA borders are depicted in black. Colored MSA areas indicate that the MSA is included in our sample, while blank MSA areas indicate that the MSA is not included. Darker colors indicate higher residual elasticity. For each MSA included in the sample, the small green dot indicates the location of the main city center within the MSA territory.

Table A.7: List of MSAs included in Compustat data

MSA name	Urb. meas.	Most popul.		Most dense	MSA name	Urb. meas.	Most popul.		Most dense
		Top 3	Top 10	Top 3			Top 3	Top 10	Top 3
Akron, OH	✓				Colorado Springs, CO	✓			
Albany-Schenectady-Troy, NY	✓				Columbia, SC	✓			
Albuquerque, NM	✓				Columbus, GA-AL				
Alexandria, LA					Columbus, IN				
Allentown-Bethlehem-Easton, PA-NJ	✓				Columbus, OH	✓			
Amarillo, TX					Corpus Christi, TX				
Ames, IA					Dallas-Fort Worth-Arlington, TX	✓		✓	
Anchorage, AK					Dalton, GA				
Anderson, IN					Davenport-Moline-Rock Island, IA-IL				
Ann Arbor, MI	✓				Dayton, OH	✓			
Appleton, WI					Decatur, AL				
Asheville, NC					Deltona-Daytona Beach-Ormond Beach, FL				
Atlanta-Sandy Springs-Marietta, GA	✓		✓		Denver-Aurora, CO	✓			
Atlantic City, NJ					Des Moines-West Des Moines, IA				
Austin-Round Rock, TX	✓				Detroit-Warren-Livonia, MI	✓			
Bakersfield, CA	✓				Dover, DE				
Baltimore-Towson, MD	✓				Dubuque, IA				
Baton Rouge, LA	✓				Duluth, MN-WI				
Battle Creek, MI					Durham, NC				
Birmingham-Hoover, AL	✓				Eau Claire, WI				
Boise City-Nampa, ID					El Paso, TX	✓			
Boston-Cambridge-Quincy, MA-NH	✓		✓		Elkhart-Goshen, IN				
Boulder, CO					Erie, PA				
Bremerton-Silverdale, WA					Eugene-Springfield, OR				
Bridgeport-Stamford-Norwalk, CT					Evansville, IN-KY				
Buffalo-Niagara Falls, NY	✓				Fargo, ND-MN				
Burlington, NC					Fayetteville-Springdale-Rogers, AR-MO				
Canton-Massillon, OH					Flint, MI				
Cape Coral-Fort Myers, FL					Florence, SC				
Charleston-North Charleston, SC	✓				Fort Collins-Loveland, CO				
Charlotte-Gastonia-Concord, NC-SC	✓				Fort Smith, AR-OK				
Chattanooga, TN-GA					Fort Wayne, IN	✓			
Cheyenne, WY					Gainesville, FL				
Chicago-Naperville-Joliet, IL-IN-WI	✓	✓	✓		Grand Rapids-Wyoming, MI	✓			
Cincinnati-Middletown, OH-KY-IN	✓				Greeley, CO				
Cleveland-Elyria-Mentor, OH	✓				Green Bay, WI				
College Station-Bryan, TX					Greensboro-High Point, NC	✓			
Greenville-Mauldin-Easley, SC	✓				Madison, WI				
Harrisburg-Carlisle, PA	✓				Manchester-Nashua, NH				
Hartford-West Hartford-East Hartford, CT	✓				Mankato-North Mankato, MN				
Hickory-Lenoir-Morganton, NC					Mansfield, OH				

(to be continued on the next page)

Table A.7 (continued from previous page)

MSA name	Urb. meas.	Most popul.		Most dense	MSA name	Urb. meas.	Most popul.		Most dense
		Top 3	Top 10	Top 3			Top 3	Top 10	Top 3
Holland-Grand Haven, MI					Medford, OR				
Honolulu, HI					Memphis, TN-AR-MS	✓			
Houston-Sugar Land-Baytown, TX	✓		✓		Miami-Fort Lauderdale-Pompano Beach, FL MSA	✓		✓	
Huntington-Ashland, WV-KY-OH					Midland, TX				
Huntsville, AL					Milwaukee-Waukesha-West Allis, WI	✓			
Indianapolis-Carmel, IN	✓				Minneapolis-St. Paul-Bloomington, MN-WI	✓			
Iowa City, IA					Mobile, AL	✓			
Jackson, MI					Monroe, LA				
Jackson, MS					Monroe, MI				
Jacksonville, FL	✓				Naples-Marco Island, FL				
Janesville, WI					Nashville-Davidson-Murfreesboro-Franklin, TN	✓			
Joplin, MO					New Haven-Milford, CT	✓			
Kalamazoo-Portage, MI					New Orleans-Metairie-Kenner, LA	✓			
Kansas City, MO-KS	✓				New York-Northern New Jersey-Long Island, NY-NJ-PA	✓	✓	✓	
Kennewick-Richland-Pasco, WA					Niles-Benton Harbor, MI				
Kingsport-Bristol-Bristol, TN-VA					Ocala, FL				
Knoxville, TN	✓				Ogden-Clearfield, UT				
Kokomo, IN					Oklahoma City, OK	✓			
Lafayette, IN					Omaha-Council Bluffs, NE-IA	✓			
Lafayette, LA					Orlando-Kissimmee, FL	✓			
Lakeland, FL					Oshkosh-Neenah, WI				
Lancaster, PA					Oxnard-Thousand Oaks-Ventura, CA	✓			
Lansing-East Lansing, MI					Palm Bay-Melbourne-Titusville, FL				
Las Vegas-Paradise, NV	✓				Panama City-Lynn Haven, FL				
Lexington-Fayette, KY					Peoria, IL				
Lincoln, NE					Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	✓		✓	
Little Rock-North Little Rock-Conway, AR	✓				Phoenix-Mesa-Scottsdale, AZ	✓			
Longview, TX					Pittsburgh, PA	✓			
Los Angeles-Long Beach-Santa Ana, CA	✓	✓	✓	✓	Pittsfield, MA				
Louisville-Jefferson County, KY-IN	✓				Port St. Lucie, FL				
Lubbock, TX					Portland-South Portland-Biddeford, ME				
Lynchburg, VA					Portland-Vancouver-Beaverton, OR-WA	✓			
Macon, GA MSA					Poughkeepsie-Newburgh-Middletown, NY				
Providence-New Bedford-Fall River, RI-MA	✓				Seattle-Tacoma-Bellevue, WA	✓			
Provo-Orem, UT					Sioux Falls, SD				
Racine, WI					Spartanburg, SC				
Raleigh-Cary, NC	✓				Spokane, WA				
Rapid City, SD					Springfield, MA	✓			
Reading, PA					Springfield, MO				
Reno-Sparks, NV					St. George, UT				
Richmond, VA	✓				St. Louis, MO-IL	✓			
Riverside-San Bernardino-Ontario, CA	✓				Stockton, CA	✓			

(to be continued on the next page)

Table A.7 (continued from previous page)

MSA name	Urb. meas.	Most popul.		Most dense	MSA name	Urb. meas.	Most popul.		Most dense
		Top 3	Top 10	Top 3			Top 3	Top 10	Top 3
Roanoke, VA					Syracuse, NY	✓			
Rochester, MN					Tampa-St. Petersburg-Clearwater, FL	✓			
Rochester, NY	✓				Toledo, OH	✓			
Sacramento-Arden-Arcade-Roseville, CA					Topeka, KS				
Salem, OR					Trenton-Ewing, NJ				
Salt Lake City, UT	✓				Tucson, AZ	✓			
San Antonio, TX	✓				Tulsa, OK	✓			
San Diego-Carlsbad-San Marcos, CA	✓				Utica-Rome, NY				
San Francisco-Oakland-Fremont, CA	✓			✓	Virginia Beach-Norfolk-Newport News, VA-NC	✓			
San Jose-Sunnyvale-Santa Clara, CA	✓			✓	Washington-Arlington-Alexandria, DC-VA-MD-WV	✓		✓	
San Juan-Caguas-Guaynabo, PR					Wausau, WI				
San Luis Obispo-Paso Robles, CA					Weirton-Steubenville, WV-OH				
Sandusky, OH					Wichita, KS	✓			
Santa Barbara-Santa Maria-Goleta, CA					Winchester, VA-WV				
Santa Cruz-Watsonville, CA					Winston-Salem, NC				
Santa Fe, NM					Worcester, MA				
Santa Rosa-Petaluma, CA					York-Hanover, PA				
Sarasota-Bradenton-Venice, FL	✓				Youngstown-Warren-Boardman, OH-PA	✓			
Savannah, GA									

Note: This table reports the list of MSAs in our Compustat data.

Table A.8: Additional city-level measures

MSA name	MSA main urban center	Polygon area (1,000 km ²)	High-level Index (%)	Developed Index (%)	Popul. density (pop/km ²)	Num. observ.
Akron, OH MSA	Akron, OH	18.33	2.29	19.90	300.2	300
Albany-Schenectady-Troy, NY MSA	Albany, NY	16.11	0.00	7.46	120.3	196
Albuquerque, NM MSA	Albuquerque, NM	15.26	3.72	2.20	37.7	55
Allentown-Bethlehem-Easton, PA-NJ MSA	Allentown, PA	16.53	4.95	14.54	220.1	176
Ann Arbor, MI MSA	Ann Arbor, MI	19.99	3.64	22.59	193.9	151
Atlanta-Sandy Springs-Marietta, GA MSA	Atlanta, GA	26.01	1.68	16.04	260.1	2922
Austin-Round Rock, TX MSA	Austin, TX	13.54	2.53	11.50	179.9	1248
Bakersfield, CA MSA	Bakersfield, CA	15.24	0.00	11.84	41.3	55
Baltimore-Towson, MD MSA	Baltimore, MD	13.11	3.85	22.48	409.6	663
Baton Rouge, LA MSA	Baton Rouge, LA	10.00	0.00	8.87	81.0	166
Birmingham-Hoover, AL MSA	Birmingham, AL	25.53	1.80	7.59	78.3	182
Boston-Cambridge-Quincy, MA-NH MSA	Boston, MA	12.28	9.39	28.38	522.9	6659
Buffalo-Niagara Falls, NY MSA	Buffalo, NY	9.00	0.49	7.24	279.0	609
Charleston-North Charleston, SC MSA	Charleston, SC	8.72	1.88	12.08	111.9	95
Charlotte-Gastonia-Concord, NC-SC MSA	Charlotte, NC	13.02	0.29	1.59	302.1	1345
Chicago-Naperville-Joliet, IL-IN-WI MSA	Chicago, IL	5.14	15.82	69.07	509.4	5569
Cincinnati-Middletown, OH-KY-IN MSA	Cincinnati, OH	16.94	1.92	15.40	191.1	757
Cleveland-Elyria-Mentor, OH MSA	Cleveland, OH	11.03	3.37	24.65	397.4	1264
Colorado Springs, CO MSA	Colorado Springs, CO	13.48	0.00	9.58	98.7	92
Columbia, SC MSA	Columbia, SC	22.53	1.62	11.64	83.1	36
Columbus, OH MSA	Columbus, OH	19.58	2.21	9.13	194.8	935
Dallas-Fort Worth-Arlington, TX MSA	Dallas, TX	21.98	6.07	25.81	299.3	4811
Dayton, OH MSA	Dayton, OH	19.36	2.94	11.31	181.2	111
Denver-Aurora, CO MSA	Denver, CO	19.26	3.44	15.64	128.2	2225
Detroit-Warren-Livonia, MI MSA	Detroit, MI	12.35	4.68	26.64	425.0	1478
El Paso, TX MSA	El Paso, TX	10.50	6.70	3.86	317.6	144
Fort Wayne, IN MSA	Fort Wayne, IN	21.85	0.54	3.71	113.4	110
Grand Rapids-Wyoming, MI MSA	Grand Rapids, MI	19.61	2.69	8.80	141.7	233
Greensboro-High Point, NC MSA	Greensboro, NC	19.11	1.26	9.03	143.8	233
Greenville-Mauldin-Easley, SC MSA	Greenville, SC	11.89	3.21	11.97	166.3	128
Harrisburg-Carlisle, PA MSA	Harrisburg, PA	13.70	1.14	8.51	133.8	180
Hartford-West Hartford-East Hartford, CT MSA	Hartford, CT	15.01	3.18	14.23	308.9	346
Houston-Sugar Land-Baytown, TX MSA	Houston, TX	16.91	8.79	35.99	282.9	3758
Indianapolis-Carmel, IN MSA	Indianapolis, IN	16.33	1.79	12.52	196.6	536
Jacksonville, FL MSA	Jacksonville, FL	9.66	3.58	19.36	172.6	463

(to be continued on the next page)

Table A.8 (continued from previous page)

