Earthquake Risk Embedded in Property Prices: Evidence from Five Japanese Cities

Data Documentation*

Masako Ikefuji
School of Social and International Studies
University of Tsukuba and ISER, Osaka University

Roger J. A. Laeven†
Department of Quantitative Economics
University of Amsterdam, CentER, and EURANDOM

Jan R. Magnus
Department of Econometrics and Data Science
Vrije Universiteit Amsterdam and Tinbergen Institute

Yuan Yue
Department of Quantitative Economics
University of Amsterdam and Tinbergen Institute

March 12, 2021

*This research was supported in part by the Japan Society for the Promotion of Science (JSPS) under grant 16K03565 (Ikefuji) and the Netherlands Organization for Scientific Research (NWO) under grant VIDI (Laeven). Queries concerning the data should be addressed to Yuan Yue. Email addresses: `ikefuji.masako.gn@u.tsukuba.ac.jp` (Ikefuji), `R.J.A.Laeven@uva.nl` (Laeven), `jan@janmagnus.nl` (Magnus), `Y.Yue@uva.nl` (Yue).

†Corresponding author. Mailing Address: PO Box 15867, 1001 NJ Amsterdam, The Netherlands. Phone: +31 (0) 20 525 4219.

1 Introduction

We are interested in the effect of earthquake risk on property prices in major cities in Japan. We select five Japanese cities/areas for our purpose (see Section 2): Tokyo Metropolitan Area (23 special wards), Osaka City, Nagoya City, Fukuoka City and Sapporo City. We shall refer to the Tokyo Metropolitan Area as a city, although officially it is an area, not a city.

Each city is divided into wards and each ward is divided into districts. Certain information is available per ward, which can affect the attractiveness of buying a property in that ward. For example, population characteristics, information about schools and medical facilities, shopping, safety, etc. These 'attractiveness' characteristics are described in Section 3.

We distinguish between three types of properties: 'residential land (land and building)', 'residential land (land only)', and condominium. Sales prices and property characteristics are available for each of these types in each of the five cities. These are described in Section 4. We do not know the exact location of a property, but we do know in which district the property lies and we also know the distance to the nearest station and the name of that station.

Some macro variables are relevant and affect house prices nationally. These variables are described in Section 5.

We next come to the earthquake and risk data. Historical earthquake data are described in Section 6. Details about the parameterization, estimation and simulation within the ETAS earthquake model are contained in Section 7. Japan is geographically split up in meshes of varying size. The largest (first mesh) is 80×80 km, the smallest (quarter mesh) is 250×250 meter. The data on these meshes are described in Section 8. While historical earthquake data are described in Section 6, earthquake *risk* data are described in Section 9. Finally, we describe how to link stations and districts to these meshes in Section 10.

2 Cities

Japan has twelve cities with populations of more than one million people. Almost 100 million Japanese, or 78% of the country's total population of 127.4 million, live in urban areas. The total population of Japan's largest 103 cities amounts to 63.9 million or just over half of all the country's residents. Tokyo, with almost nine million inhabitants, is often referred to as a city, but is officially known and governed as a 'metropolitan prefecture'. With a population of 3.7 million, Yokohama, south of Tokyo, is Japan's second largest city. It is the country's largest port and a manufacturing and ship building centre. Japan's third-largest city is Osaka with 2.7 million inhabitants. It is the country's third most important seaport and home to many leading Japanese manufacturers. Nagoya (2.3 million inhabitants) is the center of the Chukyo Metropolitan Area and is home to the Mitsubishi Aircraft Company and the Toyota factory. Eight cities have between one and two million inhabitants: Sapporo, Kobe, Fukuoka, Kyoto, Kawasaki, Saitama, Hiroshima, and Sendai.

From these twelve cities we selected five: Sapporo, Tokyo, Nagoya, Osaka, and Fukuoka. These five cities provide a good representation of the major cities in Japan, in terms of geographical spread (Sapporo in the North, Fukuoka in the South) and earthquake risk (Tokyo highest, then Osaka and Nagoya, then Sapporo and Fukuoka).

We excluded Yokohama, Kawasaki, and Saitama, because they are located in the same metropolitan area as Tokyo, the 'Kanto' area. Similarly, we excluded Kobe and Kyoto, because they are located in the same metropolitan area as Osaka, the 'Keihanshin' area. Thus, each of the three major metropolitan areas is represented: the greater Tokyo area (Tokyo, Yokohama, Kawasaki, Saitama) by Tokyo, the Kansai region (Osaka, Kobe, Kyoto) by Osaka, and the Chukyo metropolitan area by Nagoya.

To obtain a representative geographical spread we added Sapporo, the largest city in the North, and Fukuoka, the second largest city in the West after Osaka.

Hiroshima was excluded because the metro system is not sufficiently dense to identify the properties, and Sendai because it lies in the 2011 Fukushima disaster area and property prices there are completely distorted.

3 Wards and attractiveness characteristics

A designated city is a Japanese city that has a population greater than 500,000 and has been designated as such by order of the Cabinet of Japan. Designated cities are delegated many of the tasks normally performed by prefectural governments, such as public education, social welfare, sanitation, business licensing, and urban planning. Designated cities are required to subdivide themselves into wards ('ku'), each of which has a ward office conducting various administrative functions for the city government. The 23 special wards of Tokyo are not part of this system, as Tokyo is a prefecture, and its wards are effectively independent cities. The five cities together contain 80 wards: 24 in Osaka, 23 in Tokyo, 16 in Nagoya, 10 in Sapporo, and 7 in Fukuoka.

When considering to buy a property in a given city we are likely to be interested in certain characteristics of these wards, in particular characteristics that make one ward more or less attractive than another. Information about wards can be downloaded from

`https://www.e-stat.go.jp/SG1/estat/eStatTopPortalE.do`,

the portal site of official statistics of Japan, under the category 'Regional Statistics'. The English version of this website provides the statistics for one year (not the same year for each variable), while the Japanese version provides 'time-series' data. Data are available in 11 categories: A. Population and households; B. Natural environment; C. Economic base; D. Administrative base; E. Education; F. Labor; G. Culture and sports; H. Dwelling; I. Health and medical care; J. Welfare and social security; and K. Safety. More detailed information on these categories (in Japanese) is available from

`https://www.e-stat.go.jp/SG1/chiiki/FileStream.do?file=koumoku.html`.

Some variables (such as unemployment and the number of traffic accidents) are updated (or adjusted) annually, while others (such as population and households) are only updated after each 5-year census. The data cover the period 2007–2015, so in order to obtain a general indicator that reflects the attractiveness over the whole sample period of our housing price data (2006–2015), a simple average is calculated. In case of a missing record the nearest data available are used as a proxy; for example, data for year 2006 are assumed to be the same as data for 2007. In this way we construct *one* (time-independent) value for each item in each ward.

For our purpose we selected the variables listed in Table 2. Below we explain how these items were calculated from the available data.

Table 2: Attractiveness characteristics by ward

Category	Item
Population	% younger than 15 years
	% older than 65 years
	% immigrants
	% emigrants
	% foreigners
	% private households
	% nuclear families
	% one-person households
Schools, culture, & welfare	number of daycare nurseries
	number of schools (kindergarten)
	number of schools (primary)
	number of schools (junior/senior high)
	number of students per teacher (primary school)
	number of homes for the aged
	number of community halls
	number of libraries
Medical facilities	number of general hospitals
	number of physicians
	number of dentists
	number pharmacists
Safety	number of traffic accidents
	number of criminal offenses
Shopping	number of large-scale retail stores
	number of department stores
	annual sales of commercial goods
Housing	% privately owned houses
	habitable land area (% of total area)
Employment	unemployment ratio
	% of self-employed
	% of executives

3.1 Population

3.1.1 % younger than 15 years:

Table 1: Summary statistics PctU15

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.101	0.076	0.091	0.100	0.109	0.143	0.017
Osaka	0.115	0.062	0.110	0.119	0.126	0.166	0.023
Nagoya	0.125	0.073	0.109	0.122	0.145	0.165	0.025
Fukuoka	0.132	0.101	0.117	0.137	0.145	0.157	0.021
Sapporo	0.123	0.103	0.116	0.124	0.128	0.143	0.011

The percentage of population younger than 15 years is obtained as the ratio of 'population younger than 15 years old' to 'total population' of each ward. The data are constant for years 2007–2011 and for years 2012–2015. Since we only want a time-independent indicator, simple averaging and extrapolation is used, so that for each ward

$$\text{percentage U15 population} = [6 \times \text{PctU15}_{2007} + 4 \times \text{PctU15}_{2012}]/10,$$

where PctU15_t denotes the percentage in year t, the first term represents the percentage for the first 6 years (2006–2011) and the second for the last 4 years (2012–2015).

3.1.2 % older than 65 years:

Table 3: Summary statistics PctO65

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.192	0.161	0.183	0.190	0.202	0.230	0.017
Osaka	0.209	0.148	0.183	0.204	0.226	0.310	0.033
Nagoya	0.197	0.156	0.184	0.198	0.216	0.236	0.024
Fukuoka	0.161	0.145	0.157	0.161	0.164	0.176	0.010
Sapporo	0.187	0.172	0.175	0.183	0.194	0.226	0.016

Obtained as the ratio of 'population aged 65 years or more' to 'total population'. The data is constant for years 2007–2011 and for years 2012–2015. The indicator is thus averaged in the same pattern as the item '% younger than 15 years'.

3.1.3 % immigrants:

Table 4: Summary statistics PctImmi

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.077	0.043	0.059	0.075	0.085	0.133	0.024
Osaka	0.063	0.031	0.047	0.051	0.071	0.138	0.028
Nagoya	0.063	0.039	0.049	0.059	0.070	0.112	0.019
Fukuoka	0.079	0.065	0.066	0.068	0.087	0.110	0.019
Sapporo	0.065	0.050	0.054	0.062	0.072	0.103	0.016

Obtained as the ratio of 'number of immigrants from other municipalities' to 'total population'. The numbers are varying each year. We assume that data for years 2005 and 2006 are the same as year 2007.

3.1.4 % emigrants:

Table 5: Summary statistics PctEmmi

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.068	0.039	0.054	0.065	0.081	0.112	0.019
Osaka	0.050	0.032	0.039	0.043	0.056	0.098	0.018
Nagoya	0.060	0.041	0.049	0.057	0.069	0.098	0.015
Fukuoka	0.077	0.060	0.066	0.070	0.086	0.106	0.018
Sapporo	0.058	0.043	0.054	0.055	0.065	0.081	0.012

Obtained as the ratio of 'number of emigrants to other municipalities' to 'total population'. The numbers are varying each year. The data is missing in 2007–2011 for wards in Sapporo and in 2008–2011 for wards in Nagoya. For Sapporo we use year 2012 data as a substitute for the missing values. For Nagoya we use year 2007 data to substitute missing values in 2008–2009 and the year 2012 data to substitute those in 2010–2011.

3.1.5 % foreigners:

Table 6: Summary statistics PctForeign

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.029	0.014	0.020	0.025	0.030	0.064	0.014
Osaka	0.037	0.015	0.018	0.024	0.035	0.208	0.040
Nagoya	0.023	0.014	0.016	0.022	0.027	0.050	0.009
Fukuoka	0.011	0.006	0.007	0.009	0.016	0.019	0.005
Sapporo	0.003	0.002	0.002	0.003	0.004	0.006	0.001

Obtained as the ratio of 'number of foreigners' to 'total population'. The numbers are constant for years 2007–2011 and for years 2012–2015.

3.1.6 % private households:

Table 7: Summary statistics PctPrivHouse

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.983	0.930	0.975	0.994	0.998	0.999	0.021
Osaka	0.978	0.850	0.983	0.993	0.996	0.999	0.037
Nagoya	0.985	0.925	0.983	0.994	0.996	0.998	0.020
Fukuoka	0.985	0.962	0.979	0.992	0.994	0.996	0.013
Sapporo	0.997	0.987	0.998	0.999	0.999	0.999	0.004

Obtained as the ratio of 'number of private households' to 'number of households'. The numbers are constant for years 2007–2011 and for years 2012–2015.

3.1.7 % nuclear families:

Table 8: Summary statistics PctNuclear

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.448	0.338	0.406	0.435	0.495	0.571	0.072
Osaka	0.481	0.249	0.441	0.504	0.560	0.658	0.107
Nagoya	0.518	0.288	0.459	0.533	0.567	0.682	0.096
Fukuoka	0.480	0.334	0.404	0.513	0.549	0.610	0.107
Sapporo	0.583	0.410	0.514	0.576	0.665	0.721	0.100

Obtained as the ratio of ‘number of nuclear family households’ to ‘number of households’. The numbers are constant for years 2007–2011 and for years 2012–2015.

3.1.8 % one-person households:

Table 9: Summary statistics PctSingle

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.473	0.350	0.416	0.471	0.539	0.591	0.077
Osaka	0.434	0.269	0.366	0.417	0.475	0.649	0.095
Nagoya	0.388	0.230	0.341	0.370	0.470	0.580	0.093
Fukuoka	0.440	0.296	0.377	0.405	0.518	0.592	0.109
Sapporo	0.349	0.185	0.271	0.361	0.424	0.525	0.107

Obtained as the ratio of ‘number of one-person households’ to ‘number of households’. The numbers are constant for years 2007–2011 and for years 2012–2015.

3.2 Schools, culture, welfare

3.2.1 number of daycare nurseries (per inhabitable area):

Table 10: Summary statistics Ndaycare

city	mean	min	25%	50%	75%	max	sd
Tokyo	1.868	0.541	1.572	1.827	2.251	2.844	0.569
Osaka	1.667	0.360	1.212	1.764	2.087	2.924	0.626
Nagoya	0.994	0.501	0.775	0.982	1.117	1.658	0.308
Fukuoka	0.815	0.511	0.681	0.897	0.958	1.019	0.198
Sapporo	0.492	0.136	0.349	0.518	0.652	0.748	0.203

The variable is obtained as the number of daycare nurseries per square km of inhabitable area.

3.2.2 number of schools (kindergarten) (per inhabitable area):

Table 11: Summary statistics Nkindergtn

city	mean	min	25%	50%	75%	max	sd
Tokyo	1.419	0.811	1.057	1.333	1.658	2.476	0.441
Osaka	1.072	0.244	0.676	0.936	1.384	2.504	0.569
Nagoya	0.713	0.197	0.506	0.743	0.899	1.226	0.273
Fukuoka	0.704	0.281	0.397	0.693	1.009	1.140	0.353
Sapporo	0.399	0.176	0.291	0.343	0.501	0.625	0.153

The numbers of kindergartens per square km of inhabitable area are missing in years 2006–2007 for all wards in Sapporo, so we approximate numbers of these years with those of 2008. The numbers are missing in year 2010 for all wards in Nagoya, so we approximate this as the mean numbers from years 2009 and 2011.

3.2.3 number of schools (primary) (per inhabitable area):

Table 12: Summary statistics NprimSchool

city	mean	min	25%	50%	75%	max	sd
Tokyo	1.568	0.945	1.348	1.573	1.860	2.294	0.364
Osaka	1.550	0.487	1.254	1.578	1.970	2.418	0.483
Nagoya	0.934	0.438	0.828	0.966	1.068	1.298	0.212
Fukuoka	0.747	0.434	0.572	0.752	0.913	1.073	0.241
Sapporo	0.546	0.218	0.490	0.553	0.670	0.694	0.147

The number of primary schools is divided by the area of inhabitable land.

3.2.4 number of schools (junior/senior high) (per inhabitable area):

Table 13: Summary statistics Nhighsch

city	mean	min	25%	50%	75%	max	sd
Tokyo	1.668	0.708	1.052	1.569	1.944	4.668	0.859
Osaka	1.286	0.395	0.808	1.026	1.514	4.500	0.873
Nagoya	0.752	0.219	0.494	0.612	0.817	2.076	0.472
Fukuoka	0.635	0.373	0.442	0.549	0.784	1.073	0.263
Sapporo	0.428	0.157	0.367	0.406	0.491	0.762	0.155

The total number of junior and senior high schools is divided by the area of inhabitable land. For senior high schools, the numbers for Nagoya and Fukuoka are missing in year 2010; we use the average of 2009 and 2011 data as an proxy.

3.2.5 number of students per teacher (primary school):

Table 14: Summary statistics NstudentPrim

city	mean	min	25%	50%	75%	max	sd
Tokyo	18.040	15.846	16.914	17.589	19.435	20.822	1.603
Osaka	17.460	9.436	16.863	17.910	19.083	20.865	2.438
Nagoya	19.048	13.028	18.231	19.158	20.504	21.707	2.081
Fukuoka	20.604	19.031	20.148	20.714	21.268	21.651	0.926
Sapporo	19.509	16.541	19.230	19.687	20.012	21.199	1.213

The ratio of students to teachers for primary schools.

3.2.6 number of homes for the aged (per 65+ population):

Table 15: Summary statistics Nagedhome

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.178	0.122	0.153	0.174	0.197	0.290	0.040
Osaka	0.211	0.072	0.176	0.216	0.244	0.329	0.063
Nagoya	0.231	0.098	0.193	0.222	0.272	0.395	0.082
Fukuoka	0.294	0.176	0.279	0.292	0.312	0.407	0.068
Sapporo	0.227	0.175	0.179	0.203	0.219	0.428	0.078

The number of homes for the aged divided by 1/1000 times the population over 65 years old.

