Are the Income Gains from a Development Project Consumed or Saved?

Shaohua Chen and Martin Ravallion[1]

Development Research Group, World Bank
March 15, 2003

Abstract: We use a matched difference-in-difference design to estimate the time profile of the impacts on household consumption and income of an aid-financed poor-area development project in rural China. Sampled households in project and matched comparison villages were followed up annually over the project's disbursement period. We find that beneficiaries saved about half the total income gain attributed to the project — well above their average saving rate. This is consistent with the high year-to-year variability we find in the project's impact, which would have made it hard for participants to infer the gain in permanent income.

Keywords: Household consumption, savings, development projects, poverty, evaluation

JEL: D91, H43, I32, O22

[1] The work reported in this paper would not have been possible without the excellent survey data collection done by the Rural Household Surveys Team of China's National Bureau of Statistics (NBS). The support of the World Bank's Research Committee (under RPO 681-39) is gratefully acknowledged. Helpful comments were received from Quy-Toan Do, Emanuela Galasso, Alan Piazza, Gene Tidrick and seminar participants at NBS, Beijing, and MIT. These are the views of the authors, and need not reflect those of the World Bank.

1. Introduction

The development project studied in this paper aimed to greatly reduce absolute poverty in the targeted villages of southwest rural China. However, when judged by one widely used measure of poverty, the project appears to have had a disappointing impact. Initially, 58% of people in sampled project villages lived in households with consumption expenditure per person less than $1 per day at 1993 Purchasing Power Parity. The poverty rate subsequently fell by seven percentage points over a five-year period. However, over the same period, the poverty rate fell by two points in sampled non-project villages. This suggests that the project was only responsible for a five percentage-point decline in poverty.

Before concluding that the project failed to greatly reduce poverty, one needs to consider how that objective should be measured and how impact should be assessed. One issue is where the poverty line should be set; the Government's own poverty line is closer to $0.70 per day. Possibly there were higher impacts at lower poverty lines. It is not difficult to check this.

A second issue is whether the sampled non-project villages are indicative of what would have happened in the project villages in the absence of the project. Possibly the above calculation has underestimated the impact of the project because the targeted villages had intrinsically lower growth prospects.[2] With poor infrastructure, for example, the counter-factual for the project areas may entail lower subsequent income growth rates than found in better-endowed areas. It is well recognized that impact estimates need to assure that treatment and comparison units are initially similar.

A third issue has received less attention in the past in the context of project evaluation and will be the main focus of this paper. That issue is whether one should measures poverty

[2] See Ravallion (1998) and Jalan and Ravallion (1998) (using data for the same region of China).

impact in terms of consumption or income. This begs the question of how participants responded to the project. It is often assumed that poor people tend to consume the current income gains from a successful development project. However, this can be questioned. Poor people are unlikely to be especially myopic; indeed, there is now a large body of evidence in development economics consistent with the view that poor people think about the longer-term implications of their current consumption and savings choices given the uncertainties they face.[3]

Indeed, strong conditions are required for the welfare gains from a development project to be fully evident in current living standards. If the income gains are known to be permanent, and markets work well, then the consumption gains would be revealed within the project cycle. This would happen even when the project was short lived, as long as it created assets that yield long-term income gains. However, if the income gains are seen to be transient then they will be saved, rather than currently consumed. High savings from the current income gains might also arise from uncertainty about future income gains, or from positive program effects on the returns to saving, given credit market failures.

In addition to what it can tell us about the inter-temporal behavior of participants, the extent of their saving from a project's current income gains is of relevance to impact assessments. Evaluation designs rarely extend much beyond the life of the project, particularly for development projects lasting many years. Lack of impact on current consumption could either reflect a project failure (arising from faults in design or implementation) or delayed impact arising from inter-temporal behavior. In appraising a development project it is clearly important to know which explanation is closer to the truth. It is also common for ex-post evaluations to entail only one follow-up survey, at the end of the project cycle. Yet aggregate impacts may well

[3] For reviews of the theory and evidence see Deaton (1992) and Besley (1995).

vary over time. If this variability is not adequately captured in the evaluation design then one could end up with a distorted picture of the project's impact.

The development project we study is the Southwest Poverty Reduction Program (SWPRP). This began in 1995 with financial and technical support from the World Bank. The project aimed to reduce poverty by augmenting the private and (local) public capital stock of farm-households in poor areas. The program's disbursements spanned a period of seven years, ending in 2001. The finite duration of SWPRP was apparently well known to participants; the program funded well-defined sub-projects of fixed duration. However, SWPRP's income gains were clearly uncertain to participants at the time.

It cannot be presumed that simply targeting external resources to poor areas will reduce poverty in those areas, in the short term or longer-term. The external resources might displace existing domestic funding sources, with no net gain in the short-term or long-term. The central and provincial governments in China have their own poor-area programs, which have been a key instrument of anti-poverty policies in China since the mid-1980s (Leading Group, 1988; World Bank, 1992, 1997; Park et al., 2002). The World Bank's project was only available to counties that were in the set of centrally-designated "national poor" counties that were already receiving help from the government's own program. The extra funding from the Bank may have led the provincial or central governments to decrease their own support to the targeted poor areas. Or there may have been a commensurate net gain in resources, but this displaced private investment, with little longer-term gain. Or the short-term income gains may simply be unsustainable much beyond the project cycle without an injection of further funding.

We employ impact evaluation methods to assess the counter factual of what income and consumption gains could have been expected in the absence of the project. We use survey data

collected for the purpose of evaluating SWPRP to compare the changes in mean income and consumption for project villages with those found in a set of comparable non-project villages in the declared national-poor countries — giving the widely-used "difference-in-differences" estimator of impact. All such estimates assume that the non-program units are representative of what would have happened in the program units without SWPRP. We use propensity-score matching methods to assure their similarity in terms of observed characteristics at the baseline. Latent heterogeneity will still leave a bias to the extent that it interacts with exposure to the project. (Additive time-invariant heterogeneity will not of course bias a DD estimator.)

The following section describes the setting and program. In section 3 we turn to our data, while section 4 outlines our method for identifying impacts on income and consumption. Section 5 theoretical arguments as to why the income gains from a project might not be evident in current living standards. Section 6 then presents our empirical results and discusses their implications. Section 7 concludes.

2. The program

It is widely acknowledged that many inland rural areas have been lagging in China's overall economic success over the last two decades. Wide geographic disparities have emerged, notably between the coast and remote resource-deficient inland areas (Jian et al., 1996; Khan and Riskin, 1998; World Bank, 1992,1997). Partly in response to this problem, anti-poverty policies in China have emphasized poor-area development (World Bank, 1992, 1997). Local infrastructure is improved and credit is provided for private (farm and non-farm) investments.