MSA name	MSA main urban center	Polygon area (1,000 km ²)	High-level Index (%)	Developed Index (%)	Popul. density (pop/km ²)	Num. observ.
Kansas City, MO-KS MSA	Kansas City, MO-KS	19.33	2.21	16.47	102.2	624
Knoxville, TN MSA	Knoxville, TN	11.84	1.16	6.60	173.4	142
Las Vegas-Paradise, NV MSA	Las Vegas, NV	9.53	8.66	8.54	102.0	969
Little Rock-North Little Rock-Conway, AR MSA	Little Rock, AR	19.83	0.77	6.35	68.5	155
Los Angeles-Long Beach-Santa Ana, CA MSA	Los Angeles, CA	5.95	42.60	46.68	1043.8	4469
Louisville-Jefferson County, KY-IN	Louisville, KY	16.67	1.48	6.90	115.8	528
Memphis, TN-AR-MS MSA	Memphis, TN	16.12	1.56	9.83	112.5	305
Miami-Fort Lauderdale-Pompano Beach, FL MSA	Miami, FL	12.58	11.90	9.40	445.5	2106
Milwaukee-Waukesha-West Allis, WI MSA	Milwaukee, WI	10.79	5.07	13.60	415.4	1304
Minneapolis-St. Paul-Bloomington, MN-WI MSA	Minneapolis, MN	16.09	2.93	24.92	221.4	2950
Mobile, AL MSA	Mobile, AL	15.69	2.07	11.06	134.9	70
Nashville-Davidson-Murfreesboro-Franklin, TN MSA	Nashville, TN	18.81	1.61	18.41	121.6	1293
New Haven-Milford, CT MSA	New Haven, CT	8.47	5.65	21.69	548.6	245
New Orleans-Metairie-Kenner, LA MSA	New Orleans, LA	12.05	2.67	14.57	152.2	216
New York-Northern New Jersey-Long Island, NY-NJ-PA MSA	New York, NY	10.29	23.85	27.50	1102.7	10255
Oklahoma City, OK MSA	Oklahoma City, OK	20.39	2.09	6.36	94.1	200
Omaha-Council Bluffs, NE-IA MSA	Omaha, NE	12.01	1.28	8.38	80.4	428
Orlando-Kissimmee, FL MSA	Orlando, FL	19.02	3.00	18.19	264.1	588
Oxnard-Thousand Oaks-Ventura, CA MSA	Ventura, CA	10.15	9.79	72.11	175.7	437
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD MSA	Philadelphia, PA	16.84	5.92	24.81	504.2	3325
Phoenix-Mesa-Scottsdale, AZ MSA	Phoenix, AZ	12.98	18.51	17.64	120.7	1762
Pittsburgh, PA MSA	Pittsburgh, PA	12.82	2.70	20.19	171.6	1357
Portland-Vancouver-Beaverton, OR-WA MSA	Portland, OR	12.33	2.09	13.43	135.9	1131
Providence-New Bedford-Fall River, RI-MA MSA	Providence, RI	12.47	6.98	22.20	388.9	514
Raleigh-Cary, NC MSA	Raleigh, NC	15.76	0.97	7.09	229.0	659
Richmond, VA MSA	Richmond, VA	19.49	1.94	13.02	83.8	593
Riverside-San Bernardino-Ontario, CA MSA	Riverside, CA	7.80	46.27	43.55	63.0	94
Rochester, NY MSA	Rochester, NY	9.52	1.18	10.01	141.8	415
Salt Lake City, UT MSA	Salt Lake City, UT	13.49	8.41	32.36	47.0	866
San Antonio, TX MSA	San Antonio, TX	20.74	1.12	8.64	123.7	410
San Diego-Carlsbad-San Marcos, CA MSA	San Diego, CA	6.81	24.94	30.21	298.7	2486
San Francisco-Oakland-Fremont, CA MSA	San Francisco, CA	7.11	23.10	23.76	716.9	5214
San Jose-Sunnyvale-Santa Clara, CA MSA	San Jose, CA	7.62	29.80	25.87	279.2	4789
Sarasota-Bradenton-Venice, FL MSA	Sarasota, FL	6.12	5.64	30.59	223.1	171
Seattle-Tacoma-Bellevue, WA MSA	Seattle, WA	7.31	19.30	20.00	243.1	1801
Springfield, MA MSA	Springfield, MA	10.64	3.76	14.46	145.9	110

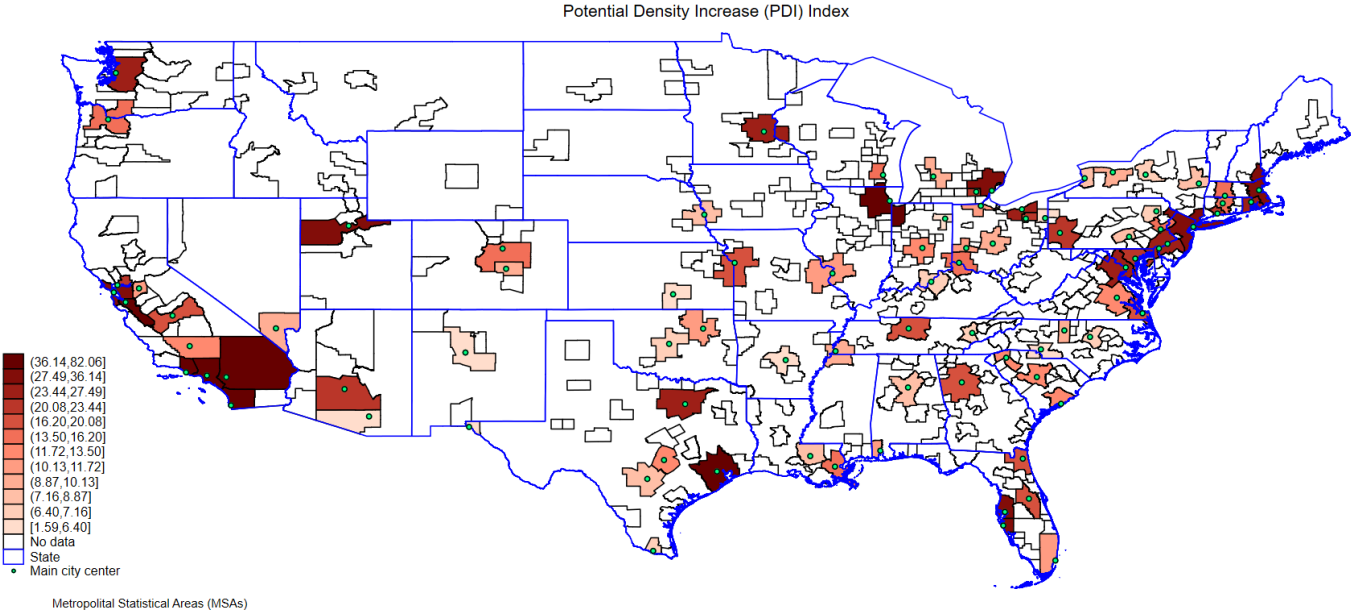
(to be continued on the next page)

Table A.8 (continued from previous page)

MSA name	MSA main urban center	Polygon area (1,000 km ²)	High-devel. Index (%)	Developed Index (%)	Popul. density (pop/km ²)	Num. observ.
St. Louis, MO-IL MSA	St. Louis, MO	21.46	2.04	11.39	125.0	1249
Stockton, CA MSA	Stockton, CA	15.42	2.25	11.09	193.9	21
Syracuse, NY MSA	Syracuse, NY	17.86	0.45	6.59	106.4	65
Tampa-St. Petersburg-Clearwater, FL	Tampa, FL	11.83	4.84	29.16	449.4	945
Toledo, OH MSA	Toledo, OH	15.43	2.01	11.39	154.2	275
Tucson, AZ MSA	Tucson, AZ	14.52	3.11	5.10	42.5	84
Tulsa, OK MSA	Tulsa, OK	17.45	1.97	11.49	59.8	450
Virginia Beach-Norfolk-Newport News, VA-NC MSA	Virginia Beach, VA	9.39	4.55	15.89	256.2	116
Washington-Arlington-Alexandria, DC-VA-MD-WV MSA	Washington, DC	14.29	3.85	25.91	413.4	3212
Wichita, KS MSA	Wichita, KS	17.37	2.04	3.24	58.9	59
Youngstown-Warren-Boardman, OH-PA MSA	Youngstown, OH	14.82	1.56	4.21	124.8	56

Note: This table reports the 82 MSAs in our Compustat data for which we calculate the urbanization measures. The table reports the value of our urbanization measures not reported in Table 1, the value of population density (2010-2019 average), and the number of observations included in our sample, by city.

Figure A.6: Potential Density Increase (PDI) Index by MSA



Note: This figure shows the Potential Density Increase (PDI) Index by MSA.

Table A.9: Additional results – Real estate supply elasticity components

	Real estate supply elasticity					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Geographic constraints	-0.024*** (0.006)	-0.018*** (0.006)	-0.023*** (0.006)	-0.017*** (0.006)	-0.013 (0.009)	-0.011** (0.005)
Regulatory constraints	-0.464*** (0.099)	-0.050 (0.172)	-0.410*** (0.105)	-0.008 (0.169)	-0.311*** (0.113)	-0.191 (0.149)
Population in 2010			-0.000** (0.000)	-0.000*** (0.000)	0.000 (0.000)	0.000 (0.000)
Polygon area					0.000* (0.000)	0.000*** (0.000)
Population density					-0.012 (0.120)	-0.005 (0.121)
GDP					-0.268* (0.155)	-0.262*** (0.093)
Elevation					0.509** (0.205)	0.409* (0.230)
Slope					-0.521** (0.205)	-0.386 (0.303)
Seacoast $\leq$ 80 km					0.425* (0.221)	0.331* (0.186)
Constant	2.580*** (0.131)	2.472*** (0.184)	2.714*** (0.106)	2.566*** (0.162)	3.129* (1.837)	3.633** (1.619)
State FE	No	Yes	No	Yes	No	Yes
Robust std. error	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	82	64	82	64	81	63
Adjusted R^2	0.457	0.744	0.489	0.771	0.653	0.858

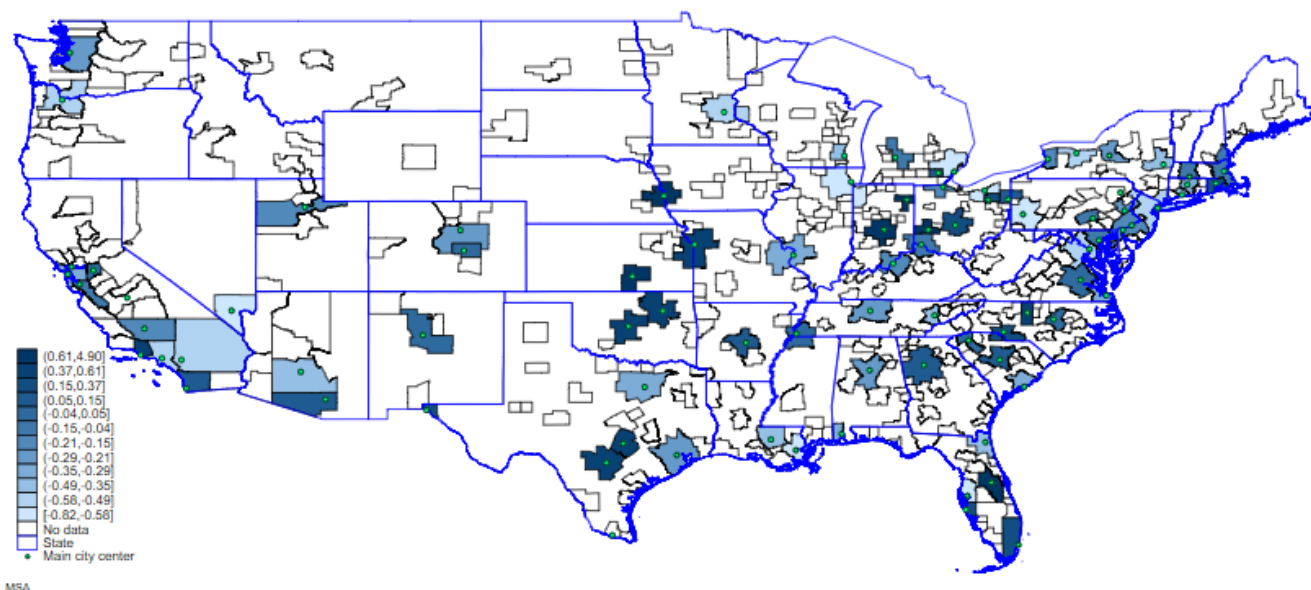
Note: This table reports additional estimation results. The unit of observation is an MSA c in state s . For all estimation models, the dependent variable is the MSA's real-estate supply elasticity, as reported by [3]. In columns (1) and (2), the model is estimated using the two main components of the supply elasticity: the MSA's geographic constraints and the MSA's regulatory constraints (share of undevelopable land and Wharton Regulation Index, respectively; obtained from [3]). In columns (3) and (4), we add the MSA population in 2010 as an additional control. In columns (5) and (6), we add all other geographic controls: the one-hour drive polygon area; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. In columns (2), (4), and (6), we include state fixed effects in the regression model. Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Residual supply elasticity by MSA

MSA Name	Elast.	MSA Name	Elast.
Akron, OH	0.195	Memphis, TN-AR-MS	0.015
Albany-Schenectady-Troy, NY	-0.372	Miami-Fort Lauderdale-West Palm Beach, FL	0.261
Albuquerque, NM	-0.024	Milwaukee-Waukesha-West Allis, WI	-0.353
Allentown-Bethlehem-Easton, PA-NJ	-0.219	Minneapolis-St. Paul-Bloomington, MN-WI	-0.501
Ann Arbor, MI	0.083	Mobile, AL	-0.314
Atlanta-Sandy Springs-Marietta, GA	0.080	Nashville-Davidson-Murfreesboro-Franklin, TN	-0.228
Austin-Round Rock, TX	0.379	New Haven-Milford, CT	-0.452
Bakersfield, CA	-0.184	New Orleans-Metairie-Kenner, LA	-0.583
Baltimore-Towson, MD	-0.092	New York-Northern New Jersey-Long Island, NY-NJ-PA	-0.566
Baton Rouge, LA	-0.427	Oklahoma City, OK	0.596
Birmingham-Hoover, AL	-0.209	Omaha-Council Bluffs, NE-IA	0.709
Boston-Cambridge-Quincy, MA-NH	-0.132	Orlando-Kissimmee, FL	0.716
Buffalo-Niagara Falls, NY	-0.194	Oxnard-Thousand Oaks-Ventura, CA	0.607
Charleston-North Charleston, SC	-0.333	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	-0.166
Charlotte-Gastonia-Concord, NC-SC	0.374	Phoenix-Mesa-Scottsdale, AZ	-0.358
Chicago-Naperville-Joliet, IL-IN-WI	-0.819	Pittsburgh, PA	-0.627
Cincinnati-Middletown, OH-KY-IN	-0.147	Portland-Vancouver-Beaverton, OR-WA	-0.501
Cleveland-Elyria-Mentor, OH	-0.681	Providence-New Bedford-Fall River, RI-MA	0.235
Colorado Springs, CO	0.019	Raleigh-Cary, NC	0.018
Columbia, SC	0.066	Richmond, VA	0.051
Columbus, OH	0.031	Riverside-San Bernardino-Ontario, CA	-0.501
Dallas-Fort Worth-Arlington, TX	-0.291	Rochester, NY	-0.491
Dayton, OH	0.923	Salt Lake City, UT	-0.149
Denver-Aurora, CO	-0.266	San Antonio, TX	0.377
Detroit-Warren-Livonia, MI	-0.739	San Diego-Carlsbad-San Marcos, CA	0.146
El Paso, TX	0.230	San Francisco-Oakland-Fremont, CA	-0.320
Fort Wayne, IN	4.897	San Jose-Sunnyvale-Santa Clara, CA	-0.113
Fresno, CA		Sarasota-Bradenton-Venice, FL	0.336
Grand Rapids-Wyoming, MI	-0.041	Scranton-Wilkes-Barre-Hazleton, PA	
Greensboro-High Point, NC	0.459	Seattle-Tacoma-Bellevue, WA	-0.245
Greenville-Mauldin-Easley, SC	-0.003	Springfield, MA	-0.088
Harrisburg-Carlisle, PA	-0.134	St. Louis, MO-IL	-0.298
Hartford-West Hartford-East Hartford, CT	-0.304	Stockton, CA	0.048
Houston-Sugar Land-Baytown, TX	-0.268	Syracuse, NY	-0.224
Indianapolis-Carmel, IN	1.110	Tampa-St. Petersburg-Clearwater, FL	-0.702
Jacksonville, FL	-0.415	Toledo, OH	-0.188
Kansas City, MO-KS	0.380	Tucson, AZ	0.089
Knoxville, TN	-0.425	Tulsa, OK	0.556
Las Vegas-Paradise, NV	-0.755	Vallejo-Fairfield, CA	
Little Rock-North Little Rock-Conway, AR	0.138	Virginia Beach-Norfolk-Newport News, VA-NC	-0.297
Los Angeles-Long Beach-Santa Ana, CA	-0.487	Washington-Arlington-Alexandria, DC-VA-MD-WV	-0.497
Louisville-Jefferson County, KY-IN	-0.159	Wichita, KS	2.357
McAllen-Edinburg-Mission, TX		Youngstown-Warren-Boardman, OH-PA	0.121

Note: This table reports the residual of the housing supply elasticity from [3], obtained by regressing it on its two main components: the MSA geographic/physical constraints and the regulatory constraints. MSA without a value of the residual housing supply are not included in our database since no firm from Compustat is located in that MSA area.