3.2.7 number of community halls (per capita):

Table 16: Summary statistics Ncommhall

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.001	0.000	0.000	0.000	0.000	0.014	0.003
Osaka	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Nagoya	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fukuoka	1.069	0.875	0.985	1.054	1.163	1.256	0.137
Sapporo	0.005	0.000	0.000	0.000	0.000	0.048	0.015

The number of community halls is divided by 1/10000 of the total population. The data is missing for Sapporo in 2007, which we approximate using 2008 data.

3.2.8 number of libraries (per capita):

Table 17: Summary statistics Nlib

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.320	0.156	0.231	0.282	0.361	0.994	0.176
Osaka	0.107	0.050	0.075	0.101	0.142	0.192	0.042
Nagoya	0.095	0.054	0.065	0.093	0.122	0.156	0.032
Fukuoka	0.078	0.041	0.067	0.078	0.097	0.099	0.022
Sapporo	0.062	0.036	0.048	0.058	0.076	0.096	0.021

The number of libraries is divided by 1/10000 of the total population. The data is missing for Sapporo in 2007, which we approximate using 2008 data.

3.3 Medical facilities

3.3.1 number of general hospitals (per capita):

Table 18: Summary statistics Nhosp

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.643	0.257	0.372	0.460	0.685	3.707	0.690
Osaka	0.779	0.326	0.566	0.670	0.937	1.412	0.300
Nagoya	0.627	0.212	0.443	0.592	0.740	1.664	0.332
Fukuoka	0.739	0.544	0.578	0.640	0.902	1.027	0.200
Sapporo	0.991	0.738	0.825	0.932	1.032	1.769	0.298

The number of general hospitals is divided by 1/10000 times the total population.

3.3.2 number of physicians (per capita):

Table 19: Summary statistics Nphysician

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.576	0.102	0.172	0.243	0.479	3.508	0.794
Osaka	0.365	0.112	0.178	0.213	0.420	1.001	0.290
Nagoya	0.324	0.090	0.153	0.223	0.362	1.155	0.284
Fukuoka	0.360	0.169	0.220	0.268	0.530	0.584	0.180
Sapporo	0.288	0.138	0.187	0.213	0.269	0.854	0.212

The number of physicians is divided by 1/100 times the total population.

3.3.3 number of dentists (per capita):

Table 20: Summary statistics Ndentist

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.312	0.066	0.089	0.106	0.209	3.427	0.693
Osaka	0.133	0.054	0.073	0.085	0.108	0.918	0.174
Nagoya	0.106	0.048	0.059	0.073	0.110	0.328	0.082
Fukuoka	0.133	0.063	0.084	0.119	0.188	0.209	0.061
Sapporo	0.095	0.056	0.064	0.069	0.074	0.223	0.062

The number of dentists is divided by 1/100 times the total population.

3.3.4 number pharmacists (per capita):

Table 21: Summary statistics Npharmacist

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.802	0.166	0.192	0.272	0.634	7.256	1.558
Osaka	0.460	0.115	0.185	0.228	0.332	3.911	0.767
Nagoya	0.287	0.111	0.165	0.213	0.290	1.165	0.252
Fukuoka	0.256	0.154	0.189	0.218	0.318	0.404	0.092
Sapporo	0.231	0.117	0.171	0.187	0.219	0.661	0.156

The number of pharmacists is divided by 1/100 times the total population.

3.4 Safety

3.4.1 number of traffic accidents (per capita):

Table 22: Summary statistics Naccident

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.725	0.313	0.421	0.483	0.635	3.703	0.716
Osaka	0.712	0.365	0.448	0.561	0.641	2.565	0.497
Nagoya	0.870	0.631	0.669	0.754	0.852	2.436	0.437
Fukuoka	1.025	0.705	0.794	0.866	1.146	1.726	0.391
Sapporo	0.504	0.354	0.419	0.452	0.558	0.894	0.156

The number of traffic accidents is divided by 1/100 times the total population. The numbers are unknown for years 2012–2015, which we approximate with the numbers from year 2011.

3.4.2 number of criminal offenses (per capita):

Table 23: Summary statistics Ncrime

city	mean	min	25%	50%	75%	max	sd
Tokyo	2.544	1.275	1.447	1.671	3.453	11.267	2.124
Osaka	3.600	1.817	2.242	2.649	3.088	14.794	2.927
Nagoya	3.098	1.979	2.050	2.621	3.106	9.615	1.872
Fukuoka	2.354	1.743	1.864	1.992	2.733	3.546	0.765
Sapporo	1.415	0.842	1.221	1.338	1.517	2.537	0.454

The number of criminal offenses is divided by 1/100 times the total population. The numbers are unknown for years 2012–2015, which we approximate with the numbers from year 2011.

3.5 Shopping

3.5.1 number of large-scale retail stores (per capita):

Table 24: Summary statistics Nlargeretail

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.305	0.094	0.112	0.119	0.248	2.132	0.451
Osaka	0.208	0.070	0.109	0.141	0.181	1.022	0.207
Nagoya	0.196	0.100	0.126	0.155	0.213	0.638	0.130
Fukuoka	0.176	0.082	0.106	0.135	0.228	0.344	0.108
Sapporo	0.204	0.159	0.175	0.184	0.208	0.353	0.056

The number of large-scale retail stores is divided by 1/1000 times the total population.

3.5.2 number of department stores (per capita):

Table 25: Summary statistics Ndepstore

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.239	0.074	0.112	0.175	0.222	1.286	0.264
Osaka	0.179	0.000	0.078	0.109	0.174	1.112	0.223
Nagoya	0.276	0.076	0.180	0.249	0.307	0.741	0.162
Fukuoka	0.191	0.023	0.098	0.213	0.242	0.419	0.133
Sapporo	0.198	0.079	0.128	0.181	0.216	0.474	0.113

The number of department stores is divided by 1/10000 times the total population.

3.5.3 annual sales of commercial goods (per capita):

Table 26: Summary statistics SalesComm

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.073	0.002	0.003	0.008	0.019	0.958	0.210
Osaka	0.023	0.001	0.003	0.004	0.011	0.243	0.053
Nagoya	0.018	0.002	0.004	0.005	0.012	0.106	0.029
Fukuoka	0.010	0.001	0.002	0.002	0.013	0.035	0.014
Sapporo	0.005	0.001	0.002	0.002	0.003	0.022	0.006

The annual sales of commercial goods are divided by the total population. The unit is in 1000 yen.

3.6 Housing

3.6.1 % privately owned houses:

Table 27: Summary statistics PctOwnedHou

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.426	0.328	0.396	0.413	0.468	0.506	0.051
Osaka	0.415	0.186	0.362	0.438	0.470	0.536	0.086
Nagoya	0.452	0.317	0.410	0.439	0.486	0.612	0.072
Fukuoka	0.377	0.241	0.309	0.399	0.421	0.534	0.098
Sapporo	0.524	0.399	0.424	0.510	0.595	0.756	0.118

Obtained as the ratio of privately owned houses per dwelling building. It is constant for years 2011–2015 so we use this number to approximate the missing data for years before 2011.

3.6.2 habitable land area (% of total area):

Table 28: Summary statistics PctHabitLand

city	mean	min	25%	50%	75%	max	sd
Tokyo	1.000	1.000	1.000	1.000	1.000	1.000	0.000
Osaka	1.000	1.000	1.000	1.000	1.000	1.000	0.000
Nagoya	0.971	0.865	0.941	1.000	1.000	1.000	0.044
Fukuoka	0.775	0.352	0.738	0.837	0.887	0.983	0.214
Sapporo	0.691	0.165	0.510	0.671	0.988	1.000	0.295

Obtained as the ratio of inhabitable land area to total area. It is constant for years 2007–2012 and for years 2013–2015.

3.7 Employment

3.7.1 unemployment ratio (per labor supply population):

Table 29: Summary statistics PctUnemploy

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.054	0.029	0.047	0.054	0.061	0.072	0.012
Osaka	0.102	0.074	0.085	0.098	0.111	0.202	0.026
Nagoya	0.055	0.045	0.047	0.055	0.060	0.074	0.009
Fukuoka	0.068	0.063	0.066	0.067	0.067	0.078	0.005
Sapporo	0.071	0.064	0.069	0.070	0.073	0.079	0.004

Obtained as the ratio of unemployed population to the population of labor supply.

3.7.2 % of self-employed (per labor supply population):

Table 30: Summary statistics PctSelfemploy

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.101	0.081	0.091	0.103	0.107	0.127	0.012
Osaka	0.107	0.081	0.093	0.105	0.119	0.146	0.017
Nagoya	0.087	0.071	0.079	0.086	0.092	0.110	0.011
Fukuoka	0.086	0.069	0.081	0.090	0.092	0.094	0.009
Sapporo	0.064	0.048	0.059	0.064	0.069	0.082	0.010

Obtained as the ratio of self-employed (including those with and without employee) to the population of labor supply.

3.7.3 % of executives (per labor supply population):

Table 31: Summary statistics PctExec

city	mean	min	25%	50%	75%	max	sd
Tokyo	0.094	0.061	0.070	0.080	0.105	0.181	0.033
Osaka	0.064	0.032	0.047	0.056	0.072	0.127	0.024
Nagoya	0.072	0.048	0.060	0.067	0.082	0.116	0.017
Fukuoka	0.052	0.044	0.045	0.048	0.053	0.075	0.011
Sapporo	0.058	0.050	0.052	0.055	0.057	0.084	0.010

Obtained as the ratio of number of executives to the population of labor supply.

4 Property prices and determinants

4.1 The MLIT data set

In our study we shall work with sales prices rather than with rental prices, because we believe sales are more permanent than rentals and therefore the effect of earthquake risk on choosing the property will be more informative.

Nakagawa *et al.* (2009) use land prices over various years (from 1980 onwards) and describe the data in their Section 3 (for the Tokyo area). Their data are based on the Koji-Chika data set published by the Ministry of Land, Infrastructure, Transport, and Tourism. The well-known Koji-Chika set provides *fictional* sales prices (as produced by 'experts') and they are only available at annual intervals, which we consider to be too long.

Thus we shall use a different data set, which provides self-reported transaction prices at three-months intervals. This data set known as the 'Real estate transaction-price information' and is provided by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT); see

`https://www.land.mlit.go.jp/webland_english/servlet/MainServlet`,

The information in this data set is based on the results of a questionnaire survey of persons involved in real estate transactions conducted by MLIT, compiled and published quarterly.

The Real Estate Transaction Questionnaire Survey was conducted for government ordinance-designed major cities of three metropolitan areas (including the 23 special wards of Tokyo, Osaka and Nagoya) starting at the 3rd quarter of 2005. The survey region expanded to cover prefectural capitals (including Sapporo and Fukuoka) starting at the 2nd quarter of 2006. After 2007 2nd quarter, the prefectural office location cities of the whole country were included in the survey region.

In our analysis we use the data from 2nd Quarter 2006 to 3rd Quarter 2015, where all five cities are surveyed. Thus, in total, we have 38 quarters of observations.

We distinguish between three types of properties: (1) 'residential land (land and building)', hereafter 'land & building'; (2) 'residential land (land only)', hereafter 'land only'; and (3) 'pre-owned condominiums', hereafter 'condos'. We have data on 362658 properties of which approximately 44% are condo's, 34% are land & buildings, and 22% are land only.

In Table 32 we provide a list of the available variables in the housing data set, together with a short description.

Table 32: List of variables, LandMLIT data set: Name and description

Variable name	Description
Type	condos/ land only/ land & building
Region	residential/ commercial/ industrial…
City/Town/Ward/Village code	postcode
Prefecture	
City/Town/Ward/Village	which ward is the property in
District	which district is the property in
Nearest station name	
Nearest station distance minutes	80m/min, accurate to minute for 0–30 min
Transaction price, total	price in 10,000 yen
Layout	number of rooms, stories etc.
Area m^2	(floor) area of the property, accurate to 5 or 10 m^2
Transaction price unit price m^2	
Land shape	
Frontage	length of land in contact of front road
Total floor area m^2	
Year of construction	
Building structure	steel / concrete / wood…
Building use	family / office / factory…
Purpose of use	similar as above
Frontage road direction	
Frontage road classification	city road / prefectural road…
Frontage road breadth	
City planning	plans for the district
Max bldg coverage ratio	
Max floor area ratio	
Transaction period	date of contract
Remarks	other transaction-related issues

4.2 Included/excluded variables and availability

Not all these variables are selected to be included in our study.

Table 33: Included variables, LandMLIT data set

Variable name	Availability
Type	all types
City/Town/Ward/Village	all records
District	all records
Nearest station name	most records
Nearest station distance (minutes)	most records
Transaction price, total	all records
Area m^2	all records
Total floor area m^2	'land & building' only
Year of construction	unknown for 'land only'
Building structure	unknown for 'land only'
Max bldg coverage ratio	all records
Max floor area ratio	all records
City planning for land use	all records
Transaction period	all records

The selected variables are provided in Table 33. Obviously the transaction price is included since its logarithm is our dependent variable. The variable 'Max bldg coverage ratio' is important in the literature.

The variables that are excluded from our data set are listed in Table 34. In the variables 'Layout' and 'Building use' there are too many categories, and in 'Purpose of use' the data are only available for too short a period.

4.3 Sample selection

For all five cities and the whole sample period (2006Q2–2015Q3), there are 362,658 records (before sample selection). In choosing the sample, the following criteria are applied:

- We exclude all records where walking time to nearest station is longer than 30 minutes or nearest station is unknown.
- We exclude records with living area larger than 2000 square meters.
- In cases of 'pre-owned condominiums' and 'residential Land (land and building)', we exclude properties built before the war (1945).

Table 34: Not included variables, LandMLIT data set

Variable name	Availability
Region	unknown for condos
City/Town/Ward/Village code	all records
Prefecture	all records
Layout	condos only
Transaction price unit price m^2	'land only' only
Land shape	unknown for condos
Frontage	unknown for condos
Building use	unknown for 'land only'
Purpose of use	only after 2013
Frontage road direction	unknown for condos
Frontage road classification	unknown for condos
Frontage road breadth	unknown for condos
Remarks	some records

Table 35: Summary of number of wards, districts, properties per type, and stations

City	Wards	Areas	Buildings	Land	Condos	Metro stations
Tokyo	23	898	57568	33991	92518	482
Osaka	24	564	21064	6901	21855	220
Nagoya	16	1379	14640	13110	11029	159
Fukuoka	7	318	7847	5660	12475	75
Sapporo	10	551	11763	9461	11461	86
Total	80	3710	112882	69123	149338	1022

After selection we are left with 91.4% of the original data, that is, 331,390 records. In addition, 47 records are apparently wrongly coded because the location information given by the 'district' and 'nearest station' do not match. We manually checked these records and decided that the information may not be accurate, so we exclude these from our sample. More information on how we verified the location information can be found in Section 10.5.

This leads to the summary statistics provided in Table 35. We emphasize that we do not know the exact location of a property. We only know two things about the location, namely the district in which the property lies and the name of and distance to the nearest station. In the five cities together there are 3710 districts and 1022 stations after applying the sample selection criteria mentioned above. So, in order to identify the location of a property, the district information is more accurate than the station information.

4.4 Property prices

For all records in our data, the total transaction value (unit: 10,000 yen) excluding overhead costs (such as agents commission) is provided. Figures are rounded to two decimal places by the provider, but no other numerical adjustments were made.

Table 36: Total price, quantiles ($\times$ 10 million yen)

City	5%	25%	50%	75%	95%	n
Land & Building						
Tokyo	1.5	3.7	5.0	7.5	34.0	57568
Osaka	0.5	1.6	3.1	4.5	26.0	21064
Nagoya	1.0	2.7	3.7	4.8	16.0	14640
Fukuoka	1.0	2.1	3.2	5.3	26.0	7847
Sapporo	0.7	1.6	2.6	3.8	14.0	11763
Land only						
Tokyo	1.2	3.1	4.9	8.3	27.0	33991
Osaka	0.6	1.6	3.0	6.4	25.0	6901
Nagoya	0.8	1.8	2.7	4.5	13.0	13110
Fukuoka	0.6	1.5	2.2	4.5	16.0	5660
Sapporo	0.4	0.9	1.3	2.3	7.4	9461
Condo						
Tokyo	0.7	1.6	2.5	3.8	7.0	92518
Osaka	0.4	1.0	1.6	2.2	3.6	21855
Nagoya	0.3	0.9	1.5	2.3	3.6	11029
Fukuoka	0.2	0.5	1.1	1.8	3.1	12475
Sapporo	0.2	0.6	1.1	1.7	2.7	11461

The main quantiles of the distribution of the total transaction price per city and per type are given in Table 36. Property prices are highly skewed with the median well below the mean. Not surprisingly, Tokyo is the most expensive city, followed by Osaka and Nagoya. Cheapest are Fukuoka and Sapporo.

