How to Evaluate Special Zones for Structural Reform: On a Perspective of Econometric Approach

Wataru SUZUKI*

(Assistant Professor, Department of Education, Tokyo Gakugei University)

(Senior Researcher, Japan Center for Economic Research)

1. Introduction

According to the promulgation of the Law on Special Zones for Structural Reform in December 2002, special zones for structural reform were established for which preferential measures for deregulation have been allowed to promote economic revitalization and structural reform.

For these special zones for structural reform, 324 applications were approved in total for parts 1 through 4 (117 applications for part 1, 47 for part 2, 72 for part 3, and 88 for part 4). Thus, deregulation has been underway since 2003 at a faster pace than expected. The proposals applied are diverse and include: 1) in the educational area, more diverse public education curricula, establishment and management of educational institutions by entities other than existing school corporations, cooperation among kindergartens and daycare centers, and deregulation of construction areas for universities; 2) in the agricultural area, encouragement to companies to engage in agricultural businesses in rural areas, preferential measures for agricultural land to be used for allotment gardens, and approval for acquisition of agricultural land; 3) in the welfare and medical care area, solicitation of private entities for management of special nursing homes for the elderly established by public institutions, introduction of the Private Finance Initiative (PFI), and solicitation of doctors from overseas; 4) in the international exchange and logistics area, around-the-clock customs clearance services, deregulation of approval for comprehensive free trade zones, lending of harbor facilities to private entities, and deregulation of visa-related procedures; and 5) in the IT and new industry creation-related areas, less restrictions on national university faculties to participate in additional jobs, promotion of industry-academia cooperation by opening national university facilities to the public, and deregulation of next-generation energy and recycling.

As pointed out by Hamada (2004), special zones for structural reform are very unique in solely promoting deregulation without any financially preferential measures such as tax reductions and subsidies. Thus, effective implementation of special zones for structural reform should contribute to economic revitalization without any costly investment, and be supposed to promptly promote nationwide deregulation. In this sense, it can be defined as a nationwide economic measure, rather than that only for regional economic revitalization. Deregulation promoted by establishment of special zones for structural reform will be subject to open policy evaluation to make decisions on national development. Such a clarified relationship between policy evaluation and operation is

absolutely unique and innovative, compared to other economic measures in Japan.

Making decisions on the national development of deregulation by policy evaluation means that the persons engaged in evaluation play important roles as they have the power of control on deregulation. Accordingly, evaluations of special zones for structural reform should be conducted through objective and scientific measures that all citizens can accept, based on transparent and fair policy evaluation process. At present, "the Evaluation Committee" established within the Headquarters for the Promotion of Special Zones for Structural Reform makes comprehensive evaluation on the special zones, and reports its opinions to the Headquarters, which will decide subsequent measures. To keep its neutral position, the Evaluation Committee comprises 10 members selected from among academic experts and private business operators, three of whom are recruited from the public. The Committee gives appropriate evaluations through its eight ad-hoc groups, dealing with areas such as medical care, welfare and labor; education; rural area revitalization; energy and safety; national land and logistics; industrial promotion; international exchange; and regional revitalization. The evaluation process is as follows[1]: Relevant regulatory agencies in charge of deregulation measures carried out in special zones shall investigate drawbacks caused by deregulation by gathering specific data or making questionnaire surveys, and report the results to Headquarters. Alongside this investigation, the Evaluation Committee will, by itself, make a survey on the influence of the national development of preferential measures for deregulation, verify the drawbacks, and make evaluations on any special concerns. Three alternative evaluations shall be given by the Evaluation Committee as follows:

(A) The said preferential measures for deregulation shall be carried out nationwide without any regional restrictions;

(B) The said preferential measures for deregulation shall be continuingly carried out in relevant regions that have specific regional characteristics; and

(C) The said preferential measures for deregulation shall be abolished.

Evaluation (A) shall be applicable when:

a) it is perceived that there are no drawbacks;

b) even if there are some drawbacks, prevention of such drawbacks can be ensured by reviewing requirements and procedures for the preferential measures for deregulation, and it is perceived that verification of the revised preventive measures is not necessary in the special zones; or

c) drawbacks are relatively small, and it is perceived that the effect of nationally developed preferential measures for deregulation is far more significant compared to such drawbacks.

In the meantime, Evaluation (B) will be applied when, even if there are some drawbacks, preventive measures against such drawbacks can be ensured by reviewing requirements and procedures for the preferential measures for deregulation, and it is perceived that verification of the corrected or added preventive measures is necessary in the special zones. Evaluation (C) will be applied when there are some drawbacks and it is perceived that it will be difficult to take preventive measures against such drawbacks by reviewing requirements and procedures for the preferential measures for deregulation.