Figure A.7: Residual Supply Elasticity by MSA



Note: This figure shows the residual real estate supply elasticity by MSA. The residual of the supply elasticity is obtained by regressing the real estate supply elasticity on its two main components: the MSA geographic/physical constraints and the MSA regulatory constraints, as reported in [3].

A.4 Estimation of firms' productivity

A possible threat to our instrumental variable approach is the non-random location of firms in space [?, e.g.,] behrens2014productive,gaubert2018firm. If firms with different levels of productivity sort into metropolitan areas according to the potential for density increase, our measure of PDI would simply reflect the differences in productivity on firms' excess returns, rather than the effect of city characteristics on firms' performance in the stock market. To address this concern, we estimate the productivity of firms and include it as an additional control in our estimating equation. In this way, we assess whether firms' sorting explains our main result or whether our measure of PDI remains statistically significant after controlling for productivity.

We estimate firm's productivity as the total factor productivity (TFP), which relates to the overall effectiveness with which capital and labor are used in the production process.⁶ Theoretically, the firm's production function can be described as:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \nu_{it} \quad (1)$$

where y_{it} is the logarithm of value added of firm i in period t ; l_{it} and k_{it} are the logarithm of firms' labor and capital, respectively; ω_{it} is productivity; and ν_{it} is the error term. Empirically, the productivity term ω_{it} cannot be observed and needs to be estimated. We employ the semi-parametric procedure suggested by [6] to estimate the relation between a firm's value added and its labor and capital.⁷ Once we estimate the parameters of the production function $\hat{\beta}_0$, $\hat{\beta}_l$, and $\hat{\beta}_k$, we can obtain the TFP at the firm-year level as the residual:⁸

$$\hat{\omega}_{it} = y_{it} - \hat{\beta}_0 - \hat{\beta}_l l_{it} - \hat{\beta}_k k_{it} \quad (2)$$

We use Compustat firm-level data to estimate firm productivity.⁹ Value added is computed as revenues net of the cost of materials. The cost of materials is calculated as total expenses minus labor expenses or, equivalently, as revenues minus operating income before depreciation and minus total staff expenses. Therefore, we can calculate the value added simply as the sum of operating income before depreciation and total staff expenses [10]. We deflate this measure of value added using the US GDP implicit price deflator.¹⁰ The stock of labor at the firm-year level is calculated as the number of employees. Capital stock is calculated as the total value of property, plant, and

⁶Productivity, as measured by Total Factor Productivity (TFP), offers a more comprehensive assessment of a company's performance than other traditional metrics, such as labor productivity or firm profitability [?, e.g.,] lieberman2008measure.

⁷Other approaches can also be implemented to estimate ω (for instance, see [7, 6] for an approach based on the generalized method of moments; [8] for an approach based on intermediate consumption). However, the main advantage of the semi-parametric approach is that it solves some econometric issues related to the estimation of the production function [6]. First, input choices are likely to be correlated with producer productivity, as more efficient producers –those with high productivity– are, all else equal, likely to hire more (and better) workers and use more capital [9]. The semi-parametric estimation controls for selection and simultaneity biases. Second, there is the possible selection bias when using a panel database, since less efficient producers –those with low productivity– are more likely to exit from the sample. The semi-parametric approach also deals with the within-firm serial correlation in productivity.

⁸To obtain a more precise result, we estimate productivity also including industry-year fixed effects in the model, as in [10].

⁹The list of Compustat variables used to calculate firm-level measures is reported in Table A.2

¹⁰GDP Implicit Price Deflator in the United States obtained from the FRED Economic Data, St. Louis Federal Reserve. Retrieved in July 2023.

equipment, deflated using the private domestic investment price deflator [11].¹¹ We take the natural logarithm of value added, stock of labor and stock of capital, and estimate productivity according to equations (1) and (2).¹² In the estimation, we also include industry-year fixed effects to net out the effect of industry-wide, year-specific shocks.

The summary statistics of our estimated productivity measure are reported in Table A.3; the correlation between the firm’s productivity and other firm-level variables is reported in Table A.11. Moreover, Figure A.8 shows the relationship between PDI and firms’ productivity in the raw data, which does not suggest a clear pattern between productivity and location characteristics. The correlation between the two variables is -0.032.

In Section 4.2 of the main text, we include the estimated firm’s productivity measure as an additional control in the regression, to address the possible concern that our main result is driven by the sorting of firms into MSA with different potential density increase according to their productivity level. Columns 9–10 of Table 5 report the results. The magnitude of the PDI coefficient declines relative to the baseline specification. This pattern suggests that some sorting of more productive firms into urban areas with growth-related characteristics may be taking place. However, the relationship between PDI and firms’ stock excess returns remains negative and statistically significant after controlling for firm productivity. This finding indicates that our main result is not fully driven by productivity-based sorting. It thus supports the validity of our IV strategy.

¹¹Gross private domestic investment (implicit price deflator) obtained from the FRED Economic Data, St. Louis Federal Reserve. Retrieved in July 2023.

¹²In cases of negative value-added, logarithms cannot be defined, and thus these observations are not included [12].

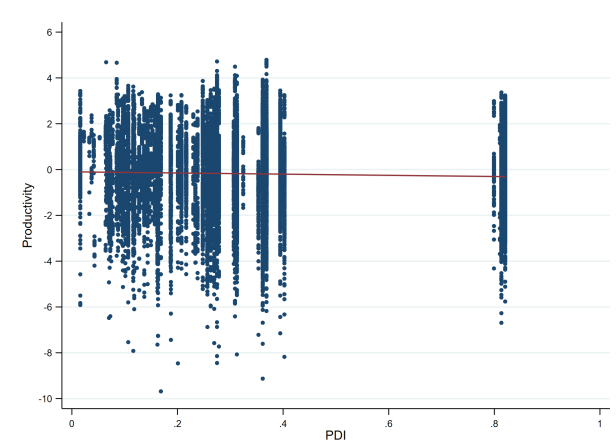
Table A.11: Correlation matrix – Firm productivity and other firm-level variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Productivity	1														
(2) Value added	0.953	1													
(3) Capital stock	0.817	0.859	1												
(4) Labor stock	0.633	0.797	0.847	1											
(5) Excess returns	0.007	-0.004	-0.013	-0.003	1										
(6) Real estate ratio	-0.196	-0.159	-0.140	-0.078	0.030	1									
(7) Leverage	0.308	0.297	0.345	0.251	0.034	-0.060	1								
(8) Market capitaliz.	0.821	0.884	0.796	0.765	-0.013	-0.131	0.207	1							
(9) Market seniority	0.256	0.277	0.429	0.376	-0.006	0.008	0.031	0.304	1						
(10) R&D expendit.	-0.084	-0.097	-0.186	-0.160	0.007	-0.018	-0.050	-0.058	-0.091	1					
(11) ROE	0.018	0.019	0.030	0.036	0.006	-0.006	-0.002	0.036	0.018	-0.004	1				
(12) ROA	0.253	0.252	0.396	0.423	-0.020	-0.080	0.028	0.352	0.226	-0.217	0.051	1			
(13) Mkt-to-book value	-0.015	-0.012	-0.011	-0.007	0.029	-0.001	-0.015	0.001	0.004	0.004	-0.330	-0.005	1		
(14) Asset growth	0.016	0.014	-0.046	-0.021	-0.031	-0.059	0.005	0.083	-0.080	0.050	0.000	0.090	0.003	1	
(15) Fixed invest.	0.071	0.059	0.085	0.037	0.005	-0.037	0.049	0.070	-0.051	-0.011	0.001	0.027	-0.002	0.021	1

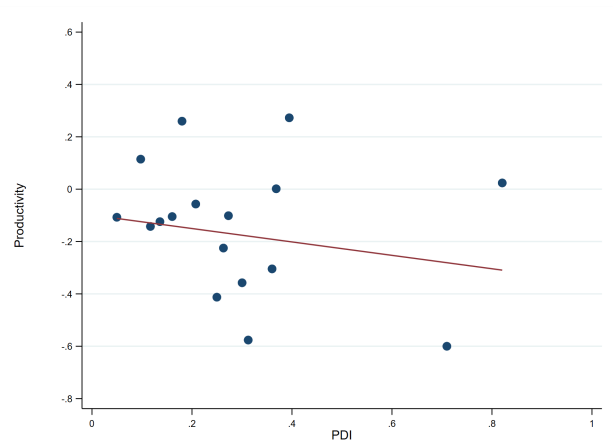
Note: This table reports the correlation among firm's productivity and other firm-level variables.

Figure A.8: Correlation between firms' productivity and PDI

(a) All observations



(b) Binned scatterplot



Note: These figures show relationship between PDI and firms' productivity, in the raw data. Panel A.8a shows all the observations, while Panel A.8b shows the binned scatterplot. The correlation between the two variables in the raw data is -0.0199.

A.5 Additional figures

This section includes some figures that provide useful additional information to complement the narrative in the main text.

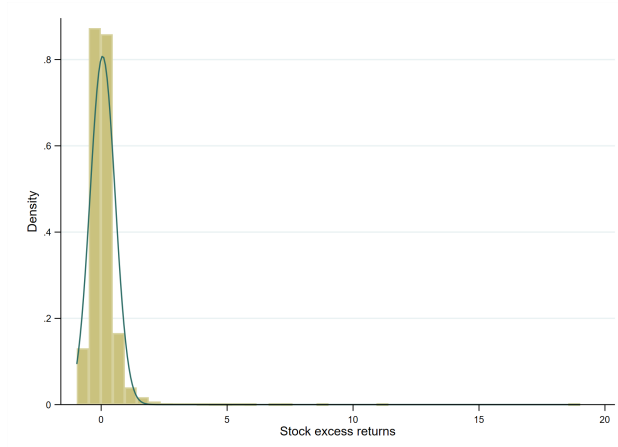
Figure A.9 shows the distribution of the firms' excess returns in the stock market in the full sample and in a sub-sample that excludes the top 1% of the distribution of the excess returns distribution. In all our empirical models, we trim the excess returns distribution to the 99th percentile.

Figure A.10 depicts the time-series of the average 30-year mortgage rate in the United States, between January 2000 and December 2010, as well as the trend line during this period. This figure supports the logic of the first-stage result of the instrumental variable (IV) approach.

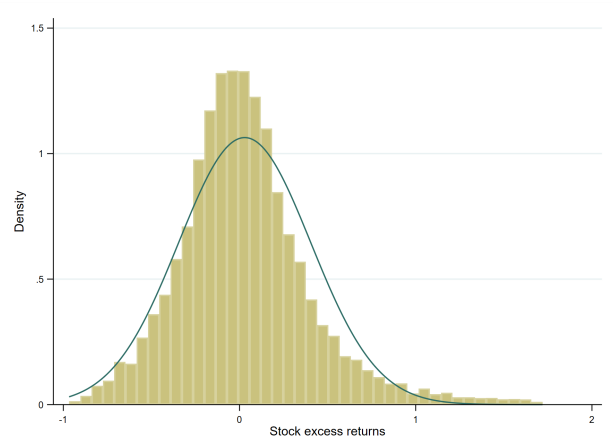
Figure A.11 shows the distribution of the firms' excess returns in the stock market for the firms included in the sub-samples we use in the second part of the empirical analysis (large versus small firms; senior and junior firms in the stock market; by industrial sector; firms with high versus low reliance on physical capital).

Figure A.9: Distribution of firms' stock returns in the data (I)