The cheapest property is a condo built in 1984 in the Nagayoshinagahara district of Hirano Ward, Osaka. It is a 1DK room of $40m^2$ and it was sold in 2012 for 530 yen (about 5 dollars and 30 cents). This is obviously a symbolic prize and one may make up an explanatory story, but we don't know the background. Such extremely cheap properties are rare in our sample. Out of our 331,343 sample records, there are 524 properties (0.16%) with a sales price under one million yen (about $10,000). This subsample of 524 properties are mostly small properties, but there are no apparent patterns in terms of location, distance to nearest station, region, city zone, or transaction period.

We attempted to find out a little more about these ‘outliers’. The ‘Land Economy and Construction and Engineering Industry Bureau’ of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) told us that their questionnaire involves people involved in real estate transactions, not real estate agencies or organizations. The information totally relies on the answers in the questionnaire. The National Tax Agency (Osaka Region) told us that it is legally possible for properties to be sold for such low prices. But there are fiscal restrictions: if the estimated value of a property and the realized deviate too much (according to the tax authority), then this sale may be subject to gift or inheritance tax. Finally, two private real-estate agencies (one in Fukuoka, one in Osaka) told us that there might be special issues with these properties. For example, the owner went bankrupt and the creditor placed a mortgage on the property; or the property suffers from a psychologically defect (such as criminal homicide or suicide, in Japan such information must be provided by the seller); or that the deal includes the right of property (house) or land lease; or renovation is very expensive so that the previous owner sold the house at low price possibly to the real estate broker who would then renovate and sell for a higher price. We marked these 524 properties in our data set, so that we are able to do the analysis with and without these ‘outliers’.

The most expensive property is a building in Ginza, Chuo Ward, Tokyo. It has a land area of $1200m^2$ and a total floor area larger than $2000m^2$. This property was sold in 2013 for 24,000 million yen (about $240 million).

The distribution of the property prices is clearly highly skewed. Hence, we also present the log-prices in Table 37. The log-prices are more symmetric, which is one reason why we choose log-prices to be our dependent variable. To analyze the symmetry of the log-prices we also present the ‘skew’ in Table 37. The skew is defined in terms of the quartiles as: $(Q3-Q2)-(Q2-Q1)$, where $Q1$, $Q2$, and $Q3$ denote the first, second, and third quartile, respectively. If this number is positive, then we say there is positive skew. Of the 15 items there is 1 without skew. Of the remaining 14, 7 have negative skew and 7 positive skew. The assumption of symmetry of log-prices therefore does not seem unreasonable.

We emphasize one other point. The legality of property sales works differently in different countries. In many countries, there are two contracts: the first when you agree on a price, the second when you actually exchange. Once the second contract is signed, the first becomes obsolete. There may be several months between the two contracts. In the first contract it would say that A intends to buy from B the following property for such and such a price, but under condition that a mortgage can be obtained, a property inspector will not find major faults, etc. So, it is binding under certain conditions. The buyer typically pays a percentage (like 10%) of the price as a guarantee. The second contract is signed when all is

Table 37: Total log-price, quantiles

City	5%	25%	50%	75%	95%	Skew	n
Land & Building							
Tokyo	16.5	17.4	17.7	18.1	19.6	0.1	57568
Osaka	15.4	16.6	17.2	17.6	19.4	−0.3	21064
Nagoya	16.1	17.1	17.4	17.7	18.9	−0.1	14640
Fukuoka	16.1	16.9	17.3	17.8	19.4	0.1	7847
Sapporo	15.8	16.6	17.1	17.5	18.8	−0.1	11763
Land only							
Tokyo	16.3	17.2	17.7	18.2	19.4	0.1	33991
Osaka	15.5	16.6	17.2	18.0	19.3	0.1	6901
Nagoya	15.8	16.7	17.1	17.6	18.7	0.1	13110
Fukuoka	15.6	16.5	16.9	17.6	18.9	0.3	5660
Sapporo	15.1	16.0	16.4	17.0	18.1	0.1	9461
Condo							
Tokyo	15.7	16.6	17.0	17.5	18.1	−0.0	92518
Osaka	15.1	16.1	16.6	16.9	17.4	−0.2	21855
Nagoya	14.9	16.0	16.5	17.0	17.4	−0.1	11029
Fukuoka	14.6	15.5	16.2	16.7	17.2	−0.2	12475
Sapporo	14.5	15.5	16.2	16.6	17.1	−0.3	11461

in order, the money is with the solicitor, and the house is empty. It is the second contract which is the official document and its date is the official date, even though the actual price has been negotiated and decided (much) earlier.

Fortunately, in Japan it works differently. There is only *one* purchase contract, which is signed after the price has been agreed on. If the buyer cancels the purchase after signing the contract, he/she loses the deposit, which is typically 10% of the price but can be lower (sometimes negotiable). In the case of a condo that is typically sold before completion, the deposit is usually much lower (less than 5% typically). Banks provide preliminary review services before signing a purchase contract. There are non-negligible cases where they eventually decide not to provide loans, but the probability of this happening is low.

Important in our case is that the purchase date given in our data set corresponds to the moment when the price was agreed, not to the moment that the exchange of property/money takes place.

4.5 Housing characteristics as explanatory variables

4.5.1 Type

According to the LandMLIT website:

> *Real estate is divided into the following types: residential land, pre-owned condominiums, etc., agricultural land, and forest land. Residential land is further divided into two types of residential land (land only) and residential land (land and building). Transactions for residential land (land only) indicate the transactions for land only. Transactions for residential land (land and building) mean the package transactions for the land and buildings, etc.. Transactions for pre-owned condominiums, etc. are the transactions for condominium units (apartments, etc.).*

For pre-owned condos, the ward, district, nearest station, distance to nearest station, floor area, year of construction, building structure, building use, city planning, building coverage ratio, and floor area ratio are provided.

For residential land (land only), the region, ward, district, nearest station, distance to nearest station, area in square meters, unit land price, land shape, frontage of land, frontage road width/direction/classification, city planning, building coverage ratio, and floor area ratio are provided.

For residential land (land and building), the region, ward, district, nearest station, distance to nearest station, area in square meters, total floor area of building, frontage road width/direction/classification, year of construction, building structure, building use, city planning, building coverage ratio, and floor area ratio are provided.

4.5.2 Location information

For each record, the city, ward, and district where the record is located are specified. In addition, the name and walking distance of the nearest station are given. These are the only two measures of location of a given record available to us.

4.5.3 Time distance to nearest station

According to the LandMLIT website:

> *For the residential land (land only), residential land (land and building), and pre-owned condominiums, etc., the name of the nearest train station and the time distance (minute) from the location of the property to the nearest train station (for subway, to the ground entrance) are displayed. A time distance less than 30 minutes is displayed in minutes. A time distance greater than 30 minutes is displayed in the following time periods: 30 minutes to 59 minutes, 1 hour to 1 hour 29 minutes, 1 hour 30 minutes to 1 hour 59 minutes, and 2 hours or more.*

The time distance by walking is calculated following laws and regulations concerning advertisement. Ordinance for 'Enforcement of the fair competition codes concerning indication of real estate', Chapter 5, Article 10, in accordance with the 'Fair competition codes', Article 15, assigns a walking rate formula. This formula states that one-minute walking on a road is equal to a distance of 80 meters. There are 1955 properties that are '0' minutes walking distance to the nearest station.

Table 38: Summary statistics distance (0–29min)

City	mean	25%	50%	75%	sd	n
Tokyo	8.2	4	7	11	5.04	184077
Osaka	6.8	4	6	9	4.21	49820
Nagoya	10.8	6	9	15	6.73	38779
Fukuoka	11.0	6	9	15	6.73	25982
Sapporo	11.2	6	10	15	7.00	32685

Table 38 summarizes the distance in minutes for each city. We provide the mean, three quantiles, and the standard deviation. Osaka has the densest railway structure, followed by Tokyo. Fukuoka, Nagoya, and Sapporo have a somewhat less dense railway/metro system.

4.5.4 Area and total floor area (in square meters)

From the official description:

> *For each of the residential land (land only), residential land (land and building), agricultural land, and forest land, the surveyed area (m^2) obtained from a survey of persons involved in transactions or the registered area (m^2) specified in a register if the surveyed area is unknown is provided. For pre-owned condominiums, etc., the floor area (m^2) of the exclusively owned area registered in a register (the area measured inside walls or other partitions) is provided. For all land types, data for small properties with an area less than 10 m^2 are not published. For properties with an area of less than 200 m^2, the area data are displayed in 5 m^2 intervals, while for properties with an area of 200 m^2 or greater, the data are displayed after rounding the figures to the first two digits from the left. For transactions for land with an area of 2,000 m^2 or greater, the data are displayed as '2,000 m^2 or greater'.*

For 'land only' types, the variable 'area' refers to the area of the land; for condos it refers to the floor area.

Another variable is the total floor area of the building (m^2). We have:

> *For buildings on residential land (land and building), the total floor area (m^2) is provided. For cases where the floor area is less than 200 m^2, the data are displayed in 5 m^2 intervals, and for cases where the floor area is 200 m^2 or greater, figures are rounded to two decimal places. For large transactions where the building floor area is 2,000 m^2 or greater, the data are displayed as '2,000 m^2 or greater' whereas for small transactions where the floor area is less than 10 m^2, the data are displayed as 'less than 10 m^2'.*

We set the total floor area of properties where the building floor area is 2,000 m^2 or greater to 2000 and those where the floor area is less than 10 m^2 to 10.

4.5.5 Year of construction and age

For properties built before 1945, the construction year data are displayed as 'before the war'. We discard these data and further categorize the year of construction into '1946–1981', '1982–2000', and '2001–now'. This categorization is due to the establishment of 'The shin-taishin, or New Earthquake Resistant Building Standard Amendment'; see

`https://japanpropertycentral.com/real-estate-faq/earthquake-building-codes-in-japan/`,

Table 39: Summary statistics area (m^2)

City	mean	25%	50%	75%	sd	n
Land & Building						
Tokyo	128.5	65.0	90.0	125.0	148.0	57568
Osaka	134.8	55.0	75.0	130.0	180.7	21064
Nagoya	186.1	110.0	135.0	180.0	178.0	14640
Fukuoka	259.4	140.0	180.0	270.0	240.6	7847
Sapporo	252.5	160.0	200.0	260.0	206.3	11763
Land Only						
Tokyo	165.7	70.0	105.0	175.0	193.0	33991
Osaka	218.0	75.0	125.0	250.0	257.9	6901
Nagoya	240.6	120.0	170.0	270.0	224.1	13110
Fukuoka	325.4	150.0	220.0	370.0	302.7	5660
Sapporo	287.8	165.0	220.0	300.0	257.2	9461
Condo						
Tokyo	46.7	20.0	45.0	65.0	27.9	92518
Osaka	53.8	30.0	60.0	70.0	26.9	21855
Nagoya	65.6	60.0	70.0	80.0	23.7	11029
Fukuoka	53.4	25.0	60.0	75.0	27.7	12475
Sapporo	69.1	60.0	70.0	85.0	22.8	11461

Table 40: Summary statistics total floor area (m^2)

City	mean	25%	50%	75%	sd	n
Land & Building						
Tokyo	195.0	85.0	95.0	145.0	302.2	57568
Osaka	255.6	90.0	105.0	200.0	384.6	21064
Nagoya	214.1	100.0	110.0	155.0	312.8	14640
Fukuoka	282.2	100.0	125.0	230.0	399.9	7847
Sapporo	286.3	110.0	140.0	300.0	351.5	11763

which is a real estate agency website, or equivalently

`https://www.uncrd.or.jp/hyogo/hesi/pdf/expmeeting/otani.pdf`,

from the United Nations website.

Table 41: Summary statistics building age

City	mean	min	25%	50%	75%	max	sd	n
Tokyo	14.4	−2	1	11	24	67	13.5	184077
Osaka	18.1	−2	4	17	30	69	14.8	49820
Nagoya	14.9	−1	0	13	25	69	13.9	38779
Fukuoka	17.2	−2	7	18	25	66	12.3	25982
Sapporo	18.1	−1	8	18	27	65	12.4	32685

We also include the numerical 'age' of the property, i.e. transaction year minus the year of construction. This is displayed in Table 41. This can be a negative number, namely if the property was sold before construction.

4.5.6 Building structure

Building structure can be 'Steel frame reinforced concrete', 'Reinforced concrete', 'Steel frame', 'Light steel structure', 'Concrete block', 'Wooden', or combinations of these structures. This leads to a large number of building structures. We summarize these in five categories: 'contains steel frame reinforced concrete' (SRC), 'contains reinforced concrete but not steel frame reinforced concrete' (RC), 'contains steel but not reinforced concrete' (S), 'contains wood but not steel or reinforced concrete' (W), and 'NA'. Table 42 contains a summary of the available information.

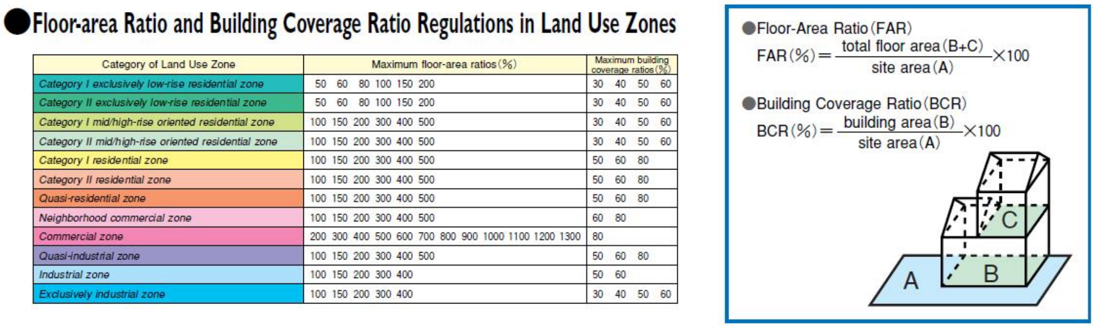
Table 42: Number of properties per building structure

Building structure	*Land & Building*	*Condo*
SRC	2288	54904
RC	12581	92804
S	17619	1135
W	75017	14
NA	5377	481
Total	112882	149338

4.5.7 Building coverage ratio and floor area ratio

For all three types the *designated* maximum building coverage ratio (%) and maximum floor-area ratio (%) are provided. These ratios are legally allowed maxima, different for each piece of land. Usually buildings with larger designated ratios are more expensive.

The building coverage ratio is the percentage of the site area to the building area. The floor area ratio is the percentage of the total floor area to the site area.



●Floor-area Ratio and Building Coverage Ratio Regulations in Land Use Zones

Category of Land Use Zone	Maximum floor-area ratios (%)	Maximum building coverage ratios (%)
Category I exclusively low-rise residential zone	50 60 80 100 150 200	30 40 50 60
Category II exclusively low-rise residential zone	50 60 80 100 150 200	30 40 50 60
Category I mid/high-rise oriented residential zone	100 150 200 300 400 500	30 40 50 60
Category II mid/high-rise oriented residential zone	100 150 200 300 400 500	30 40 50 60
Category I residential zone	100 150 200 300 400 500	50 60 80
Category II residential zone	100 150 200 300 400 500	50 60 80
Quasi-residential zone	100 150 200 300 400 500	50 60 80
Neighborhood commercial zone	100 150 200 300 400 500	60 80
Commercial zone	200 300 400 500 600 700 800 900 1000 1100 1200 1300	80
Quasi-industrial zone	100 150 200 300 400 500	50 60 80
Industrial zone	100 150 200 300 400	50 60
Exclusively industrial zone	100 150 200 300 400	30 40 50 60

Figure 1: Urban land use planning system in Japan (Ministry of Land, Infrastructure and Transport)

For different city planning zones, there are different limits on these ratios, as shown Figure 1.

Table 43: Summary statistics Building Coverage Ratio

City	mean	min	25%	50%	75%	max	sd	n
Tokyo	65	30	60	60	80	80	11	184077
Osaka	75	40	80	80	80	80	9	49820
Nagoya	63	30	60	60	80	80	12	38779
Fukuoka	64	30	60	60	80	80	12	25982
Sapporo	62	30	60	60	80	80	13	32685

Table 44: Summary statistics Floor Area Ratio

City	mean	min	25%	50%	75%	max	sd	n
Tokyo	302	60	200	300	400	1300	157	184077
Osaka	341	80	200	300	400	1300	172	49820
Nagoya	240	50	200	200	200	1100	130	38779
Fukuoka	238	50	150	200	400	1000	134	25982
Sapporo	214	50	200	200	200	900	104	32685

The relevant data in our data set are summarized in Table 43 and Table 44.

4.5.8 City planning

For all three types of records, the use of districts designated by the City Planning Act is provided. A detailed explanation can be found at

`https://www.mlit.go.jp/crd/city/plan/tochiriyou/pdf/reaf_e.pdf`.

