Proceedings of the 5th International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/2026.38714

The Factory That was Never Built: Evidence from a Survey Experiment on Compensation in China

Songjie Wu

Haidian Kaiwen Academy, Beijing, China
wusongjie2008@outlook.com

Abstract. Conventional welfare accounting assigns the same present value to cash flows that occur at the same times, yet third-party compensation assessments may depend on whether the affected activity has already started. This paper randomizes a survey to 600 adults from mainland China. The author asks participants to assess one of two things that are the same in all other respects: a) a factory that stops running after running for one year, and b) a factory that got the approval, fund and site allocation, but was cancelled before the start of construction. In both cases, the facility is expected to generate an annual local net income of CNY 48 million for ten years beginning the following year, and the government considers the same ten-year compensation plan in each case. The result is that people give higher compensation after the start of operations: the average is CNY 74.4 million for the after start of the operations, compared to CNY 58.4 million for the before start of the construction, that is a difference of CNY 16.0 million (which is 33% of the designated income). This difference is small for the low compensation amounts but becomes larger in the upper half of the distribution, where the responses are also more variable. It is slightly smaller at the very top. These results show that the operational status matters for the third-party determinations about the compensation, not only for the equivalence in cash flow. Therefore, the compensation schemes could differentiate between the income substitution documented and a different evidential evaluation of the confirmed transition obligations and not put simply an automatic premium depending on being in an early operational status.

Keywords: environmental compensation, regulatory takings, just compensation, loss aversion, survey experiment

1. Introduction

From the perspective of the cost-benefit analysis, the same financial valuations could be assigned to two different cases: in the first one a community loses its revenues as a consequence of the closure of a factory, and in the second one the same earnings are forgone because the study stopped a proposed facility before the construction [1, 2]. Under the Kaldor–Hicks compensation criterion, those who gain from a change could in principle compensate those who lose and still have a net gain, under the Hicks criterion, those who would be harmed could pay the beneficiaries to block the change [1, 2]. Coase shows how the distribution of the resources is determined by the property rights and the costs of transactions [3]. These frameworks give some standards to accounting and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.38714

negotiation, but they do not answer if someone that is outside of the situation, see as more established an existing income flow than a planned stream of equal certainty.

Understanding from behavioral science and legal economics highlights the importance of realized loss in both the time and the structure attached to interrupted flows of revenues. It is found that opportunity costs are often neglected in many fields of study, from individual consumer decisions [4] to the distribution of resources of public policies [5] and wider experiments [6]. Also, the perceived value of an asset may be changed by being or not owned, anticipatory expectancies [7, 8] and realized losses are often of more legal and social importance than missed gains [9]. Studies with third-party observers find that people tend to favor compensatory measures after inequitable distributions [10] and are sensitive to loss-oriented framing in redistribution contexts [11]. Against this theoretical background, the study examine levels of public compensation while holding constant the start date of cash flows, monetary magnitude, certainty, duration, underlying policy rationale, and the identities of the payer and recipients.

In order to answer this question, the paper designs a randomized survey to adults in mainland (for the purpose of this study, adults are only living in mainland). In the survey, each respondents are asked to judge from two situations, i.e. a manufacturing plant which has stopped after one year of operation, or a project which has obtained the full approval and the full funding and the location, but it is terminated just before the beginning of the construction. The main result that runs is the intention-to-treat effect for the full scenario description on the reported numbers of the annual compensation. Using randomized methods, the paper looks at whether the operation status matters for the third-party awards, and at the same time controlling for the timing of the income receipt and the structure of the compensation packages.

2. Interpretive framework and competing predictions

In the realized-loss context (R120) the manufacturing plant which has finished the administrative procedure and is running without interruption for a year should stop its operations because the new environmental regulations are imposed. In the forgone-gain situation (F120), the project for which permits, construction capital, and land were secured is abandoned before construction begins. In either case the facility was expected to bring a guaranteed net local income of CNY 48 million a year, from the next year and for one decade. In both cases the authorities looked at an equal arrangement of ten successive yearly contributions to a community fund to support the public services and the employment transitions.

