October 2021

Estimation of nonlinear functions using coarsely discrete measures in panel data: A case of a relationship between land prices and earthquake risks in the Tokyo Metropolitan District

Tao Gu (Daito Bunka University, gutao2002@hotmail.com),
Masayuki Nakagawa (Nihon University, nakagawa.masayuki@nihon-u.ac.jp),
Makoto Saito (Nagoya University, saito@soec.nagoya-u.ac.jp), and
Hisaki Yamaga (Kuwansei Gakuin University, hisaki.yamaga@gmail.com)

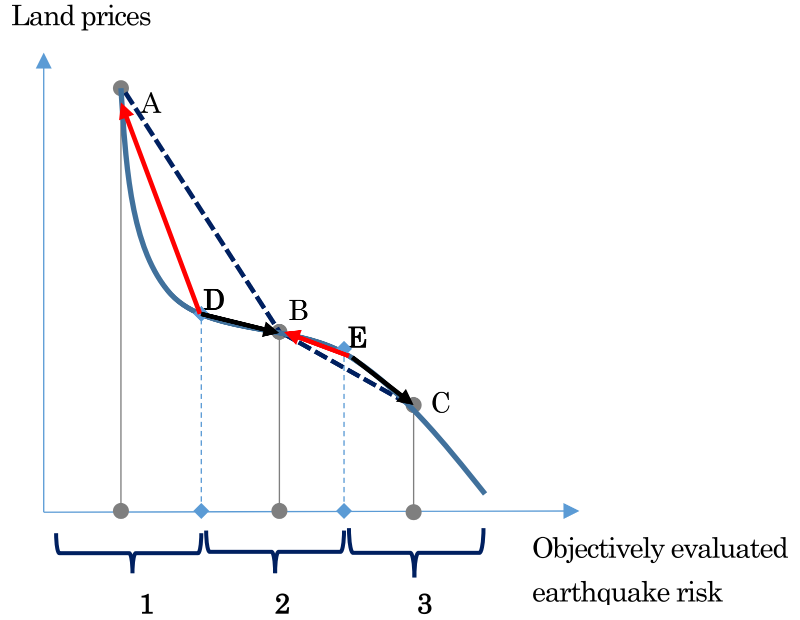
Abstract: This paper proposes a simple method to estimate a nonlinear function using only coarsely discrete explanatory variables in panel data. Here, the intervals of the concerned explanatory factor are represented by a small number of discrete indexes. A basic idea behind the proposed econometric specification is to carefully distinguish between two types of the discrete variable, assuming that if the variable changes between two points of time, it increases from near the upper bound of one rank below, or that it decreases from near the lower bound of one rank above. Then, dynamic pricing behavior at the boundary between two consecutive ranks is approximated as properly as possible. Applying the proposed method, this paper estimates a nonlinear relationship between land prices and earthquake risks, the latter of which is assessed only on a coarsely discrete scale. The panel datasets of land prices and earthquake risks are recorded for around 2,000 or more fixed places over time in the Tokyo Metropolitan District. The estimated nonlinearity of land pricing functions is then interpreted along prospect theory in behavioral economics.