The planning can be: Category 1 Exclusive Low Rise Residential District, Category 1 Exclusive Mid-high Rise Residential District, Category 1 Residential District, Category 2 Exclusive Low Rise Residential District, Category 2 Exclusive Mid-high Rise Residential District, Category 2 Residential District, Commercial District, Exclusive Industrial District, Industrial District, Near-commercial District, Outside Urban Planning Area, Semi-industrial District, Semi-residential District, Semi-urban Planning Area, or Urbanization Control Area.

We may further categorize them into subclasses: Residential, Commercial, Industrial, and Other ('Urbanization control area', 'Non-divided city planning area', 'Quasi-city planning area', and 'Outside city planning area'). The number of properties for each land use category is summarized in Table 45.

Table 45: Number of properties for each land use category

Land Use	*Land & Building*	*Land Only*	*Condo*
Residential area	76303	48011	50178
Commercial area	20979	12324	71237
Industrial area	15339	8467	26260
Urbanization control area	140	231	11
NA	121	90	1652
Total	112882	69123	149338

4.6 Information ignored in our analysis

4.6.1 Region

For all condos, the region information is unknown. For 'land only' and 'land & building', the region can be one of the following: residential area, commercial area, potential residential area, or industrial area.

4.6.2 Layout

The layout information is only known for condos. This can take the following values: NA, 3LDK, 4LDK, 5LDK, 2LDK, 1K, 2DK, 1R, 3DK, 1LDK, 1DK, Open Floor, 2LDK+S, 2K, 3LDK+S, 2DK+S, 3LD, 1LDK+S, 1DK+S, 4DK, 4LDK+S, 2L, 3LK, 5LDK+S, Studio Apartment, 2LK, Duplex, 3K, 3DK+S, 4K, 2K+S, 1R+S, 3K+S, 1LK, 5DK, 1L, 6DK, 1K+S, 7LDK, 6LDK, 4DK+S, 2LK+S, 3LD+S, 2LD+S, 8LDK, 4L, 4LDK+K, 4L+K, 3D, 6LDK+S.

We do not use this information since it is difficult to categorize and already we have floor area in our selected data.

4.6.3 Unit land price

According to the MLIT website, the unit price (10,000 yen) per m^2 is provided only for 'land only'. This unit price is obtained by dividing the total transaction value of each plot of land by the land area (m^2). The variable is not available for 'land & building' and condos, so we do not use this information.

4.6.4 Land shape

Only for 'land only' and 'land & building' the general shape of land is provided. The land shape can take the following forms: NA, rectangular shaped, semi-square shaped, semi-rectangular shaped, irregular Shaped,semi-trapezoidal shaped, trapezoidal shaped, semi-shaped, square-shaped, flag-shaped, etc.

4.6.5 Frontage of land / frontage road

For 'land only' and 'land & building', the frontage/width of land (in m) is provided, that is, the length of land in contact with a frontage road, as well as the width (in m), type, and direction of the road in contact with the land.

4.6.6 Building use

For the buildings on residential land (land and building) and exclusively owned areas of pre-owned condominiums, the current usage is provided. This can be:

house, shop, other, office, housing complex, parking lot, factory, warehouse, workshop, or any combination of these. We do not include this variable since there are too many interactions and it would be difficult to categorize.

4.6.7 Purpose of use

Purpose of use is provided only for records where the transaction period is after 1st quarter 2013. For these records, purpose of use can be: NA, house, shop, other, office, warehouse, factory, among which 'house' is the majority (45,373 among 53,596 records where purpose of use is known). We do not use this information because of its limited availability.

4.6.8 Additional remarks that might impact housing price

According to the MLIT website, remarks are provided when there is additional information that may have impact on transaction prices. These are provided only when relevant additional information is obtained via a questionnaire survey. Out of 158,474 records where remarks are provided, 137,490 are condos. The remarks can be: NA, dealings of non-redecorating real estate, dealings of redecorated real estate, dealings with auction or arbiter participation, dealings including private road, dealings between related objects, dealings including special circumstances, dealings of adjacent land, dealings of real estate that includes damage, dealings of real estate with mortgage issues, dealings including a valueless house, or a combination of these items.

5 Macro-economic variables

Housing prices are affected by general economic conditions. In order to incorporate possible effects of these economic conditions, we include the following macro-economic indicators as explanatory variables: GDP, CPI, interest rate and stock price.

Table 46: Macro-economic variables

Name	Description	Frequency	Source
GDP	Nominal (not seas. adj.)	Quarterly	Cabinet office
CPI	All items, 2015-base	Monthly	Statistics Bureau
Interest rate	Basic discount rate	Quarterly	Bank of Japan
TOPIX	Tokyo Stock Price Index	Monthly	Cabinet Office

GDP figures are provided by the cabinet office website

`https://www.esri.cao.go.jp/index-e.html`.

We use the nominal GDP series, not seasonally adjusted. The reason that we not adjust for quarter is that we include many other controls in the analysis that vary over quarters. If we would take seasonally adjusted GDP series, then the determinants of the seasonal adjustment and our own control variables would become confounded.

The monthly CPI data can be downloaded from

`https://www.stat.go.jp/english/data/cpi/index.htm`.

We use the version released on August 26, 2016, which is 2015-based. Since our housing price data are per quarter, we integrate the monthly data into quarterly data using simple averages.

Interest rate is one of the most important factors that have an impact on house prices. We use quarterly time-series data of the 'basic discount rate',

`https://www.stat-search.boj.or.jp/index_en.html#`,

provided by the Bank of Japan.

Stock prices reflect business conditions, which might affect the housing market. We download monthly data of the Tokyo Exchange Tokyo Price Index (TOPIX) from the ESRI website

`https://www.esri.cao.go.jp/en/stat/di/di-e.html`.

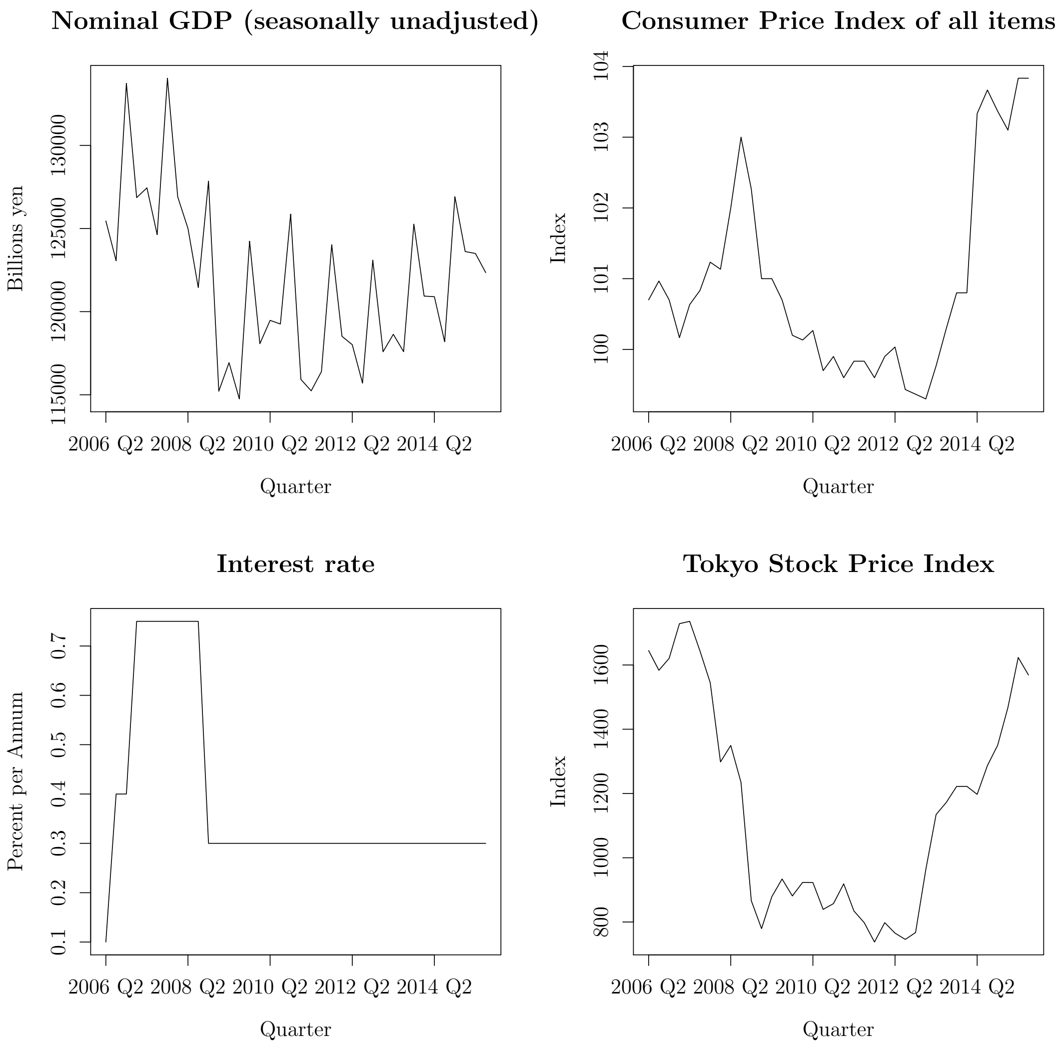


Figure 2: Japan macro-economic data series, 2006 Q2 – 2015 Q3

6 Historical earthquakes

6.1 Data source

The Japan Meteorological Agency (JMA) employs a seismic intensity scale to measure the intensity of earthquakes. It is measured in units of ‘shindo’ (seismic intensity). The JMA scale differs from the more common Richter scale (and the Moment Magnitude Scale), which measure magnitude, that is, the energy released by the earthquake. In contrast, the JMA scale describes the degree of shaking. The intensity of an earthquake is not completely determined by its magnitude, but varies with the event’s depth and the distance from the event. For example, an earthquake may be described as shindo 4 in Tokyo, shindo 3 in Yokohama, and shindo 2 in Shizuoka. The JMA operates a network of 180 seismographs and 627 seismic intensity meters and provides real-time earthquake reports to the media and on the Internet.

The JMA data can be downloaded from

`https://www.data.jma.go.jp/svd/eqev/data/bulletin/shindo_e.html`.

Each year’s earthquake data are stored in separate `.dat` files, which provide record entries of two possible types:

(1) Hypocenter record; and

(2) Seismic intensity and acceleration data record.

A type (1) record contains the information we need for modeling the earthquake process: date, time, exact location, depth, magnitude, intensity, etc. Following each type (1) record there are one or more type (2) records, which contain descriptions from seismic detection stations about this earthquake.

The data are available for the years 1923–2015. A subset of the data can be selected with dropdown menus from

`https://www.data.jma.go.jp/svd/eqdb/data/shindo/index.php`.

These interactive tables display the hypocenter records that are natural earthquakes (specified as ‘1’ in column ‘Subsidiary information’) with known maximum intensity and known hypocenter (specified in column ‘Identifiers’, which cannot be ‘N: Hypocenter unknown’).

6.2 Description

The description of all the earthquake parameters can be found on `https://www.data.jma.go.jp/svd/eqev/data/bulletin/data/shindo/format_e.txt`.

In each type (1) record, the following information is available:

- Record type identifier (A: hypocenter record, B: hypocenter record (for two or more spatio-temporally close earthquakes whose seismic intensity data cannot be separated), D: hypocenter record (for two or more temporally close earthquakes whose seismic intensity data cannot be separated)
- Year, month, day, hour, minute, second of origin time (Japan Standard Time = UTC + 9 hours)
- Standard error (of origin time, in seconds)
- Latitude of hypocenter (degrees and minutes)
- Standard error for latitude (minutes)
- Longitude of hypocenter (degrees and minutes)
- Standard error for longitude (minutes)
- Depth in kilometers
- Standard error for depth (kilometers)
- Magnitude 1
- Magnitude type (1),

  JMA magnitudes: J: MJ - Local Meteorological Office magnitude; D: MD - Displacement magnitude; d: Md - As per MD, but for two stations; V: MV - Velocity magnitude; v: Mv - As per MV, but for two or three stations.

  Moment magnitudes: W: MW - Moment magnitude based on the JMA centroid moment tensor solution.

  Other organizations' magnitudes: B: mb - USGS body wave magnitude; S: MS - USGS surface wave magnitude.
- Magnitude 2
- Magnitude type 2 (see magnitude type 1)
- Travel time table type

- Hypocenter location precision (1: Depth-free method; 2: Depth-slice method; 3: Fixed depth; 4: Based on depth phase; 5: Based on S-P time; 7: Poor solution; 8: Undetermined or not accepted)

- Subsidiary information on event (1: Natural earthquake; 2: Insufficient number of JMA stations; 3: Artificial event; 4: Noise; 5: Low-frequency earthquake)

- Maximum intensity (1: One; 2: Two; 3: Three; 4: Four; 5: Five (until September 1996); 6: Six (until September 1996); 7: Seven; A: Five lower; B: Five upper; C: Six lower; D: Six upper; R: Remarkable earthquake (shock felt over 300 km away) (until 1977); M: Moderate earthquake (shock felt over 200 km away but not over 300 km away) (until 1977); S: Small earthquake (shock felt over 100 km away but not over 200 km away) (until 1977); L: Local earthquake (shock felt less than 100 km away) (until 1977); F: Felt earthquake (until 1984); X: Shock felt by some people but not by JMA observers (until September 1996))

- Damage class (after Utsu) (1: Slight damage (cracks on walls and ground); 2: Light damaged (broken houses, roads, etc.); 3: 2–19 fatalities or 2–999 houses destroyed; 4: 20–199 fatalities or 1,000–9,999 houses destroyed; 5: 200–1,999 fatalities or 10,000–99,999 houses destroyed; 6: 2,000–19,999 fatalities or 100,000–999,999 houses destroyed; 7: 20,000+ fatalities or 1,000,000+ houses destroyed; X: Injury or damage of unclear scale (until 1988); Y: Injury and damage included in the grade for the preceding or following event (until 1988))

- 1929–1988 Tsunami class (after Utsu)

- Number of epicenter location district

- Number and name of epicenter location region

- Number of shocks felt

- Identifiers (K: JMA hypocenter identified with high precision; S: JMA hypocenter identified with low precision; N: Hypocenter unknown (first observation point used); U: USGS hypocenter; I: ISC hypocenter; R: Preliminary hypocenter (included only in district observatory databases); H,D,M: Exact observation time unknown)

6.3 JMA intensity scale

The tables detailing the JMA intensity scales can be found in

`https://www.jma.go.jp/jma/en/Activities/inttable.html`.

An excerpt of the description of each intensity level is shown below.

Human perception and reaction, indoor situation, outdoor situation

Seismic intensity	Human perception and reaction	Indoor situation	Outdoor situation
0	Imperceptible to people, but recorded by seismometers.	-	-
1	Felt slightly by some people keeping quiet in buildings.	-	-
2	Felt by many people keeping quiet in buildings. Some people may be awoken.	Hanging objects such as lamps swing slightly.	-
3	Felt by most people in buildings. Felt by some people walking. Many people are awoken.	Dishes in cupboards may rattle.	Electric wires swing slightly.
4	Most people are startled. Felt by most people walking. Most people are awoken.	Hanging objects such as lamps swing significantly, and dishes in cupboards rattle. Unstable ornaments may fall.	Electric wires swing significantly. Those driving vehicles may notice the tremor.
5 Lower	Many people are frightened and feel the need to hold onto something stable.	Hanging objects such as lamps swing violently. Dishes in cupboards and items on bookshelves may fall. Many unstable ornaments fall. Unsecured furniture may move, and unstable furniture may topple over.	In some cases, windows may break and fall. People notice electricity poles moving. Roads may sustain damage.
5 Upper	Many people find it hard to move; walking is difficult without holding onto something stable.	Dishes in cupboards and items on bookshelves are more likely to fall. TVs may fall from their stands, and unsecured furniture may topple over.	Windows may break and fall, unreinforced concrete-block walls may collapse, poorly installed vending machines may topple over, automobiles may stop due to the difficulty of continued movement.
6 Lower	It is difficult to remain standing.	Many unsecured furniture moves and may topple over. Doors may become wedged shut.	Wall tiles and windows may sustain damage and fall.
6 Upper	It is impossible to remain standing or move without crawling. People may be thrown through the air.	Most unsecured furniture moves, and is more likely to topple over.	Wall tiles and windows are more likely to break and fall. Most unreinforced concrete-block walls collapse.
7		Most unsecured furniture moves and topples over, or may even be thrown through the air.	Wall tiles and windows are even more likely to break and fall. Reinforced concrete-block walls may collapse.

Wooden houses

Seismic intensity	Wooden houses	
	High earthquake resistance	Low earthquake resistance
5 Lower	-	Slight cracks may form in walls.
5 Upper	-	Cracks may form in walls.
6 Lower	Slight cracks may form in walls.	Cracks are more likely to form in walls. Large cracks may form in walls. Tiles may fall, and buildings may lean or collapse.
6 Upper	Cracks may form in walls.	Large cracks are more likely to form in walls. Buildings are more likely to lean or collapse.
7	Cracks are more likely to form in walls. Buildings may lean in some cases.	Buildings are even more likely to lean or collapse.