The designated annual income, set at V = 4,800 units of 10,000 CNY, is a yearly replacement standard, rather than the present value of the ten-year income flow. Since both the incomes have the same timing and the same magnitudes, they have the same present value for any uniform discount rate. An annual compensation equal to V is exactly the declared yearly income while amounts different from V are either reduced or increased annual transfer. Such comparison with thresholds finds the points of the two compensation distributions where they are different, without giving any specific psychological meaning for every figure that is bigger than V.

For every participant, the compensation ratio is the yearly award divided by V. For any threshold, the empirical survival curve shows what fraction of individuals choose an amount greater or equal to that threshold. Since awards are non-negative, the area underneath this curve normalized by V is the average compensation, area divided by V to obtain that ratio. Thus, the mean difference, the fraction attaining V, and higher-threshold proportions are different summaries of the same result, depending on whether the difference is uniform or concentrated at certain award levels.

DOI: 10.54254/2754-1169/2026.38714

For R120, three mechanisms can lead to high award levels. Possibly there are operational reasons for which one has a legitimate claim, the impacted labor and capital assets are more real, or one has in a logical way some deductions about the pledged obligations and the transition costs. If there is a uniformly perceived burden, then there is a very wide parallel adjustment, if one has varied judgments that lead to high awards, one can increase the divergence and the spread in the upper part of the distribution. These possibilities are distinguished by examining both the average value and the overall distributional shape, post-treatment ratings are analyzed separately and are not used as mediating variables.

The zero-difference accounting standard is defined by equal amounts and dates, not by reference-dependent preferences. On the other hand, the expectations-based reference point then would admit nonzero differences when the operation changes the rational expectations [8]. Besides, the approvals that are completed, the financing arrangements, and land that is acquired, which are the unbuilt project, support predictions for either equality or reversal. Therefore, equality, reversal, a homogeneous R120 shift, and a concentration in the upper half are different empirical results.

3. Design and measurement

3.1. Design diagnostics before the main study

Two pretests were done before the main study was taken, but neither is an adequate test of the hypotheses of the study. They were meant to be only a test of the measurement instrument, and their data were not included in the aggregations of the main sample. Here is the description of these diagnostic procedures in table 1.

Table 1. Descriptive instrument diagnostics and factual recall

Stage	Versions	N	Correct amount	Correct status	Correct certainty	Inferential role
Pretest 1	Separate operating / pre-construction	40	62.5%	25.0%	Not comparable	Complexity diagnostic
Pretest 2	Separate A / B fielding	42	47.6%	35.7%	11.9%	Usability diagnostic
Main study	Respondent-level randomization	600	78.7%	78.7%	82.3%	Causal evidence

3.2. Main experiment

Participants in mainland China were recruited by posting on the online dissemination: WeChat, Douban and author's personal networks. Participants had to be at least 18 years old, able to read simplified Chinese, currently living in mainland China, and willing to continue participation. The system randomly assigned 600 participants to two groups: R120 (304 participants) and F120 (296 participants). All randomized data were kept in the study, so there is no missing value for either the outcome measures or the covariates.

The two scenarios were both about a drinking-water regulation that remained valid no matter if there was any compensation. They set the level of income from the following year for a ten-year time starting the following year, with the policy as cause, the government as payer, and the community fund as receiver, and ten successively annual compensation payments. The survey presented one scenario first and then asked the participants to indicate the amount of each annual payment on a box from 0 to 14,400 (0 to 3V), in units of 10,000 CNY, the middle point of the range

DOI: 10.54254/2754-1169/2026.38714

is at 7,200 (1.5V). It then measured perceived entitlement, reliance and sacrifice with each four-point question, asked the main justification for the award, evaluated the recall about the status of the factory, the amount and the certainty, and noted the gender, age group and educational background. Compensation is the main outcome variable.