(a) All observations

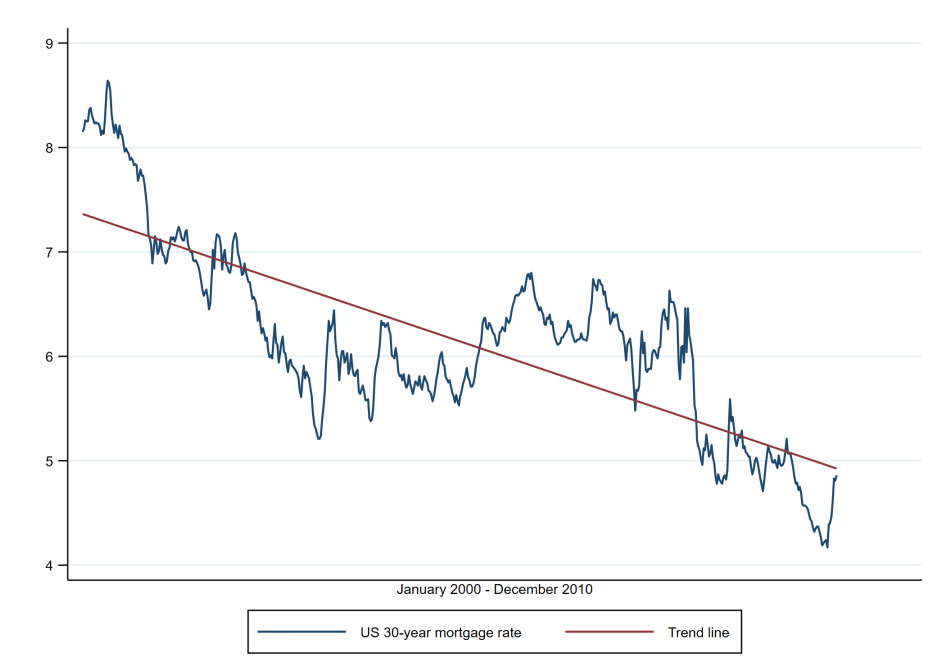


(b) Excluding top 1% of the distribution



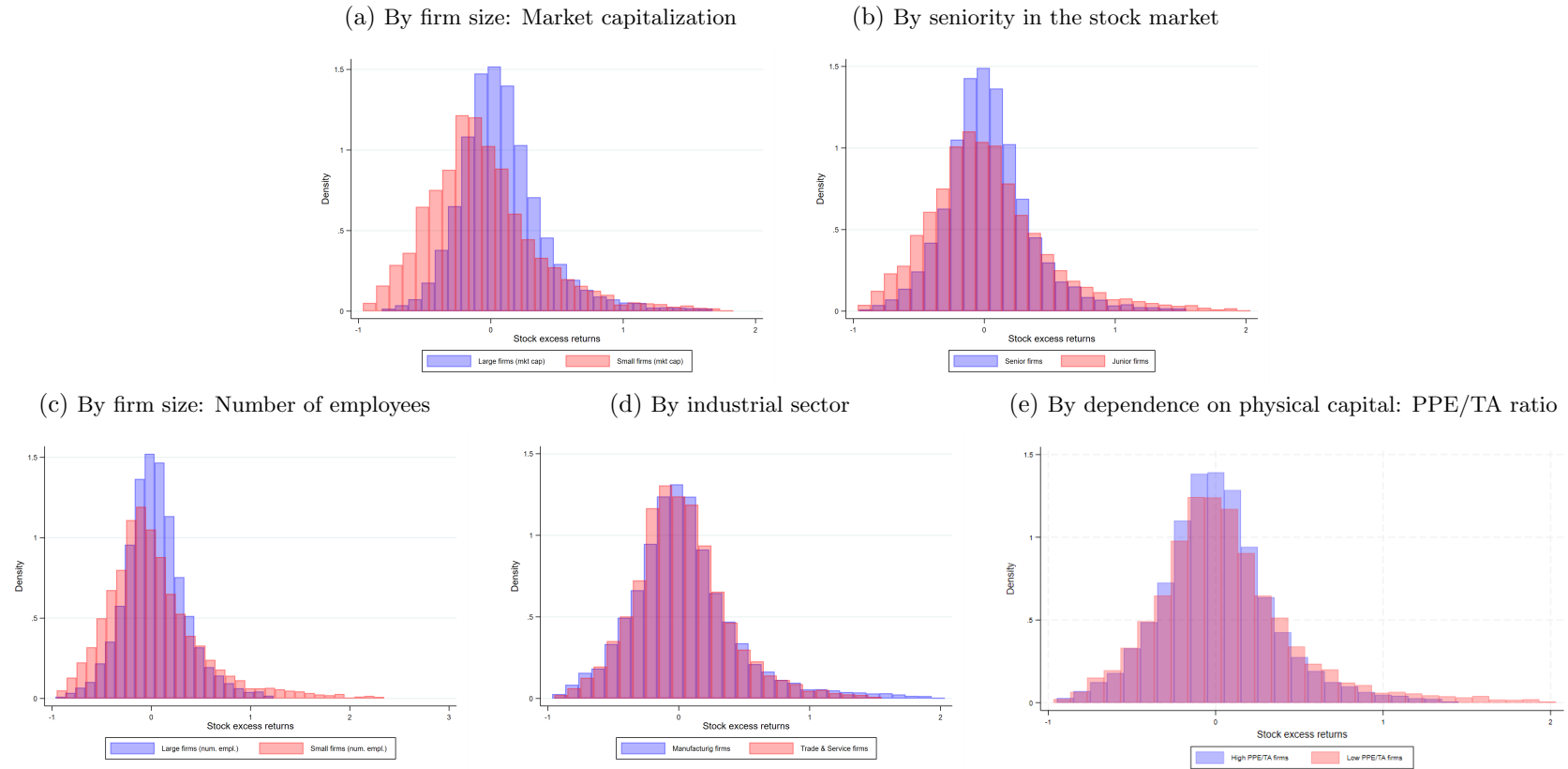
Note: This figure shows the distribution of the stock excess returns variable in the database. Figure A.9a includes all observations and shows a considerably right-skewed distribution. Figure A.9b excludes the top 1% of the distribution of the distribution and shows an approximately normal distribution. In the estimation, we always include excess returns up to the 99th percentile of the distribution.

Figure A.10: Average 30-year mortgage rate in the United States, 2000-2010



Note: This figure shows the evolution of the average 30-year mortgage rate in the United States between 2000 and 2010. The red line indicates the trend line across the period.

Figure A.11: Distribution of firms' stock returns in the data – Sub-samples



Note: This figure shows the distribution of the stock excess returns variable in the database, by sub-samples. Figure A.11a shows the distribution by firm size defined as market capitalization, distinguishing between large and small firms (i.e., above and below the median market capitalization). Figure A.11b shows the distribution by firm seniority in the stock market, distinguishing between senior and junior firms (i.e., above and below the median average years of seniority in the stock market during the period under study). Figure A.11c shows the distribution by firm size defined as number of employees, distinguishing between large and small firms (i.e., above and below the median number of employees). Figure A.11d shows the distribution of the excess returns by industrial sector, distinguishing between manufacturing and trade & services firms. Figure A.11e shows the distribution by firm dependence on physical capital, proxied by the PPE/TA ratio. All excess returns are reported, excluding the top 1% of the distribution of the distribution in the corresponding sub-sample.

A.6 Additional results and robustness checks

This section reports additional results, and robustness checks that complement the main tables and discussion in the article. We also provide a more detailed discussion on the magnitude of some key results reported in the main text, under the robustness checks (Section 4.2) and the heterogeneity analysis (Section 4.3).

Urbanization measures. Table A.12 shows additional results using the share of high-developed and developed land. Columns 1–2 report the results using each share in separate regressions. Overall, they confirm that the share of developed land (with potential for rapid building density increase) is more relevant for the perceived riskiness of firms in the stock market, since the effect of the developed index is statistically significant and larger in magnitude. A one-standard deviation increase in the high-developed share is associated with a decrease of 1.21 percentage points in the excess returns (not statistically significant), while an increase of the same magnitude of the developed share is associated with a statistically significant decrease of 2.04 percentage points.

In columns 3–4, we report the estimation of the model with the high-developed and developed indices when excluding the largest MSAs. When we exclude the largest cities, both coefficients are negative and statistically significant; the magnitude of the effect of the developed share is slightly higher. When excluding the three most populous cities (column 3), a one-standard-deviation increase in the high-developed share leads to a decrease in excess returns of 1.95 percentage points, while the same increase in the developed share is associated with a decrease of 2.02 percentage points. When excluding the three most densely populated cities (column 4), a one-standard-deviation increase in the high-developed share leads to a decrease in excess returns of 1.92 percentage points, while the same increase in the developed share is associated with a decrease of 2.10 percentage points. Overall, the share of developed land remains relevant in explaining the perceived risk of firms in the stock market. In a robustness check, we include the share of high-developed land as an additional control in the regression and show that our main results remain valid (see Table 5).

Columns 5–6 show the results excluding large firms in terms of market capitalization and number of employees. Our main conclusions remain valid: the share of high-developed land has a statistically insignificant effect, while the share of developed land has a negative and statistically significant effect. A one-standard-deviation increase in the developed share is associated with a decrease in excess returns of 3.37 and 2.56 percentage points, respectively.

IV estimation using residual supply elasticity as instrument. Table A.13 shows the results of employing the Residual Supply Elasticity as an instrument for PDI, instead of the regular real estate supply elasticity. This residual elasticity is obtained from regressing Housing Supply Elasticity on its two main components: the MSA-level geographic constraints (undevelopable land) and regulatory constraints (Wharton Regulation Index), as discussed in [3]¹³. The OLS and IV estimates in columns 1 and 2 are the same as in our baseline result in Table 2. The OLS estimate in column 3 is obtained from excluding the MSA-level geographic constraints (undevelopable land) and regulatory constraints (Wharton Regulation Index) as controls. A one-standard deviation in-

¹³See detailed explanation in Appendix A.3

crease in PDI is associated with a decrease in excess returns of 2.12 percentage points. The IV model in column 4 does not include the MSA-level geographic constraints (undevelopable land) and regulatory constraints (Wharton Regulation Index) as controls, and the Residual Supply Elasticity is used as an instrument for PDI. The coefficient is negative and statistically significant. A one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 6.3 percentage points. These results are consistent with our main conclusions.

Exclusion of large metropolitan areas. In columns 4–7 of Table 2, we report the OLS and IV estimation results excluding the three most populous US MSAs (New York, Los Angeles, and Chicago) and the three most densely populated US MSAs (Los Angeles, San Francisco, and San Jose). Excluding the three most populous cities, a one-standard-deviation increase in PDI is associated with a decrease in excess returns of 2.21 percentage points (column 4); excluding the three most densely populated cities, a one-standard-deviation increase in PDI is associated with a decrease in excess returns of 2.29 percentage points (column 6). The instrument is valid in both cases, with a statistically significant negative effect of housing supply elasticity on PDI in the first stage, with a partial F-statistics above the threshold of 10. Excluding the three most populous cities, a one-standard-deviation increase in PDI leads to a decrease in excess stock returns of 3.19 percentage points in the second stage (column 5); excluding the three most densely populated cities, a one-standard-deviation increase in PDI leads to a 4.79 percentage point decrease in excess returns (column 7).

Figure A.12 shows the estimated coefficients of the PDI measure excluding one large city at the time (OLS in Panel A.12a and IV in Panel A.12b). In both panels, the first estimate is for the full sample, including all available MSAs (columns 2–3 of Table 2). The second and third coefficients are estimated excluding the three most populous and three most densely populated MSAs in the US, respectively (columns 4–7 of Table 2). Each of the following coefficients comes from a separate regression on the sub-sample that excluded the corresponding reported city. We exclude the ten most populous MSAs in the US: Atlanta, Boston, Chicago, Dallas, Houston, Los Angeles, Miami, New York, Philadelphia, and Washington. We also exclude San Francisco and San Jose, which are the most densely populated MSAs together with Los Angeles.¹⁴ The coefficients are stable across the different sample specifications and the main conclusions of Table 2 are confirmed even when we do not consider the major metropolitan areas.

Exclusion of large firms. In columns 1–4 of Table 3, we report the estimation results when we exclude the largest firms. In particular, we consider both the market capitalization and the number of employees and exclude the firms in the top 25th percentile of the respective firm size distribution. Focusing on OLS estimates, excluding the largest firms in terms of market capitalization, a one-standard-deviation increase in PDI leads to a decrease in excess stock returns of 2.77 percentage points (column 1); excluding the largest firms in terms of employees, a one-standard-deviation increase in PDI leads to a 2.49 percentage point decrease in excess returns (column 3). IV estimates are larger in magnitude than our main results; a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 6.76 and 6.74 percentage points, focusing on firms size in terms of market capitalization and labor stock, respectively (columns 2 and 4). The

¹⁴The list is the same in 2010 and in 2020, although the positions in the ranking change over time. Source: US Census Bureau. Retrieved in July 2023. These are also the cities most represented in our data (see Appendix Table A.8).

larger magnitude of the results and the more pronounced biased-toward-zero of the OLS estimates when we exclude large firms suggest that our relationship of interest –between potential density growth and perceived firm risk in the stock market– is more relevant for medium and small firms. Moreover, the OLS estimates show an increased bias towards zero when we omit the largest firms, as the possible rise in input costs due to density growth is likely more pronounced for medium- and small-sized firms.

Proxy for PDI of multi-establishment firms. In columns 5–8 of Table 3, we report the estimation results using the measures of “average PDI” described in Section 4.2. This approach attempts to address one of the main limitations of Compustat data: the lack of information regarding firms’ location of branches and establishments, beyond the headquarters.

Results in columns 5–6 are obtained from re-estimating our main regressions using the firm-level average PDI and housing supply elasticity. The OLS coefficient is negative but statistically indistinguishable from zero (column 5). The coefficient of the IV estimate is negative and statistically significant at the 10% level. The result indicates that a one-standard-deviation increase in firms’ average PDI leads to a decrease in excess stock returns of 1.65 percentage points (column 6). These measures are constructed as averages across MSAs and are therefore likely subject to measurement error. As a result, the relationship between the firm’s average PDI and excess returns is estimated less accurately than in the baseline model.

Results in columns 7–8 show that our main conclusions are robust to controlling for broader geographic exposure. Here, we estimate the baseline model, in which PDI is measured at the firm’s headquarters location, and include the firm’s average PDI as an additional control. Particularly, the OLS estimation shows that a one-standard-deviation increase in PDI leads to a decrease in excess stock returns of 1.68 percentage points (column 7), whilst the IV estimation shows that the same increase in PDI leads to a decrease in excess stock returns of 4.93 percentage points (column 8). The coefficient for average PDI are statistically insignificant for both estimations. These results confirm the validity of our main conclusions.

Alternative time frames. In Table 4, we report the results of the analysis using different time frames for firm-level financial data to account. Results in columns 1–4 are estimated using a later starting period of time to ensure that our results are not affected by the final phases of the 2008 financial crisis. In columns 1–2 we use firms data between 2012 and 2019, while in columns 3–4 we use firm data between 2013 and 2019. All the estimated coefficients are consistent with our baseline results in terms of sign, magnitude, and statistical significance. The OLS estimations show that a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 1.87 and 1.70 percentage points, respectively (columns 1 and 3). The IV estimation shows that a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 4.33 and 5.56 percentage points, respectively (columns 2 and 4).

In columns 5–8 of Table 4, we instead use firm-level data for different time frames around 2015—the year of our urban density measures. In columns 5–6, we consider a shorter window, with firm data between 2012 and 2017 (3-year window around 2015). In columns 7–8, we consider a longer window, with firm data between 2005 and 2024 (10-year window around 2015). The OLS results show that a one-standard-deviation increase in PDI is associated with a decrease in excess stock

returns of 1.99 and 1.45 percentage points, respectively (columns 5 and 7). For comparison, in our baseline OLS results using firm-level data between 2010 and 2019, we find that a one-standard-deviation increase in PDI is related to a decrease in firm excess returns on the stock market by 1.80 percentage points (column 2 of Table 2). The IV estimation shows that a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 5.84 and 4.98 percentage points, respectively (columns 6 and 8). For comparison, in our baseline IV results using firm-level data between 2010 and 2019, we find that a one-standard-deviation increase in PDI is related to a decrease in firm excess returns on the stock market by 4.57 percentage points (column 3 of Table 2).

The coefficients remain significant across these different time frames, confirming the validity of the main conclusions and the choice of time frame. The results are consistent with our explanation: we find stronger effects for the shorter time window, in which we use financial data closer to the date of our geographical measures. Results are instead weaker for the longer time window, when we also include financial data further away in time from our geographical measures. These findings confirm that our conclusions are not sensitive to the specific sample window and support our baseline choice of a time frame for firm-level data from 2010 to 2019 (5-year window around 2015).

Placebo test. In columns 9–10 of Table 4, we report the results of the placebo test, in which we employ firm-level data for a time period between 2001 and 2009. This time window is before 2015—the year we built our urban density measures. This placebo test helps account for possible alternative drivers of the result, such as reverse causality or long-term trends. The coefficients are indistinguishable from zero in both the OLS and IV estimates, supporting our general conclusions.