Reinforced-concrete buildings

Seismic intensity	Reinforced-concrete buildings	
	High earthquake resistance	Low earthquake resistance
5 Upper	-	Cracks may form in walls, crossbeams and pillars.
6 Lower	Cracks may form in walls, crossbeams and pillars.	Cracks are more likely to form in walls, crossbeams and pillars.
6 Upper	Cracks are more likely to form in walls, crossbeams and pillars.	Slippage and X-shaped cracks may be seen in walls, crossbeams and pillars. Pillars at ground level or on intermediate floors may disintegrate, and buildings may collapse.
7	Cracks are even more likely to form in walls, crossbeams and pillars. Ground level or intermediate floors may sustain significant damage. Buildings may lean in some cases.	Slippage and X-shaped cracks are more likely to be seen in walls, crossbeams and pillars. Pillars at ground level or on intermediate floors are more likely to disintegrate, and buildings are more likely to collapse.

Figure 3: Explanation of the JMA Seismic Intensity Scale, source: JMA website

6.4 Sample selection

We only extract the type (1) records. In order to use the records for modeling the earthquake process, we may choose samples based on time period, hypocenter location, magnitude or intensity threshold, and depth.

We select only the records that are 'natural earthquakes'. In doing so we discard the records labeled as 'Insufficient number of JMA stations', 'Noise', or 'Low-frequency earthquake' since these records may not be reliable.

We discard the records with unknown hypocenters since the location information is inaccurate; furthermore the magnitudes for these records are also unknown.

We discard the records with unknown 'maximum intensity' since these records are earthquakes that are spatio-temporally close to another earthquake so that they cannot be separated.

Note that even with the same sample selection criteria as described in the literature, the resulting sample catalog can be quite different. This is because the JMA catalog has been updated (modified) many times as technology and knowledge have improved; see

`https://www.data.jma.go.jp/svd/eqev/data/bulletin/data/hypo/relocate.html`

for further discussion (Japanese only).

6.5 Summary statistics

There are 194,882 records in the entire 1923–2015 period, among which 105,685 records are natural earthquake with known hypocenter and known maximum intensity. Table 47 contains a summary of these records.

Table 47: Summary of JMA earthquake records for 1923–2015

Magnitude	Intensity	n
Magnitude<3	<=4	26015
3<=Magnitude<4	<=4	33464
	5/5-lower/5 upper	4
4<=Magnitude<5	<=4	22845
	5/5-lower/5 upper	57
5<=Magnitude<6	<=4	7085
	5/5-lower/5 upper	157
	6/6-lower/6 upper	7
6<=Magnitude<7	<=4	1283
	5/5-lower/5 upper	131
	6/6-lower/6 upper	24
	7	1
Magnitude>=7	<=4	109
	5/5-lower/5 upper	44
	6/6-lower/6 upper	22
	7	2
Unknown	<=4	14432
	5/5-lower/5 upper	3

In some records two magnitudes (and two magnitude types) are reported. The second magnitude is supplementary information and the difference between the two magnitudes recorded for the same earthquake is usually small. In Table 47 and in other places where magnitudes are needed, we use the first magnitude (Magnitude 1) without further clarification.

7 ETAS estimation and simulation

7.1 The ETAS model

The Epidemic Type Aftershock Sequence (ETAS) model has been introduced by Ogata (1988), and it has been widely used to capture the quiescence and activation of seismic activities. The underlying idea is that each earthquake may trigger a sequence of aftershocks like 'epidemics' and that the severity of influence diminishes over time (and distance). While there have been many space-time extensions of the Ogata model, we choose the temporal version of this model for simplicity of estimation and simulation. Indeed, with five cities, this temporal treatment is natural and simpler than first estimating a space-time version to a large area that covers all five cities and next isolating the city effects. The temporal version can be briefly described as follows.

Given the observations of earthquake occurrences at time $t_1, t_2, \ldots, t_n$ on an interval $[0, T]$ ($T \geq t_n$), the associated counting process is defined as

$$N_t = \sum_{i=1}^{n} \mathbb{1}_{t_i \leq t}.$$

The corresponding left-continuous $\mathcal{F}_t$-conditional jump intensity process λ_t describes the mean jump rate per unit of time,

$$\lambda_t = \lambda(t|\mathcal{F}_t) = \lim_{h \downarrow 0} \frac{1}{h} \ \Pr\left[N_{t+h} - N_t > 0 | \mathcal{F}_t\right].$$

In a temporal ETAS model, the conditional intensity function may be written as

$$\lambda_t = \lambda_\infty + \sum_{t_i < t} c(m_i, m_c) g(t - t_i),$$

where λ_∞ (shocks per unit of time) is the background seismicity with $\lambda_\infty > 0$. The aftershock decay (time response function) takes the form of the modified Omori (or Omori-Utsu) function,

$$g(t - t_i) = \frac{K}{(t - t_i + C)^p}.$$

The weight assigned to the aftershock decay is an exponential function of the difference between the magnitude of the earthquake and the cut-off magnitude m_c:

$$c(m_i, m_c) = \exp(\beta(m_i - m_c)).$$

The intensity consists of the background intensity and a weighted sum of all the aftershock decays, where the sum is taken over all earthquakes before time t.

7.2 Estimation and simulation

For the estimation of the ETAS model, we take the earthquake catalog of areas around the five Japanese cities in the period 1970-1-1 to 2015-12-31. The spatial windows for each city are shown below.

Table 48: Spatial window of the earthquake catalog

City	latMin	latMax	lngMin	lngMax
Tokyo	34.0	37.0	138.0	141.0
Osaka	33.5	36.5	134.0	137.0
Nagoya	33.5	36.5	135.5	138.5
Fukuoka	32.0	35.0	129.0	132.0
Sapporo	41.5	45.5	138.5	143.5

Range of housing data

City	latMin	latMax	lngMin	lngMax
Tokyo	35.5	35.8	139.6	139.9
Osaka	34.6	34.8	135.4	135.6
Nagoya	35.0	35.3	136.8	137.0
Fukuoka	33.5	33.7	130.2	130.5
Sapporo	42.9	43.2	141.2	141.5

Table 48 contains the range of latitudes and longitudes for the spatial windows within which we employ the earthquake records to estimate the ETAS model. We also display the range of latitudes and longitudes where the properties in the housing data set are located.

The spatial windows have been chosen such that: (1) the spatial windows amply encompass all the property locations in our data set and the properties are located around the center of the spatial windows for each city; (2) the number of observations within each spatial window is sufficiently large, in the sense that they are roughly comparable across cities and that they yield meaningful results for the ETAS estimation procedure; and (3) the estimation of ETAS models for each city is appropriate, in the sense that they converge and pass the Berman (1983) test of residuals, where we consider whether the transformed inter-arrival times are iid exponential random variables with unit mean.

We see that the spatial windows are relatively large compared to the location of the properties in our housing data set. This can also be seen in the map of Figure 4, where the squares with black borders are the spatial windows for

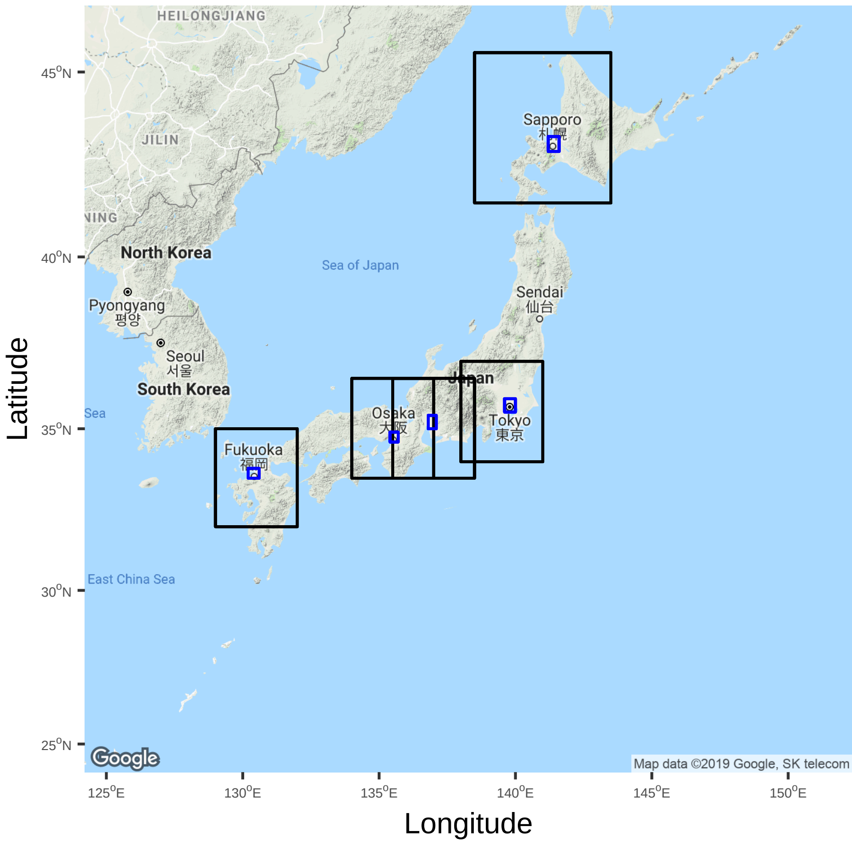


Figure 4: Spatial windows encapsulating each city

selecting the records to enter the estimation, while the squares with blue borders are the windows that closely encapsulate the location of the housing properties.

Due to tectonic and other differences, the areas within each spatial window might not be homogeneous in terms of the intensity functions, but as we interpret the objective short-run probabilities as measures of short-term seismic risk driving people's perceptions of seismic risk in a broader living environment, the potential variation is averaged out. Indeed, it is quite natural to take the living environment (i.e., the spatial window) larger than just the city of interest, as seismic activity just outside the city of interest may also impact the perception of seismic risk. Furthermore, this also helps in reducing the bias of the ETAS parameters due to seismicity originating outside of the extent of the spatial window.

There may be some concern that earthquakes occurring close to the center of the city may affect people's perception of earthquake risk in the near future more severely than earthquakes occurring within but near the border of the spatial window. This effect is captured at least to some extent because earthquakes at the center of the city have an 'impact radius' that overlaps more with the spatial window, so that aftershocks of these earthquakes are included more completely in the spatial window.

To further study the potential variation within the spatial window, we analyze the effect of earthquakes originating from different directions of the spatial window. We subdivide each spatial window into four equal-sized parts: South West (SW), South East (SE), North West (NW), and North East (NE). Then we consider the major earthquakes (i.e., the earthquakes with magnitude > 5.5) within each of the fours parts, and summarize the corresponding average number of major earthquakes (with magnitude > 6.0 for Tokyo and > 5.5 for all other cities) occurring in the (full) spatial window for each city within three months after the initial major shock.

Table 49: Average number of major earthquakes in the full spatial window for each city within 90 days after an initial major shock, categorized by direction of the initial major shock.

City	Direction of origin	No. of initial major quakes	No. of major quakes within 90 days	Average no. of major quakes within 90 days
Fukuoka	NW	2	1	**0.50**
	NE	3	0	**-**
	SE	6	3	**0.50**
	SW	2	0	**-**
Nagoya	NW	2	0	**-**
	NE	5	3	**0.60**
	SE	7	1	**0.14**
	SW	4	1	**0.25**
Osaka	NW	1	0	**-**
	NE	2	0	**-**
	SE	4	1	**0.25**
	SW	8	1	**0.13**
Sapporo	NW	0	0	**-**
	NE	3	0	**-**
	SE	20	3	**0.15**
	SW	6	6	**1.0**
Tokyo	NW	7	0	**-**
	NE	43	19	**0.44**
	SE	19	0	**-**
	SW	34	40	**1.1**

As can be seen from Table 49, the average numbers of major shocks throughout the spatial domain of the city, following an initial major shock within three months, are roughly similar for each direction of origin in a city.

We show below the p-values corresponding to Berman's (1983) Kolmogorov-Smirnov (K-S) test statistics, the number of observations (N) as well as the model parameter estimates for each city. The estimation threshold is chosen to be magnitude 4.5 for Fukuoka, Nagoya, Osaka, and Sapporo; and 5 for Tokyo. This means that in each city, only earthquake records above the corresponding thresholds are used for estimation.

Table 50: ETAS estimation results (with threshold 4.5 for Osaka, Nagoya, Fukuoka, Sapporo and 5 for Tokyo)

city	K-S	N	λ_∞	K	C	p	β
Tokyo	0.764	370	0.0076	0.0351	0.0155	1.0072	0.9253
Osaka	0.738	154	0.0073	0.0014	0.0064	1.1913	2.4854
Nagoya	0.940	177	0.0071	0.0066	0.0009	0.9616	1.7250
Fukuoka	0.982	102	0.0037	0.0098	0.0035	1.0330	1.4390
Sapporo	0.874	486	0.0193	0.0039	0.1044	1.2091	2.4424

The estimation was conducted by employing the 'PtProcess' `R` package (`https://www.jstatsoft.org/article/view/v035i08`), which uses maximum likelihood estimation. With a suitable adaptation of the code, the Hessian can also be obtained, and the standard errors of the parameters are then computed as the square roots of the diagonal elements of the inverted Hessian. The output of the estimation procedure are five parameter estimates, for: $\log(\mu)$, $\log(A)$, $\log(\alpha)$, $\log(C)$, and $\log(p)$. The final parameters that appear in the ETAS model are a transformation: $\lambda_\infty = \mu$, $K = A \cdot C^p$, C, p, and $\beta = \alpha$.

Table 51: Standard errors of the ETAS parameters

city	N	$\log(\lambda_\infty)$	$\log(A)$	$\log(\beta)$	$\log(C)$	$\log(p)$
Tokyo	370	-4.881	0.845	-0.078	-4.165	0.007
(st.err.)		0.179	0.432	0.248	0.603	0.043
Osaka	154	-4.918	-0.560	0.910	-5.045	0.175
(st.err.)		0.096	0.872	0.095	0.935	0.098
Nagoya	177	-4.954	1.677	0.545	-6.965	-0.039
(st.err.)		0.189	0.927	0.190	1.185	0.090
Fukuoka	102	-5.591	1.228	0.364	-5.669	0.032
(st.err.)		0.180	0.886	0.214	1.082	0.084
Sapporo	486	-3.948	-2.803	0.893	-2.259	0.190
(st.err.)		0.069	0.451	0.051	0.456	0.059

We show the standard errors of each model in Table 51. We see that the parameters λ_∞, β, and C have relatively small standard errors in nearly all cases, while A and p have somewhat larger standard errors. Note that $\log(p)$ is not significantly different from 0 for Fukuoka, Nagoya, and Tokyo. In fact, this is a well-known phenomenon. Ogata (1988) shows that the best ETAS model fitted to his data, among several possible restrictions, is to set $p = 1$. (This is called a pure Omori's law.) We did not impose this restriction in our estimation, but if we had it would not have made much difference.

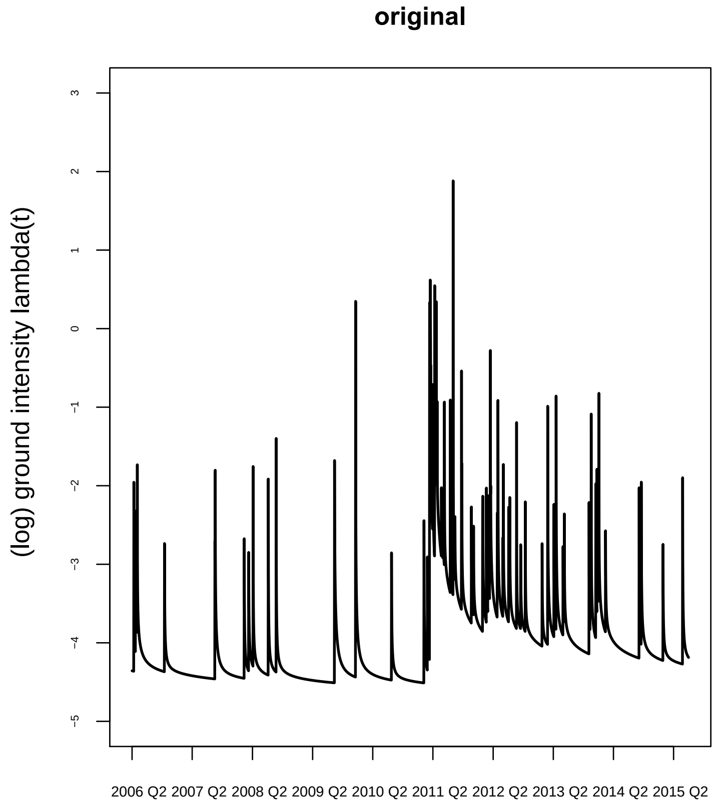
Moreover, and perhaps even more importantly from the perspective of the regression design, we analyze in Figures 5–7 the impact of changes in the ETAS parameter values on the object of interest, the (log) intensity function λ_t. (We focus attention on Tokyo; conclusions for the other cities are similar.) In the figure we draw 11 lines: one with the estimated parameter values and then for each of the five parameters one standard deviation away from its estimated value. We conclude that the variation in the log intensity function due to reasonable perturbations of the estimated parameters is relatively small, which is reassuring for the robustness of the estimation results.