In the Basic Policies for Economic and Fiscal Management and Structural Reform 2003 (Cabinet Decision of June 27, 2003), it is stipulated that "the deregulation measures for special zones shall be immediately subject to national development if it is judged that there is no noticeable problem." Thus, the evaluation process for special zones is characterized by national development being a "default" applied "but for any noticeable problems". Policies requiring budgetary measures are usually evaluated by verification of their "efficiency". In other words, it is unprecedented to accept those policies only without any problems, and the system is thus extremely convenient to promote deregulation[2]. Surprisingly, it is a responsibility of the relevant regulatory agencies to verify if there are

---

1) Sources: "Evaluation Policy for Special Zones for Structural Reform (draft)" (a reference for the Evaluation Committee, December 18, 2003), "Basic Policies for the Promotion of Special Zones for Structural Reform," etc.

2) One fair reason for this may be that deregulation basically involves no budget allocation. Of course, it does not mean that its effectiveness is not evaluated. The Evaluation Committee verifies the effectiveness through its own surveys.

any noticeable harms in deregulation[3]. As shown in medical or pollution lawsuits, any party bearing the burden of proof is more disadvantageous in general. Therefore, it also significantly works for deregulation. A shorter evaluation period (the first evaluation is to be conducted in the first half of fiscal 2004) is another advantage for deregulation.

Now, what kinds of evaluation methods should be taken? Can they be expected to give scientific and objective evaluations? Unfortunately, as far as I can see from the materials and schemes of the evaluation methods disclosed on the Cabinet Office's website[4] by the Evaluation Committee, I cannot help being concerned about them. At present, 1) a survey plan for the first half of FY 2004 (in draft)[5], 2) the regulatory agencies' survey plans[6], and 3) (a sample of) evaluation survey methods that are supposed to be examined independently by the Evaluation Committee (materials prepared by a private investigation company)[7] are made available for each of ad-hoc groups. Although they are not final and may be subject to change, they are, in general, too far from the standard methods for program policy evaluation, and seem to have many problems. For example, although policy evaluation should usually require investigations not only on the regions where the policies are carried out but also on other regions where they are not carried out for comparison, the survey plans seem to only cover special zones. Indicators for policy evaluation, too, are mainly based on the number of cases in which deregulation is conducted, while there are sometimes no performance indicators set for policy goals. Even if such performance indicators are set, a number of projects are often evaluated by extremely unobjective indicators, such as the results of satisfaction surveys. The Evaluation Committee itself is supposed to investigate the effects of national development of deregulation, but it must first overcome various methodical problems including General Equilibrium Treatment Effects, as mentioned later, to judge whether measurement results in a specific area can be applicable nationwide. However, at least with reference to the materials above, I am not convinced that the Committee is trying to address such problems. The evaluation of deregulation is politically challenging because, in most cases, some parties are in fact deprived of their vested rights as a result of deregulation. As such, it is necessary to compare the loss incurred by any vested right holder with a new benefit generated by deregulation. However, no survey method to make a direct comparison between the cost and benefit seems to have been devised yet. As a matter of fact, in the evaluation "(A) The said preferential measures for deregulation shall be carried out nationwide without any regional restrictions" above, there could be many cases to which a judgment standard of "c) drawbacks are relatively small, and it is perceived that the effect of nationally developed preferential measures for deregulation is far more significant compared to such drawbacks" is applicable. However, the method of comparison is still unclear, and there remains a concern about its effectiveness.

Keeping this in mind, I will, in this paper, discuss designing the evaluation of special zones based on the econometric methods of program policy evaluation used in economics. I will introduce the methods of standard program policy evaluation in Section 2; discuss the issues inherent in special zone evaluations and the investigation schemes to tackle these issues in Section 3; introduce three case studies of evaluations that have been disclosed by ad-hoc groups while addressing the relevant issues and improvements in Section 4; and come to a conclusion in Section 5.

---

3) The Basic Policies for Economic and Fiscal Management and Structural Reform 2003 specify the evaluation methods as follows: "regarding the investigation of preferential measures for deregulation, the relevant regulatory agencies shall be responsible for verifying harmful effects due to the national development of deregulation based on whether the preferential measures applied for special zones have caused any harms or not."
4) http://www.kantei.go.jp/jp/singi/kouzou2/hyouka.html
5) Reference materials for the fourth ad-hoc group meeting.
6) Reference materials for the fourth ad-hoc group meeting.
7) Reference materials for the third ad-hoc group meeting.

2. Standard Methods for Program Policy Evaluation[8]

2.1 Difference Estimator and endogeneity

An "experiment" is an ideal evaluation method for a certain policy. As a matter of fact, in the areas of medicine, psychology and natural science, experiments are conducted to evaluate the policies. For example, in a new drug evaluation, the drug and placebo are randomly distributed to the subjects to verify the differences between the two. Strict random distribution called double blind method[9] is not to disclose to the subjects or distributors which is a new drug or a placebo. Although it is difficult in economics to carry out such strict experiments, large-scale random-subject experiments called Social Experiments are often conducted in the United States: for example, the well-known Negative Income Tax Experiment; RAND Health Insurance Experiment conducted by RAND Corporation about the impact of health insurance on health care demand; and Student Teacher Achievement Ratio (STAR) Project that surveyed the effects of small-class education and subsidy funding.