3.3. Estimation

The intention-to-treat disparity in average compensation when comparing the R120 and F120 condition is the central parameter of interest. Participants are assigned to the groups before they see the vignettes, and all responses that were generated by randomization are kept in the dataset, so the raw difference in means is exactly as in the experiment. To test the sharp null hypothesis that no individual experiences a treatment effect, this paper uses a Fisher randomization test that fixes the observed outcomes and randomize R and F labels again for 100,000 times [12]. The resulting p-value based on the sharp null hypothesis should not be considered as a test of weak null hypothesis concerning a zero average effect. For estimating the average difference, Welch confidence interval allows for heterogeneity in variances, and Cohen's d is reported the effect in terms of pooled standard deviations.

The ordinary least squares framework, adjusted for the other demography, includes categorical variables for age groups, gender and education, and uses HC2 robust standard errors [13]. This specification is a sensitivity analysis about the role of composition, however, the estimate unadjusted from the randomized comparison is the main result. Showing both sets of results checks whether the results are substantially affected by observed demographic composition.

The distributional examinations explain the structure of the main mean difference. To check for quantile differences, the percentile intervals are obtained from 100,000 bootstrap resamples made in each arm. The threshold proportions at levels V, 1.5 V, 2 V, and 2.5 V are used to separate basic income replacement from progressively larger compensation awards. The empirical curve relating award thresholds plots the proportion of the participants that select at least each specific amount over the full range of awards, if normalized by V, the area under the curve is the mean compensation. An omnibus test for distributional differences is done with a Kolmogorov-Smirnov statistic based on 100,000 label permutations, and dispersion is summarized via a median-centered Levene test. The three perception related outcomes are testing secondary family for which familywise error is controlled using the Holm correction [14]. The item concerning reasons is treated as exploratory, and the recall items work only as implementation diagnostics and do not set the analytical sample.

Given the realized sample size, the minimum detectable standardized effect size at a two-sided significance level of 0.05 and 80% power is d = 0.229. In terms of the observed pooled standard deviation, it corresponds to CNY 8.07 million. This sensitivity analysis for the fixed-sample design describes the amount of the precision that is obtainable for the main comparison [15].

4. Results

The main dataset consisted of 304 observations for the group R120 and 296 for the group F120. The two experimental arms showed no obvious differences in age or education, the female share was 40.5% in R120 and 48.0% in F120. The average compensation award of the group R120 was 7,443.4 (where each unit is 10,000 CNY), while for the group F120 it was 5,841.6. The difference between R120 and F120 is 1,601.9, in units,equivalent to CNY 16.02 million per year, or 33.4% of the designated annual amount. The 95% CI for the difference is, 10.39, 21.65 (in millions of CNY)

DOI: 10.54254/2754-1169/2026.38714

and the sharp-null randomization p-value is below 0.0001 and the effect size (d) is 0.455. When the categorical age groups, gender and education levels were controlled the difference is 1,565.0 (units) (95% CI: 1,002.4, 2,127.6). This is very close to the unadjusted estimate. The half-widths of the CI of the unadjusted model and of the adjusted are almost the same, 562.7 and 562.6 (in units). So, the process of the adjustment does not change much the precision of the estimate. In table 2 it can be seen the description of the two randomized conditions, means, measures of dispersion, quantiles, and particular threshold of the compensation for both randomized conditions. Figure 1 plots the difference between the groups by plotting the complete empirical survival curve for each arm.

Table 2. Compensation distribution by randomized scenario, in 10,000 CNY per year

Outcome	Forgone gain F120	Realized loss R120	R minus F
N	296	304	n/a
Mean compensation	5,841.6	7,443.4	1,601.9
Standard deviation	2,985.4	3,974.7	n/a
Median	6,100	7,550	1,450
25th percentile	3,300	4,000	700
75th percentile	8,325	11,200	2,875
Share at or above V = 4,800	63.9%	70.1%	6.2 pp
Share at or above 7,200	35.8%	52.6%	16.8 pp
Share at or above 9,600	12.8%	35.5%	22.7 pp

Note: Differences are calculated from unrounded values, so displayed means may not subtract exactly.