In our estimation of the ETAS model we restrict attention to large earthquakes exceeding a magnitude of 4.5 for Fukuoka, Nagoya, Osaka, and Sapporo and a magnitude of 5 for Tokyo. Using a fixed high threshold for estimation can in a sense be interpreted as a very basic deterministic denoise (or even declustering) device. Separating triggering earthquakes from other earthquakes is a difficult problem, and within the temporal ETAS model that we employ it is in principle not required to explicitly distinguish between triggering earthquakes and other earthquakes. As is well-known, declustering for the identification of background earthquakes (or mainshocks) does not have a unique solution. In particular, seismicity declustering can lead to ambiguous results when using different methods. In a spatial-temporal model, rather than the purely temporal ETAS model based on high threshold exceedances we employ, there would have been a clearer need for more sophisticated declustering methods (and residual analysis) such as those of Zhuang *et al.* (2002) (see also Schoenberg, 2003, Zhuang, 2006, van Stiphout *et al.*, 2012, Zhuang *et al.*, 2012).

We also analyze the completeness of our earthquake catalogs, as another verification of the ETAS estimation. That is, we estimate a magnitude of completeness M_c for each earthquake catalog that we use for the estimation of the ETAS model. Here, M_c is defined as the lowest magnitude at which 100% of the earthquakes in a space-time volume are detected. For this analysis, we use a catalog-based method referred to as the Maximum Curvature (MAXC) technique (Wiemer and Wyss, 2000) and surveyed in Mignan and Woessner (2012).

We obtain mean estimates of M_c from 2000 bootstrap samples of the earthquake catalogs for each of the 5 cities that we consider; see Table 52. We use the `R` functions provided in Mignan and Woessner (2012).

We further illustrate in Figures 8–12 the completeness of the catalogs by plotting each event magnitude against its catalog sequential number, as in Zhang *et al.* (2012). We also display the magnitude threshold used in our estimation of the ETAS model, which is 4.5 for Osaka, Nagoya, Fukuoka, Sapporo and 5 for Tokyo. As expected in view of the relatively recent earthquake catalogs, we confirm their completeness. In particular, our analysis confirms that the estimated magnitude



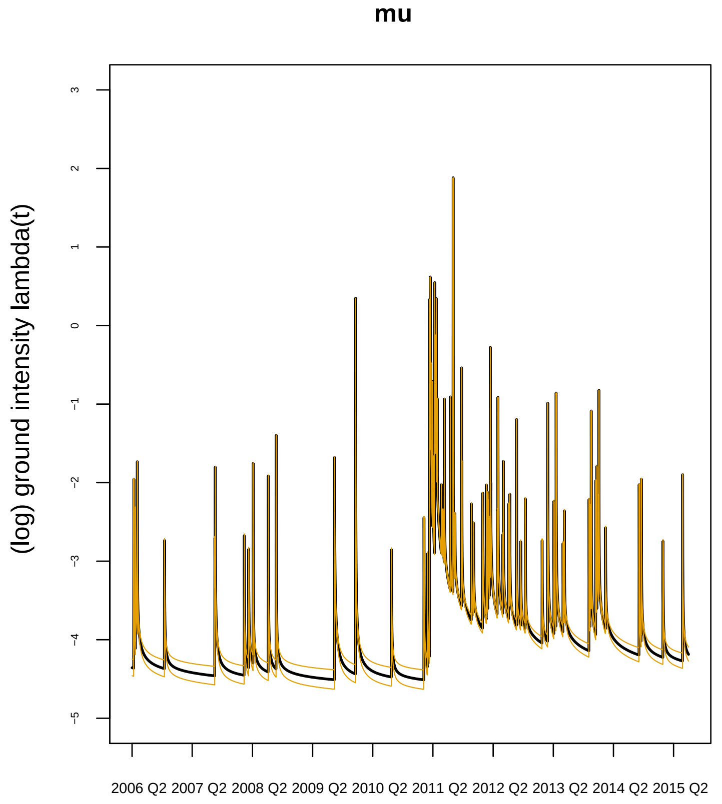
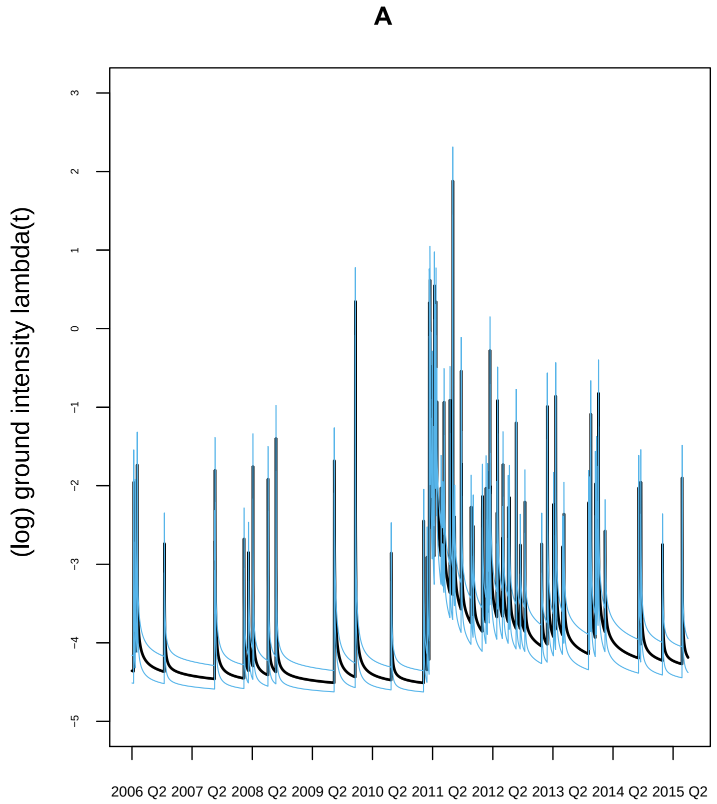


Figure 5: Variation in ETAS parameters: Tokyo



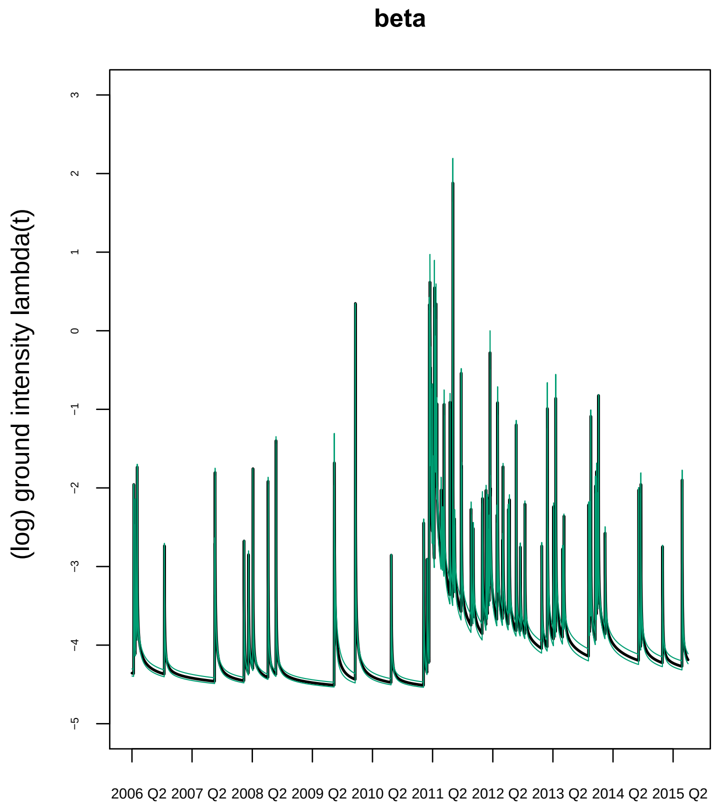
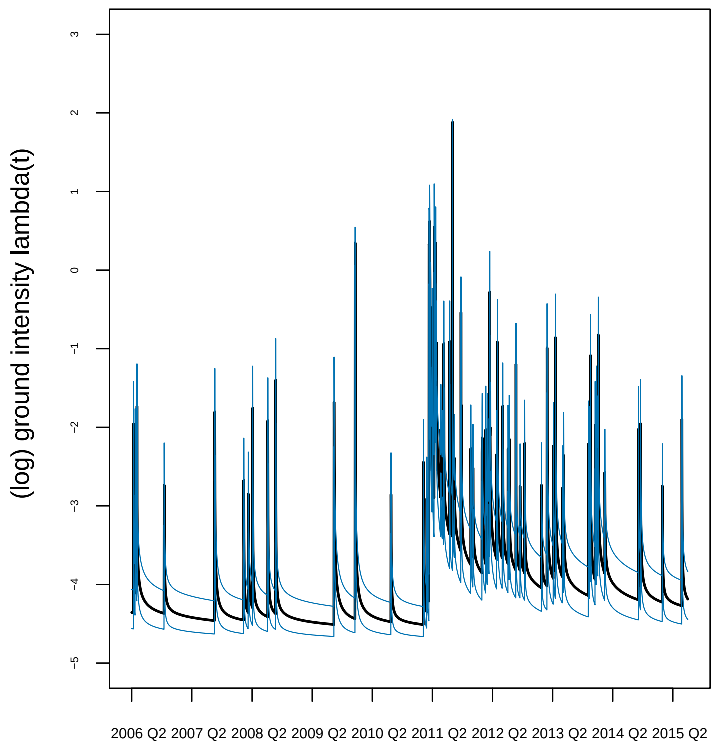


Figure 6: Variation in ETAS parameters: Tokyo (continued)

c



p

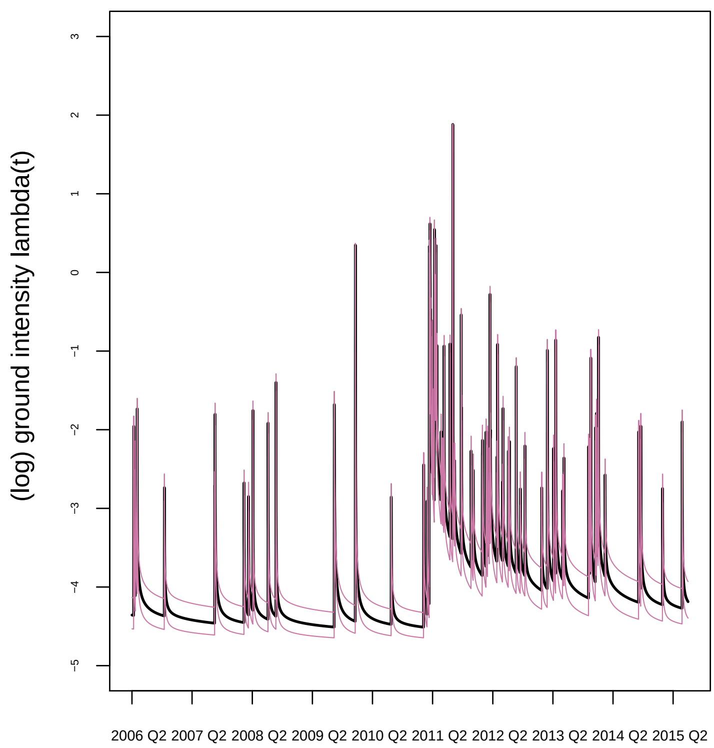


Figure 7: Variation in ETAS parameters: Tokyo (continued)

Table 52: Summary of the M_c estimates for each city

city	M_c mean	M_c std	threshold
Tokyo	2.9109	0.0576	5.0
Osaka	2.7250	0.1567	4.5
Nagoya	2.6890	0.1353	4.5
Fukuoka	2.7603	0.0954	4.5
Sapporo	3.7315	0.1091	4.5

of completeness M_c is well below the threshold we have selected to estimate the temporal ETAS model for each city.

The estimated intensity and actual earthquake events are used to simulate 90-days probabilities of an earthquake exceeding the magnitude threshold of 5.5 for each city. More specifically, with the model estimated to the data, we simulate sample paths from the model, 90-days forward, using the actual earthquake events up to and including the previous quarter. For this purpose, we follow the simulation method of Ogata (1981). The objective short-run earthquake probabilities are then obtained by averaging the simulated frequencies of exceeding the magnitude threshold of 5.5 in the next 90 days for each city.

In Figures 13 and 14 we plot the earthquake intensities along with the corresponding short-run probability series for two of the five cities, Tokyo and Nagoya. As apparent from the figures, the short-run probabilities spike up immediately after a large earthquake and die out gradually until another earthquake occurs.

