Reply to Jann

Market Premium, Social Process, and Statisticism

Yu Xie
University of Michigan

Xiaogang Wu
Hong Kong University of Science & Technology

Statistics is a powerful, yet potentially dangerous, tool. More than two decades ago, the late Otis Dudley Duncan (1984:226) warned us concerning the danger of "statisticism": "the notion that computing is synonymous with doing research, the naïve faith that statistics is a complete or sufficient basis for scientific methodology, the superstition that statistical formulas exist for evaluating such things as the relative merits of different substantive theories." Duncan's warning has long been understood as applying to sociologists who do quantitative research. His concern, however, is equally applicable to readers, even some careful readers, of quantitative research.

In our earlier work (Wu and Xie 2003), we asked the question, "Does the market pay off?" Our emphasis was on the potential heterogeneity of workers in the market sector. Using work history data, we distinguished between two types of workers in the market sector: early birds and later entrants. We were concerned with the possibility that pooling early birds and later entrants, even if they each have the same education returns as stayers in the state sector, may make the returns to education appear higher in the market sector than in the state sector. Our main empirical results, which also were confirmed by Jann, showed that later entrants, but not early entrants, have significantly higher returns to education than stayers.

The thrust of Jann's (2005) comment is that there is insufficient statistical power in the data for differentiating the education returns of early birds from those of later entrants and those of stayers.[1] This point is technically valid, although only within the narrow statistical paradigm of "null hypothesis significance testing," which has been severely criticized in recent decades (e.g., Cohen 1994). A basic problem is that *any* difference is bound to become "statistically significant" with sufficient data. The late John Tukey (1991:100) had the following to say about this paradigm: "Statisticians classically asked the wrong question—and were willing to answer with a lie, one that was often a downright lie. They asked 'Are the effects of A and B different' and they were willing to answer 'no'. . . . [We know] that the effects of A and B are always different—in some decimal place—for any A and B."

In the final analysis, our disagreement with Jann is not about the technical correctness of statistical methods, but about *how* statistical methods should be used in sociological research. We are strong believers in the viewpoint that statistical methodology should not be separated from substantive concerns in guiding research. Jann's "methodological" critique of our work is misdirected precisely because it is narrowly methodological, lacking an understanding of both the substantive research question and the underlying social processes. As we show later, the substantive research question renders irrelevant his test for the difference between early birds and later entrants.

Direct all correspondence to Yu Xie, P.O. Box 1248, 426 Thompson Street, Population Studies Center, University of Michigan, Ann Arbor, Michigan (yuxie@umich.edu) or Xiaogang Wu, Social Science Division, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (sowu@ust.hk). The authors thank Emily Greenman, Jerry A. Jacobs, and Tony Perez for comments and suggestions.

[1] It was the first author who suggested to Jann that his problem with the Wu and Xie (2003) article should be rephrased as a problem of insufficient statistical power.

To be fair, Jann should not singled out for falling into the trap of "statisticism" because such practice is so widespread in current sociology that it often makes quantitative research unappealing. Jann's comment illustrates a common temptation among quantitative sociologists: reliance on canned statistical tests rather than substantive knowledge. Thus, we take this opportunity to draw a general lesson for all of us: only when combined with a substantive understanding of the social processes involved can statistical methods result in fruitful research.

At a fundamental level, Wu and Xie's (2003) study was descriptive. We emphasized this point throughout the article by alerting readers to the danger of aggregation when workers in a single group are in fact heterogeneous. Although we questioned the prevailing wisdom that marketization per se "caused" the education returns to be higher, we never intended our statistical analysis to be more than a descriptive exercise. In such an exercise, formal statistical tests can be informative when there is sufficient statistical power. In the absence of strong statistical information, substantive knowledge should prevail.

In his comment, Jann treats the three groups under discussion—stayers, early birds, and later entrants—as though they were symmetric, like those of an experimental design. In doing so, he borrows the language of multiple-group comparisons commonly used in analysis of variance (ANOVA) associated with experimental designs. However, because we were dealing with observational data, our concerns were with the between-group and within-group heterogeneity generated by *social processes*. The earnings regimes for the three groups resulted from a *cumulative* historical process that clearly is asymmetrical (Figure 1) and thus should be treated as such in an analysis. Comparing the three groups as if they were three experimental conditions is both inappropriate and misleading.