In evaluation of such experiments, a target outcome is set first for the treatment. Examples are the scores of the Stanford Achievement Test for the STAR Project that evaluates the treatment effects on educational projects, and the data on receiving medical treatment influenced by insurance (the number of days for outpatient treatment, the number of days for hospitalization, and medical expenses) for the RAND Health Insurance Experiment. Subjects are randomly divided into the Treatment Group and the Control Group (without treatment). An evaluation will be made based on the comparison of differences in the outcome between these two groups. The easiest measurement method is a regression analysis in which the outcome is defined as Y_i, the treatment implementation level as X_i, and other observable factors that may decide the outcome as $W_{i,j}$, as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \sum_j \beta_j W_{i,j} + u_i \qquad (1)$$

For example, if an assumption is made that a new educational program is carried out, Y_i could be the school achievements; X_i is the dummy variable, in which the Treatment Group is 1 and the Control Group is 0; and $W_{i,j}$ is the determinant of the built-in achievements, such as the parents' income and academic background, enrollment in private or public schools, and attendance in after-hours cram schools. The error term u_i in this case includes an unobservable factor which decides Y_i : for example, students' academic abilities such as IQ, and enthusiasm for learning. If u_i satisfies the following condition, known as Conditional Mean Independence Assumption (CMIA), β_1 shall be estimated without bias.

$$\text{CMIA } E(u_i | X_i, W_{i,j}) = \gamma_0 + \sum_j \gamma_j W_{i,j} \qquad (2)$$

CMIA, which means that u_i has no correlation with X_i when $W_{i,j}$ (any factor except the treatment) is controlled, will be satisfied if X_i is randomly allocated for each factor $W_{i,j}$. In the case of the educational program above, it falls under the factor of enrollment in private or public schools. Although the private school dummy X_i obviously affects students' achievements, CMIA will be satisfied if the subjects are randomly sampled at private and public schools respectively[10]. Then, the Treatment Effect, or the Causal Effect, is defined as follows:

$$E(Y_i | X_i = 1, W_{i,j}) - E(Y_i | X_i = 0, W_{i,j}) = \beta_1 \qquad (3)$$

---

8) This section is a primary introduction for an econometric method of program policy evaluation conducted in economics. Good textbook examples include Heckman and Smith (1995), Wooldridge (2001), and Stock and Watson (2003). Tomioka (2004) is probably the first comprehensive survey done in Japan. Ono (2003) is an example that escapes from defining policy evaluation in a narrow sense, which is usually the case in economics, but explains it in wider terms, especially in connection with special zones.

9) If a subject knows whether he/she has been given a placebo or a new drug, such psychological influence may negatively affect the strictness of the experiment. It is also known in social experiments that the psychological effect, wherein a subject knows he/she is the subject of the experiment, may cause a bias (i.e. Hawthorne Effect).

10) Random sampling also satisfies this condition, completely regardless of $W_{i,j}$.

and it will be estimated by Ordinary Least Squares (OLS) without bias. This is called a Difference Estimator in the econometric analysis of program policy evaluation. On the contrary, the Difference Estimator will be biased if the CMIA is not satisfied. For example, if the treatment X_i is not based on random sampling but instead on participants' application for the experiment, it is thought that there is a correlation between u_i, which includes students' academic abilities and enthusiasm for learning, and X_i. In this case, the treatment effect β_1 may be either overestimated or underestimated with bias. However, in policy programs economics usually deals with, as opposed to social experiments, the treatment is usually conducted on people who wish to attend, and it is therefore extremely difficult to avoid this problem known as an endogeneity problem or selection bias. The endogeneity problem may also arise in social experiments in which subjects are chosen randomly. For instance, in a new educational program, some families may refuse to join the program due to a school transfer, etc. (Partial Compliance) or abandon the program due to a loss of enthusiasm or academic abilities (Attrition). In such cases, the actual program attendance X_i and the error term u_i correlate with each other, and the bias in the Difference Estimator may be large (Heckman and Smith, 1995).

The most typical solution for such an endogeneity problem is to use the instrumental variable. An unbiased estimate can be obtained by finding the instrumental variable $E(u_i \mid Z_i) = 0$, which correlates with X_i. For example, in the case of Partial Compliance or Attrition of a new educational program, even if the actual enrollment X_i is correlated with u_i, it can be dealt with relatively easily because the original random sampling can be used as the instrumental valuable Z_i. However, it is hard to deal with the endogeneity problem such as participation in a program by application because finding an instrumental valuable is extremely difficult in the first place[11]. Other options include an estimate using a simultaneous equation setting X_i as a model. However, as a model setting has a significant influence on the estimate results, a new problem will arise, that is, how to select appropriate models. Another option is random selection of program participants among applicants. This brings no bias because there will be no correlation between participation and unobservable factors such as academic ability and enthusiasm. However, the problem here is that the measurement of the effect obtained only covers "the applicants". In other words, if we are considering expanding the coverage of the program to include those who have not applied, as in the case with the special zones, we will discover that we are unable to obtain information about the effect of the program on those people[12].