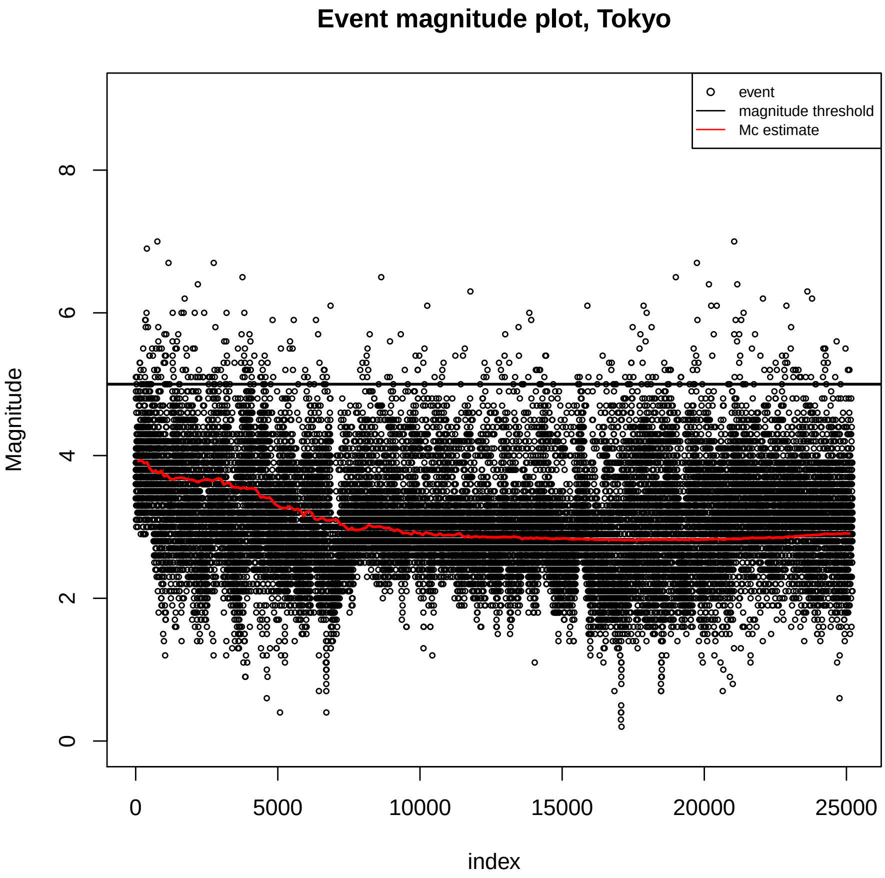


Figure 8: Plot of event magnitude against catalog sequential number, Tokyo

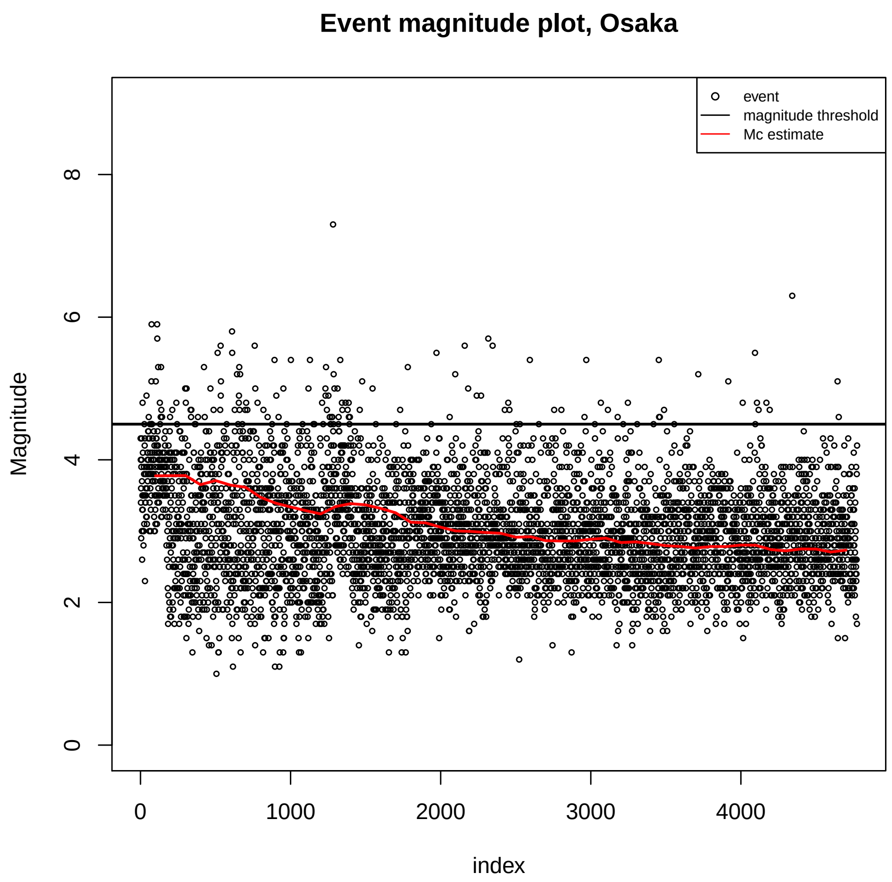


Figure 9: Plot of event magnitude against catalog sequential number, Osaka

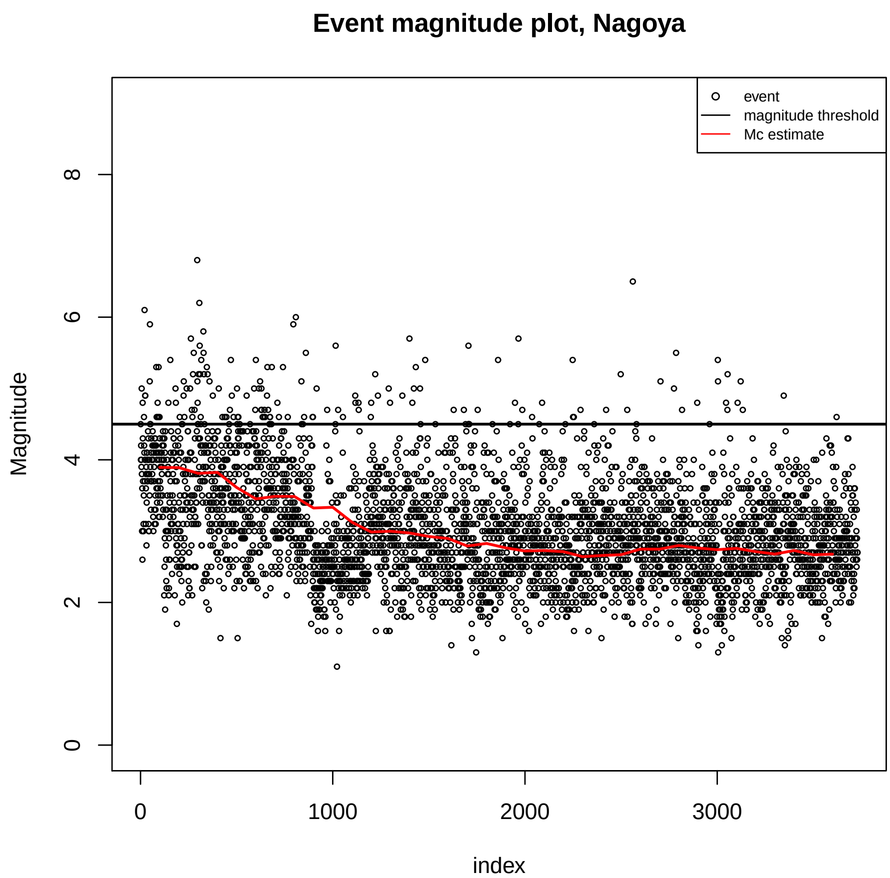


Figure 10: Plot of event magnitude against catalog sequential number, Nagoya

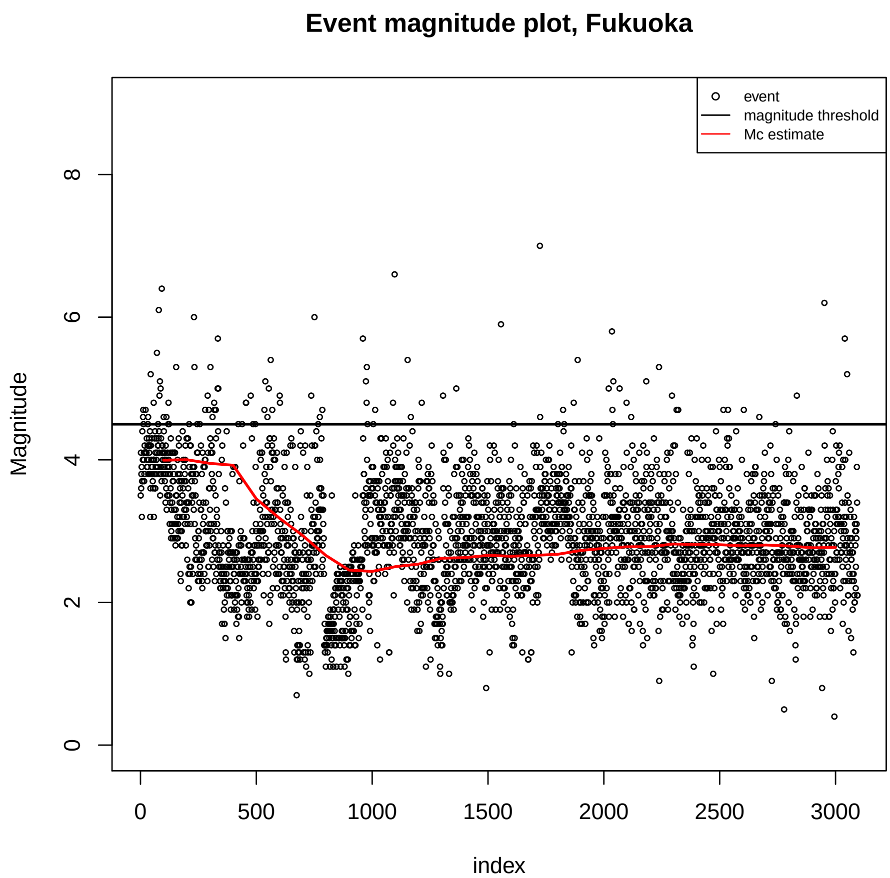


Figure 11: Plot of event magnitude against catalog sequential number, Fukuoka

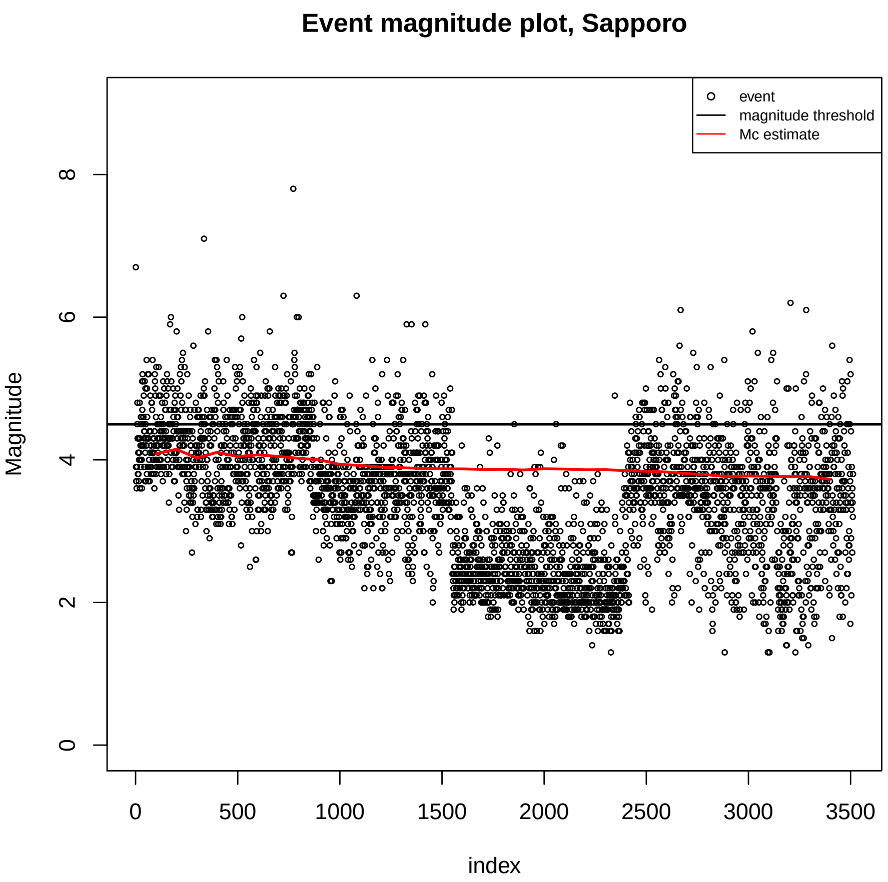


Figure 12: Plot of event magnitude against catalog sequential number, Sapporo

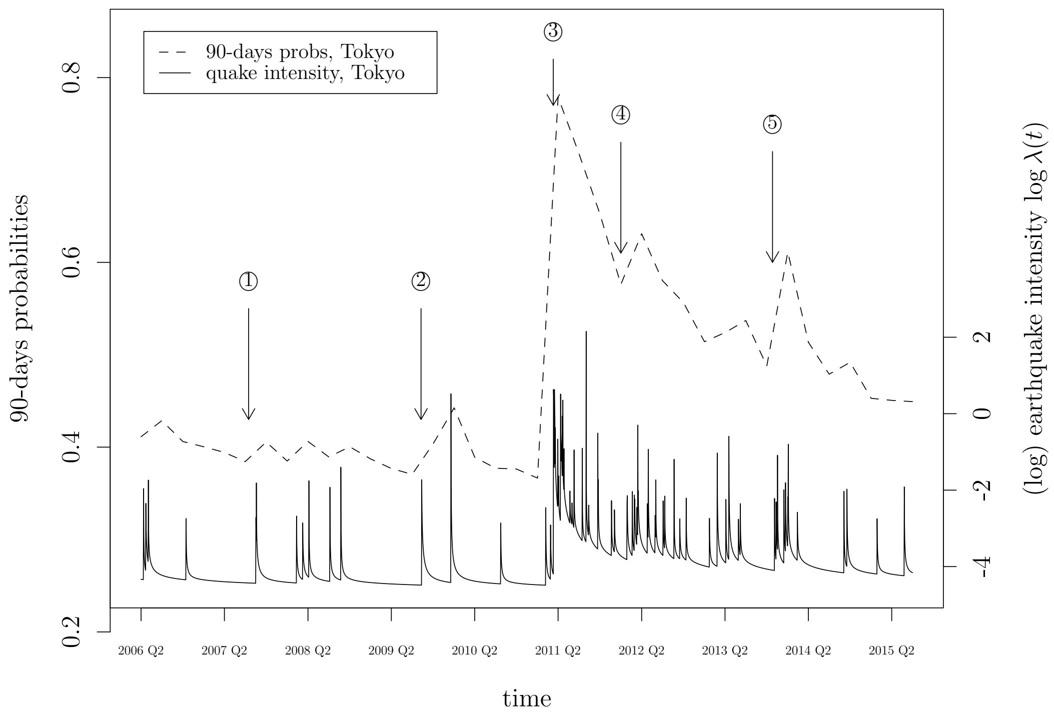


Figure 13: Short-run earthquake risk: Simulated short-run earthquake probabilities and the logarithm of the earthquake intensity series for Tokyo. Events marked in the graph: ①: 2007-07-16 Chuetsu Offshore earthquake, $M6.8$. ②: 2009-08-09 Izu Islands earthquake, $M6.8$ and 2009-08-11 Shizuoka earthquake, $M6.5$. ③: 2011-03-11 Tohoku earthquake, $M9.0$. ④: 2012-01-01 Izu Islands, $M7.0$. ⑤: 2013-10-26 Fukushima-ken oki earthquake, $M7.1$. (Source: Japan Meteorological Agency.)

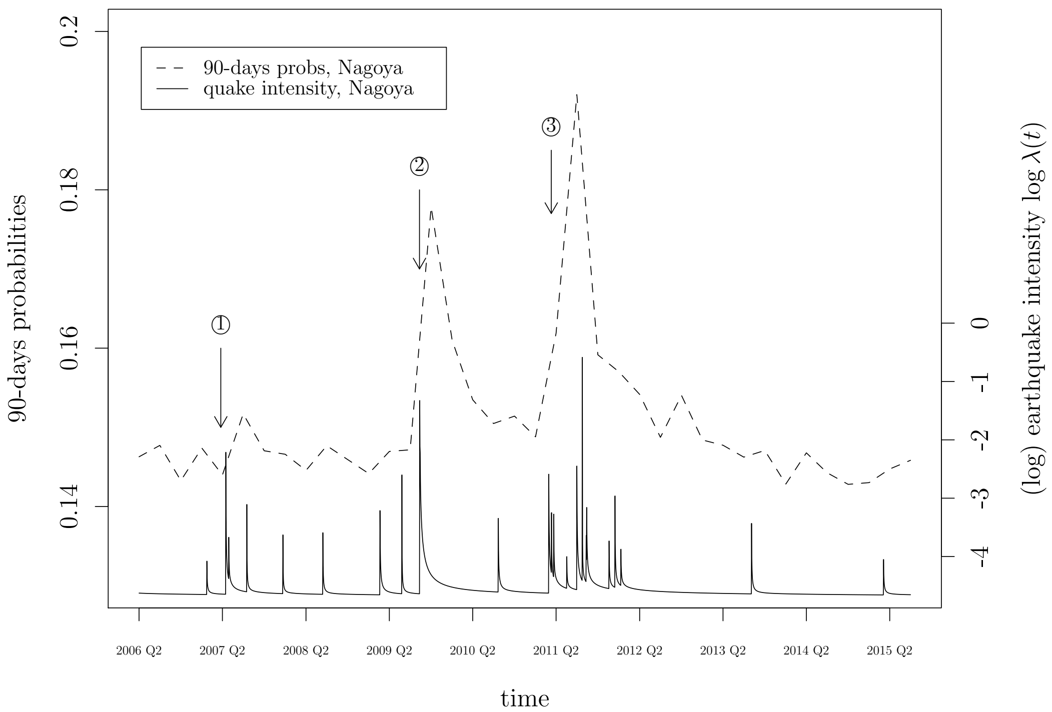


Figure 14: Short-run earthquake risk: Simulated short-run earthquake probabilities and the logarithm of the earthquake intensity series for Nagoya. Events marked in the graph: ①: 2007-03-25 Noto Hanto earthquake, $M6.9$. ②: 2009-08-11 Shizuoka earthquake, $M6.5$. ③: 2011-03-11 Tohoku earthquake, $M9.0$. (Source: Japan Meteorological Agency.)

8 Mesh codes

The Statistics Bureau of Japan explains the definition of mesh code in 'Standard Grid Square and Grid Square Code Used for the Statistics'; see the following webpage

`https://www.stat.go.jp/english/data/mesh/02.htm`.

There are five levels of precision:

- First mesh ('Primary Area Partition') is obtained by dividing the whole area of Japan into blocks measuring 1 degree of longitude and 2/3 degree of latitude.
- Second mesh ('Secondary Area Partition') is obtained by dividing first mesh areas into 8×8 squares.
- Third mesh ('Third Area Partition / Basic Grid Square') is obtained by dividing second mesh areas into 10×10 squares.
- Half/quarter mesh is obtained by dividing the third mesh grid into 2×2 or 4×4 equal parts to get half or quarter grid squares, respectively.

Table 53: Levels of mesh codes

Level	No. digits	Format	Scale
First mesh	4	XXXX	80km
Second mesh	6	XXXX–XX	10km
Third mesh	8	XXXX–XXXX	1km
Half mesh	9	XXXX–XXXX–X	500m
Quarter mesh	10	XXXX–XXXX–X–X	250m

The coding and scale of the five levels is given in Table 53.

In Figure 15 we provide a map of all first meshes.

In Figure 16 we show how mesh code numbers are calculated from the coordinates, based on

`https://www.stat.go.jp/english/data/mesh/05-1.htm`.