Figure 1 presents a schematic flow chart of the respondent types in the 1996 survey of Life Histories and Social Change in Contemporary China, the data used by Wu and Xie (2003). The y-axis represents the employment sector (state vs. market), and the x-axis represents historical time. We make the convenient assumption that the market sector is an absorbing state so that there is no reverse transition from the market sector to the state sector.[2] In 1978, at the beginning of the Chinese economic reform, 1,197 respondents worked in the state sector. By 1987, 11 percent had made the transition into the market sector ($d = 1$) and are called "early birds." Among the remaining 1,068 workers in the state sector and the 522 new entrants who started in the sector between 1978 and 1987, 16 percent made the transition into the market sector ($d = 2$) and are called "later entrants."[3] The remaining 1,337 respondents are called "stayers."

The Wu and Xie (2003) article was partly responsible for causing Jann's confusion, because his reanalysis is modeled after Wu and Xie's (2003) regressions of logged earnings on education, sector, and their interactions. Extensive discussion in Wu and Xie (2003) on the differential returns to education by sector makes it appear as if education is the causal factor, with sector as a covariate. However, our real research question concerned earnings differences by sector, with education as a confounder. Let us revisit a passage (Wu and Xie 2003:430) cited and emphasized by Jann.

> The crucial difference between the two hypotheses is the treatment of early birds. In Hypothesis 1, early birds are grouped within later entrants because they share the common feature of being in the market sector. . . . In Hypothesis 2, early birds are grouped with stayers because the two types of workers were approaching a convergence, against which later entrants were selectively recruited into the market sector.

Jann assumes that the phrase "are grouped with" here means "share the same education coefficient with." This is incorrect, as shown in Figure 1 of Wu and Xie (2003:431). It is possible, for example, that early birds and later entrants differ in earnings and education distributions but have the same education returns as stayers, yet pooling early birds and later entrants together still could yield higher returns to education

[2] That is, we excluded a small number of "market losers" from Figure 1 because of the group's small size (Wu and Xie 2003). We based the classification on Wu and Xie's comprehensive measure of the market sector.

[3] Here 275 workers, including 82 "later entrants" in Wu and Xie's (2003) original article who entered the labor force after 1987 were dropped from the analysis.

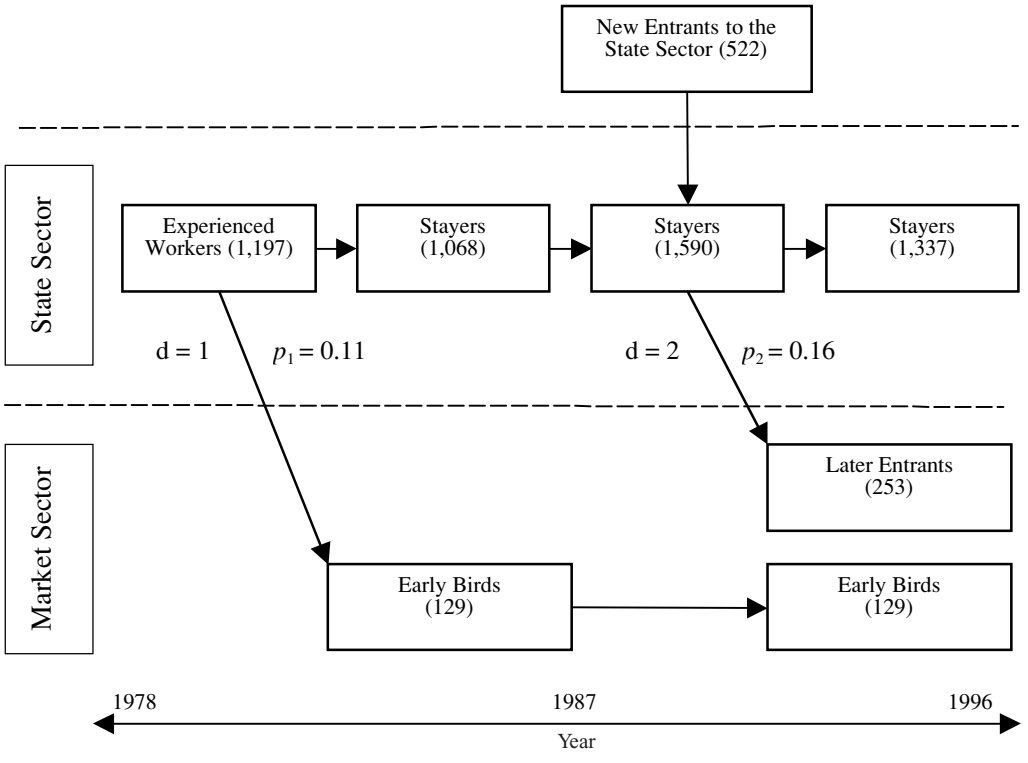

Figure 1. Flow Chart of Labor Market Transitions in China, 1978–1996.

among workers in the market sector than among stayers in the state sector.