JEL classification: R14, R30, D91

1. Introduction

Nonlinearity does matter in the field of behavior economics in general, and in the context of prospect theory in particular. Estimating nonlinear functions, which frequently emerge from applications of prospect theory, requires explanatory variables to be densely continuous, but available variables are often coarsely discrete in natural experiments. This paper presents a simple method to estimate such a nonlinear function by exploiting panel data structures of the concerned discrete measure. This method carefully distinguishes between two types of the coarsely discrete explanatory variable, assuming that if the variable changes between two points of time, it increases from near the upper bound of one rank below, or that it decreases from near the lower bound of one rank above. In addition, the earthquake risk is assumed to distribute over the interval of the rank to which the risk index increases or decreases. Then, it can properly approximate dynamic pricing behavior at the boundary between two consecutive ranks (in the neighborhood of both the lower bound of one rank above and the upper bound of one rank above). A resulting econometric specification shares the features of *nonlinear probability weighting*, *rank dependence*, and *asymmetry between gains and losses*, all of which are essential ingredients in prospect theory. In this paper, a nonlinear relationship between land prices and earthquake risks is estimated by the proposed method, when the measure for earthquake risks is assessed only on a scale of discrete measures. Then, we interpret the estimation results along various versions of prospect theory. According to the nonlinear probability weighting function (an inverted S-shaped function), which is adopted as one of major theoretical devices in prospect theory (Tversky and Kahneman, 1992; Prelec, 1998), small-sized risks (measured in terms of the objective probability of disastrous events) tend to be overweighted in subjective risk assessment, but such overweighting quickly dissolves as risks approach near-zero. Conversely, medium-sized risks are likely to be underweighted. However, such underweighting rapidly disappears as risks become large. Let us employ a simple setup where land pricing is linearly decreasing in the *subjectively* evaluated earthquake risk. But the subjective risk is not observable. All we can observe is the objectively evaluated risk. Assuming that the objective probability is distorted by nonlinear probability weighting, we estimate a nonlinear relationship between land prices and the *objective* risk. More concretely, land prices rise fasts when overweighting of the underlying risk dissolves as the risk approaches near-zero. On the other hand, land prices are relatively insensitive to the objective risk, when the medium-sized risk is underweighted. But land prices deteriorate fast when underweighting of the underlying risk disappears as the risk becomes large. As depicted by a blue solid line in **Figure 1**, a nonlinear function consequently emerges for the relationship between land prices and the objective earthquake risk.

Figure 1: A nonlinear relationship between land prices and objective earthquake risks



Given densely continuous risk measures in cross-sectional datasets, it is quite possible to estimate precisely the above nonlinear relationship. However, it is impossible to do so only with coarsely discrete risk measures at a particular point of time. Suppose that three intervals of the objective earthquake risk are represented by discrete indexes, 1 (safest), 2, and 3 (riskiest). The nonlinear function may be approximated by two thick blue dotted lines AB and BC, both of which connect the midpoints of each interval, but this approximation is never able to capture precisely a nonlinear nature of the function in question. Line AB fails to approximate either a right derivative at point A or a left derivative at point B, while line BC does not succeed in capturing either a right derivative at point B or a left derivative at point C.

A basic idea behind the estimation method proposed in this paper is quite simple. The method can compensate for the absence of continuous risk measures in cross-sectional datasets by exploiting changes in coarsely discrete measures between two points of time in panel datasets. Here, it is assumed that if the concerned discrete measure changes over time, then it decreases from near the lower bound of one rank above, or it increases from near the upper bound of one rank below. More concretely, it can approximate a right derivative at point A (B) by a red arrow DA (EB) using risk-improving observations from Rank 2 to 1 (3 to 2), and a left derivative at point B (C) by a black arrow DB (EC) using risk-deteriorating observations from Rank 1 to 2 (2 to 3).

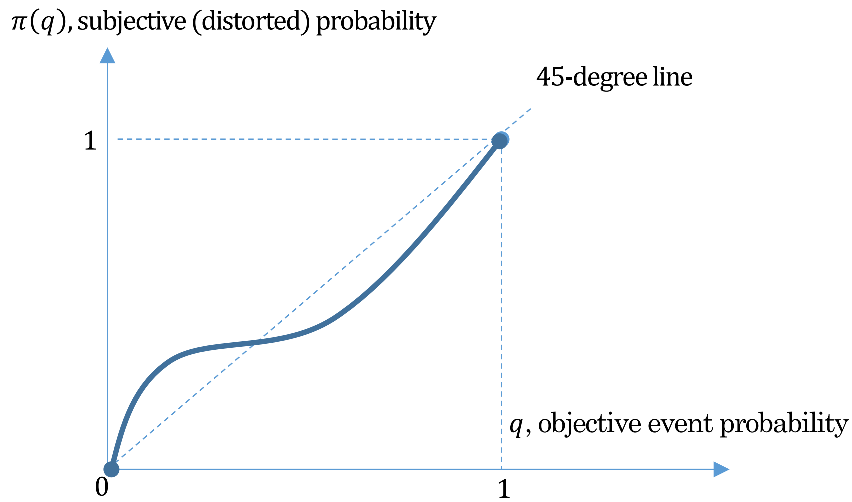
For this purpose, we have a wonderful environment of natural experiments in the Tokyo Metropolitan District (Tokyo MD). The Tokyo Metropolitan Government (Tokyo MG) evaluates earthquake risks by coarsely discrete indexes throughout the Tokyo MD every five years, though except for its western mountain region. More concretely, it ranked earthquake risks on a relative scale of one (safest) to five (riskiest) for every numbered subdivision (*cho-me* in Japanese) of all wards, cities, and towns in the Tokyo MD in 1998, 2003, 2008, 2013, and 2018. On the other hand, the Ministry of Land, Infrastructure, Transport, and Tourism (MLITT) lists the land prices which are appraised on every new year for many fixed points of location mainly in urban areas throughout Japan. For the Tokyo MD, land prices of around 2,000 or more fixed places are appraised every year.