Additional controls. In Table 5 we show that our main conclusions remain valid when we include additional controls to account for potential violations of the exclusion restriction of the IV model.

In columns 1–2, we include the MSA population in the year 2000 as an additional control to account for lasting differences in the size of the metropolitan areas and underlying features that historically could lead to the concentration of population in specific areas. The sign, magnitude and significance remain similar to our main results. In particular, when we include the MSA population in the year 2000, one-standard-deviation increase in PDI is associated with a decrease in excess returns of 2.14 percentage points from OLS, and of 4.71 percentage points from IV.

A possible violation could occur if the real estate supply elasticity affects not only the share of semi-urban area, but also the fully constructed one. In columns 3–4 of Table 5, we include in the regression the high-developed index. The significance and magnitude of the results are comparable to the main estimation. In particular, when we include the high-developed index to account for the proportion of land that is already completely developed and where it is not feasible to increase the building density on the extensive margin, a one-standard-deviation increase in PDI is associated with a decrease in excess returns of 1.78 percentage points from OLS, and of 4.89 percentage points from IV.

Another violation could arise if the local real estate supply elasticity affects housing prices in a way that could impact the excess returns of firms beyond the real-estate ratio we control for in the regression. In columns 5–6 of Table 5, we show the estimation results obtained by

including the annual percentage change in the MSA-level house price index as an additional control. The magnitude of the OLS estimate is slightly higher than the main result, with a one-standard-deviation increase in PDI leading to a 2.06 percentage point decrease in excess returns. However, the bias-toward-zero of the OLS result is attenuated in this case, since the IV result suggests that a one-standard-deviation increase in PDI is associated with a decrease in stock excess returns of 3.75 percentage points. This finding further bolsters our argument that the OLS estimate might reflect some impacts related to the possible increase in local building density, which increases the perceived risk for firms, thus diminishing the role of land availability in reducing this perceived risk.

In columns 7–8 of Table 5, we include the production of patents per MSA as a control, reflecting the R&D activity at the MSA level. The results show that, for the OLS estimation, a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 1.82 percentage points, whilst the IV estimation gives a decrease of 4.54 percentage points with respect to the same increase in PDI. These results confirm the validity of our baseline results.

In columns 9–10 of Table 5, we include the estimated firm productivity in the regression model to address the concern of the potential sorting of firms into urban areas according to urban characteristics that would confound our relationship of interest.¹⁵ The magnitude of these results is nearly half of our main estimate, indicating that the distribution of firms across metropolitan regions with varying potential for density increase based on their productivity is relevant. However, the estimated coefficient of PDI remains negative and statistically significant, suggesting that the local potential increase in building density plays a significant role in the perceived risk of firms in the stock market, beyond differences in productivity. In particular, the OLS result suggests that a one-standard-deviation increase in PDI is associated with a decrease in firms' excess returns of 0.88 percentage points (which corresponds to 2.35% of the standard deviation of the excess returns variable), when accounting for firms' productivity. The IV estimated coefficient implies that, controlling for productivity, a one-standard-deviation increase in PDI leads to a 2.95 percentage points decrease in stock excess returns (which corresponds to 7.86% of the standard deviation of the excess returns variable). The magnitude of the estimated coefficient of PDI is lower than our main results in Table 2, but it remains negative and statistically significant.

In Table A.14, we report the results that include additional pairs of these controls in the estimation. Our conclusions from the main text are confirmed.

The results in Table A.15, derive from estimating a model which includes firm's growth in revenue and EBITDA as additional controls. Columns 1–2 show the results when we control for revenue growth. The OLS results show that a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 1.92 percentage points (column 1). The IV result shows that a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 4.56 percentage points (column 2). Columns 3–4 report the results when we control for EBITDA growth. The OLS coefficient shows that a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 1.89 percentage points (column 3), and the IV coefficient shows that the same increase in PDI is associated with a decrease in excess stock returns of 4.47 percentage points (column 4). In columns 5–6, the estimation model includes both revenue and EBITDA growth. The OLS and IV results show that a one-standard-deviation increase in PDI is related to a decrease in excess returns of 1.89 and 4.47 percentage points, respectively. Overall,

¹⁵For details of firm productivity estimation, see Appendix A.4

when controlling for revenue and EBITDA growth, the estimated coefficients for PDI remain negative, statistically significant, and with a similar magnitude than our baseline results, confirming the validity of our conclusions.

Sensitivity to polygon definition. As discussed in Section 2.1, to calculate our building density measures, we define a one-hour drive polygon around the city center of the main city of the MSA. We chose this cut-off since it reflects the fact that the average willingness to drive for commuting reasons depends on the average time, rather than the average distance, and one-hour is commonly assumed as the standard time people are willing to commute between residence and workplace [13, 14, 15]. To ensure that our results are not sensitive to this definition, we conduct a systematic set of counterfactual scenarios that mimic plausible changes in land composition resulting from smaller or larger catchment areas around the center of the MSA.

First, we increased and decreased the total original polygon area by 10% and 15%. For each of these four alternative scenarios, we then calculated the amount of high-developed (HD), developed (D), and low-developed (LD) areas obtained by considering a set of assumptions. Across the different scenarios, HD land is assumed to be 0%, 1% and 2.5% of the change in total area; D is assumed to be 0%, 5%, 10%, 15%, 20%, 25%, 30%, and 50% of the change in total land; and LD is calculated as the remaining percentage of the change in total area. In other words, percentages are expressed relative to the change in total land and sum to 100% in all scenarios. Combining these assumptions, we obtain 24 different distributions of land types for each of the four alternative scenarios, for a total of 96 (see Table A.16).

We then re-estimate our main OLS and IV specifications using the “simulated PDI” under each of these 96 scenarios. A summary of the results from this exercise is reported in Figures A.13–A.14. Across all cases, our main conclusions are confirmed: the sign of the coefficients is stable, statistical significance is preserved, and the coefficient magnitudes present little variation. Overall, the results of this simulation indicate that our findings are not sensitive to reasonable deviations from the one-hour driving radius assumption.

Alternative ratios of building density. In Table A.17 we report the OLS relationship between alternative ratios of building density and firms’ excess returns. These ratios of building density are based on our definition of types of land: high-developed (HD), developed (D), and low-developed (LD). The main differentiation we make between developed (D) and low-developed (LD) is the pace at which building density increase can occur on the two types of land. Low-developed (LD) areas represent the part of the land that is sparsely built. Thus, this land is fully available for development, but the increase in building density is likely to be slow due to the lack of urban infrastructure. Instead, developed areas (D) of the city offer greater potential for rapid growth in urban density due to existing infrastructure and the availability of land for further development, compared with low-developed areas (LD).¹⁶

In column 1, we report our baseline results from Table 2, using our definition of PDI ($D/D+LD$). Our definition of PDI is intended to measure the proportion of developable land that can be developed (i.e., where the density of buildings can increase) relatively fast in the near future. Our baseline result is that a one-standard-deviation increase in PDI is associated with a decrease in

¹⁶See detailed definitions in Section 2.1.

firms' excess returns of 1.80 percentage points. In column 2, we use the share of low-developed land over the total polygon excluding geographically constrained land ($LD/HD + D + LD$). Since this measure reflects the proportion of land in the MSA that can be developed slowly, we find that a one-standard-deviation increase in this measure is associated with an increase in excess returns of 2.79 percentage points. In column 3, we use the share of low-developed land over the developable area ($LD/D + LD$), which corresponds to $1 - PDI$. Therefore, the result is the opposite of that for PDI, with a one-standard-deviation increase in PDI associated with an increase in excess returns of 1.80 percentage points. In column 4, we use the share of developed land over the sum of developed and high-developed land ($D/D + HD$). If we consider that high-developed and developed areas are those types of land with urban development, this accounts for the proportion of developed land where building density can still increase. This can be interpreted as the *current* status of urban development, but is not related to the *potential* for fast urban development in the near future (in other words, this should be related to the density increase in the past). Consistent with this argument, we find a not statistically significant relationship between the measure and firms' excess returns.

Exclusion of geographically constrained MSAs. Table A.18 shows results from estimating our model on a sample that excludes the most geographically constrained MSAs. To define the extent of the geographic constraints, we consider the share of the MSA polygon that is “undevelopable”. This measure is obtained from our land classification method and accounts for areas where construction is totally restricted, such as water bodies, natural reserves, city parks, and steeply sloped terrain.¹⁷ In columns 1–2, we exclude the MSA with a share of undevelopable land in the top 25-th percentile of the distribution. The OLS result shows that a one-standard-deviation increase in PDI is associated with a decrease in excess returns of 2.88 percentage points (column 1). The IV result shows that a one-standard-deviation increase in PDI is linked to a decrease in excess returns of 10.05 percentage points (column 2). In columns 3–4, we exclude the MSA with a share of undevelopable land in the top 10-th percentile of the distribution. From the OLS estimate, a one-standard-deviation increase in PDI is associated with a decrease in excess returns of 3.28 percentage points (column 3). From the IV estimate, a one-standard-deviation increase in PDI is associated with a decrease in excess returns of 7.75 percentage points (column 4). In columns 5–6, we exclude the MSA with a share of undevelopable land in the top 5-th percentile of the distribution. The OLS result shows that a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 1.94 percentage points (column 5). The IV result shows that a one-standard-deviation increase in PDI is associated with a decrease in excess stock returns of 4.72 percentage points (column 6). Overall, these results support our baseline conclusions, suggesting that the results are not driven by areas with highly constrained geography.

Alternative dependent variables. We re-estimate Equations (2) and (3) using alternative definitions of excess returns. First, we follow a more naive approach by defining the excess return as the simple difference between the firm's monthly stock return and the risk-free rate of the market. Second, we calculate the excess returns as residuals according to equation 1, but using additional Fama & French factors.¹⁸ The estimation results of these alternative specifications are reported in Table A.19 and are consistent with our main results in Table 2.

¹⁷See detailed definitions in Section 2.1.

¹⁸The detailed definitions are reported in Table A.2.

We also re-estimate the statistical models using the stock return, defined as the year-to-year percentage change in the average annual stock price of the firm. Although a positive relationship between risk and return lies at the heart of finance theory, this measure does not consider systemic risk, average market returns, or the influence of common factors. Therefore, simple stock returns are a less sophisticated measure of firm risk than excess returns. Results are reported in the first three columns in Table A.20. In the OLS estimation, a one-standard-deviation change in PDI is associated with a decrease in the stock returns of 1.08 percentage points, which is lower than our main results using excess returns. The effect of the IV estimate for PDI is higher than the one from the main results; a one-standard-deviation change in PDI is associated with a decrease in the stock returns of 5.15 percentage points. Our instrumental variable approach isolates the source of variation in the potential increase in building density that comes from land availability. This suggests a direct effect of the potential intensification of construction at the MSA level and the perceived risk of firms located in the urban area (conditional on a variety of firm and city controls). However, there are systemic and contextual factors that might confound the effect in the OLS estimation, biasing the estimated coefficient toward zero. This is consistent with the difference in magnitude of the estimated coefficient reported in the main results in Table 2. Overall, this confirms that the excess return is a more precise measure of the investors' perceived risk towards the firm.

Finally, to ensure that our results do not depend on differences in stock prices among firms, we also re-estimate our statistical models using the annual average stock price and the annual price volatility as dependent variables. Results are reported in columns 4–9 of Table A.20 and suggest that PDI does not have a statistically significant effect on stock prices, which do not necessarily reflect the investors' perception of the firm's risk. The IV estimate of the effect of PDI on price volatility is slightly statistically significant. This is consistent with the fact that higher stock price volatility is related to a greater degree of risk and potential losses. However, our measure of stock excess returns is more precise in capturing the stock risk.

Alternative fixed-effects. Our estimation model includes fixed effects for the state, industry sector, and year. This means that we are comparing firms located in the same US state and operating in the same industrial sector, while controlling for annual shocks that affect the whole economy. However, some shocks could have a different effect on different states or industrial sectors. To capture this, we can include in the model year-industry and year-state fixed effects. In Tables A.21 and A.22, we report our main results, re-estimated progressively including more restrictive fixed effects, and the conclusions of Table 2 remain valid.

Alternative clustering of the standard errors. In Table A.23, we show the results obtained by clustering the standard errors at the level of MSA-by-industry, where industry is defined as the 4-digit SIC code. This alternative clustering raises the number of clusters from 81 to 1,462. The statistical significance of the estimated coefficients is maintained.

Excluding MSAs with the highest PDI variation. As described in detail in Appendix A.1, our image segmentation algorithm implies some slight margin of error in the definition of the categories of types of land. To minimize the margin of error, we run the algorithm five times on each input image and our measure of PDI is the average of the PDI measure calculated using the

area of developed and low-developed land from each iteration of the algorithm. To ensure that our main results are not affected by the variation in the PDI measure, in Table A.24 we replicate the main results excluding the observations in the MSAs with the highest variation in the PDI measures. Results are stable and similar in magnitude, thus the conclusions of Table 2 remain valid.

Heterogeneity analysis: Market capitalization. In columns 1–4 of Table 6, we evaluate the relationship between the MSA potential density increase and firm stock excess returns, dividing the sample between large and small firms in terms of their market capitalization.¹⁹ For large firms, the effect of PDI on excess returns is statistically indistinguishable from zero. For small firms, we confirm a negative and statistically significant effect. For the latter, according to the OLS estimate, a one-standard-deviation increase in PDI is associated with a decrease in excess returns of 5.73 percentage points, while the IV estimate is larger in magnitude, with the same change in PDI associated with a 10.23 percentage point decrease in excess returns. As expected, these results suggest that the urban characteristics of cities play a role only in the evaluation of the risk of small firms.

Heterogeneity analysis: Seniority in the stock market. In columns 5–8 of Table 6, we report the results by dividing the sample according to firms’ seniority in the stock market.²⁰ For senior firms in the stock market, we find a statistically insignificant relationship between PDI and stock excess returns. For firms more recently listed on the stock market, we find a statistically significant and negative effect of PDI on firms’ excess returns. According to the OLS estimate, a one-standard-deviation increase in PDI is associated with a 7.21 percentage point decrease in excess returns, while the magnitude of the effect of the IV estimate is larger, and the same increase in PDI is related to an 11.52-percentage-point decrease in excess returns. As anticipated, these findings validate that the MSA’s potential to enhance building density relatively fast is significant for companies that are newer to the stock market, where factors supporting their future expansion are vital.