2.2 The Difference-in-Difference Estimator

As mentioned above, it is difficult for ordinary policy program evaluations, which cannot make social experiments, to deal with endogeneity problems. However, ordinary policy programs may carry out exogenous treatment in which individuals cannot influence implementation of the treatment. A typical example is a program covering all the persons in a special zone. In this case, the treatment is exogenous because people in the special zone may have no other choice. Such treatment is known as a Natural Experiment or a Quasi-Experiment, which can measure a non-biased effect through a method called the Difference-in-Difference Estimator which makes use of the information obtained before and after the treatment from both the Treatment Group within the special zone and the Control Group outside the zone[13]. The Difference-in-Difference Estimator gives an estimate as shown in

---

11) It is also known that a low correlation between the instrumental variable and X_i causes a bias (Staiger and Stock (1997)). Furthermore, it is pointed out that there were some problems due to insignificant instrumental variables even in some past studies (Angrist and Krueger (1991) and Card and Sullivan (1988)), which have been known to be adequately successful in setting instrumental variables by making use of natural experiments (Rosenzweig and Wolpin (2000)).

12) Another study is underway to nonparametrically control unobservable factors using observable variables that correlate with unobservable factors, and it has already produced some results. Nevertheless, I will not discuss this in this paper as it is a too advanced topic. Please refer to Heckman, Ichimura and Todd (1997,1998) for more detailed information.

13) In recent Japan, the Difference-in-Difference Estimator has been applied much more frequently. Recent examples include Abe (1997), Oishi (2003), Otake and Yamaga (2003), and Suzuki (2004).

(4) below, making use of the difference (ΔY_i) in an outcome of the same subject[14] before and after the treatment.

$$\Delta Y_i = \beta_0 + \beta_1 X_i + \sum_j \beta_j W_{i,j} + u_i \qquad\qquad (4)$$

The Difference-in-Difference Estimator compares the difference in behavior of the same individual before and after the exogenous treatment with the difference in that of a homogeneous group without treatment during the same period. Thus, β_1 will be equal to the difference in the outcome of the Treatment Group before and after the treatment ($\Delta Y_{i,treatment}$) minus that of the Control Group during the same period ($\Delta Y_{i,control}$) ($\beta_1 = \Delta Y_{i,\ treatment} - \Delta Y_{i,control}$). One of the advantages of the Difference-in-Difference Estimator is to be able to completely control the changes of society as a whole during the said period of time by removing the individual effect (including unobservable factors such as abilities and enthusiasm) through comparison between different points of time for the same individual and comparing with the Control Group. Thus, endogeneity problems that came across in the discussion about the Difference Estimator can be solved. However, if the treatment implementation is not exogenous to individuals or entities like corporate bodies, CMIA is not satisfied due to the correlation between u_i and X_i, so that bias will result. This means that bias occurs even when prior access to information about the treatment influences and changes the pre-treatment behavior. A condition of homogeneity between the Treatment Group and the Control Group for each $W_{i,j}$ is also attached by CMIA.

2.3 Other problems

As mentioned above, even if the Difference-in-Difference Estimator is obtained by making use of Natural Experiments or the Difference Estimator is obtained by appropriately setting an instrumental variable, whether or not these results can be applied to outside the specific areas is another story. For example, if an experiment (Natural Experiment) is conducted in extremely special areas or on extremely special subjects, generalization is difficult (Nonrepresentative Sample). Even in such a case, if adequate samples are available, it is possible to revise the estimate by artificially creating a sample homogenerous to generalized subjects (e.g. the nationwide) through the Bootstrap or other resampling methods. However, if only a few samples are available, it is difficult to make such revisions, and, among others, estimation by a small number of samples damages the very credibility of the estimate of specific areas.

Another serious problem is what is called General Equilibrium Treatment Effects. For example, assuming that an experiment (Natural Experiment) finds an effect that occupational training contributes to higher employment rates and wages, this effect may, in fact, only come from the depletion of employment opportunities of the people in the neighboring areas who are not the participants of the program (Displacement Effect). In this case, we need to judge the nationwide effect by taking into account the people who have been obviously deprived of their job opportunities. In the meantime, higher wages in a certain area may end up at a much lower rate than that measured because if similar occupational training programs are launched nationwide, the supply of trained personnel will increase. As policy evaluation often gives regional and short-term measurement of a partial equilibrium effect, the measured effect that results in general equilibrium seems to have more bias than expected (Heckman, Lochner and Taber, 1998). In this case, an alternative solution such as an evaluation based on the Computable General Equilibrium (CGE) analysis model can be taken. It is difficult, though, for the current CGE to have adequate accuracy for quantitative evaluation, and the quantitative modification of the measured values requires future scientific development. We should also remember the Lucus Critique issue that the measurement results are affected by the immediate change in people's expectations upon national development(Lucus, 1976).

---

14) The Difference-in-Difference Estimator may make comparisons of repeated cross section data, rather than the same individual's data, before and after the treatment.

3. How to Design Policy Evaluation of Special Zones for Structural Reform

Now I would like to discuss challenges and points we should consider when applying standard evaluation methods mentioned in Section 2 to special zones for structural reform and how we should deal with them.