Within southwest China, there is evidence that these programs have been reaching poor rural areas; by a wide range of criteria and using data for 1985-90, Jalan and Ravallion (1998)

show that the areas chosen tend to be poorer than those not picked.[4] At the same time, there are also signs from the same study of unconditional (absolute and relative) divergence over time between the areas covered by the program and those not. In the five years after these programs began (1985-90), average consumption growth rates in the areas covered in southern China were actually lower than growth rates in the areas not covered (Jalan and Ravallion, 1998).

However, substantial underestimation of the impact of such poor-area programs can be expected if one simply compares growth rates in areas targeted by the program and those not, given that whether or not an area is targeted depends on observable differences in local characteristics that are also likely to influence the growth prospects. On controlling for geographic heterogeneity in a micro consumption growth model, Jalan and Ravallion (1998) find that households living in areas targeted by the program had higher consumption growth than one would have expected. The gains from the program were enough to prevent absolute decline. But they were not enough to reverse the underlying divergent tendencies in the rural economy. Significant impacts on average incomes from the program are also found by Park et al (2002), using income growth regressions on county data over all of China. However, Park et al. find a diminished growth impact from the program in the 1990s (relative to the 1980s).

A substantial increase in external aid for poor-area development in China began with the World Bank's Southwest Poverty Reduction Program, which began in 1995. This had the explicit aim of reducing poverty by providing resources to poor farm-households and improving local infrastructure. The program was targeted to poor areas within 35 designated "national poor" counties in southwest China (Guangxi, Guizhou and Yunan). The SWPRP involved an investment of about $US400 million over 1995-2001 from both a World Bank loan and

[4] Though this is not to say that targeting was perfect. Using a county-level panel data set for all of China for the period 1981–1995, Park et al., (2002) find signs that political factors have affected targeting and that leakage to non-poor counties has increased over time while coverage has improved.

counterpart funding from the Government of China. As in other development projects financed by the Bank, there were numerous appraisal and supervision missions by Bank staff and consultants, and these missions often probed quite deeply into the project's local operations, including numerous visits to participating poor counties and villages. Both authors participated in some of these missions and worked with staff of the National Bureau of Statistics (NBS) on the design of the survey data collection done for the purpose of evaluating SWPRP (described in the next section).

The program comprised a range of income-generating activities including methods for raising grain yields, animal husbandry, and reforestation. There was also a component for off-farm employment, including voluntary rural labor mobility and support for township-village enterprises. SWPRP also included local social services and rural infrastructure initiatives, including tuition assistance to poor farmer's children, upgrading village school and health clinics, the construction of rural roads and piped water supply systems.[5] Table 1 gives the breakdown of total project investment by category. In common with other development projects, SWPRP provided the capital and technical assistance, but it did not provide insurance, and many of the project activities are likely to entail non-negligible income risk. The income gains will depend on a number of contingencies, including the vagaries of the weather (given the evident importance of agriculture in the breakdown in Table 1), uncertain demand for the new products and risk to earnings from migration.

The selection of sub-projects aimed to take account of local conditions and the expressed preferences of participants and local stakeholders. How much participation by the poor there was in practice is a moot point. We discussed this with participants, and with the sociologist

[5] There was also a component for institution building and poverty monitoring, which financed the data collection used in the present analysis.

responsible for assessing the extent of beneficiary participation during supervision missions; it was clear that the record was mixed, varying from village to village, and county to county.

Whether in fact the resources transferred to participants actually financed the identified project is also unclear. To some degree all external aid is fungible. Yes, it could be verified in supervision that the proposed sub-project was actually completed. But one cannot rule out the possibility that it would have been done otherwise. Participants and local leaders would naturally have put forward the best development option they saw, even if it was something they planned to do anyway with the resources already available; then there is some other (infra-marginal) expenditure that was really being financed by the aid.[6] Similarly, there is no way of ruling out the possibility that non-project villages benefited by a re-assignment of public spending by local authorities, thus lowering the differential impact of program participation.

3. Data for the evaluation

An initial baseline survey in 1995 was followed by five annual surveys over 1996-2000. All surveys were done by the Rural Household Survey (RHS) team of NBS. The sample size for the annual surveys was 2000 households spanning 20 project counties and 200 villages. (Notice that our sampled non-project villages also come from project counties; we return to this feature of the design below.) It was originally intended to have 100 villages in each of the project and non-project townships within the project counties. However, the assignment of project villages had not been finalized at the time the samples of villages were drawn, and it turned out that 13 of the originally sampled non-project villages did in fact get the project. So we end up with 113 project villages and 87 non-project villages in the same counties. 10 randomly sampled

[6] In a forthcoming paper we test whether there is any difference in impact across different types of project spending.

households were interviewed in each village (project and non-project). The sampling methods followed standard practices for the RHS (as described in Chen and Ravallion, 1996).

There is a serious comparability problem between the 1995 survey and the subsequent surveys. Because of delays in NBS obtaining the locations of project villages, the first survey in December 1995 had little choice but to use a one-time interview method, asking for recall over the full year. The use of this long recall period is likely to lead to underestimation of income and consumption. The subsequent surveys use the daily diary method and collect much more accurate income and consumption data. As a consequence, the rates of income and consumption growth are clearly over estimated using 1995 as the baseline.

Because of these problems in the 1995 survey, we decided to use the 1996 survey as the baseline instead of the 1995 survey. This means that our baseline is not free of contamination by the project; 16% of the program's total disbursement on projects at household level had been made by the middle of 1996, and 23% had been made by the end of 1996. (In fact about half of this was in 1995, so this is also contaminated as a baseline survey, aside from the comparability problem.) So we are likely to be underestimating the program's impact. We consider the implications of this possibility for our comparisons of consumption and income gains.

The surveys were closely modeled on NBS's Rural Household Survey, which is described in detail in Chen and Ravallion (1996). This is a good quality budget and income survey, notable in the care that goes into reducing both sampling and non-sampling errors. Sampled households maintain a daily record on all transactions plus log books on production. Local interviewing assistants (resident in the sampled village, or nearby) visit each household at roughly two weekly intervals. Inconsistencies found at the local (county-level) NBS office are checked with the respondents. The sample frame is all registered agricultural households.

The consumption expenditure aggregate we use is what is referred to as "living expenditures" in the RHS. This comprises cash spending on all goods and services and imputed values of in-kind spending. It excludes transfer payments (cash or imputed values of transfers to relatives living in urban areas, interest and insurance payments, fines, transaction costs in acquiring assets or changing land-usage), though these only account for a small share of total spending (3.7% over the whole sample in 1996). The income aggregate includes cash income from all sources and imputed values for in-kind income (household production which includes farming, forestry, animal husbandry, handicrafts etc.).