To advance this inquiry, let us reconceptualize the substantive problem with explicit counterfactuals in the language of causal inference (Heckman 2005; Holland 1986; Manski 1995; Winship and Morgan 1999). Suppose we are interested in the causal impact of the entry to the market sector on (potential) future earnings in 1996. Conceptually, there are two causal questions in this setup: (1) what is the effect of an early transition? (i.e., $d = 1$) and (2) what is the effect of a late transition? (i.e., $d = 2$). Of course, these two questions are inherently asymmetric. The second is sensible only for those workers who did not experience an early transition, whereas the first involves the counterfactual comparison between those who experienced an early transition and those who did not, regardless what happened to them later. To borrow the notation for causal inference with time-varying treatments (Brand and Xie 2005), let Y^{d}_i denote the ith person's potential outcome if the person has made a transition at time d ($d = 1, 2, \infty$), with $d = \infty$ denoting that the person has not made a transition by the end of the study (i.e., a stayer). Note that for a person who has made an early transition ($d = 1$), the counterfactual outcome should follow the principle of "forward-looking sequential expectation" (Brand and Xie 2005), a combination of a late transition ($d = 2$) and staying ($d = \infty$). We thus define the average causal effect for the first question as

$$E(Y^{d=1}) - E(Y^{d>1}) = E(Y^{d=1}) - [E(Y^{d=2})P_2 + E(Y^{d=\infty})(1 - P_2)]. \quad (1)$$

Note that the transition probabilities are conditional so that $P_2 = P(d = 2 \mid d > 1)$.

For the second question, the comparison is simpler, involving two regime-specific means:

$$E(Y^{d=2}) - E(Y^{d>2}) = E(Y^{d=2}) - E(Y^{d=\infty}) \quad (2)$$

It is never possible to compute quantities defined by equations 1 and 2 because we observe only one of the three potential outcomes for each worker. To infer causality, it is necessary to introduce the ignorability assumption, which must be taken as provisional because it is unlikely to hold in reality. The ignorability assumption

states that all systematic differences associated with the transitions can be summarized by a set of observed covariates (X) (Rosenbaum and Rubin 1984).

Given this assumption, the expected earnings can be estimated on the basis of the observed covariates, including education. As shown in equations 1 and 2, we need four conditional expectations for the causal analyses: $E(Y^{d=1} \mid X)$ and $E(Y^{d>1} \mid X)$ for the first question, and $E(Y^{d=2} \mid X)$ and $E(Y^{d>2} \mid X)$ for the second question. The ignorability assumption means that $E(Y^{d=1} \mid X)$ can be estimated among early birds, $E(Y^{d=2} \mid X)$ among later entrants, and $E(Y^{d>2} \mid X)$ among stayers. However, $E(Y^{d>2} \mid X)$, as a weighted sum of two conditional expectations, should be estimated from both later entrants and stayers. Given that P_2 is small (at .16), a crude approximation of $E(Y^{d>1} \mid X)$ can be estimated from stayers (i.e., giving a full weight to stayers). This approximation is an interpretation of Wu and Xie's (2003) analysis strategy. Because later entrants constituted only a *small* proportion of the appropriate group against which early birds are to be compared, it makes little sense to compare, as Jann recommends, later entrants directly with early birds.

To illustrate the utility of this reconceptualization, we perform propensity score analyses. Because of space limitation, we present only the most important findings in this discussion. The full results from this exercise are reported elsewhere (Xie and Wu 2005). To borrow the jargon from the causal inference literature, we consider two "treatments" in our study: an early entry to the market sector and a late entry to the market sector. For the first treatment, the "control" group consists of workers who did not make an early entry, and thus includes stayers as well as later entrants. For the second treatment, the "control" group consists of stayers only. The propensity score method allows us to summarize all the differences between the treatment and control groups with a single dimension: the probability of receiving a particular treatment. We then compute the average treatment effect on earnings within each propensity score stratum.