3.1 How to deal with endogeneity problems

The most important problem in evaluations of special zones, as opposed to social experiments, is that the zones are specified based on voluntary applications of municipalities, rather than randomly assigned. The problem is not so serious if evaluations are made on respective individuals or entities such as corporate bodies within special zones, because, for individuals and corporate bodies, deregulation promoted by establishment of special zones is regarded as exogenous structural reform free from their influences, which means the Difference-in-Difference Estimator can be used. Thus, the data of individuals and corporate bodies in special zones, as well as those of individuals and corporate bodies in non-deregulated areas outside the zones, should be gathered before and after approval of special zones, so that the formula (4) can be calculated. However, there are some points we should notice. First, as in the case like the Ota City special zone for foreign language education, for which applications from outside the zone are also acceptable, outside applicants' participation in the program obviously becomes an endogenous variable. Therefore, measurement should be taken only on the applicants within the zones. Second, we must be careful in treating the individuals and corporate bodies in special zones when they have a choice in accepting deregulation. For instance, even the people in special zones can choose whether to make enrollment in kindergarten for children under three years old, and thus a similar endogeneity problem can happen here. In this case, setting the Treatment Group in special zones as an instrumental variable, an estimate may be made by the instrumental variable method. Third, in special zones for industrial promotion, some applications for approval as special zones were submitted simply because there were corporate bodies that wanted deregulation to begin with, and thus these applications cannot be regarded as exogenous structural reform. In this case, special attention should be paid at the measurement to excluding corporate bodies with questionable endogeneity, such as the launching of plant operations right after the approval of special zones. Fourth, if the approval of special zones were expected in advance, no desirable estimate could be made on the data sampled before structural reform, even though the data covers the period before the approval. Accordingly, it is recommended to gather the data for a longer period of time going as far back to before such expectation arose.

However, endogeneity of applications for special zones can present a problem when we are obliged to observe only one subject as a whole, as shown in the special zones for harbor logistics or airports, rather than individuals and corporate bodies themselves. In this case, "participation in the program" is obviously endogenous because applicants are those municipalities with enthusiasm and abilities in deregulation of logistics. Therefore, as mentioned in Section 2.1, the solutions include devising an instrumental variable or simultaneously estimating a model to decide participation in the program itself. However, it can be extremely difficult to set realistically appropriate instrumental variables or models. The alternative for this is to randomly allocate different times for approvals among the applying municipalities and gather data from both with and without approvals. In this case, no endogeneity problem occurs because special zones are approved regardless of participants' enthusiasm or abilities. The measured effect is, however, only from municipalities "with enthusiasm and abilities". Effects by applying deregulation to unenthusiastic and incompetent municipalities, that is, deregulation nationwide, cannot be measured. Accordingly, if possible, some treatment might be required to give approvals for special zones even for municipalities which did not submit voluntary applications.

3.2 General Equilibrium Treatment Effects

As mentioned in Section 2.3 above, General Equilibrium Treatment Effects are diverse. The most serious impact on evaluations for special zones is the Displacement Effect working against neighboring areas. For example, as seen in the Utsunomiya nigiwai special zone, if an evaluation is made on only the area subject to deregulation under the Large Scale Retail Stores Law, the special zone effect is supposed to be excessive due to the customer drain from neighboring communities. In this case, as the Displacement Effect is eliminated after deregulation has spread throughout the country, the said effect should be deducted from the estimate.

Measurement can be done by gathering the data from all areas, including neighboring areas where customers possibly go. In some special zones, these areas are not limited to the neighboring areas. For example, for special zones for harbor logistics, other harbors of an equivalent scale should possibly be included even if they are far apart; in case of special zones for airports, the data for airports at a farther distance may have to be collected. This is not only a matter of deduction of the Displacement Effect but of setting the Control Group. Thus, to see the effects of special zones, data of appropriate areas that are not special zones has to be chosen as that of the Control Group, and neighboring areas of a similar size are usually chosen due to a similarity in natural conditions, etc. However, choice of any neighboring area that can be influenced by the Displacement Effect significantly lowers the Control Group's comparison effect, and it doubly results in bias that excessively measures the effects of special zones. In such cases, neighboring areas cannot be set as the Control Group.

3.3 Comparative view of the size of drawbacks and effects

No matter what kinds of deregulation are carried out, there are always vested right holder groups damaged by deregulation. In this sense, deregulation is almost impossible without implying damage. We should consider how the deregulation effect should be compared with damage to evaluate the net effect (positive effect minus negative effect). The Evaluation Committee stipulates in its reference materials that "(A) The said preferential measures for deregulation shall be carried out nationwide without any regional restrictions" if "c) drawbacks are relatively small, and it is perceived that the effect of nationally developed preferential measures for deregulation is far more significant compared to such drawbacks." However, it is not clear what specific comparative measures will be taken. As a matter of fact, in most survey plans prepared by ad-hoc groups, no direct comparison is available between surveys by the relevant regulatory agencies on harmful effects and those by ad-hoc groups or contracted private investigation companies. It is often impossible to directly compare deregulation effects with damage. In this case, some kind of "comprehensive judgment" is required, but the judgment tends to subjective and highly likely to be unpersuasive.