4. Identification strategy

The standard difference-in-difference (DD) method compares changes in measured outcomes between the sampled treatment group and a group of non-participants. In this context, we would point to two potentially important sources of bias in this method. Firstly, we define "non-participant" as a village that did not get the program but is in a county that did get the program. This raises the possibility of interference between the two groups of villages. From our field work and discussions with NBS and project staff, we came to the conclusion that the physical distances involved would not mean that geographic proximity is an important source of contamination. However, sharing a common local government could be a more serious problem. Since all project counties are automatically amongst China's nationally-designated "poor counties" they are covered by the Government's national poor-area program. This design feature is clearly needed to assure that the comparison of income and consumption gains between project and (matched) non-project villages can reveal the impact of the Bank's program. However, this is not as clean an identification strategy as it might seem at first glance. The fact that the project and non-project villages come from the same counties covered under other programs could also

generate a downward bias in our estimated impacts. This will happen if SWPRP displaced other programs in the project villages, to the benefit of the non-project villages in national poor-counties. We have no basis for assessing the extent of this possible bias.

There is a second source of bias that we can go some way toward addressing. As already noted, DD will give a biased impact estimate if the subsequent outcome changes are a function of initial conditions that also influence the assignment of the sample between the two groups. This is known to be a serious concern in this context, based on past research on poor area programs in the same region of rural China (Jalan and Ravallion, 1998). Additionally, we have the possibility that the 13 villages that had to switch from the original sample of non-project villages to the final sample of project villages were somehow purposively selected.

To deal with the observable sources of heterogeneity between our samples of project and non-project villages we use a flexible, largely non-parametric, method of controlling for initial heterogeneity, based on the propensity-score matching (PSM) method introduced by Rosenbaum and Rubin (1983). Single-difference PSM gives unbiased impact estimates as long as there is no selection bias due to latent heterogeneity. By taking the double difference after matching in the baseline survey we can eliminate any time-invariant additive selection bias. It has been argued that combining PSM with DD can greatly reduce (but not eliminate) the bias found in other non-experimental evaluations (Heckman, Ichimura and Todd, 1997; Heckman et al., 1998).

To outline the method in more formal terms, let D_i be a dummy variable taking the value unity for any participating village and zero for nonparticipants. Let $P(X_i) = \Pr(D_i = 1 | X_i)$ denote the propensity score, giving the probability of SWPRP participation for observational unit *i* conditional on a vector X_i of pre-exposure control variables. Rosenbaum and Rubin (1983) prove that if the D_i's are independent over all *i*, and outcomes are independent of participation

given X_i (i.e. unobserved differences do not influence whether or not *i* participates) then outcomes are also independent of participation given $P(X_i)$, just as they would be if participation was assigned randomly. PSM uses $P(X_i)$ to select comparison subjects for each of those treated. In effect, what the Rosenbaum-Rubin result establishes is that if no selection bias remains when controlling for X_i then no bias will remain when controlling solely for $P(X_i)$. We follow common practice in the matching literature of using a parametric binary response model to estimate the propensity score for each observation in the participant and the comparison-group samples. The comparisons are then constrained to assure that project and non-project villages that share sufficiently similar values of their observed characteristics as reflected in their propensity scores.

The possibility that some treatment villages may have to be dropped for lack of sufficiently similar comparators points to the possibility of a trade off between two possible sources of bias in the resulting estimates of the mean impact in project villages. On the one hand, there is the aforementioned need to assure comparability in terms of initial characteristics, to reduce bias in the difference-in-difference. This speaks to the importance of common support. On the other hand, it creates a new possibility of sampling bias in inferences about impact on the population of treated villages, to the extent that we lose treatment villages in achieving common support; this is a well known problem in the evaluation literature.[7] Recognizing this trade-off, we also present our estimates only eliminating non-participating villages that are outside the propensity-score range found for treatment villages, while retaining the original sample of treatment villages. For comparison purposes, we also present estimates without matching.

[7] See the discussion of non-overlapping support bias in Heckman et al. (1997, 1998).

For the reasons discussed in the introduction, we are interested in assessing the impacts on both income (Y_{it} for household *i* at date *t*) and mean consumption (C_{it}), so as to infer savings. For the purposes of the following exposition we focus on the mean, though one can simply reinterpret the following formulae for some other summary statistic of the distribution, such as the proportion of people below the poverty line.

We can write the outcome measures for income and consumption of the *i*'th treatment household ($D_i = 1$) at date *t* as:

$$(Y_{it} \mid D_i = 1) = Y_{it}^* + G_{it}^Y + e_{it}^Y \quad (i = 1,..,n; t = 0,..,T) \tag{1.1}$$

$$(C_{it} \mid D_i = 1) = C_{it}^* + G_{it}^C + e_{it}^C \quad (i = 1,..,n; t = 0,..,T) \tag{1.2}$$

where Y_{it}^* and C_{it}^* are the counter-factual income and consumption for household *i* in the SWPRP village if the program had not existed, G_{it}^Y and G_{it}^C are the corresponding gains attributable to the project and e_{it}^Y and e_{it}^C are zero-mean innovation error terms uncorrelated with program participation; these allow for measurement error in Y_{it} and C_{it}.

Indicators of the counter-factual are available from a comparison group and are given by $\hat{Y}_{it}^*$ and $\hat{C}_{it}^*$. These are noisy indicators due to miss-matching (selection bias) arising from latent heterogeneity. We make the standard assumption that the selection bias is separable and time invariant, and so it is swept away by taking differences over time. On taking the expectation over all participants, the mean differences-in-differences for income and consumption are:

$$E[(Y_{it} - \hat{Y}_{it}^*) - (Y_{i0} - \hat{Y}_{i0}^*) \mid D_i = 1] = E(G_{it}^Y - G_{i0}^Y \mid D_i = 1) \tag{2.1}$$

$$E[(C_{it} - \hat{C}_{it}^*) - (C_{i0} - \hat{C}_{i0}^*) \mid D_i = 1] = E(G_{it}^C - G_{i0}^C \mid D_i = 1) \tag{2.2}$$

(Noting that, by assumption, the differenced error terms $e_{it}^{Y} - e_{i0}^{Y}$ and $e_{it}^{C} - e_{i0}^{C}$ have zero expected value amongst participants. Equations 3.1 and 3.2 also implicitly entail averaging over the distributions of the control variables used in matching.) When period 0 is a genuine baseline prior to the intervention (and not in any way contaminated by the program assignment) we have $G_{i0}^{Y} = G_{i0}^{C} = 0$. Then the DD estimates the mean current gains in consumption and income for program participants (often referred to as the "treatment effect on the treated" in the evaluation literature.) We will consider the implications for our results of the possibility that $G_{i0}^{Y} = G_{i0}^{C} \neq 0$.