Heterogeneity analysis: Number of employees. In columns 1–4 of Table 7, we report the results for firms divided according to their number of employees.²¹ For large firms (in terms of labor stock), the OLS estimate is statistically indistinguishable from zero. The IV estimate is statistically significant and similar in magnitude to our main results; a one-standard-deviation increase in PDI corresponds to a decrease in excess returns of 4.89 percentage points. For small firms, the estimated effects are statistically significant (regardless of the estimation strategy) and larger in magnitude. The OLS estimate suggests that a one-standard-deviation increase in PDI relates to a 4.96 percentage point decrease in excess returns in the stock market. The IV estimate associates the same increase in PDI with a decrease in excess returns of 10.08 percentage points. These effects are very similar to those estimated for small firms in terms of market capitalization, confirming that the potential increase in building density at the MSA level is more relevant for decreasing the

¹⁹We divide the sample according to the firms’ market capitalization, in the year we first observe the firm in our sample. We define a firm as *large* if its initial market capitalization is above the median and as *small* if it is below.

²⁰We define a firm as *senior* in the stock market if the years publicly traded are above the median and as *junior* if they are below. Notice that the correlation between our split based on market capitalization and that on seniority in the stock market is low (0.117). Therefore, we capture different stock market characteristics of firms.

²¹We define a firm as *large (in terms of employees/labor stock)* if the number of employees at the time of the first observation in our sample is above the median and as *small (in terms of employees/labor stock)* if it is below. The correlation between our split according to market capitalization and that based on number of employees is 0.559.

perceived risk in the stock market of smaller firms in terms of employees.

Heterogeneity analysis: Physical capital intensity. In columns 5–8 of Table 7, we report the results for firms divided according to their intensity of physical capital, calculated as the ratio of Property, Plant, and Equipment (PPE) to Total Assets (TA).²² For firms that are intensive in physical capital, the OLS estimate shows that a one-standard-deviation increase in PDI is associated with a 1.77 percentage-point decrease in excess returns. The IV results are larger in magnitude: a one-standard-deviation increase in PDI is associated with a decrease in excess returns of 12.53 percentage points. For firms with low intensity in physical capital, results are distinct. The OLS coefficient of PDI is not statistically significant. From the IV estimate, we find that a one-standard-deviation increase in PDI is associated with a 4.69 percentage-point increase in excess returns. These results suggest that the MSA’s potential to enhance building density relatively fast is significant in decreasing the perceived risk in the stock market for firms with a high intensity of physical capital. For firms with low intensity of physical capital, the MSA’s potential to enhance building density relatively fast is instead associated with an increase in perceived risk in the stock market.

Heterogeneity analysis: Industrial sector. In columns 9–12 of Table 7, we present the results by dividing the sample between *manufacturing* firms and *trade and service* firms.²³ confirm the negative effect in both sectors, but with considerable differences. On one hand, for manufacturing firms, a one-standard-deviation increase in PDI is associated with a 2.15 percentage point decrease in excess returns, according to the OLS estimate. The magnitude of the effect increases with the IV strategy, where a one-standard-deviation increase in PDI is associated with a decrease in excess returns of 5.04 percentage points. Both effects are statistically significant. On the other hand, for firms in the trade and service sectors, the OLS estimate implies that a one-standard-deviation increase in PDI by one standard deviation is related to a decrease of 4.23 percentage points in excess returns. However, the IV estimate is not statistically significant. These results suggest that the MSA’s potential to enhance building density relatively fast is significant in decreasing the perceived risk in the stock market for manufacturing firms.

²²We define a firm with *high intensity in physical capital* if its PPE-to-TA ratio at the time of the first observation in our sample is above the median and with *low intensity in physical capital* if it is below.

²³We classify firms according to the US Securities and Exchange Commission classification of SIC codes. Our split based on the industry classification is weakly correlated with the other splits we take into account for the heterogeneous effects: 0.18 with the split according to the number of employees; 0.05 with the split based on market capitalization; and 0.026 with the split based on seniority in the stock market.

Table A.12: Additional results – Urbanization measures

	Excess returns					
	Excluding:					
			3 most populous MSAs (NY, LA, CHI)	3 most densely pop. MSAs (LA, SF, SJ)	Top 25-th percent. firm size: Market capitaliz.	Top 25-th percent. firm size: Num. employees
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Main Results						
High-developed share	-0.106 (0.098)		-0.217* (0.119)	-0.241** (0.106)	-0.086 (0.135)	-0.107 (0.122)
Developed share		-0.144*** (0.035)	-0.231*** (0.037)	-0.144*** (0.037)	-0.241*** (0.063)	-0.219*** (0.052)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Clust. std. error: MSA	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	11,789	11,789	9,333	9,919	7,263	7,403
Cluster N	81	81	78	78	76	76
Dep. var. mean	0.029	0.029	0.033	0.027	-0.005	0.029
Dep. var. st.dev.	0.375	0.375	0.378	0.360	0.413	0.426

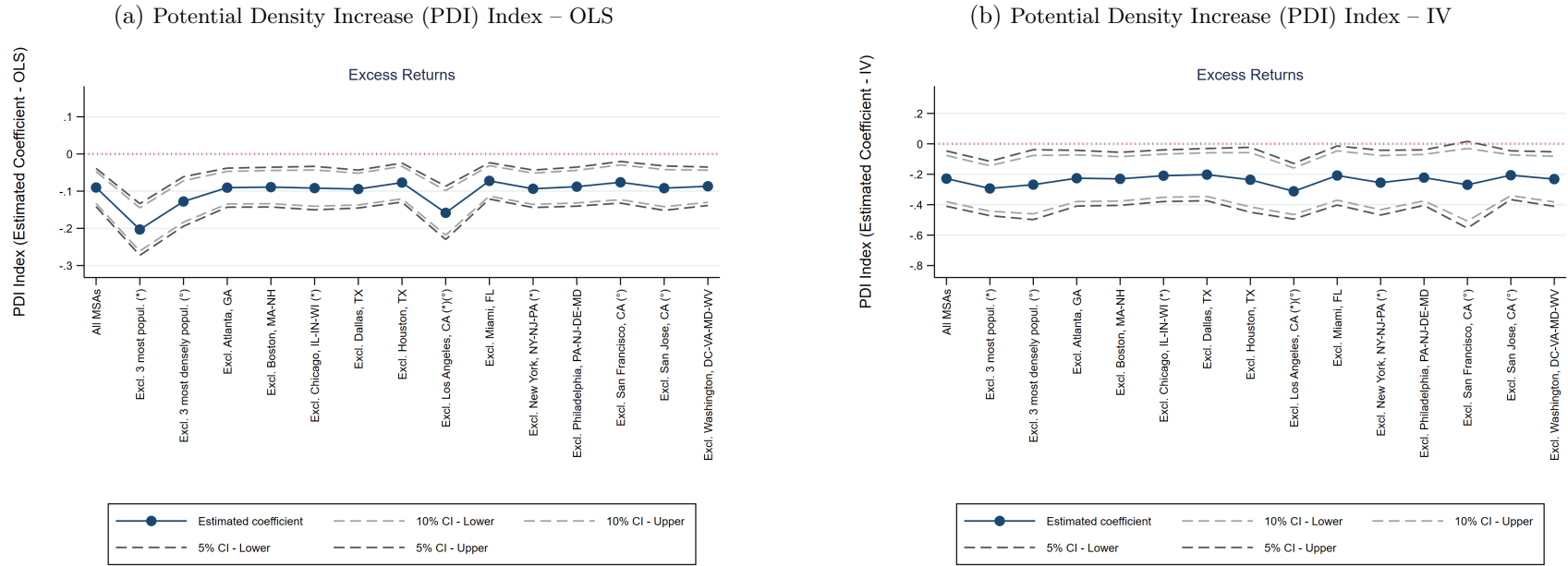
Note: This table reports additional estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors: excess market return (MKT), small-minus-big return (SMB), high-minus-low return (HML), and momentum (UMD). The sample excludes observations with excess returns in the top 1% of the distribution. In columns (1) and (2), the model is estimated using the full sample. In column (3), the sample does not include firms located in the three most populous cities in the US (both in 2010 and 2020): New York (NY), Los Angeles (CA), and Chicago (IL). In column (4), the sample does not include firms located in the three most densely populated cities in the US (both in 2010 and 2020): Los Angeles (CA), San Francisco (CA), and San Jose (CA). In column (5), the sample does not include firms with a market capitalization in the top 25-th percentile of the distribution. In column (6), the sample does not include firms with a number of employees in the top 25-th percentile of the distribution. The models include two regressors of interest: the high-developed index is the share of highly developed land over the total of buildable land; the developed index is the share of developed land over the total of buildable land. Results in columns (1)-(4) derive from ordinary least squares estimation (OLS). All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.13: Additional results – IV estimation using residual housing supply elasticity as instrument

	Excess returns			
	(1) OLS	(2) IV	(3) OLS	(4) IV
Main Results				
PDI	-0.090*** (0.026)	-0.228*** (0.091)	-0.106*** (0.027)	-0.313*** (0.099)
First Stage: PDI				
Housing Supply Elasticity		-0.157*** (0.037)		
Residual Housing Supply Elasticity				-0.146*** (0.035)
Partial F -statistics		17.858		17.881
Partial R^2		0.142		0.166
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes
Incl. components of Housing Supply Elasticity	Yes	Yes	No	No
Clust. std. error: MSA	Yes	Yes	Yes	Yes
Observations N	11,789	11,789	11,789	11,789
Cluster N	81	81	81	81
Dep. var. mean	0.046	0.046	0.046	0.046
Dep. var. st.dev.	0.369	0.369	0.369	0.369

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors: excess market return (MKT), small-minus-big return (SMB), high-minus-low return (HML), and momentum (UMD). The sample excludes observations with excess returns in the top 1% of the distribution. The regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in columns (1) and (3) derive from ordinary least squares estimation (OLS); results in columns (2) and (4) derive from instrumental variable estimation (IV). The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include baseline MSA-level controls: the one-hour drive polygon area; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Models in columns (1) and (2) also include as MSA-level controls the components of the housing supply elasticity: share of undevelopable land; Wharton Regulation Index. In the IV model in column (2), PDI is instrumented using the housing supply elasticity [3]. Models in columns (3) and (4) do not include the components of the housing supply elasticity as MSA-level controls. In the IV model in column (4), PDI is instrumented using the *residual of the housing supply elasticity*, obtained by regressing housing supply elasticity on its two main components: geographic and regulatory constraints [3]. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in the model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.12: Robustness checks – Excluding most populous and most densely populated MSAs



Note: This figure shows the estimated coefficient (including the 90% and 95% confidence intervals, respectively represented by the dashed light and dark gray lines) of the Potential Density Increase (PDI) Index. Each coefficient is obtained from a separate regression; OLS estimation in A.12a; IV (instrumental variable) estimation in Figure A.12b. Each regression is estimated using a sub-sample of the observations, specified on the x-axis. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, and UMD). Each sample excludes observations with excess returns in the top 1% of the distribution. All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level.

Table A.14: Robustness checks – Additional controls

	Excess returns											
	Including additional controls:											
	Population 2000 & high-developed		Population 2000 & house price		Population 2000 & productivity		High-developed & productivity		High-developed & house price		House price & productivity	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV	(9) OLS	(10) IV	(11) OLS	(12) IV
Main Results												
PDI	-0.104*** (0.032)	-0.232*** (0.084)	-0.112*** (0.031)	-0.200*** (0.064)	-0.051* (0.028)	-0.147* (0.074)	-0.049* (0.025)	-0.160* (0.082)	-0.096*** (0.025)	-0.196*** (0.072)	-0.071** (0.028)	-0.139** (0.063)
First Stage: PDI												
Housing Supply Elasticity		-0.150*** (0.028)		-0.191*** (0.028)		-0.150*** (0.027)		-0.138*** (0.030)		-0.183*** (0.029)		-0.212*** (0.042)
Partial F -statistics		28.706		45.620		30.112		20.924		39.679		25.397
Partial R^2		0.172		0.233		0.171		0.129		0.180		0.204
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Incl. MSA popul. in 2000	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No
Incl. high-developed share	Yes	Yes	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Incl. house price index	No	No	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Incl. firm's productivity	No	No	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Clust. std. error: MSA-level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	11,789	11,789	11,683	11,683	10,097	10,097	10,097	10,097	11,683	11,683	9,993	9,993
Cluster N	81	81	74	74	81	81	81	81	74	74	74	74
Dep. var. mean	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
Dep. var. st.dev.	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, UMD). The sample excludes observations with excess returns in the top 1% of the distribution. The regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in columns (1), (3), (5), (7), (9) and (11) derive from ordinary least squares estimation (OLS); results in columns (2), (4), (6), (8) and (10) derive from instrumental variable estimation (IV), where PDI is instrumented using the housing supply elasticity [3]. The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). In all IV models, the partial F -statistics is above the standard threshold of 10, suggesting that the instrument is relevant. All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. The models in columns (1) and (2) include the MSA's population in 2000 and the high-developed share. The models in columns (3) and (4) include the MSA's population in 2000 and the yearly change in the house price index. The models in columns (5) and (6) include the MSA's population in 2000 and the firms' productivity. The models in columns (7) and (8) include the MSA's high-developed share and the firms' productivity. The models in columns (9) and (10) include the MSA's high-developed share and the yearly change in the house price index. The models in columns (11) and (12) include the MSA's yearly change in the house price index and the firms' productivity. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.15: Robustness checks – Additional controls related to firms' growth

	Excess returns					
	Including additional controls:					
	Revenue growth		EBITDA growth		Revenue & EBITDA growth	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Main Results						
PDI	-0.095*** (0.025)	-0.226** (0.089)	-0.094*** (0.025)	-0.222** (0.089)	-0.094*** (0.025)	-0.222** (0.089)
First Stage: PDI						
Housing Supply Elasticity		-0.157*** (0.037)		-0.157*** (0.037)		-0.157*** (0.037)
Partial F -statistic		17.756		17.755		17.753
Partial R^2		0.141		0.141		0.141
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Incl. revenue growth	Yes	Yes	No	No	Yes	Yes
Incl. EBITDA growth	No	No	Yes	Yes	Yes	Yes
Clust. std. error: MSA-level	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	11,736	11,736	11,736	11,736	11,736	11,736
Cluster N	81	81	81	81	81	81
Dep. var. mean	0.046	0.046	0.046	0.046	0.046	0.046
Dep. var. st.dev.	0.368	0.368	0.368	0.368	0.368	0.368