There are two main findings from this propensity score analysis. First, the propensity model for a late transition differs from that for an early transition because the mechanisms for making transitions changed. Whereas human capital and political capital measures such as education, party membership, seniority, and cadre connection negatively predicted the probability of an early transition to the market sector, this pattern was much less pronounced for a late transition.

Second, we find the treatment effect of market entry to be very different for a late transition than for an early transition. For an early transition, we find no effects on earnings in any of the propensity score strata. For a late transition, the estimated treatment effect is relatively large and significantly different from zero for the four lowest propensity score strata. We present the results in Figure 2. If we pool the different strata together for an overall treatment effect under the homogeneous effect assumption, the estimate is 236 RMB yuan (Chinese currency), with a standard error of 54, resulting in a highly significant t value of 4.36. However, the assumption of a homogeneous treatment effect is clearly violated by the downward trend observed in Figure 2. Using a hierarchical linear model, we find that the size of the treatment effect strongly and negatively depends on the propensity score, with a unit change in stratum rank (i.e., crossing a propensity score stratum) associated with a reduction of 94 RMB yuan in the treatment effect (a significant relationship, with $t = -3.6$). That is, the benefit of a late transition into the market sector is greatest among those who were least likely to make the transition and diminishes with the propensity for making the transition.

To the question "does the market pay off?" these new results yield no simple answer. We do not find a generic market effect on earnings. Rather, the effects vary across two dimensions. First, confirming Wu and Xie's (2003) earlier results, we again find no evidence of a premium for an early market transition, whereas a late transition into the market sector is associated with higher earnings. Furthermore, we show that even among later entrants, the benefit of working in the market sector sharply decreases with the propensity of having made the transition. Hence, the summary finding of our reanalysis is that the market premium is limited to only later entrants who otherwise had a low likelihood of making a transition to the market sector. Who are they?

In all likelihood, these low-propensity later entrants are workers doing especially well in the

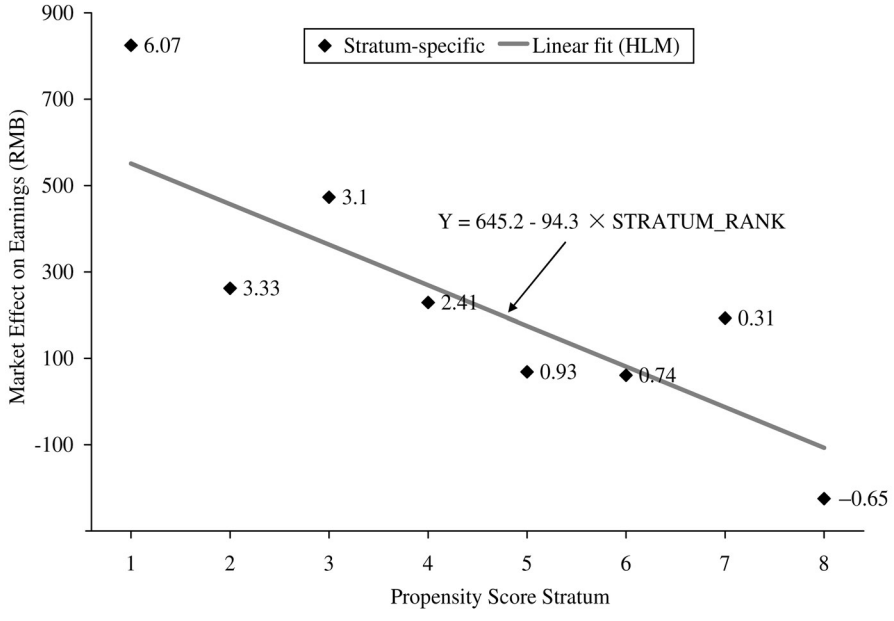

Figure 2. Market Treatment Effect on Earnings by Propensity Stratum: Later Entrants versus Stayers.