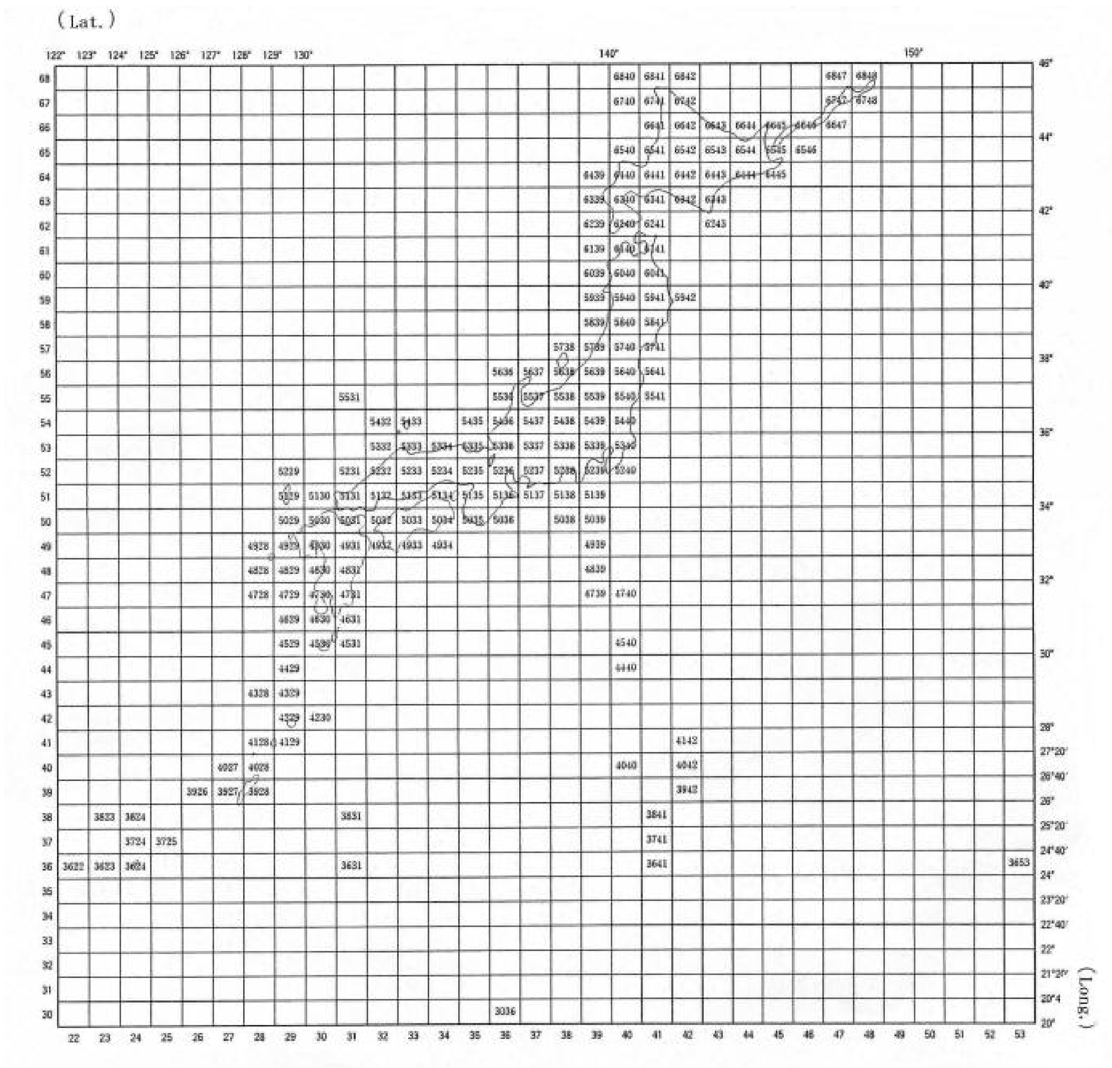


Figure 15: Map of first meshs

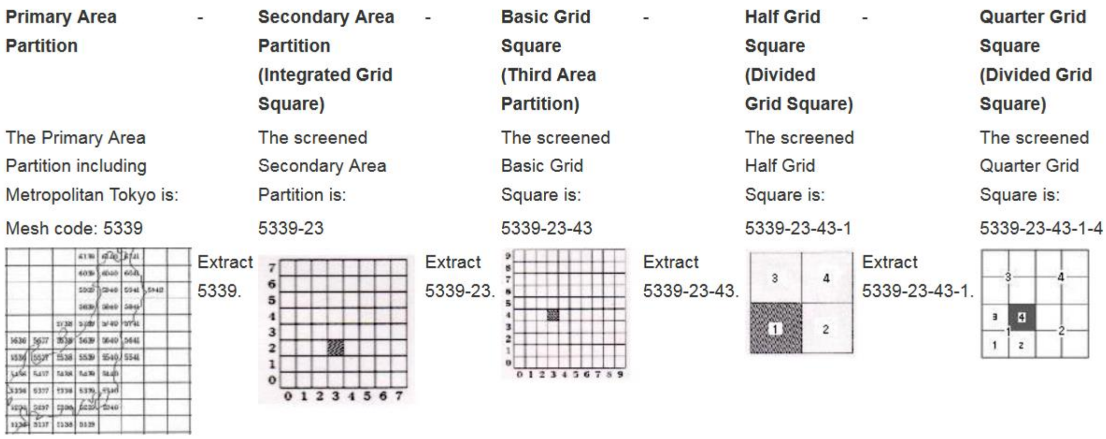


Figure 16: Method of calculating mesh code numbers from coordinates.

9 Predicted earthquake risks

9.1 The J-SHIS data set

While we considered data on actual earthquakes in the previous section, the current section deals with data on earthquake risk. Earthquake risk has several dimensions. What matters is not only how likely it is that there is an earthquake, but also how much damage the resulting fire will cause. Houses (since 1981) are essentially earthquake-proof, but they may not be fire-proof. When in a Japanese text it says 'earthquake risk' the meaning can be ambiguous. It may mean only earthquake, but more likely it means 'earthquake and related risks combined'.

The Japan Seismic Hazard Information Station (J-SHIS) was established to help prevent and prepare for earthquake disasters The seismic hazard maps serve as a sharing platform of seismic hazard information by regarding the maps as a group of information incorporating the underlying data used in the evaluation process, such as the seismic activity models, seismic source models, and subsurface structure models, rather than mere maps as final products. Operations started in May 2005.

Four years later, in 2009, it was decided to incorporate the latest technology and a new J-SHIS system was launched. The new J-SHIS manages various data in an integrated manner. It includes the new National Seismic Hazard Maps for Japan which consist of the Probabilistic Seismic Hazard Maps (PSHM) for Japan with a 250 m mesh resolution and the Scenario Earthquake Shaking Maps (SESM) based on detailed strong motion estimation of earthquakes occurring at major active fault zones, as well as the deep subsurface structure models for Japan and 250 m mesh geomorphological land classification models used for the required calculations. The new J-SHIS also provides these data in a user-friendly manner by superposing them on background maps. The new J-SHIS is a web mapping system based on open source software which allows general users to easily view various data on their Internet browsers. Especially, the notable new functions have enabled the users to overlap the seismic hazard maps on Google maps including the layer transparency function, to freely move and zoom in and out the maps, to view the seismic hazard maps with a 250 m mesh resolution, to search a precise location by addresses and postal codes, to select and show a source fault on the browser, and to display attribute values for each mesh. The new system has been in operation since July 2009.

Responding to the 2011 off the Pacific Coast of Tohoku Earthquake, studies are being made on improvement of J-SHIS.

9.2 Data source

The relevant J-SHIS data is the 'probability of exceedance'. According to

`http://www.j-shis.bosai.go.jp/en/glossary`,

this term means:

> *The probability that shakes will exceed a certain level of intensity at a point for a certain time period (over the next 30 or 50 years, in this guidebook). For example, the 'Map of ground motions of seismic intensity for a 3% probability of exceedance occurring within 30 years from the present' means the probability that each point is affected by shakes exceeding its seismic intensity shown on the map is 3% within 30 years from the present.*

The probabilities are either 'average case' or 'maximum case'.

'Average case' means:

> *In the long-term evaluation for the 98 major active fault zones, there are many cases where both mean recurrence interval and the latest event have been evaluated with a range of values. In the preparation of the seismic hazard map, the result of calculating the probability of occurrence by using the mid-values of individual ranges is called 'Average case',*

while 'Maximum case' means:

> *The result of calculating the probability of occurrence by using the smallest value of mean recurrence interval and the oldest value of the latest event is the highest probability.*

Two analytical periods are typically given: 30 years (T30) and 50 years (T50). In our data set we confine ourselves to the 'average case' and an analytical period of 30 years.

The J-SHIS data are available in two different formats for two subperiods. For 2008–2014, the probability of earthquakes happening in the coming 30 or 50 years exceeding certain intensity thresholds can be downloaded using `Python` from

`http://www.j-shis.bosai.go.jp/map/JSHIS2/download.html?lang=en`

under tab 'Dataset/Probablistic Seismic Hazard Maps/Seismic Hazard Map'. The records come in separate `.csv` files for each year and each first mesh. We downloaded 8 first meshes: Tokyo `5339`; Osaka `5135` and `5235`; Nagoya `5235`, `5236`, and `5237`; Fukuoka `4930` and `5030`; Sapporo `6441`. Each entry corresponds with a 1/4 mesh area (250m), and the thresholds are 5-lower (denoted as I45), 5 upper (denoted as I50), 6-lower (I55) and 6 upper (I60). There can be multiple models in one year. In 2012 there are data for two models and in 2013 there are data for three models. We use model 1 by default in both years.

For 2005–2008, the PSHM map data can be downloaded from

`http://wwwold.j-shis.bosai.go.jp/j-shis/index_en.html`,

one zip file each year for the whole of Japan. The thresholds can be chosen as 6-lower or 5-lower. Each entry corresponds with a third mesh area (1km). The probabilities are only available for an analytical period of thirty years. There is some overlapping in the two sources: both the new and the old system provides data for year 2008. Although in different formats, these data coincide (maximum difference is 5e-7, which is just rounding error). For further analysis we use the new data source for the year 2008.

9.3 Summary statistics

In summary, there are annual data for two subperiods: 2005–2008 and 2008–2016. (The year 2008 appears in both sets.) In the first period each mesh is about 1 square km (third mesh); in the second period each mesh is about 250 square meters (quarter mesh).

Seismic intensity is given on a scale from 0 to 7, where 5 and 6 are further divided into 5-lower, 5-upper, 6-lower, and 6-upper. In the second period the data provide:

- probability of exceedance larger than 5-lower (30 years);
- probability of exceedance larger than 5-upper (30 years);
- probability of exceedance larger than 6-lower (30 years);
- probability of exceedance larger than 6-upper (30 years).

In the first period, only two of these probabilities are provided, namely 5-lower and 6-lower. Summary statistics of the hazard probabilities for all the stations in the housing sample are provided below.

Table 54: Summary statistics of seismic hazard probabilities (for each unique district in the housing sample), averaged over 2005–2014

City	mean	min	25%	50%	75%	max	sd	#districts
Exceeding intensity level '5 lower'								
Tokyo	1.00	0.99	0.99	1.00	1.00	1.00	0.00	898
Osaka	0.93	0.90	0.92	0.94	0.95	0.97	0.02	564
Nagoya	0.96	0.91	0.94	0.97	0.98	0.98	0.02	1379
Fukuoka	0.39	0.06	0.30	0.42	0.48	0.56	0.12	318
Sapporo	0.33	0.05	0.21	0.33	0.44	0.51	0.12	551
Exceeding intensity level '6 lower'								
Tokyo	0.35	0.16	0.22	0.28	0.49	0.59	0.13	898
Osaka	0.37	0.22	0.30	0.39	0.44	0.52	0.09	564
Nagoya	0.56	0.21	0.41	0.61	0.67	0.77	0.14	1379
Fukuoka	0.03	0.00	0.02	0.03	0.03	0.05	0.01	318
Sapporo	0.01	0.00	0.01	0.01	0.02	0.03	0.01	551

Generally speaking, Tokyo, Nagoya and Osaka are all highly prone to small earthquakes. The probability of an earthquake in Tokyo (in a 30 year period) with intensity more severe than '5-lower' is close to 1. Nagoya is more likely to have larger earthquakes than Tokyo and Osaka. The variation of probabilities of severe earthquakes within each of these three big cities is also much larger than the other two. Fukuoka and Sapporo seem less risky to have larger earthquakes, but there is still considerable probability of small earthquakes. These tables show that there is sufficient variation in the 'riskiness' of the five cities and that the risk measures under both thresholds (intensities 5-lower and 6-lower) are important in characterizing a distribution of earthquake risk.

10 Districts versus stations

10.1 Two options

From the property price data set (as described in Section 4), the district, ward, city and the nearest station of each property is provided. Since we do not know the exact location of a property, we need to use a proxy. There are two options. The first option is to identify a property with the district in which it lies. A city consists of wards, and wards consist of districts. So there are many districts within one city.

The second option is to identify a property with its nearest station. In our analysis we confine the data to properties within 30 minutes of a station (walking distance) and we know the name of this nearest station.

We shall investigate both options.

10.2 Station coordinates

Since we confine our analysis to properties that are within 30 minutes walking distance from a station, it is not unreasonable to identify the location of the property by the location of that metro station. There are fewer stations than there are districts, so this division will be somewhat less accurate.

We find the location information of each station, given the name of the station and the city it is located in. It is necessary to specify the city as well since in some cases the same station name appears in different cities (such as Nakanoshima Station in Sapporo and Osaka).

10.3 District coordinates

Since districts are often irregularly shaped, the coordinates are approximate and not necessarily at the exact geometric center of a district.

10.4 From coordinates to mesh codes

Using either the district or the nearest station, we are able to obtain the rough location of each property. The location information can then be used to find the time-varying seismic risk probabilities associated with each property. In the JSHIS risk data set, the seismic hazard information is stored for each mesh grid.

Given the coordinates, we calculate the 10 digit mesh codes using the method described in Figure 16. The result is double-checked by feeding the coordinates to a geocoding converter website

`http://www.geosense.co.jp/map/tool/geoconverter.php`.

We note that the Geoconverter can also take station or district names directly as an input, but when something goes wrong (e.g. location cannot be found) it produces inaccurate results without warning, so we have to obtain the coordinates first.

A brief summary of the number of stations and the number of sample records for each first mesh code is shown below, assuming that we use the nearest station as proxy for the property location.

Table 55: Number of districts and stations for each first mesh

City	first mesh	#districts	#stations	#samples
Tokyo	5339	898	482	184077
Osaka	5235	304	115	30107
	5135	260	105	19713
Nagoya	5236	1295	150	36552
	5237	84	9	2227
Fukuoka	5030	318	75	25982
Sapporo	6441	551	86	32685

10.5 Cross validation

In the procedures of finding coordinates using location names and finding mesh codes using coordinates, we used multiple sources in order to minimize the risk of associating properties with wrong locations. However, it is still possible that either or both of the location proxies were wrongly reported since the property price data set is obtained from surveys. We thus need to check the validity of the location information of each record.

For a given record in the sample, we have obtained the coordinates of the nearest station; we also know the approximate center of the district where the property is located in. Since we have chosen the walking distance to nearest stations to be less than 30 minutes, the distance from the nearest station to the district center should not be too large. If the distance is above a certain threshold, then this record is suspicious and manual check is in order. We have chosen this threshold to be 6.22 km which is the 99% quantile of the station to district center distances of all the unique station-district pairs.

In the end we narrowed down the list of questionable records to 45 station-district combinations that seemed invalid. These correspond to 47 records in the sample and have been excluded.

Related literature

Berman, M. (1983). Comment on "Likelihood analysis of point processes and its applications to seismological data" by Ogata. *Bulletin of the International Statistical Institute*, 50, 412–418.

Brookshire, D. S., Thayer, M. A., Tschirhart, J., and Schulze, W. D. (1985). A test of the expected utility model: Evidence from earthquake risks, *Journal of Political Economy*, 93, 369–389.

Gu, T., Nakagawa, M., Saito, M., and Yamaga, H. (2011). On asymmetric effects of changes in regional risk rankings on relative land prices in the Tokyo metropolitan area: A test of implications implied by prospect theory using market equilibrium prices (in Japanese), *Journal of Behavioral Economics and Finance*, 4, 1–19.

Hanaoka, C., Shigeoka, H., and Watanabe, Y. (2015). Do risk preferences change? Evidence from panel data before and after the Great East Japan Earthquake, preprint.

Hidano, N., Hoshino, T., and Sugiura, A. (2015). The effect of seismic hazard risk information on property prices: Evidence from a spatial regression discontinuity design, *Regional Science and Urban Economics*, 53, 113–122.

Mignan, A. and Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*. doi:10.5078/corssa-00180805. Available at http://www.corssa.org.

Nakagawa, M., Saito, M., and Yamaga, H. (2007). Earthquake risk and housing rents: Evidence from the Tokyo metropolitan area, *Regional Science and Urban Economics*, 37, 87–99.

Nakagawa, M., Saito, M., and Yamaga, H. (2009). Earthquake risks and land prices: Evidence from the Tokyo metropolitan area, *The Japanese Economic Review*, 60, 208–222.

Naoi, M., Seko, M., and Ishino, T. (2012). Earthquake risk in Japan: Consumers' risk mitigation responses after the Great East Japan Earthquake, *Journal of Economic Issues*, 46, 519–530.

Naoi, M., Seko, M., and Sumita, K. (2009). Earthquake risk and housing prices in Japan: Evidence before and after massive earthquakes, *Regional Science and Urban Economics*, 39, 658–669.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27, 23–31.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical Association*, 83, 9–27.

Schoenberg, F. P. (2003). Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98, 789–795.

van Stiphout, T., Zhuang, J., and Marsan, D. (2012). Seismicity declustering. *Community Online Resource for Statistical Seismicity Analysis*. doi:10.5078/corssa-52382934. Available at http://www.corssa.org.

Wiemer, S. and Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan. *Bulletin of the Seismological Society of America*, 90, 859–869.

Yamaga, H., Nakagawa, M., and Saito, M. (2002). Earthquake risks and land pricing: The case of the Tokyo metropolitan area (in Japanese), *Journal of Applied Regional Science*, 7, 51–62.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97, 369–380.

Zhuang, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society B*, 68, 635–653.

Zhuang, J., Harte, D., Werner, M. J., Hainzl, S., and Zhou, S. (2012). Basic models of seismicity: Temporal models. *Community Online Resource for Statistical Seismicity Analysis*. doi:10.5078/corssa-79905851. Available at http://www.corssa.org.