The above earthquake risk measure released by the Tokyo MG is not cardinal, but ordinal. However, it is still significant for our empirical purpose because the econometric method focuses on not how much the risk differs among various points of location in a static context, but in which direction the risk is revised near the boundary between two consecutive ranks (in the neighborhood of both the lower bound of one rank above and the lower bound of one rank below) in a dynamic context. Combining the panel data of the discrete earthquake risk measures with that of land prices for each fixed place, we thus estimate a nonlinear relationship between land prices and earthquake risks by the method proposed in this paper. Then, we explore whether the estimated nonlinear function is interpretable consistently along prospect theory.

2. Econometric specification

Let us present a simple land pricing model in the presence of earthquake risks. Here, land prices are assumed to be discounted by the expected earthquake damage qD, where q and D denote the objective event probability and the damage from an earthquake.[1] In prospect theory, not the objective, but subjective probability is employed. The objective probability is distorted according to the nonlinear probability weighting function $\pi(q)$, which is often depicted as an inverted S-shape by a blue solid line in **Figure 2** (Tversky and Kahneman, 1992; Prelec, 1998). The small objective probability is overweighted, but such overweighting dissolves as the probability approaches zero. On the other hand, the middle-sized objective probability is underweighted, but such underweighting disappears as the probability is quite large.

Figure 2: An inverted S-shape of the probability weighting function

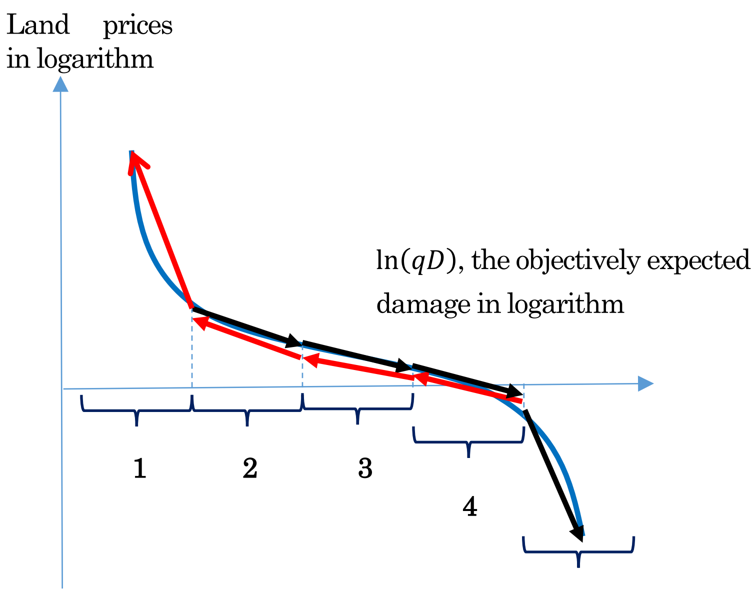


Here, the standard setup of prospect theory is modified slightly. The weighting function $\pi(\quad)$ is still applied to the event probability q. The damage D is standardized in some way. Thus, the expected damage is under or overestimated according to $\pi(q)D$. Then, $\ln[\pi(q)D]$ is subtracted from a logarithmic land price ($\ln P$) after adjusted by other important factors in land pricing. Accordingly, a nonlinear land pricing function of the objectively expected damage ($\ln(qD)$) is depicted by a blue solid line in **Figure 3**, which is a mirror image of **Figure 2** in a vertical direction.

However, we do not have any continuous measure for earthquake risks, $\ln(qD)$. All we have for earthquake risks is the earthquake risk measure that is assessed only on a discrete scale of one (safest) to five (riskiest). Here, it is assumed that the smaller (larger) q accompanies the smaller (larger) D, and that the ranks of q and D correspond to that of qD.