Notes: Numbers in the scatterplot are *t* values for an earnings comparison between later entrants (treatment group) and stayers (control group); $t < 1.96$ indicates that there is no significant difference in earnings between the treatment and control groups within a propensity score stratum. The linear fit is based on the hierarchical linear model estimates (level 2 model with slopes from the level 1 model as outcomes regressed on propensity stratum rank); effect of propensity stratum rank is statistically significant ($t = -3.6$).

state sector. A cost–benefit analysis suggests that for a person to make a transition from the state sector to the market sector, the benefit of voluntarily entering the market must exceed that of staying in the state sector. Workers who do well in the state sector and are unlikely to lose their jobs have a good incentive to stay put. For them, the attraction of the market sector must be sufficiently large to more than compensate for the advantages they already enjoy in the state sector. Therefore, only those with the best market opportunities actually make the transition. We highlighted these individuals graphically in Wu and Xie (2003:435), referring to them as voluntary later entrants. These results illustrate a classic violation of the ignorability assumption, the problem of endogeneity. Individuals select their "treatment" on the basis of the anticipated outcome, which is *not* homogeneous across workers. This kind of insight into social processes can never be produced by analyses such as Jann's. His critique focuses on an inappropriate comparison between early birds and later entrants, a result of his reliance on uninformative statistical tests at the expense of substantive knowledge. As such, Jann's statistical exercise contributes little to the understanding of the social processes underlying the empirical patterns reported by Wu and Xie (2003).

Yu Xie *is Otis Dudley Duncan Professor of Sociology and Statistics at the University of Michigan; a Research Professor at the Population Studies Center and the Survey Research Center of the Institute for Social Research; and a Faculty Associate at the Center for Chinese Studies. His main interests are social stratification, demography, statistical methods, Chinese studies, and sociology of science. His published works include* Statistical Methods for Categorical Data Analysis *(Academic Press 2000) with Daniel Powers,* Women in Science: Career Processes and Outcomes *(Harvard University Press 2003) with Kimberlee Shauman, and* A Demographic Portrait of Asian Americans *(Russell Sage Foundation and Population Reference Bureau 2004) with Kimberly Goyette.*

Xiaogang Wu *is Assistant Professor of Social Science at the Hong Kong University of Science & Technology (HKUST). His research interests include social stratification and mobility, labor markets and economic*

sociology, and organization. He has published in American Sociological Review, Social Forces, *and* Demography. *His current research applies causal models to examine the relationship between stratification processes and structural inequalities in China. He recently received a 3-year grant, funded by the Hong Kong Research Grants Committee (RGC), to study social inequality in access to educational opportunities in reform-era China.*

REFERENCES

Brand, Jennie E. and Yu Xie. 2005. "Time-Varying Treatments, Time-Varying Effects: Causal Effects in a Longitudinal Setting." Annual Winter/Spring Meeting of Sociological Methodology Section, American Sociological Association, April 22, Chapel Hill, NC.

Cohen, Jacob. 1994. "The Earth Is Round ($p < .05$)." *The American Psychologist* 49:997–1003.

Duncan, Otis Dudley. 1984. *Notes on Social Measurement, Historical and Critical.* New York: Russell Sage Foundation.

Heckman, James J. 2005. "The Scientific Model of Causality." Department of Economics, University of Chicago, Chicago, IL. Unpublished manuscript.

Holland, Paul W. 1986. "Statistics and Causal Inference" (with discussion). *Journal of American Statistical Association* 81:945–70.

Jann, Ben. 2005. "Earnings Returns to Education in Urban China: A Note on Testing Difference among Groups." *American Sociological Review* 70:860–864.

Manski, Charles. 1995. *Identification Problems in the Social Sciences.* Boston, MA: Harvard University Press.

Rosenbaum, Paul R. and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of American Statistical Association* 79:516–24.

Tukey, John W. 1991. "The Philosophy of Multiple Comparisons." *Statistical Science* 6:100–16.

Winship, Christopher and Stephen L. Morgan. 1999. "The Estimation of Causal Effects from Observational Data." *Annual Review of Sociology* 25:659–707.

Wu, Xiaogang and Yu Xie. 2003. "Does the Market Pay Off? Earnings Returns to Education in Urban China." *American Sociological Review* 68:425–42.

Xie, Yu and Xiaogang Wu. 2005. "Market Premium, Social Process, and Statistical Naivety: Further Evidence on Differential Returns to Education in Urban China." PSC Research Report 05-578, Population Studies Center, University of Michigan, Ann Arbor, Michigan.