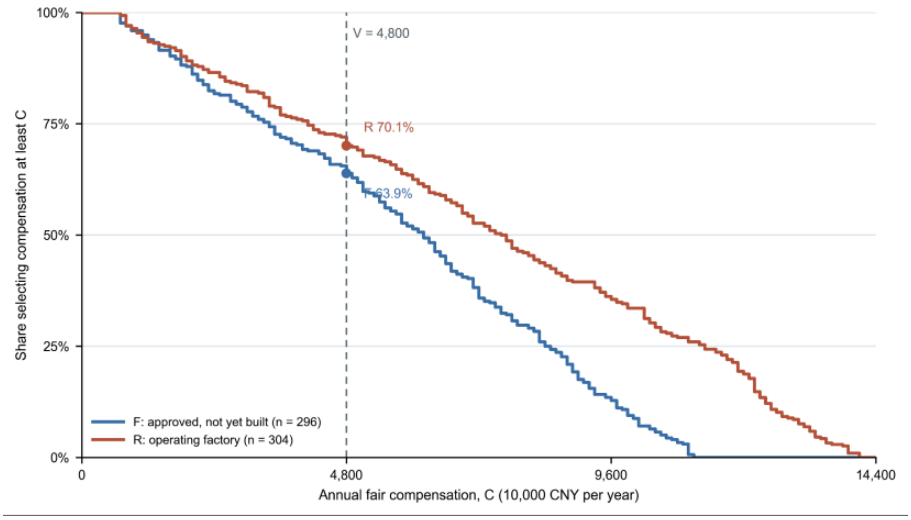


Figure 1. Empirical Survival Curves in the randomized main study

The unsmoothed curves use all 296 F120 and 304 R120 responses and report the share selecting at least compensation C. The vertical line marks the stipulated annual net income V = 4,800 units of 10,000 CNY.

These are the raw, unsmoothed, trajectories which contain the whole sample of 296 responses of the group F120 and 304 of the group R120, of the proportion of those that have chosen a level of compensation at least C. A vertical line marks the stipulated annual net income V = 4,800 units (one unit = 10,000 CNY). Pre-designed awards ranged from 0 to 14,000 units. Observed awards ranged

DOI: 10.54254/2754-1169/2026.38714

from 600 to 14,000 units. The difference at the 10th percentile was 200 units (95% bootstrap confidence interval: -370 to 830), at the 25th percentile it was 700 units (range: -200 to 1,775), at the median it was 1,450 units (interval: 600 to 2,300), at the 75th percentile it was 2,875 units (range: 1,725 to 3,600), at the 90th percentile it was 2,770 units (interval: 2,300 to 3,400). Hence, the gap was quite small for lower level of awards but became substantially more and more big from the median onwards, the gap is observed at the 90th percentile is a bit smaller than that at the 75th percentile.

Specially, if considering the standard account value V, 70.1% of people who were in the R120 group, and 63.9% in the F120 group give awards equal to or greater than V, so the gap is 6.2 percentage points. At the 1.5V the gap is 16.8 points, and it is the maximum of 22.7 points at 2V of the pre-designed scale. It then decreases to 18.8 percentage points at 2.5V, 18.8% of R120 participants give an award at or above this level, compared with no F120 participants doing so. The empirical distributions are much different statistically (KS = 0.265, randomization $p < 0.0001$). The variance that shows in the R120 is 1.77 times bigger than the F120 (Levene $p < 0.001$). These tests show differences in distribution and dispersion, but the Welch interval is used to infer the mean difference.

In both arms of the experimental group most of the participants chose the options corresponding to at least V. The difference between the two arms was mainly at values above V, but it is not monotonic at the highest thresholds up to 2.5V. The midpoint at 1.5V is close to the R120 mean of 7,443 and median of 7,550, and the upper limit of 3V is near the R120 maximum of 14,000. Then the absolute magnitudes are understood in the context of implemented bounded scale, the contrast, which was obtained between arms, is the one of concentration in the upper half and of more dispersion, and not of the acceleration during longer time near the ceiling of the scale.