Fortunately, we have the panel datasets of

Figure 3: Estimation of a nonlinear land pricing function by discrete earthquake risk measures



[1] Rigorously, in the presence of risk aversion on the consumer's side, the expected damage should be further adjusted by a marginal rate of substitution between a current safe state and a forthcoming disastrous state. Suppose that wealth is equal to W at a current safe state, and W declines by uninsured damages Z on the occurrence of an earthquake. Given that a utility function $U(\quad)$ is increasing, concave, and differentiable, the expected damage pD should be adjusted by $\frac{U'(W-Z)}{U'(W)}$. Here, it is assumed that the curvature of $U(\quad)$ is quite small as documented in the context of the equity premium puzzle (Mehra and Prescott, 1985). Accordingly, $\frac{U'(W-Z)}{U'(W)}$ is close to one. Thus, the risk aversion part is ignored in the current specification.

this discrete risk measure for every numbered subdivision of all wards, cities, and towns in the Tokyo MD. At the same time, we have a panel dataset of land prices for around 2,000 or more fixed points of location in the Tokyo MD. By exploiting these panel datasets of the earthquake risk measure and the land prices, we propose the following econometric specification to estimate a nonlinear land pricing function.

As assumed above, the ranks of q and $\ln D$ accord with that of $\ln(qD)$. Then, $-[\ln[\pi(q)]+\ln D]$ is specified in a *rank-dependent* manner. To begin with, it is formulated by a stepping function of the discrete risk measure $\sum_{i=2}^{r_{n,t}} a_{i,t}$, where $r_{n,t}$ denotes a discrete risk rank, 2, 3, 4, or 5, and $a_{i,t}<0$ represents risk sensitivity for each rank of earthquake risk. Then, a logarithmic land price of location n in year t ($\ln P_{n,t}$) is approximated by this stepping function together with other explanatory variables:

$$\ln P_{n,t} = p_t(r_{n,t}, x_{j,n,t}, f_n) = \sum_{i=2}^{r_{n,t}} a_{i,t} + \sum_{j=1}^{J} b_j x_{j,n,t} + f_n + const_t, \tag{1}$$

where $x_{j,n,t}$ represents a time-varying factor, and f_n denotes a fixed factor for location n. In addition to the earthquake risk factor, $x_{j,n,t}$ and f_n play an important role in determining land prices.

The *gain/loss asymmetric* nature is further introduced into equation (1) as follows. If a land price in logarithm ($\ln P_t$) is decreasing in land risks ($r_{n,t}$), then $a_{i,t}<0$ for $i=2,3,4,$ and 5. As discussed in the introduction, the interpretation of parameter $a_{i,t}$ in the step function is quite subtle. For example, $a_{2,t}$ cannot be interpreted as either a right derivative at Rank 1 or a left derivative at Rank 2. Without densely continuous risk measures for $r_{n,t}$, it is impossible to estimate derivatives at different points properly by cross-sectional datasets only. However, it is possible to approximate the two derivatives separately by distinguishing among risk-improving, risk-deteriorating, and risk-invariant observations in panel datasets. In other words, the panel data structure allows us to not only to eliminate fixed effects (f_n) as usual, but also to differentiate between the two derivatives.

For this purpose, equation (1) is further specified as

$$p_t^+(r_{n,t}, x_{j,n,t}, f_n) = \sum_{i=2}^{r_{n,t}-1} a_{i,t}^+ + a_{r_{n,t},t}^+ + \sum_{j=1}^{J} b_{j,t} x_{j,n,t} + f_n + const_t^+, \tag{2}$$

for risk-deteriorating observations (from $r_{n,t}-1$ to $r_{n,t}$) that are assumed to increase from near the upper bound of one rank below. On the other hand, equation (1) is specified as

$$p_t^-(r_{n,t}, x_{j,n,t}, f_n) = \sum_{i=2}^{r_{n,t}+1} a_{i,t}^- - a_{r_{n,t}+1,t}^- + \sum_{j=1}^{J} b_{j,t} x_{j,n,t} + f_n + const_t^-, \tag{3}$$