5. Saving out of the income gains from a development project

By separately estimating the income and consumption gains, the above formulation of the evaluation problem allows for saving out of the current income gains. Before turning to the empirical results it is of interest to ask: why might we find that the income gains are saved?

As a benchmark model, consider Friedman's (1957) Permanent Income Hypothesis (PIH). This assumes that consumption is directly proportional to permanent income, which is the annuity value of life-time wealth. In our case, permanent income has a counter-factual component (in the absence of the program) and a component due to the program (which is zero in the absence of the program). The contribution of the program to permanent income is denoted G_{it}^{YP} and it is assumed that the full impact on income can be written as:

$$G_{it}^{Y} = G_{it}^{YP} + G_{it}^{YT} \qquad (3)$$

where G_{it}^{YT} is a transient component. By construction, the counter-factual is independent of participation in the program, and we assume that this is also true of any measurement error or transient component to consumption.

We focus initially on the special case in which there is no saving from permanent income (though we relax this below). Thus we have the following model for consumption with and without the program:

$$(C_{it}|D_i = 1) = Y_{it}^{*P} + G_{it}^{YP} + \nu_{it} \tag{4.1}$$

$$(C_{it}|D_i = 0) = Y_{it}^{*P} + \nu_{it} \tag{4.2}$$

in which we also allow for a zero-mean innovation error term, ν_{it}. Comparing (4.1) with (1.2) it is plain that $G_{it}^{YP} = G_{it}^{C} + C_{it}^{*} + \varepsilon_{it}^{C} - Y_{it}^{*P} - \nu_{it} = G_{it}^{C}$ since $(C_{it}|D_i = 0) = Y_{it}^{*P} + \nu_{it} = C_{it}^{*} + \varepsilon_{it}^{C}$. Thus $G_{it}^{YP} = G_{it}^{C}$, i.e., *the consumption gain from the program identifies the permanent income gain.* It follows that positive saving from the income gain ($G_{it}^{Y} > G_{it}^{C}$) reveals that some of that gain is thought to be transient by program participants.

This benchmark model makes a number of strong assumptions, most notably that permanent income is entirely consumed, there are no constraints on borrowing and there are no transaction costs or sources of lumpiness in consumption.[8] As the following discussion will illustrate, more general models suggest other reasons why the current income gains from a development project might be saved.

One reason is uncertainty about how much of the gain is in fact permanent and how much is transient. Participants may then save as a hedge against this income uncertainty. This will be the case if the marginal utility of consumption is a convex function of consumption. Then, by Jensen's inequality, a mean-preserving increase in uncertainty about future incomes will increase the marginal utility of future consumption; current savings will then rise to preserve equilibrium

[8] As originally formulated, the PIH also assumes that labor supply is exogenous and that preferences are homothetic. For further discussion in the context of more realistic models of consumption see Deaton (1992) and Besley (1995).

(Gersovitz, 1988). There is evidence of such precautionary saving in the same setting as the SWPRP (Jalan and Ravallion, 2001).

Introducing borrowing constraints into the PIH can also generate savings from even permanent income gains due to a development project. The PIH assumes perfect credit and risk markets, which does not appear to be realistic.[9] Assume instead that households can save but not borrow. The anticipation of future borrowing constraints when negative income shocks are experienced may well lead program participants to save from an increase in permanent income, as a contribution to their buffer stock. Lumpiness in the consumption choice set, in the presence of borrowing constraints, could also distort the empirical relationship between the permanent income gains from a development project and current consumption. Small income gains will be saved to overcome the constraint.

Yet a further reason for high savings from the project's income gains posits that the program's investments raised the marginal product of private capital — that the program's inputs are cooperant in production with private capital — and that private capital is geographically immobile, so that the marginal product of capital is equalized with a local rate of interest, that varies geographically. (This is the type of model outlined in more formal terms in Jalan and Ravallion, 2002, who find supporting evidence for this region of rural China.) Under these conditions, the program can induce higher saving through its effect on the marginal product of capital in the participating localities.

All these modifications to the PIH will tend to create lags between the program's income gains and the impacts on consumption. Higher living standards might not then be evident until

[9] See Jalan and Ravallion (1998) who provide evidence for this same region of rural China that rural households are not well insured against income shocks, and that this insurance failure is more severe for the asset-poor.

after SWPRP's completion. By tracking annual income and consumption gains over time we can look for signs of lagged impacts on consumption.

Political economy suggests yet a further explanation for low impacts on living standards despite significant income gains. Possibly the direct income gains were not saved at all (at least within the project villages) but were somehow expropriated by higher-level (county or provincial) authorities and diverted to other uses, possibly benefiting non-project villages elsewhere. Recall that our consumption aggregates exclude transfer payments. We will check if transfer payments responded positively to the project, consistent with some form of expropriation. The dynamics of income and consumption impacts will also offer clues as to the plausibility of this political economy explanation. If the local income gains were being siphoned off by a higher level of government then one would expect to see little sign of lagged consumption gains after an income gain due to the project. An expropriation model would also lead one to expect declining income gains, through disincentive effects of the taxation. We will look for these features in the income profile over time of consumption and income gains attributed to SWPRP.

6. Results

Table 2 gives the sample mean income and consumption by year for both the project villages and the non-project villages. Project villages started off slightly poorer on average than non-project villages in the same county, in terms of both income and consumption. By the end of the period, the project villages had caught up in mean income, but not consumption. This is suggestive of saving from the project's income gains. But before drawing that conclusion we need to consider the possibility of selection bias arising from the initial differences between project and non-project villages arising from purposive targeting of the program.

6.1 Matching methods and impact on poverty

The first step is to estimate the propensity scores. The sampled project and non-project villages are pooled and we run a probit regression for the village assignment to these two groups. We include as explanatory variables virtually all the village level variables for 1995 that could be constructed from the data set. Table 3 gives the results.

We find a number of significant covariates of program participation. The SWPRP villages tend to be in more mountainous remote areas, are less likely to have electricity, less likely to have a school in the village or nearby, though more likely to have a health clinic within the village relative to nearby. The project villages also tend to have higher populations, with lower mean income and more land per capita, reflecting lower population density. It is evident from Table 3 that the project villages tend to be poorer than other villages within the project counties.