To facilitate the comparison between the effect and damage resulting from deregulation, it is necessary to replace all effect/damage indicators with comparable "comprehensive indicators". The best prospect is monetary value as economic benefit (or drawback). If deregulation is reflected in cost reduction and added values, its effect can be converted to monetary values, so, too, is the loss of vested right holders. The advantage (or disadvantage) of consumers and vested right holders can be also converted to benefit (or loss) by a stated preference method called the Contingent Valuation Method (CVM). CVM, which has been developing mainly in environmental economics, public economics and experimental economics, asks respondents to estimate monetary values of assets that are not traded in markets and virtual projects. Typical CVM examples are Willingness to Pay (WTP), Willingness to Accept (WTA), and Conjoint Analysis (CA).[15] Now, there is a problem. As it is called "strategic bias," once the respondents learn that the estimated amount will have an actual influence on policy decision-making, those who will receive benefits may overstate the effect, while those who will incur losses may understate the effect. However, if a system can be established to use the stated preference values within special zones only for applying the deregulation of the special zones to other municipalities or throughout the country and not for revising or abolishing these zones, CVM can be investigated and designed to exclude strategic bias, because it will be freed at least from the interests of vested right holders and beneficiaries within the zones. As CVM can be investigated even in the areas where deregulation has not been actually carried out, it is possible to make surveys in other areas for a more accurate calculation of the effect of nationwide applications.[16]

15) Among quite a number of papers about CVM, those specifying medical care applications include Tsuji, Suzuki, et al. (2002) and Suzuki and Ohkusa (2000).

16) This is called the possibility of Benefit Transfer. CVM, including WTP, makes evaluations for the benefit amount in all the areas (where questionnaire surveys are not available) by the Benefit Transfer function calculated from sample areas, after checking the possibility of Benefit Transfer from the survey results of multiple sample areas. Special zones may also need such processes.

4. Evaluation Method Review

Finally, I would like to show some examples of evaluation plans scheduled to be carried out in the first half of FY2004 as case studies, point out the challenges, and seek solutions. As already mentioned in Section 1, the Evaluation Committee has disclosed its reference materials as follows: 1) a survey plan for the first half of FY 2004 (in draft); 2) the regulatory agencies' survey plans; and 3) (a sample of) evaluation survey methods (materials prepared by a private investigation company). Among the programs equipped with all of the above materials, I would like to mention three programs: (i) employment placement program through the establishment of the government-industry joint counter (ad-hoc group on medical care, welfare and labor); (ii) commercial revitalization program in core downtown areas (ad-hoc group on regional revitalization); and (iii) trade promotion program through the reduction of overtime customs clearance charges (ad-hoc group on national land and logistics).

4.1 Employment placement program through the establishment of the government-industry joint counter (ad-hoc group on medical care, welfare and labor)

This is the program, approved as part 1, conducted in "the special zone for employment creation by use of human resource businesses" in Adachi-ku, Tokyo. Hello Work Adachi and a placement service company jointly open a counter in the Adachi-ku Office and work cooperatively on employment placement.

The Evaluation Committee carries out "the surveys on effects of the program", while the relevant regulatory agencies conduct "the surveys on its harmful effects". The former comprises "surveys through local public entities" conducted regularly by the secretariat and "direct surveys on service providers, demanders and consumers" carried out as required. As "surveys through local public entities" are typical surveys, which are in keeping with almost all of the other programs, respondents should describe three points: 1) progress indicators; 2) self-evaluation of the progress; and 3) specific achievements. Specifically, as for 1) progress indicators, the number of visitors to the joint counter and the number of employment agreements concluded[17] should be specified as of April 2004 and theretofore (if any comparative figures are available). As for 2) self evaluation of the progress, the five-grade evaluation shall be made as follows: A) the objective is almost achieved or steady progress is being made; B) even though problems exist, progress is being made; C) the program has just been launched and it will take a while to get some results; D) the program is in a preparatory stage with specific efforts on standby; and E) progress has come to a standstill due to damage incurred. As for 3) specific achievements, which should be described, no specific requirements for descriptions[18] are made. As for "direct surveys on service providers, demanders and consumers", it is specified in the reference material, "(a sample of) evaluation survey methods (materials prepared by a private investigation company)," that, starting from the FY2004 first quarter, surveys shall be conducted annually on 1) the number of job applicants who have made use of the program; 2) the employment rate of the job applicants stated in 1); and 3) satisfaction level of service providers and job applicants for this program. In the meantime, in "the surveys on harmful effects of the program" by the relevant regulatory agencies, questionnaires and hearings are given to the counter attendants of private service providers and Hello Work, as well as to job applicants, employers, and local public entities. Survey items include the status of job placement, complaints, response of visitors, privacy measures, waiting time, requests from visitors, financial burdens, merits and demerits.

Based on the discussions in Section 2.3 above, the following problems and improvements can be pointed out for these surveys. First, the status of employment placement of Adachi-ku as a whole including private service providers and Hello Work should be identified, rather than collecting the data only from the joint counter at the Adachi-ku Office. As for comparative data before the special zone status was approved, data (on the job placement status in Adachi-ku as a whole and the success rate of placement) should be collected for one or two years prior to the first expectation of the approval. Second, a survey on the same items should be made of other areas than

---

17) Respondents are instructed to list other appropriate indicators for the effect, if any.
18) One example for description: Economic benefit by preferential measures.