As shown in Table 1, the aggregated correct-recall rates were 78.7% (amount), 78.7% (status) and 82.3% (certainty). At the arm-specific level, these were between 78.0% and 83.4%. The differences between the arms for all items were very small, the p values from the randomization tests were all equal to, or greater than, 0.520. All participants were included in the intention-to-treat analysis.

5. Treatment effects on post-treatment perceptions

The R120 condition caused an elevation on all three measured perceptual metrics. In the case of the scores evaluated on the four-point scales, the scores for entitlement increased by 0.654, for reliance by 0.545, and for sacrifice by 0.746. Standardized effect sizes ranged from d = 0.537 to 0.733, and all Holm-corrected randomization p-values were below 0.001. These results that are described after the intervention are the results that are represented in Table 3 and Figure 2. These sources describe the variation of the reported classifications after the randomized historical exposures. But they do not explain the way that these perceptions are linked to the final compensation figures.

Table 3. Randomized effects on post-treatment perceptions

Post-treatment measure (1-4)	Forgone gain	Realized loss	Difference	Cohen's d	Holm p
Income counted as due	2.287	2.941	0.654	0.634	<0.001
Reasonable reliance	2.304	2.849	0.545	0.537	<0.001
Economic sacrifice	2.274	3.020	0.746	0.733	<0.001

DOI: 10.54254/2754-1169/2026.38714

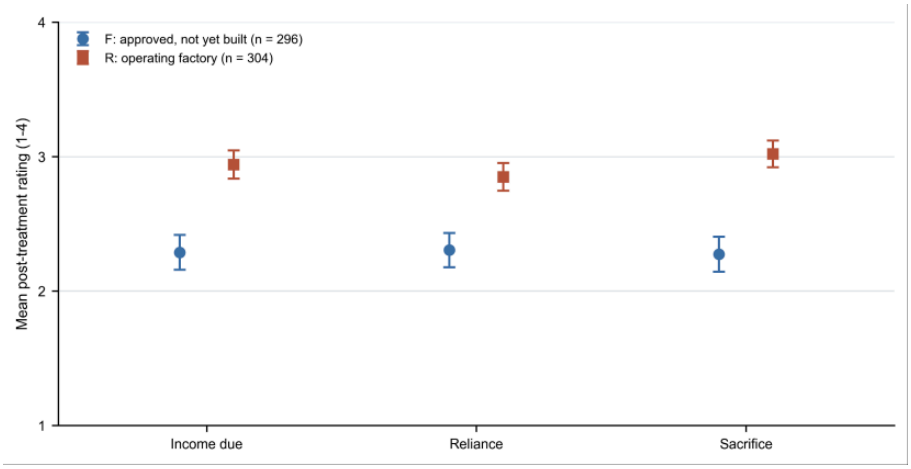


Figure 2. Mean post-treatment perception scores with 95% t-based confidence intervals by randomized scenario

The scale runs from 1 to 4, F120 has n = 296 and R120 has n = 304. Differences between arms are randomized effects on reported perceptions.

The measurement scale is from 1 to 4, the size of the sample is n = 296 for F120 and n = 304 for R120. The differences among the experimental arms are the randomized effects on the reported perceptions of the participants. The visual differences are present for all three of the results, and they are more pronounced in the case of the economic sacrifice, the inferential statistics for these observations are given in the Holm-adjusted randomization tests in Table 3.

The reason-based item included both considered directional effects. The choice of the actual operation was made by 31.6% of the participants in the group R120 and 21.3% in the group F120, while the choice of the forgone development opportunity was made by 7.6% and 12.2% of the respondents, respectively. In both the experimental arms, the given amount per year was the most commonly mentioned reason, with prevalence rates 37.2% and 41.6%. Correlations between these perception measures and compensation were small in each arm, with absolute Spearman rho no greater than 0.09. Since these items were given after the amount judgment and were not randomized in the sequence, they are not considered as mediators. The reason item is used just to record what considerations have invoked the respondents, and not to find a cause for the award.