The next step is to match the project and non-project villages. Figure 1 gives the frequency distribution of the propensity scores for project and non-project villages. It can be seen that there are regions of non-overlapping support. We consider two methods of matching. In the first, all matches must be within the outer bounds of the region of common support for the propensity scores; we refer to this as "outer-support matching". In the second method, comparisons are only permitted if the absolute difference in propensity scores is within pre-determined caliper bounds; we call this "caliper-bound matching." Project and non-project villages outside the caliper bounds are discarded. This method clearly gives the closest matching of treatment and control villages, but it can do so at a cost to sample size and representativeness. We set the tolerance levels for the caliper at 0.01. The choice of this tolerance is somewhat arbitrary. However, we found that too many villages were lost when the tolerance went much

below 0.01. If one was relying on single difference matching then one might want closer matches than even our 0.01 absolute difference in scores. However, here we can exploit the fact that we have multiple observations to "difference-out" any (time-invariant) errors due to miss-matching. With 0.01 tolerance level, we end up with only 63 of the original sample of project villages to be matched with 34 non-project villages.

The final step is to calculate the DD estimates. Given that the project's main aim was poverty reduction, we begin by calculating the impact on poverty incidence in the final year of the study period. We use probably the most common measure of absolute poverty in developing countries, namely the proportion of the population living in households with consumption per person below the international poverty line of $1/day at 1993 Purchasing Power Parity (Chen and Ravallion, 2001); this is equivalent to 808 Yuan per year per person at 1995 prices.

Table 4 gives the results. We find reductions in the incidence of poverty due to the program, though the magnitude varies by matching method and poverty line. The biggest difference is not between the unmatched DD and matched DD, but rather between the two methods of matching. The unmatched DD and matched DD using the outer-support criterion indicate that the poverty rate by the end of the study period had fallen by 5-6 percentage points due to the project. However, using caliper-bound matching, we find no impact on poverty.

To test robustness to the choice of poverty line, Figure 2 gives our estimate of impact over the whole distribution. The figure gives the difference between the empirical cumulative distribution function of consumption for the treatment villages and the counter-factual comparison group. (The results are similar for unmatched DD as for outer-support matching, so we only give results for matched DD to make the figure easier to read.) For caliper-bound matching we find that the negligible poverty impact for the $1/day line is not robust to the choice

of poverty line, with more sizable impacts emerging amongst the poorest and least poor in the project villages. (The impacts become statistically significant ate about 6 percentage points.)

6.2 Behavioral responses through savings

Table 5 gives the unmatched DD estimates, while Table 6 gives the matched DD estimates using both matching methods described above. We give the annual impacts, the two-year moving average of the annual impact and the cumulative impacts.

Let us focus first on the results for the final year of the study period, 2000. While we find sizeable income gains over time in the project villages, this is not evident for the counter-factual comparison group. It should be recalled that 1996 was a particularly good year for rural incomes given that the government had substantially increased the overall level of its procurement prices for foodgrains at this time; the change was short-lived however. So the small counter-factual gain that we find is not too surprising. (This nicely illustrates the importance of differencing out the changes in the comparison group; in the absence of the project one would have expected a similar income decline in the treatment villages.)

Taking account of both the changes over time and the differences between the treatment and comparison villages, the estimated double difference for 2000 indicates an income gain attributable to the project of around 17-21% of initial mean income (depending on the matching method).[10] However, we find little or no impact on consumption; indeed, we cannot reasonably reject the null hypothesis that the consumption impact over the whole period is zero. The vast bulk of the income gain in 2000 was saved.

Recall that we are measuring consumption by what is termed "living expenditures" in the RHS. So our definition of "savings" implicitly includes transfer payments as well as flows into

[10] Note that the baseline means differ for caliper-bound matching, given the change in the number of project villages used for the analysis. The 1996 mean income for the 63 project villages used for the caliper-bound matching is 968.75 Yuan.

the stock of financial and physical assets. One can question whether some of these transfer payments should be included as savings. However, transfer payments do not account for the high savings out of the project's income gains. Indeed, mean transfer payment actually fell slightly in the project villages over 1996-2000, and we found that the DD estimate was negative though not significantly so.

As noted in section 2, there are likely to have been impacts in 1996. On the assumption that these gains would have initially impacted on incomes rather than consumptions, we will have underestimated the true income impact and underestimated the extent of saving from the current income gains. As we will see below, the inter-temporal pattern of income and consumption impacts within the evaluation period offers support for this conclusion.

To see the impact of this high savings rate on the poverty measures, we re-calculated the DD estimates using incomes. For the unmatched DD and the matched DD using outer-support criterion, the impacts on income poverty were 11.5% points (t = -4.03) and 11.3% (t = -3.65) respectively (instead of 5.0 and 6.3% for consumption poverty). The impact is greater using caliper-bound matching; instead of the very small 0.6 percentage point impact on consumption poverty using the $/day line by the caliper-bound matching, we find that the income poverty rate fell by 15.7% points (t = -4.41). Figure 3 gives the impacts on income poverty over the whole distribution. Comparing Figures 2 and 3 it is evident that the largest divergence between the income and consumption impacts tends to be in the middle of the distribution.

We have seen that the results for 2000 suggest that virtually all of the aggregate income gain was saved. Let us now turn to the results for the three intervening years, 1997-99, as also given in Tables 5 and 6. We will focus on the results for outer-support matching, noting any marked differences with the results for the other two methods.

Mean income was higher in all years due to the project and significantly so in all years except 1999. The gains were lower in the second and third years than the first and last. Despite the large income gain in the first year, there was negligible impact on consumption in that year. Appreciably higher consumption only emerged in the second year (1998). The relatively low income gain in 1999 was followed by a lower impact on consumption in 2000. By the end of the study period, 50% of the cumulative income gain attributed to the project had been saved. Caliper-bound matching gives an even higher savings rate, of 58%.

While one should be cautious with only four yearly observations, there is a pattern in Tables 5 and 6 that is suggestive of lagged consumption impacts from income gains. The high income gains attributed to the project in 2000 may then be expected to be reflected in higher future consumption, beyond the study period. Neither the signs of lagged consumption impacts nor the fact that the highest income gains were in the last year are supportive of the existence of some hidden form of expropriation of the project's income gains.

Comparing the three evaluation methods, the most notable difference is that caliper-bound matching (entailing the tightest matching in terms of initial characteristics) tends to give lower impact estimates than the other two methods. This is not consistent with the expectation discussed in section 2 that the relatively poorer villages targeted by such a program would tend to have intrinsically lower growth prospects; if anything we find the opposite, though the difference is small. However, it should be recalled that our comparison villages were chosen from the same (poor) counties as the project villages. The bias in unmatched comparisons might well only emerge when making comparisons across project and non-project counties, given that there can be large inter-county differences in initial conditions relevant to growth prospects (Jalan and Ravallion, 1998).