Adachi-ku as the Control Group. Given that the Displacement Effect may show up, it is recommended to gather data from both neighboring and distant areas. Third, satisfaction surveys are insignificant unless comparison can be made between before and after the use of service or with other channels than the joint counter. Fourth, some possible ideas to enable comparison between the effect and damage in monetary values are to make surveys on the WTA (Willingness to Accept) values damaged due to a breach of privacy, the increase in personnel expenses caused by inappropriate allocations of Hello Work, and the income of those who have succeeded in finding jobs through the joint counter.

4.2 Commercial revitalization program in core downtown areas

This program aims at soliciting large-scale retail stores to move to the Utsunomiya downtown and the area around JR Utsunomiya west exit, Tochigi Prefecture, by simplifying the procedures pursuant to the Large Scale Retail Stores Law. "The surveys through local public entities" conducted by the Evaluation Committee secretariat cover items such as 1) the number of stores as an indicator of the progress made; and 2) a five-grade self-evaluation of the progress made. In the meantime, according to "(a sample of) evaluation survey methods (materials prepared by a private investigation company)," "the direct surveys on service providers, demanders and consumers" should be made of chambers of commerce and industry on items such as 1) the establishment status of large scale retail stores (name, establishment time, size, and effect); 2) influence on daily life environment in the neighborhood; and 3) whether preferential measures to solicit large scale retail stores have promoted enough awareness of the retailers for construction of such stores. In "the surveys on harmful effects of the program" by the relevant regulatory agencies, the chambers of commerce and industry, as well as other business associations in the neighborhood, are asked, "Has there been any damage to the daily life environment in the neighborhood that is identified in the requirements for the large scale retail store operators under the Large Scale Retail Stores Law? If so, are you able to identify the cause for such damage? What do you think is necessary to prevent the damage?" (Respondents can give answer freely.)

The most important concern about these evaluation methods is that the evaluation time (i.e. the first half of FY2004) may be too soon. No strict evaluation can be expected in the awareness survey for large scale retail store operators. Utsunomiya applied for the special zone after experiencing the successive withdrawal of large-scale retail stores. This means that the starting point is very low and thus special attention should be paid because the effect of the large-scale retail store launch may be overly valued.[19] With this in mind and in light of discussions in Section 2.3, first, the survey should also be conducted on other comparable areas as the Control Group for Utsunomiya. Similarly, as the Displacement Effect is expected in the neighboring areas, they should also be surveyed separately. Second, for direct comparison between damage and effect, objective indicators need to be evaluated, such as sales and list prices of large-scale retail stores, the decrease in sales of local retail stores and their list prices. It is recommended that local residents be given the WTP survey because the indicators of awareness and the number of successful stores cannot show final policy goal evaluations and are difficult to compare with damage considered especially serious.

4.3 Trade promotion program through the reduction of overtime customs clearance charges

This program has been approved for Yokohama and Kobe ports and aims at promoting trade traffic as well as establishing the overtime customs clearance system. "The surveys through local public entities" conducted by the Evaluation Committee secretariat cover items such as 1) the number of applications for overtime customs clearance services as an indicator of progress made and 2) a five-grade self-evaluation of the progress. In the meantime, according to "(a sample of) evaluation survey methods (materials prepared by a private investigation company)", "the direct surveys on service providers, demanders and consumers" should cover the following items: for the customs, 1) increase of overtime customs clearance due to preferential measures; 2) the state of costs,

---

19) Such an effect, known as the Ashenfelter's dip, is also shown in occupational training programs.

income and balance influenced by preferential measures; for trading firms, 3) the status of and reasons for the use of customs clearance; 4) changes in costs and lead time due to preferential measures; 5) the conditions and willingness to facilitate more frequent use of overtime customs clearance; for customs brokers, 6) the status of and reasons for the use of customs clearance; 7) the conditions and willingness to facilitate more frequent use of overtime customs clearance; and 8) the state of costs, income and balance influenced by preferential measures. In "surveys on harmful effects of the program" by the relevant regulatory agencies, the only survey item is the decrease in clearance charges due to the application of discounts.

For this program, evaluation indicators are clear as the increase in quantity and amount passing through customs and the increase in trade volume, while the damage against them can be directly compared with the decrease in clearance charges, which makes evaluation fairly easy. Surveys, however, should also be conducted on harbors in other areas, in addition to the special zones where implementation of this program is approved. It may be appropriate to make surveys on other countries, for instance, South Korea. The Displacement Effect is expected to exist beyond the neighboring areas, too.