for risk-improving observations (from $r_{n,t}+1$ to $r_{n,t}$) that are assumed to decrease from near the lower bound of one rank above. Given the above two specifications, $a_{r_{n,t},t}^+$ ($a_{r_{n,t},t}^-$) can be interpreted as a left derivative (a right derivative). Finally, equation (1) is specified as

$$p_t^0(r_{n,t}, x_{j,n,t}, f_n) = \sum_{i=2}^{r_{n,t}} a_{i,t}^0 + \sum_{j=1}^{J} b_{j,t} x_{j,n,t} + f_n + const_t^0, \tag{4}$$

for risk-invariant observations. Note how equations (2) and (3) are specified in the same gain/loss asymmetric manner as how a probability weight is defined in cumulative prospect theory (Tversky and Kahneman, 1992).[2]

---

[2] By equations (2) and (3), a change in the land price $\Delta P_{n,t}$ due to a change in the risk index is determined by an increment in a probability weighting part $-\Delta[\ln[\pi(q)]+\ln D]$. In turn, $-\Delta[\ln[\pi(q)]+\ln D]$ depends on whether the underlying risk deteriorates or improves. For risk deterioration from $r_{n,t}-1$ to $r_{n,t}$, $\Delta P_{n,t} = \sum_{i=2}^{r_{n,t}} a_{i,t}^+ - \sum_{i=2}^{r_{n,t}-1} a_{i,t}^+ = \sum_{i=r_{n,t}}^{5} a_{i,t}^+ - \sum_{i=r_{n,t}+1}^{5} a_{i,t}^+ = a_{r_{n,t},t}^+$, and for risk improvement from $r_{n,t}+1$ to $r_{n,t}$, $\Delta P_{n,t} = -\left(\sum_{i=2}^{r_{n,t}+1} a_{i,t}^- - \sum_{i=2}^{r_{n,t}} a_{i,t}^-\right) = -a_{r_{n,t}+1,t}^-$. In cumulative prospect theory, on the other hand, a probability weight ω_i is

Putting equations (2), (3), and (4) in a panel data setup, it is assumed that the risk sensitivity $a_{i,t}^s$ ($s =$ $+, -,$ or 0) may change between time t_0 and t_1 in a rather restrictive manner as follows:

$$a_{i,t_1}^s = a_{i,t_0}^s + c^s. \tag{5}$$

That is, the overall risk sensitivity may change over time for equations (2), (3), and (4).

For an observation whose risk measure increases from r_{n,t_0} to r_{n,t_1} by one rank, a first difference in land prices is expressed as

$$\ln P_{n,t_1}^+ - \ln P_{n,t_0}^+ = \sum_{i=2}^{r_{n,t_1}-1} \left(a_{i,t_1}^+ - a_{i,t_0}^+\right) + a_{r_{n,t_1},t_1}^+ + \left(\sum_{j=1}^{J} b_{j,t_1} x_{j,n,t_1} - \sum_{j=1}^{J} b_{j,t_0} x_{j,n,t_0}\right) + \left(const_{t_1}^+ - const_{t_0}^+\right)$$
$$= c^+ r_{n,t_1} + a_{r_{n,t_1},t_1}^+ + \textstyle\sum_{j=1}^{J}\left(b_{j,t_1} - b_{j,t_0}\right)x_{j,n,t_1} + \sum_{j=1}^{J} b_{j,t_0}\left(x_{j,n,t_1} - x_{j,n,t_0}\right) + \left(const_{t_1}^+ - const_{t_0}^+ - 2c^+\right). \tag{6}$$

For an observation whose risk measure decreases from r_{n,t_0} to r_{n,t_1} by one rank, on the other hand, it is specified as

$$\ln P_{n,t_1}^- - \ln P_{n,t_0}^- = \sum_{i=2}^{r_{n,t_1}+1} \left(a_{i,t_1}^- - a_{i,t_0}^-\right) - a_{r_{n,t_1}+1,t_1}^- + \left(\sum_{j=1}^{J} b_{j,t_1} x_{j,n,t_1} - \sum_{j=1}^{J} b_{j,t_0} x_{j,n,t_0}\right) + \left(const_{t_1}^- - const_{t_0}^-\right)$$
$$= c^- r_{n,t_1} - a_{r_{n,t_1}+1,t_1}^- + \textstyle\sum_{j=1}^{J}\left(b_{j,t_1} - b_{j,t_0}\right)x_{j,n,t_1} + \sum_{j=1}^{J} b_{j,t_0}\left(x_{j,n,t_1} - x_{j,n,t_0}\right) + \left(const_{t_1}^- - const_{t_0}^-\right). \tag{7}$$