6.3 Implications

Our estimated income gains from the SWPRP can be interpreted as the output returns from the project's investments within the disbursement period. Let I_t denote the project's real investment in period *t* and let N_t denote the number of beneficiaries in that year. Given a period *t* rate of return from the project of r_t, the income impact can be written as:

$$G_t^Y = r_t \sum_{j=1}^{t} I_j / N_j \quad (t=1,..,T) \tag{5}$$

From the project documents we calculated the total investment by year (by the World Bank and the Government). By the end of the project this was 1120 Yuan per person per year in 1995 prices, averaged over the population of project villages. This is the cumulative investment over the project cycle per beneficiary. Table 7 gives the corresponding numbers by year.[11] The table also gives the values of r_t from equation (9) using the income gains from Tables 5 and 6.

We find average rates of return of 9-10%. This could be an underestimate, to the extent that the Bank's program displaced other programs in the project villages, to the benefit of the non-project villages. (Recalling that project and non-project villages come from national-poor areas covered under other poor-area programs, as discussed in section 3.)

The fact that the project and comparison villages were drawn from the same national-poor counties covered by the Government's pre-exiting programs also means that the rates of return in Table 7 should be interpreted as incremental returns from the Bank's program on top of the Government's programs. Jalan and Ravallion (1998) estimated an average rate of return of

[11] These were calculated from the project documents using the cumulative total project investments (deflated to 1995 prices) normalized by the cumulative number of beneficiaries. However, the project documents only give the number of households covered by the project. To obtain the per capita disbursements we used mean household size in the full sample of project villages by province and year.

12% for the Government's poor area development program in the same region of China over 1985-90. Using different methods, Park et al., (2002) also estimate a rate of return to the Government's national poor-area program of 12% in the period 1992-95.[12] So the compound rate of return from the SWPRP and the Government's own program is 22-23%.

However, it can also be seen from Table 7 that the annual returns varied substantially from year to year, though disbursement per beneficiary did not. So the considerable volatility that we find in the income gains from the project is not due to variability in the cumulative program investments but is due to fluctuations in the return on that investment. A simple way to gauge the importance of the inter-temporal variability in returns to the variation in project impact is to ask what the range (maximum minus minimum) in impact estimates would have been at the time-mean rate of return (Table 7). We find that this simulated range in impacts accounts for less than one tenth of the actual range (9% without matching and 1% and 6% for the two matching methods respectively).

The income gains from the program would appear to be more variable over time than other income sources. From Table 2, the range of annual mean incomes is about 16% of the overall mean in the project villages while the range of the project's income impacts is about 150% of the mean impact. And this difference appears to be reflected in the savings rates. The baseline data indicate that 16-17% of income was saved in the project villages (Table 2). As already noted, the baseline year was a good year for agriculture, due to unusually high foodgrain procurement prices set by the government. So presumably the average saving rate in that year was, if anything, higher than normal. Yet the average saving rate we find from the income gains during the life of the project was 50%.

[12] The Jalan and Ravallion (1998) method was described earlier in this paper. Park et al., (2002) used regional growth regressions, estimated at county level.

With such variability in the income gains from the project, one can conjecture that project participants would have had a hard time inferring the project's impact on permanent income. This is consistent with the argument that the high saving rate out of the income gains implied by our results for the evaluation period as a whole reflects transience or uncertainty in the project's income gains. Furthermore, none of the other possible explanations for high saving from the project's income gains appear to be as plausible in the light of our empirical findings. Explanations that posit that the project increased the returns to saving (to overcome borrowing constraints) would appear to have a hard time explaining the variability over time that we find in the savings rate from the project's income gains. The facts that the high aggregate savings rate is not attributable to measured transfer payments, and that income gains do not fall over time, are not supportive of the expropriation model discussed in section 5.

While these observations may not be conclusive (for example the variability could be due in part to time-varying measurement errors) they are at least suggestive that the transience and uncertainty in the project's income impacts is a plausible explanation of our findings.

The variability in returns has implications for the design of evaluations. Single follow-up designs can clearly be deceptive. Suppose for example that the design had relied on only two surveys, one in 1996 (just after the project began) and one in 1999 (just before it finished). This evaluation design would have considerably underestimated the average annual income gain from the project, and overestimated the consumption gain, given the time path of the underlying income gains. Or suppose that one only knew the income gains in the last year (as given in Tables 5 and 6) one would conclude that the rate of return was 18%. However, the estimate for the last year is hardly indicative of other years (Table 7).

7. Conclusions

We have studied the income and consumption impacts of a rural development project in China over the bulk of its disbursement cycle. On comparing income changes in project villages with those in matched non-project villages, we find that the project resulted in an average income gain over five years of around 10% of baseline mean income, representing an average return on the project's disbursements of about 9-10%, on top of the impact of the Government's pre-existing assistance to poor areas.

However, we find that half of the cumulative income gain was saved, so that the project's impact is far less evident in participants' consumptions. Indeed, on comparing the final year of the study period with the first, we find little or no impact on mean consumption or on consumption poverty using an international "$/day" poverty line, though the poverty impact depends critically on the poverty line used; there are indications of significant impacts on consumption poverty for lower poverty lines.

We also find large year-to-year differences in impact; for example, the estimated income gain in the final year was 23% of baseline income (an 18% return on the project's total disbursement) and virtually all of this was saved. The impact variability was primarily due to variability in the return to the program's investments rather than the level of that investment.

Our results clearly reject the seemingly commonly held view that poor people tend to rapidly consume the income gains from a public program. Indeed, we find a high saving rate. When interpreted in terms of the simplest Permanent Income Hypothesis, our results imply that participants felt that a large share of the income gains were likely to be transient. Uncertainty about future incomes and future borrowing possibilities can also lead to high saving out of the income gains from such a program. The considerable variability that we find in the programs'

income returns suggests that participants would have had a hard time assessing the program's impact on permanent income.

Finding that even poor participants may choose to save a large share of the current income gains from external aid has an important implication for assessments of the efficacy of anti-poverty programs, given their finite time horizons and that it is common to study poverty impacts within a relatively short period of time — often no more than the period of the disbursement cycle. A large share of the impact on peoples' living standards may occur beyond the life of the project.

This does not necessarily mean that credible evaluations will need to track welfare impacts over much longer periods than is typically the case, raising concerns about feasibility. But it does suggest that evaluations need to look carefully at impacts on partial intermediate indicators of longer-term impacts — such as incomes in our case — even when good measures of the welfare objective — consumption in our case — are available within the project cycle. The choice of such indicators will need to be informed by an understanding of participants' behavioral responses to the program.

Our results also warn against evaluation designs that only do one follow-up survey (normally at the end of the project cycle). With short-lived projects and/or good respondent recall this will not be a problem. However, for many development projects, including the one studied here, these conditions do not hold, and finer observations over time are needed to have any hope of understanding the project's impacts and participants' responses.