5. Conclusion

In this paper, I have discussed about evaluations of special zones to start in the first half of FY2004, in terms of econometric analysis of program evaluation which is showing rapid development. Special zone evaluations are collectively conducted by "the Evaluation Committee" established within the Headquarters for the Promotion of Special Zones for Structural Reform, and their frameworks are extremely favorable for deregulation. However, I am skeptical that evaluations will be carefully, scientifically, and objectively conducted. From the latest disclosed reference materials, it can be pointed out that 1) sometimes the choice of final evaluation indicators to evaluate special zones is inappropriate; 2) it is necessary to gather, in addition to the indicators of special zones, the indicators of the areas where the Control Group is located; 3) the influence on neighboring areas should be surveyed if the Displacement Effect is expected; and 4) for comparison between damage and effect, it is recommended to replace evaluation indicators with monetary values. Special zones are established to promote deregulation throughout the country, rather than only for regional revitalization. Decision on national development depends on the evaluation results, and, in this sense, the evaluation is more important than ordinary policy evaluations. I earnestly hope that deregulation will proceed as a result of a national consensus, without excessively depending on the opinions of some vested right holders. To this end, objective and scientific evaluations to persuade the whole nation are required.

[Bibliography]

Ono, Tatsuya, "Requirements of Special Zones for Structural Reform to Revitalize the Japanese Economy and Society – Special Zones for Structural Reform as a Social Experiment," Fujitsu Research Institute (FRI) Report (2003), No. 177.

Tsuji, M., Suzuki, W., Taoka, F., Tejima M., "Applicability of the Contingent Valuation Method (CVM) for Medical Technology Evaluation: Analysis of an Alienation Factor between WTP and WTA by the Survey Data," Medicine and Society (The Health Care Science Institute, 2002), Vol. 12 No. 2, pp. 107-119.

Tomioka, Jun, "Econometrics of Program Evaluation: The Survey," Administrative Evaluation Study Group Report (Japan Center for Economic Research, 2004).

Abe, Yukiko, "Elderly Male Labor Supply and an Old Age Pension System for Those Still Working in 1980-1990," JCER Economic Journal (1997), No. 36, pp. 1-32.

Akiko, Oishi, "Labor Supply of Married Women, Taxation and Social Security System," Quarterly of Social Security Research (2003), Vol. 39 No. 32, pp. 86-300.

Suzuki, Wataru, "Analysis of Japan's Medical Demands Using the Health Insurance Claim Form Data and Reverification of Effects of Medical System Reform," Japan Medical Association Research Institute Working

Paper (2004), No. 97.

Otake, Fumio and Yamaga, Hisaki, "An Old Age Pension System for Those Still Working and Elderly Male Labor Supply," Social Security in the Era of Choice (University of Tokyo Press, 2003), edited by Seike, Atsushi and Tajika, Eiji.

Hamada, Koichi, "Economic Significance of Special Zones," Japan Law and Economics Association Newsletter (2004), Vol. 1, No. 1.

Yashiro, Naohiro, "Structural Reform: A Legal and Economic Approach" (Yuhikaku Publishing Co., 2003).

Angrist, J., "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administration Records," American Economic Review (1990), 80(3), pp. 313-335.

Angrist, Joshua and Krueger, Alan, B., "Does Compulsory School Attendance Affect Schooling and Earnings?," Quarterly Journal of Economics (1991), 106, pp. 979-1014.

Card, David and Sullivan, Daniel, "Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment," Econometrica (1988), 56, pp. 497-530.

Heckman, J., "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," Journal of Human Resources (1997), 32(3), pp. 441-462.

Heckman, J., "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," The Quarterly Journal of Economics (2000), 115(1), pp. 45-97.

Heckman, J., "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," Journal of Political Economy (2001a), 109(4), pp. 673-748.

Heckman, J., "Accounting for Heterogeneity, Diversity and General Equilibrium in Social Policy Evaluation," Economic Journal (2001b), 111, F654-699.

Heckman, J., Lalonde, R., and Smith, J., "The Economics and Econometrics of Active Labor Market Programs", Handbook of Labor Economics (Amsterdam: Elsevier Science, 1999), Volume 3, edited by Ashenfelter, A. and D. Card.

Heckman, J., Lochner, L. and Taber, C., "General-Equilibrium Treatment Effects: A Study of Tuition Policy," American Economic Review (1998), 88(2), pp. 381-386.

Heckman, J. and Smith, J., "Assessing the Case for Randomized Social Experiments," Journal of Economic Perspectives (1995), 9(2), pp. 85-110.

Heckman, J.J., H. Ichimura and P. Todd, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," Review of Economic Studies (1997) 64, pp. 605-654.

Heckman, J.J., H. Ichimura and P. Todd, "Matching as an Econometric Evaluation Estimator," Review of Economic Studies (1998) 65, pp. 261-294.

Lucas, R., "Econometric Policy Evaluation: A Critique," Carnegie-Rochester Conference Series on Public Policy (1976), 1, pp. 19-46.

Staiger, D. and J.H. Stock, "Instrumental Variables Regression with Weak Instruments," Econometrica (1997), pp. 557-586.

Stock, J.H. and M. Watson, "Introduction to Econometrics" (Addison-Wesley, 2003).

Rosenzweig, Mark and Wolpin, Kenneth, I., "'Natural Experiments' in Economics," Journal of Economic Literature (2000), 38, pp. 827-874.

Wooldridge, J., "Econometric Analysis of Cross Section and Panel Data" (Cambridge, MA: MIT Press, 2001).