For an observation without any change in the risk measure, it is derived as

$$\ln P_{n,t_1}^0 - \ln P_{n,t_0}^0 = \sum_{i=2}^{r_{n,t_1}} \left(a_{i,t_1}^0 - a_{i,t_0}^0\right) + \left(\sum_{j=1}^{J} b_{j,t_1} x_{j,n,t_1} - \sum_{j=1}^{J} b_{j,t_0} x_{j,n,t_0}\right) + \left(const_{t_1}^0 - const_{t_0}^0\right)$$
$$= c^0 r_{n,t_1} + \textstyle\sum_{j=1}^{J}\left(b_{j,t_1} - b_{j,t_0}\right)x_{j,n,t_1} + \sum_{j=1}^{J} b_{j,t_0}\left(x_{j,n,t_1} - x_{j,n,t_0}\right) + \left(const_{t_1}^0 - const_{t_0}^0 - c^0\right). \tag{8}$$

Putting equations (6), (7), and (8) together, an empirical specification takes the following form for observation n whose risk measure changes from r_{n,t_0} by one rank in year t_1.

$$\ln P_{n,t_1} - \ln P_{n,t_0} = c^0 r_{n,t_1} + (c^+ - c^0) r_{n,t_1} D^+ + (c^- - c^0) r_{n,t_1} D^- + a_{r_{n,t_1},t_1}^+ D^+ - a_{r_{n,t_1}+1,t_1}^- D^-$$
$$+ \textstyle\sum_{j=1}^{J}\left(b_{j,t_1} - b_{j,t_0}\right)x_{j,n,t_1} + \sum_{j=1}^{J} b_{j,t_0}\left(x_{j,n,t_1} - x_{j,n,t_0}\right) + const + const^+ D^+ + const^- D^-, \tag{9}$$

where D^+ (D^-) represents a dummy variable for deteriorating (improving) observations. Note that a negative sign appears in front of $a_{r_{n,t_1}+1,t_1}^- D^-$.

When the risk measure changes by two ranks, equation (9) is respecified as

$$\ln P_{n,t_1} - \ln P_{n,t_0} = c^0 r_{n,t_1} + (c^+ - c^0) r_{n,t_1} D^+ + (c^- - c^0) r_{n,t_1} D^-$$
$$+ \left(a_{r_{n,t_1},t_1}^+ + a_{r_{n,t_1}-1,t_1}^+\right) D^+ - \left(a_{r_{n,t_1}+1,t_1}^- + a_{r_{n,t_1}+2,t_1}^-\right) D^-$$
$$+ \textstyle\sum_{j=1}^{J}\left(b_{j,t_1} - b_{j,t_0}\right)x_{j,n,t_1} + \sum_{j=1}^{J} b_{j,t_0}\left(x_{j,n,t_1} - x_{j,n,t_0}\right) + const + const^+ D^+ + const^- D^-. \tag{10}$$

---

defined by a change in weighting function $\Delta\pi(q)$. In turn, $\Delta\pi(q)$ depends on whether outcome x_i, paired with probability q_i, increases or decreases. Suppose $x_1 < \cdots < x_{\bar{n}} < r < x_{\bar{n}+1} < \cdots < x_N$. For outcome gains from a reference point r, $\omega_n = \pi^+\left(\sum_{i=n}^{N} q_i\right) - \pi^+\left(\sum_{i=n+1}^{N} q_i\right)$, and for outcome losses from r, $\omega_n = \pi^-\left(\sum_{i=1}^{n} q_i\right) - \pi^-\left(\sum_{i=1}^{n-1} q_i\right)$.