Table 1: Composition of spending under SWPRP

	% of total investment
Education	8.60
Health	5.37
Labor mobility	9.74
Rural infrastructure	17.24
Agriculture	43.05
Rural enterprise development	11.52
Institution building	1.69
Project and poverty monitoring	2.78
Total	100.00

Table 2: Mean household income and consumption per capita by year

		Project villages		Non-project villages	
		Mean	Std. dev.	Mean	Std. dev.
1996	Income	992.74	713.47	1155.47	603.45
	Consumption	841.13	468.63	943.66	444.38
1997	Income	1084.86	658.14	1148.86	628.80
	Consumption	874.72	441.08	954.57	512.99
1998	Income	1108.91	603.27	1189.28	680.96
	Consumption	937.01	541.27	951.11	497.81
1999	Income	1182.23	681.62	1285.25	807.03
	Consumption	1002.91	658.89	1050.27	591.22
2000	Income	1259.47	913.70	1225.22	669.92
	Consumption	943.09	579.15	1023.31	696.10

Note: Household-size weighted means at 1995 prices using Provincial Rural CPI. Sample sizes: 1130 households in project villages and 870 households in non-project villages (10 households per village in both cases).

Table 3: Probit regression of village participation in the SWPRP

	Coefficient	Z score
Village on the plains	Reference	
Hills	4.6023	2.651
Mountainous	2.6301	1.616
Whether village has electricity	-0.8272	-1.722
... telephones	-0.1088	-0.248
... road passing through it	0.4085	0.971
... radio transmitters	0.4683	0.972
Whether village can receive TV transmission	0.2141	0.531
Located <5 km from the nearest market	0.3084	0.364
...5 -10 km from the nearest market	-0.3476	-0.406
...10 –20 km from the nearest market	1.1554	1.167
…> 20 km	Reference	
# of days in a cycle during which the market assembles	-0.0888	-0.662
County town within 5 km	Reference	
Distance from village to county town is 5-10 km	1.1096	1.230
...10-20 km	-0.6387	-0.842
...>20 km	-0.4168	-0.596
Township=village	Reference	
Distance from village to township is within 5 km	0.5466	0.609
...5 –10 km	0.7836	0.877
...10-20 km	-1.0477	-1.141
Main mode of transportation used by the villager: bicycle	-0.5539	-1.026
...bus	-0.1329	-0.415
...other automobile	0.6948	1.440
…walking	Reference	
Nearest train station is within 5 km	-0.1729	-0.192
...5-10 km	1.1186	1.137
...10-20 km	0.4978	0.429
…>20 km	Reference	
Nearest bus station is within 5 km	-0.0173	-0.050
...5-10 km	0.2013	0.432
...10-20 km	0.3736	0.718
…> 20 km	Reference	
Whether village has a day-care center	0.5773	0.848
Elementary school is in village	Reference	
Nearest elementary school is within 5 km	0.0520	0.128
...5-10 km	0.5050	0.900
Middle school is in village	Reference	
Nearest middle school is within 5 km	0.8846	1.871
...5-10 km	-0.0652	-0.142
...10-20 km	1.6566	2.416
...>20 km	1.3317	1.847

Medical clinic in village	Reference	
Nearest medical clinic is within 5 km	-1.0271	-2.322
...5-10 km	-0.2405	-0.518
...10-20 km	-0.8605	-1.290
...>20 km	-0.5790	-0.581
Total population of the village	0.0004	2.097
Elevated land (mu)	-0.0016	-2.653
Forest land (mu)	0.0000	-1.160
# of people work in TVE over # of labor.	0.0845	1.135
Whether village has TVE	-0.4689	-1.027
Output of grain per capita (kg/person)	0.0019	1.732
Net income per capita	-0.0033	-3.349
(End of year) # of pigs per person	0.7031	1.274
(End of year) # of cows per person	0.3248	0.267
(End of year) # of sheep, goat per person	0.6432	1.034
(End of year) # of poultry per person	0.4133	2.608
(End of year) # of honey bee per person	-5.1474	-1.765
Workforce per capita	0.0463	1.506
Average household size	-0.0785	-0.992
Share of workforce female	-0.1132	-1.875
Cultivated land per capita (mu).	1.3591	2.685
Grassland per capita (mu)	2.5915	1.926
Guangxi	1.4329	2.198
Guizhou	1.1390	1.656
Intercept	-4.2891	-1.649
Pseudo-R^2	0.3130	

Note: The village is the unit of observation (n=200) and all explanatory variables are pre-intervention (1995).

Table 4: Impacts of SWPRP on poverty in 2000

1996 poverty incidence (*H*) in project villages (%)	(1) Change in *H* in project villages	(2) Change in *H* in comparison villages	Double difference (1)-(2)
No matching (113 project villages compared to 87 non-project villages)			
57.86	-6.66	-1.63	-5.03 (-1.75)
Outer-support matching (113 villages matched with 71 comparison villages)			
57.86	-6.66	-0.33	-6.33 (-2.07)
Caliper-bound matching (63 project villages matched with 34 comparison villages)			
59.72	-4.00	-3.39	-0.61 (-0.17)

Note: Poverty line =808 Tuan per year per person (1995) prices, equivalent to $1.08 per day at 1993 consumption PPP. 1130 sampled households in project villages; 870 in non-project villages. T-ratios for the null hypothesis that DD=0 in parentheses.

Table 5: Unmatched difference-in-difference estimates

	(1) Gain in project villages	(2) Gain in comparison villages	Difference-in-difference (1)-(2)		
			Annual	Two-year moving average	Cumulative
1997					
Income	92.12	-6.61	98.72 (3.07)	n.a.	n.a.
Consumption	33.59	10.91	22.68 (1.07)	n.a.	n.a.
Saving	58.53	-17.51	76.04 (2.34)	n.a.	n.a.
1998					
Income	116.17	33.81	82.36 (2.63)	90.54	181.08
Consumption	95.88	7.45	88.43 (3.77)	55.56	111.12
Saving	20.29	26.36	-6.07 (-0.18)	34.98	69.97
1999					
Income	189.48	129.78	59.70 (1.65)	71.03	240.79
Consumption	161.77	106.61	55.16 (1.93)	71.80	166.28
Saving	27.71	23.17	4.54 (0.13)	-0.77	74.51
2000					
Income	266.73	69.76	197.97 (5.14)	128.34	437.75
Consumption	101.96	79.65	22.31 (0.81)	38.74	188.59
Saving	164.77	-9.89	174.66 (4.49)	89.60	249.17

Note: Household-size weighted means at 1995 prices with all 113 sampled project villages compared to 87 sampled non-project villages. T-ratios for the null hypothesis that DD=0 in parentheses.