6. Conclusion

The findings about means and about distributions answer different parts of the question. Although longer operational history led to a higher average award, this effect was not a uniform markup for all participants: lower awards remained more clustered, while upper-half awards were more dispersed, though less so at the highest threshold. So, the randomized result is a result of a heterogeneous distributional answer to the full package of operating-status scenarios, and it does not indicate that it is a constant markup for all the part based on the operating stage or that there is a verified mediation mechanism.

The observed distributional characteristics are a driver of two arms. The standard baseline could match the recorded annual income flow, whereas an evidence-based assessment mechanism could examine confirmed payroll data, active contractual agreements, sunk investments, and transition-related obligations. The operational status only is not enough to active a multiplier. The architecture makes clear what is the base for any other possible element, instead of hiding it in one of the

DOI: 10.54254/2754-1169/2026.38714

discretionary pieces. The criteria of eligibility should be based on a prescribed threshold before enforcements, requires complying with the rules of the game and exclude the spent when to make the construction or the recruitment faster when the rules are predictable. Thus, the income substitution and the help in the transition can be checked in a stand-alone way with the document that has been saved before the prescribed cutoff. In this kind of configuration, the real commitments are recognized, incentive is not wished in a strategic way to premature operations.

Payments for environmental services use the same opportunity-cost idea both to frame participants and to structure contracts. In this study there is one novelty to have in the experimental design, because it is more asking if the verified commitments change when an activity is going from state of ready to the operation. By being documented these commitments before the review of awards and by removing the automatics of the compensation, the criteria of the compensation can be made more clearer and better in accordance with the evidence of the randomized judgement.

References

[1] Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal, 49*(195), 549–52. doi: 10.2307/2224835

[2] Hicks, J. R. (1939). The foundations of welfare economics. *Economic Journal, 49*(196), 696–712. doi: 10.2307/2225023

[3] Coase, R. H. (1960). The problem of social cost. *Journal of Law and Economics, 3*, 1–44. doi: 10.1086/466560

[4] Frederick, S., Novemsky, N., Wang, J., Dhar, R., & Nowlis, S. (2009). Opportunity cost neglect. *Journal of Consumer Research, 36*(4), 553–61. doi: 10.1086/599764

[5] Persson, E., & Tinghog, G. (2020). Opportunity cost neglect in public policy. *Journal of Economic Behavior & Organization, 170*, 301–12. doi: 10.1016/j.jebo.2019.12.012

[6] Maguire, A., Persson, E., & Tinghog, G. (2023). Opportunity cost neglect: A meta-analysis. *Journal of the Economic Science Association, 9*(2), 176–192. doi: 10.1007/s40881-023-00134-6

[7] Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98*(6), 1325–1348. doi: 10.1086/261737

[8] Koszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics, 121*(4), 1133–1165. doi: 10.1093/qje/121.4.1133

[9] Cohen, D., & Knetsch, J. L. (1992). Judicial choice and disparities between measures of economic values. *Osgoode Hall Law Journal, 30*(3), 737–770. doi: 10.60082/2817-5069.1722

[10] van Doorn, J., Zeelenberg, M., & Breugelmans, S. M. (2018). An exploration of third parties' preference for compensation over punishment: Six experimental demonstrations. *Theory and Decision, 85*(3–4), 333–351. doi: 10.1007/s11238-018-9665-9

[11] Keinprecht, M. (2025). Third party loss aversion reduces spectator redistribution. Department of Economics Working Paper. 382. WU Vienna University of Economics and Business. doi: 10.57938/58eea2dd-d9f5-404c-bc58-b3a31fbeea25

[12] Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd.

[13] MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics, 29*(3), 305–25. doi: 10.1016/0304-4076(85)90158-7

[14] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

[15] Lakens, D. (2022). Sample size justification. *Collabra: Psychology, 8*(1), 33267. doi: 10.1525/collabra.33267