If the risk measure changes by three ranks, $\left(a^{+}_{r_{n,t_1},t_1} + a^{+}_{r_{n,t_1}-1,t_1}\right) D^{+}$ and $-\left(a^{-}_{r_{n,t_1}+1,t_1} + a^{-}_{r_{n,t_1}+2,t_1}\right) D^{-}$ in equation (10) are modified as $\left(a^{+}_{r_{n,t_1},t_1} + a^{+}_{r_{n,t_1}-1,t_1} + a^{+}_{r_{n,t_1}-2,t_1}\right) D^{+}$ and $-\left(a^{-}_{r_{n,t_1}+1,t_1} + a^{-}_{r_{n,t_1}+2,t_1} + a^{-}_{r_{n,t_1}+3,t_1}\right) D^{-}$.

The above econometric specification shares the features of *nonlinear probability weighting*, *rank dependence*, and *gain/loss asymmetry*, all of which are essential ingredients in prospect theory. The specification can be described graphically using **Figure 3**. In this figure, a nonlinear land pricing function at time t_1 is depicted by a blue solid line. Here, equation (6) for risk-deteriorating observations is represented by a black arrow, while equation (7) for risk-improving observations is expressed by a red arrow. When black and red arrows simultaneously approximate the same nonlinear function depicted by a blue solid line, $a^{+}_{2,t_1} = a^{-}_{3,t_1}$, $a^{+}_{3,t_1} = a^{-}_{4,t_1}$, and $a^{+}_{4,t_1} = a^{-}_{5,t_1}$ are required as additional restrictions in order to connect equation (6) with equation (7) at the boundary between Rank 1 and 2, 2 and 3, 3 and 4, and 4 and 5. Then, the following estimations results are expected:

i. If strong nonlinearity occurs at Rank 2 and 4, then $a^{-}_{2,t_1} < a^{+}_{2,t_1}$, and $a^{-}_{5,t_1} > a^{+}_{5,t_1}$ are predicted. If nonlinearity is weak except between 1 and 2, and 4 and 5, then $a^{-}_{3,t_1} \approx a^{+}_{3,t_1}$, and $a^{-}_{4,t_1} \approx a^{+}_{4,t_1}$ are likely to hold.
ii. If strong nonlinearity occurs at Rank 3, and 4, then the result changes to $a^{-}_{3,t_1} < a^{+}_{3,t_1}$, and $a^{-}_{5,t_1} > a^{+}_{5,t_1}$. In addition, $a^{-}_{2,t_1} \approx a^{+}_{2,t_1}$, and $a^{-}_{4,t_1} \approx a^{+}_{4,t_1}$ tend to hold for weakly nonlinear parts.
iii. If strong nonlinearity arises at Rank 2 and 3, then the result is revised as $a^{-}_{2,t_1} < a^{+}_{2,t_1}$, and $a^{-}_{4,t_1} > a^{+}_{4,t_1}$. In addition, $a^{-}_{3,t_1} \approx a^{+}_{3,t_1}$, and $a^{-}_{5,t_1} \approx a^{+}_{5,t_1}$ tend to hold for weakly nonlinear parts.

Here is a final remark on the above econometric framework. As explained in detail in Section 3, the earthquake risk measure released by the Tokyo MG is not cardinal, but ordinal. But it is still significant for this empirical method, because the current specification focuses not on how much the risk differs among various points of location in a static context, but in which direction the risk is revised at the boundary between two consecutive ranks in a dynamic context. Then, we compare between estimates of $a^{-}_{i,t}$ and $a^{+}_{i,t}$ as two-way risk sensitivities from the identical boundary between Rank $i-1$ and i $(i = 2,3,4,\text{and } 5)$. The additional restriction $a^{+}_{i,t} = a^{-}_{i+1,t}$ $(i = 2,3,\text{and } 4)$ also neutralizes the effect of ordinality. That is, the risk sensitivity $a^{+}_{i,t}$ and $a^{-}_{i,t}$ except for $a^{-}_{2,t}$ and $a^{+}_{5,t}$ are standardized such that the price impact is indifferent between risk deterioration from Rank i to $i+1$ and risk improvement from Rank $i+1$ to i.

References:

Mehra, Rajnish, and Edward C. Prescott, 1985, "The equity premium: A puzzle," *Journal of Monetary Economics*, 15 (2), 145-161.
Prelec, Drazen, 1998, "The probability weighting function," *Econometrica* 32 (1/2), 122-136.
Tversky, Amos, and Kahneman, Daniel, 1992, "Advances in prospect theory: Cumulative presentation of uncertainty," *Journal of Risk and Uncertainty*, 5 (4), 297-323.