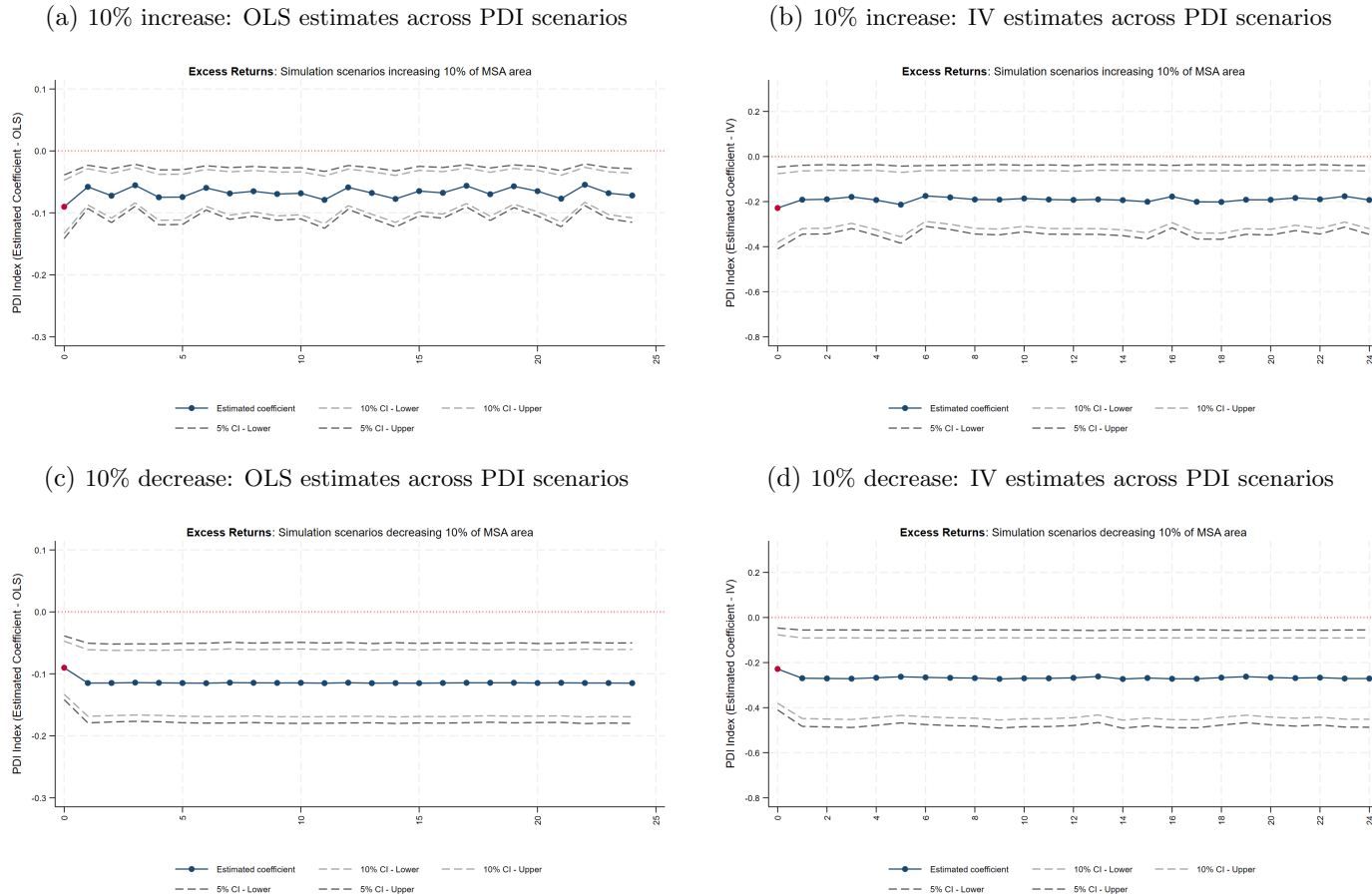
Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, UMD). The sample excludes observations with excess returns in the top 1% of the distribution. The regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in columns (1), (3), and (5) derive from ordinary least squares estimation (OLS); results in columns (2), (4), and (6) derive from instrumental variable estimation (IV), where PDI is instrumented using the housing supply elasticity [3]. The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). Models also include additional firm-level controls to account for company growth: models in columns (1) and (2) control for revenue growth; models in columns (3) and (4) control for EBITDA growth; models in columns (5) and (6) control for both. All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.16: Sensitivity to polygon definition – Land allocation scenarios under alternative assumptions in increase/decrease of total land from original polygon

Change in total land from original polygon: +10%; -10%; +15%; -15%					
Scenario #	HD (%)	D (%)	HD+D (%)	LD (%)	Sum (%)
1	0.00	0.00	0.00	100.00	100
2	0.00	5.00	5.00	95.00	100
3	0.00	10.00	10.00	90.00	100
4	0.00	15.00	15.00	85.00	100
5	0.00	20.00	20.00	80.00	100
6	0.00	25.00	25.00	75.00	100
7	0.00	30.00	30.00	70.00	100
8	0.00	50.00	50.00	50.00	100
9	1.00	0.00	1.00	99.00	100
10	1.00	5.00	6.00	94.00	100
11	1.00	10.00	11.00	89.00	100
12	1.00	15.00	16.00	84.00	100
13	1.00	20.00	21.00	79.00	100
14	1.00	25.00	26.00	74.00	100
15	1.00	30.00	31.00	69.00	100
16	1.00	50.00	51.00	49.00	100
17	2.50	0.00	2.50	97.50	100
18	2.50	5.00	7.50	92.50	100
19	2.50	10.00	12.50	87.50	100
20	2.50	15.00	17.50	82.50	100
21	2.50	20.00	22.50	77.50	100
22	2.50	25.00	27.50	72.50	100
23	2.50	30.00	32.50	67.50	100
24	2.50	50.00	52.50	47.50	100

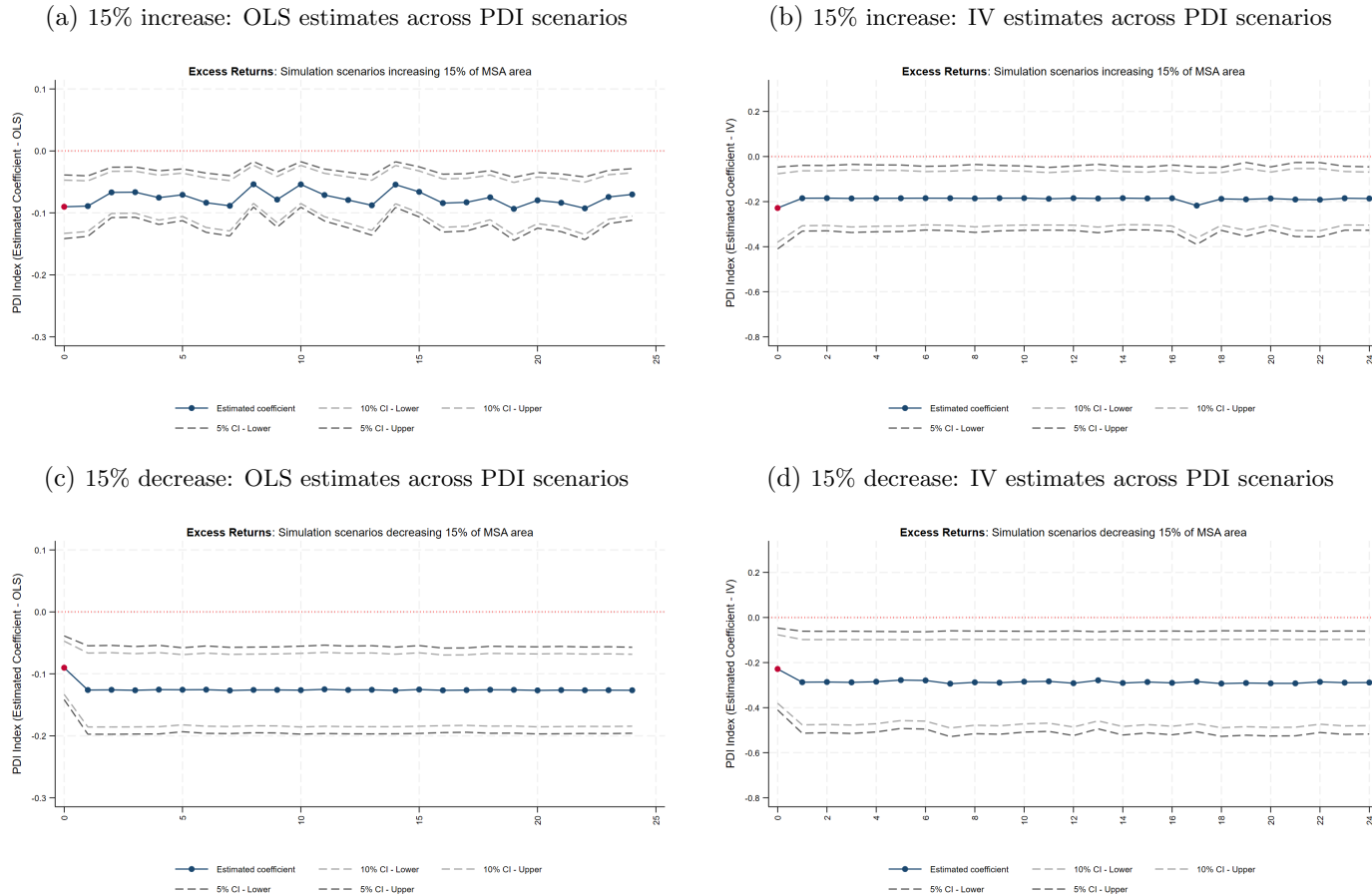
Notes: The table reports hypothetical land allocation scenarios under alternative total land assumptions: $\pm 10\%$ and $\pm 15\%$ in total land from the original polygon around the main city center of the MSA. Each scenario reports an alternative allocation of land types in the change in total area that leads to the newly simulated polygon. HD denotes highly developed land, D developed land, and LD low-developed land. Across the different scenarios, HD is assumed to be 0%, 1% and 2.5% of the change in total area; D is assumed to be 0%, 5%, 10%, 15%, 20%, 25%, 30%, and 50% of the change in total land; and LD is calculated as the remaining percentage of the change in total area. Percentages are expressed relative to the change in total land and sum to 100% in all scenarios.

Figure A.13: Sensitivity to polygon definition – Simulation of different scenarios of PDI from 10% change in total land from original polygon



Note: This figure shows the estimated coefficient (including the 90% and 95% confidence intervals, respectively represented by the dashed light and dark gray lines) of the Potential Density Increase (PDI) Index. Each coefficient is obtained from a separate regression. Panels (a) and (c) report OLS estimates, while panels (b) and (d) report IV (instrumental variable) estimates. Each regression is estimated for a different scenario of Table A.16, specified on the x-axis. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, and UMD). Each sample excludes observations with excess returns in the top 1% of the distribution. All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level.

Figure A.14: Sensitivity to polygon definition – Simulation of different scenarios of PDI from 15% change in total land from original polygon



Note: This figure shows the estimated coefficient (including the 90% and 95% confidence intervals, respectively represented by the dashed light and dark gray lines) of the Potential Density Increase (PDI) Index. Each coefficient is obtained from a separate regression. Panels (a) and (c) report OLS estimates, while panels (b) and (d) report IV (instrumental variable) estimates. Each regression is estimated for a different scenario of Table A.16, specified on the x-axis. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, and UMD). Each sample excludes observations with excess returns in the top 1% of the distribution. All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level.

Table A.17: Robustness check – Alternative ratios of building density

	Excess returns			
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
Main Results				
$\text{PDI} \left(\frac{D}{D+LD} \right)$	-0.090***			
	(0.026)			
$\frac{LD}{HD+D+LD}$		0.126***		
		(0.030)		
$\frac{LD}{D+LD}$			0.090***	
			(0.026)	
$\frac{D}{D+HD}$				-0.052
				(0.040)
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes
Clust. std. error: MSA	Yes	Yes	Yes	Yes
Observations N	11,789	11,789	11,789	11,789
Cluster N	81	81	81	81
Dep. var. mean	0.029	0.046	0.046	0.046
Dep. var. st.dev.	0.375	0.369	0.369	0.369

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors: excess market return (MKT), small-minus-big return (SMB), high-minus-low return (HML), and momentum (UMD). The sample excludes observations with excess returns in the top 1-st percentile. In column (1), the regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. In column (2), the regressor of interest is an alternative definition given by the ratio between low-developed land and the sum of high-developed, developed and low-developed land. In column (3), the regressor of interest is an alternative definition given by the ratio between low-developed land and the sum of developed and low-developed land. In column (4), the regressor of interest is an alternative definition, given by the ratio between developed land and the sum of developed and high-developed land. Results derive from ordinary least squares estimation (OLS). All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.18: Robustness checks – Excluding most geographically constrained MSAs

	Excess returns					
	Excluding MSAs with highest share of geographically constraint land:					
	Top 25-th percentile		Top 10-th percentile		Top 5-th percentile	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Main Results						
PDI	-0.154** (0.069)	-0.537*** (0.170)	-0.152*** (0.050)	-0.358** (0.146)	-0.094*** (0.033)	-0.228** (0.093)
First Stage: PDI						
Housing Supply Elasticity		-0.093*** (0.017)		-0.108*** (0.021)		-0.163*** (0.041)
Partial F -stat.		31.351		26.593		15.866
Partial R^2		0.198		0.212		0.199
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Clust. std. error: MSA-level	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	8,277	8,277	9,732	9,732	10,785	10,785
Cluster N	60	60	73	73	77	77
Dep. var. mean	0.040	0.040	0.037	0.037	0.043	0.043
Dep. var. st.dev.	0.349	0.349	0.352	0.352	0.371	0.371

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, UMD). The sample excludes observations with excess returns in the top 1% of the distribution. The regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in columns (1), (3), and (5) derive from ordinary least squares estimation (OLS); results in columns (2), (4), and (6) derive from instrumental variable estimation (IV), where PDI is instrumented using the housing supply elasticity [3]. The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). Results in this table exclude MSAs with the highest share of undevelopable land (i.e., geographically constrained areas, such as water bodies, high-slope mountains, parks, and so on). Models in columns (1) and (2) exclude the top 25-th percentile of most geographically constrained MSAs; models in columns (3) and (4) exclude the top 10-th percentile of most geographically constrained MSAs; models in columns (5) and (6) exclude the top 5-th percentile of most geographically constrained MSAs. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.19: Robustness check – Alternative dependent variables (I)

	Return – Risk-free			Excess Return (5FF)			Excess Return (5FF)			Excess Return (6FF)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
Main Results												
High-developed Index	-0.142			-0.064			-0.078			-0.072		
	(0.101)			(0.093)			(0.092)			(0.092)		
Developed Index	-0.144***			-0.140***			-0.140***			-0.143***		
	(0.036)			(0.033)			(0.033)			(0.033)		
PDI		-0.102***	-0.299**		-0.087***	-0.244**		-0.090***	-0.228**		-0.090***	-0.246**
		(0.038)	(0.120)		(0.026)	(0.099)		(0.026)	(0.091)		(0.026)	(0.098)
First Stage: PDI												
Housing Supply Elasticity			-0.158***			-0.158***			0.157***			-0.158***
			(0.037)			(0.037)			(0.037)			(0.037)
Partial F -statistics			17.907			17.899			17.859			17.901
Partial R^2			0.142			0.142			0.142			0.142
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clust. std. error: MSA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	11,781	11,781	11,781	11,789	11,789	11,789	11,789	11,789	11,789	11,788	11,788	11,788
Cluster N	81	81	81	81	81	81	81	81	81	81	81	81
Dep. var. mean	0.076	0.076	0.076	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
Dep. var. st.dev.	0.422	0.422	0.422	0.373	0.373	0.373	0.375	0.375	0.375	0.373	0.373	0.373

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . In columns (1)-(3), the dependent variable is the annualized returns, defined as the monthly stock return net of the risk-free rate. In columns (4)-(6), the dependent variable is excess returns, defined as the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, and UMD) plus the profitability factor (RMW). In columns (7)-(9), the dependent variable is excess returns, defined as the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, and UMD) plus the investment factor (CMA). In columns (10)-(12), the dependent variable is excess returns, defined as the annualized residuals of the regression of monthly return net of the risk-free rate on all six standard Fama and French factors (MKT, SMB, HML, UMD, RMW, and CMA). All samples exclude observations with excess returns in the top 1% of the distribution, regardless of the definition of the dependent variable. The models in columns (1), (4), (7), and (10) include two regressors of interest: the high-developed index is the share of highly developed land over the total of buildable land; the developed index is the share of developed land over the total of buildable land. In columns (2)-(3), (5)-(6), (8)-(9), and (11)-(12), the regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in columns (1)-(2), (4)-(5), (7)-(8), and (10)-(11) come from ordinary least squares estimation (OLS); results in columns (3), (6), (9) and (12) come from instrumental variable estimation (IV), where PDI is instrumented using the housing supply elasticity [3]. The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). In all IV models, the partial F -statistics is above the standard threshold of 10, suggesting that the instrument is relevant. All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.20: Robustness check – Alternative dependent variables (II)

	Stock returns (% change in price)			Average price (Log)			Price volatility (Log)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
Main Results									
High-developed	-0.100 (0.095)			-1.618*** (0.350)			-1.196*** (0.388)		
Developed	-0.080** (0.039)			0.433** (0.173)			0.363* (0.204)		
PDI		-0.054* (0.032)	-0.257*** (0.097)		0.102 (0.166)	0.467 (0.397)		0.144 (0.162)	0.710* (0.422)
First Stage: PDI									
Housing Supply Elasticity			-0.155*** (0.037)			-0.158*** (0.037)			-0.158*** (0.037)
Partial F -statistics			17.715			17.901			17.916
Partial R^2			0.140			0.143			0.143
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clust. std. err.: MSA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	11,483	11,483	11,483	11,899	11,899	11,899	11,872	11,872	11,872
Cluster N	81	81	81	81	81	81	81	81	81
Dep. var. mean	0.077	0.077	0.077	3.007	3.007	3.007	2.151	2.151	2.151
Dep. var. st.dev.	0.313	0.313	0.313	1.178	1.178	1.178	1.121	1.121	1.121

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors: excess market return (MKT), small-minus-big return (SMB), high-minus-low return (HML), and momentum (UMD). The sample excludes observations with excess returns in the top 1% of the distribution. The models in columns (1), (4), and (7) include two regressors of interest: the high-developed index is the share of highly developed land over the total of buildable land; the developed index is the share of developed land over the total of buildable land. In columns (2)-(3), (5)-(6), and (8)-(9), the regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in columns (1)-(2), (4)-(5), and (7)-(8) come from ordinary least squares estimation (OLS); results in columns (3), (6), and (9) come from instrumental variable estimation (IV), where PDI is instrumented using the housing supply elasticity [3]. The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). In all IV models, the partial F -statistics is above the standard threshold of 10, suggesting that the instrument is relevant. All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.21: Robustness check – Alternative fixed-effects (I)

	Excess returns								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
Main Results									
High-developed	-0.047 (0.087)			-0.058 (0.082)			-0.073 (0.087)		
Developed	-0.108*** (0.035)			-0.130*** (0.034)			-0.104*** (0.036)		
PDI		-0.068** (0.028)	-0.156* (0.080)		-0.082*** (0.026)	-0.212** (0.093)		-0.069** (0.029)	-0.171* (0.088)
First Stage: PDI									
Housing Supply Elasticity			-0.156*** (0.039)			-0.158*** (0.038)			-0.156*** (0.040)
Partial F -statistics			15.794			17.449			15.440
Partial R^2			0.139			0.142			0.139
Year FE	No	No	No	No	No	No	No	No	No
State FE	Yes	Yes	Yes	No	No	No	No	No	No
Industrial sector FE	No	No	No	Yes	Yes	Yes	No	No	No
State $\times$ Year FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Year FE	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clust. std. error: MSA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	11,215	11,215	11,215	11,769	11,769	11,769	11,195	11,195	11,195
Cluster N	81	81	81	80	80	80	80	80	80
Dep. var. mean	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
Dep. var. st.dev.	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . In columns (1)-(3), the dependent variable is the stock return, simply calculated as the year-to-year percentage change in the stock price. In columns (4)-(6), the dependent variable is the annual average stock price. In columns (7)-(9), the dependent variable is the annual stock price volatility. All samples exclude observations with a value of the dependent variable in the top 1% of the distribution, regardless of the definition of the dependent variable. The models in columns (1), (4), and (7) include two regressors of interest: the high-developed index is the share of highly developed land over the total of buildable land; the developed index is the share of developed land over the total of buildable land. In columns (2)-(3), (5)-(6), and (8)-(9), the regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in columns (1)-(2), (4)-(5), and (7)-(8) come from ordinary least squares estimation (OLS); results in columns (3), (6), and (9) come from instrumental variable estimation (IV), where PDI is instrumented using the housing supply elasticity [3]. The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). In all IV models, the partial F -statistics is above the standard threshold of 10, suggesting that the instrument is relevant. None of the models includes year fixed effects. The models in columns (1)-(3) include state fixed effects and industrial sector times year fixed effects. The models in columns (4)-(6) include industrial sector fixed effects and state times year fixed effects. The models in columns (7)-(9) include industrial sector times year fixed effects and state times year fixed effects. In all specifications, the industrial sector is defined by the 4-digit SIC code. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.22: Robustness check – Alternative fixed-effects (II)

	Excess returns								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
Main Results									
High-developed	-0.047 (0.087)			-0.056 (0.082)			-0.071 (0.087)		
Developed	-0.108*** (0.035)			-0.129*** (0.034)			-0.104*** (0.036)		
PDI		-0.068** (0.028)	-0.156* (0.080)		-0.081*** (0.026)	-0.210** (0.093)		-0.068** (0.028)	-0.169* (0.088)
First Stage: PDI									
Housing Supply Elasticity			-0.156*** (0.039)			-0.158*** (0.038)			-0.156*** (0.040)
Partial F -statistics			15.782			17.441			15.415
Partial R^2			0.139			0.142			0.139
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State $\times$ Year FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Year FE	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA-level controls.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clust. std. error: MSA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	11,215	11,215	11,215	11,769	11,769	11,769	11,195	11,195	11,195
Cluster N	81	81	81	80	80	80	80	80	80
Dep. var. mean	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029	0.029
Dep. var. st.dev.	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375	0.375

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . In columns (1)-(3), the dependent variable is the stock return, simply calculated as the year-to-year percentage change in the stock price. In columns (4)-(6), the dependent variable is the annual average stock price. In columns (7)-(9), the dependent variable is the annual stock price volatility. All samples exclude observations with a value of the dependent variable in the top 1% of the distribution, regardless of the definition of the dependent variable. The models in columns (1), (4), and (7) include two regressors of interest: the high-developed index is the share of highly developed land over the total of buildable land; the developed index is the share of developed land over the total of buildable land. In columns (2)-(3), (5)-(6), and (8)-(9), the regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in columns (1)-(2), (4)-(5), and (7)-(8) come from ordinary least squares estimation (OLS); results in columns (3), (6), and (9) come from instrumental variable estimation (IV), where PDI is instrumented using the housing supply elasticity [3]. The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). In all IV models, the partial F -statistics is above the standard threshold of 10, suggesting that the instrument is relevant. All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. The models in columns (1)-(3) also include industrial sector times year fixed effects. The models in columns (4)-(6) also include state times year fixed effects. The models in columns (7)-(9) also include industrial sector times year fixed effects and state times year fixed effects. In all specifications, the industrial sector is defined by the 4-digit SIC code. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.23: Robustness check – Alternative clustering of standard errors

	Excess returns	
	(1) OLS	(2) IV
Main Results		
PDI	-0.090** (0.041)	-0.228** (0.114)
First Stage: PDI		
Saiz elasticity		-0.157*** (0.019)
Partial F -statistics		69.853
Partial R^2		0.142
Year FE	Yes	Yes
State FE	Yes	Yes
Industrial sector FE	Yes	Yes
Firm-level controls	Yes	Yes
MSA-level controls	Yes	Yes
Clust. std. error: MSA $\times$ 4-digit SIC code	Yes	Yes
Observations N	11,789	11,789
Cluster N	1,462	1,462
Dep. var. mean	0.046	0.046
Dep. var. st.dev.	0.369	0.369

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, UMD). The sample excludes observations with excess returns in the top 1% of the distribution. The regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in column (1) derive from ordinary least squares estimation (OLS); results in column (2) derive from instrumental variable estimation (IV), where PDI is instrumented using the housing supply elasticity [3]. The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA $\times$ SIC level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.24: Robustness check – Excluding MSAs with the highest variation in the PDI measure

	Excess returns							
	Excluding MSAs with the highest PDI standard deviation:							
	Top 5-th percentile		Top 10-th percentile		Top 25-th percentile		Above mean	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Main Results								
PDI	-0.079*** (0.026)	-0.271** (0.114)	-0.064** (0.029)	-0.261** (0.126)	-0.080** (0.040)	-0.251** (0.106)	-0.092** (0.041)	-0.243*** (0.090)
First Stage: PDI								
Housing Supply Elasticity		-0.144*** (0.041)		-0.134*** (0.040)		-0.156*** (0.053)		-0.180*** (0.050)
Partial F -statistics		12.497		11.113		8.495		12.847
Partial R^2		0.110		0.094		0.112		0.160
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industrial sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clust. std. error: MSA-level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations N	10,709	10,709	10,304	10,304	9,170	9,170	8,785	8,785
Cluster N	76	76	73	73	62	62	56	56
Dep. var. mean	0.029	0.029	0.028	0.028	0.028	0.028	0.028	0.028
Dep. var. st.dev.	0.373	0.373	0.374	0.374	0.369	0.369	0.369	0.369

Note: This table reports our estimation results. The unit of observation is firm-year: a firm i in year t , which operates in industry k and is located in a city (MSA) c in state s . For all estimation models, the dependent variable is the firm's excess returns in the stock market. Excess returns are the annualized residuals of the regression of monthly return net of the risk-free rate on the four standard Fama and French factors (MKT, SMB, HML, UMD). The sample excludes observations with excess returns in the top 1% of the distribution. The regressor of interest is the potential density increase (PDI) index, defined as the ratio between developed land and the sum of developed and low-developed land. Results in columns (1), (3), (5) and (7) derive from ordinary least squares estimation (OLS); results in columns (2), (4), (6), and (8) derive from instrumental variable estimation (IV), where PDI is instrumented using the housing supply elasticity [3]. The upper panel reports the main results: the OLS estimates and the second-stage estimates for the IV estimation; the bottom panel reports the information about the first stage of the IV regression (estimated coefficient, partial F -statistics, and partial R^2). Results in this table exclude the MSAs with the highest PDI standard deviation, which is due to the fact that the image segmentation algorithm has some margin of error in classifying the types of land according to the level of building density (see Appendix A.3). Models in columns (1) and (2) exclude the MSAs whose PDI standard deviation falls in the top 5-th percentile; models in columns (3) and (4) exclude the MSAs whose PDI standard deviation falls in the top 10-th percentile; models in columns (5) and (6) exclude the MSAs whose PDI standard deviation falls in the top 25-th percentile; and models in columns (7) and (8) exclude the MSAs whose PDI standard deviation is above the mean. All models include year, state, and industrial sector (defined at the 4-digit SIC code) fixed effects. All models include firm-level controls, lagged one year: real estate ratio; leverage; logarithm of market capitalization; R&D ratio; ROE; ROA; market-to-book value; fixed investment ratio; asset growth; and years since first listing (not lagged). All models include MSA-level controls: the one-hour drive polygon area; share of undevelopable land; Wharton Regulation Index; logarithm of population density; logarithm of GDP; logarithm of mean elevation; logarithm of mean slope; and an indicator for the presence of seacoast within 80 km from the city center. Standard errors are reported in parentheses; the standard errors are clustered at the MSA level. The number of clusters in each model is reported in the table. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

References

- [1] De Bellefon MP, Combes PP, Duranton G, Gobillon L, Gorin C. Delineating urban areas using building density. *Journal of Urban Economics*. 2021;125:103226.
- [2] Arribas-Bel D, Garcia-López MÀ, Viladecans-Marsal E. Building (s and) cities: Delineating urban areas with a machine learning algorithm. *Journal of Urban Economics*. 2021;125:103217.
- [3] Saiz A. The geographic determinants of housing supply. *The Quarterly Journal of Economics*. 2010;125(3):1253–1296.
- [4] Gyourko J, Saiz A, Summers A. A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index. *Urban Studies*. 2008;45(3):693–729.
- [5] Rappaport J, Sachs JD. The United States as a coastal nation. *Journal of Economic Growth*. 2003;8:5–46.
- [6] Pakes A, Olley S. A limit theorem for a smooth class of semiparametric estimators. *Journal of Econometrics*. 1995;65(1):295–332.
- [7] Arellano M, Bond S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*. 1991;58(2):277–297.
- [8] Levinsohn J, Petrin A. Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*. 2003;70(2):317–341.
- [9] Marschak J, Andrews WH. Random simultaneous equations and the theory of production. *Econometrica*. 1944;p. 143–205.
- [10] Imrohoroglu A, Tuzel S. Firm-level productivity, risk, and return. *Management Science*. 2014;60(8):2073–2090.
- [11] Hall BH. The stock market’s valuation of R&D investment during the 1980’s. *The American Economic Review*. 1993;83(2):259–264.
- [12] Francis DC, Karalashvili N, Maemir H, Rodriguez Meza J. Measuring total factor productivity using the Enterprise Surveys. A methodological note; 2020.
- [13] Clark WA, Huang Y, Withers S. Does commuting distance matter?: Commuting tolerance and residential change. *Regional Science and Urban Economics*. 2003;33(2):199–221.
- [14] Rouwendal J. Search theory and commuting behavior. *Growth and Change*. 2004;35(3):391–418.
- [15] Uchida H, Nelson A. Agglomeration index: Towards a new measure of urban concentration; 2009.