Table 6: Matched difference-in-difference estimates

	(1) Gain in project villages	(2) Gain in comparison villages	Difference-in-difference (1)-(2) Annual	Two-year m.a.	Cumulative
Outer-support matching (113 villages matched with 71 comparison villages)					
1997					
Income	92.12	-9.02	101.14 (2.90)	n.a.	n.a.
Consumption	33.59	17.16	16.44 (0.71)	n.a.	n.a.
Saving	58.53	-26.18	84.70 (2.43)	n.a.	n.a.
1998					
Income	116.17	46.29	69.88 (2.06)	85.51	171.02
Consumption	95.88	7.90	87.98 (3.50)	52.21	104.42
Saving	20.29	38.39	-18.10 (-0.51)	33.30	66.60
1999					
Income	189.48	146.95	42.53 (1.09)	56.21	213.55
Consumption	161.77	84.83	76.94 (2.55)	82.46	181.36
Saving	27.71	62.12	-34.41 (-0.92)	-26.26	32.19
2000					
Income	266.73	69.11	197.62 (4.77)	120.08	411.17
Consumption	101.96	78.47	23.49 (0.80)	50.22	204.85
Saving	164.77	-9.36	174.13 (4.17)	69.86	206.32
Caliper-bound matching (63 project villages matched with 34 comparison villages)					
1997					
Income	110.70	15.35	95.35 (2.37)	n.a.	n.a.
Consumption	47.79	30.36	17.43 (0.63)	n.a.	n.a.
Saving	62.91	-15.00	77.92 (1.92)	n.a.	n.a.
1998					
Income	113.47	31.68	81.79 (2.19)	88.57	177.14
Consumption	99.26	18.87	80.38 (2.86)	48.91	97.82
Saving	14.22	12.81	1.41 (0.03)	39.66	79.32
1999					
Income	187.81	179.49	8.32 (0.16)	45.05	185.46
Consumption	148.52	93.95	54.57 (1.61)	67.48	152.39
Saving	39.29	85.54	-46.25 (-0.88)	-22.42	33.07
2000					
Income	178.66	-22.36	201.02 (4.55)	104.67	386.48
Consumption	85.60	75.94	9.66 (0.27)	32.12	162.05
Saving	93.06	-98.30	191.36 (4.21)	72.55	224.43

Note: Household-size weighted means at 1995 prices. T-ratios for the null hypothesis that DD=0 in parentheses.

Table 7: Cumulative investment and returns by year

	Cumulative investment per project participant (Yuan/person; 1995 prices)	Year-specific rate of return (%)		
		Unmatched DD	Outer-support matched DD	Caliper-bound matched DD
1997	1087	9.1	9.3	8.8
1998	1060	7.8	6.6	7.7
1999	998	6.0	4.3	0.8
2000	1120	17.7	17.6	17.9
Average	1066	10.2	9.5	8.8

Figure 1: Histograms of the propensity scores

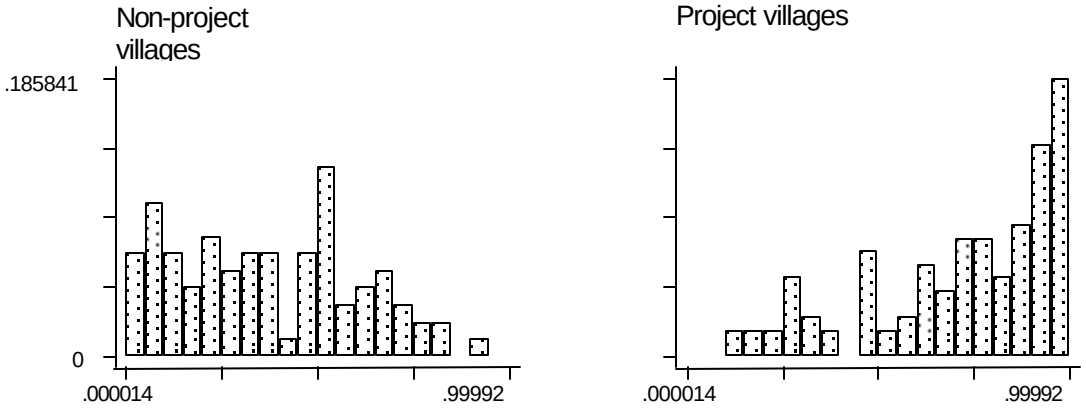


Figure 2: Impacts on consumption poverty

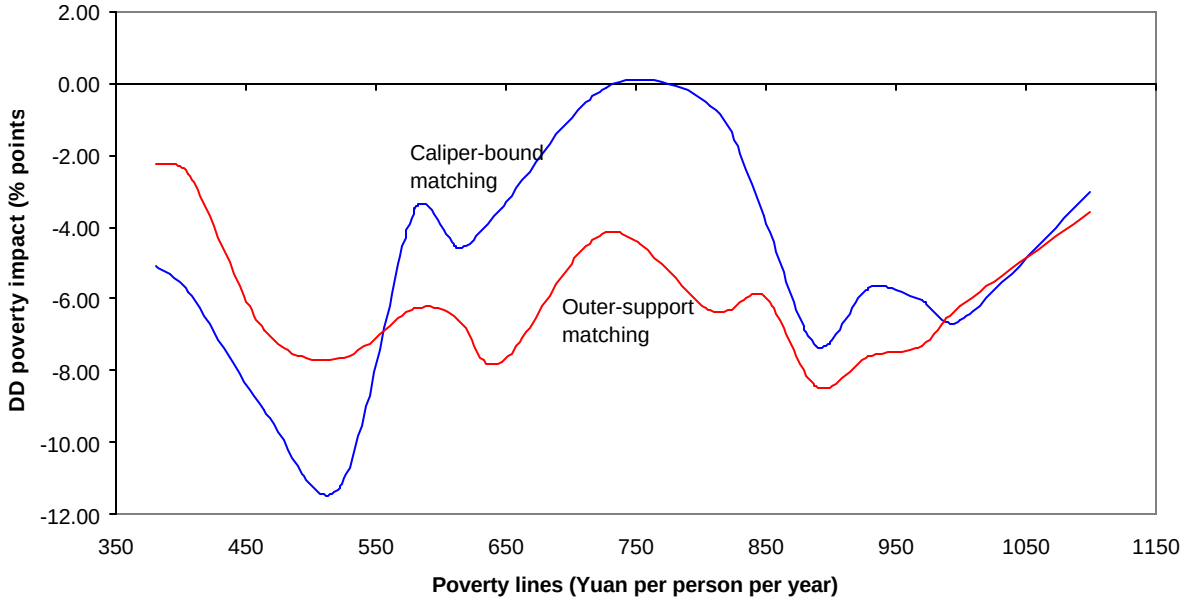
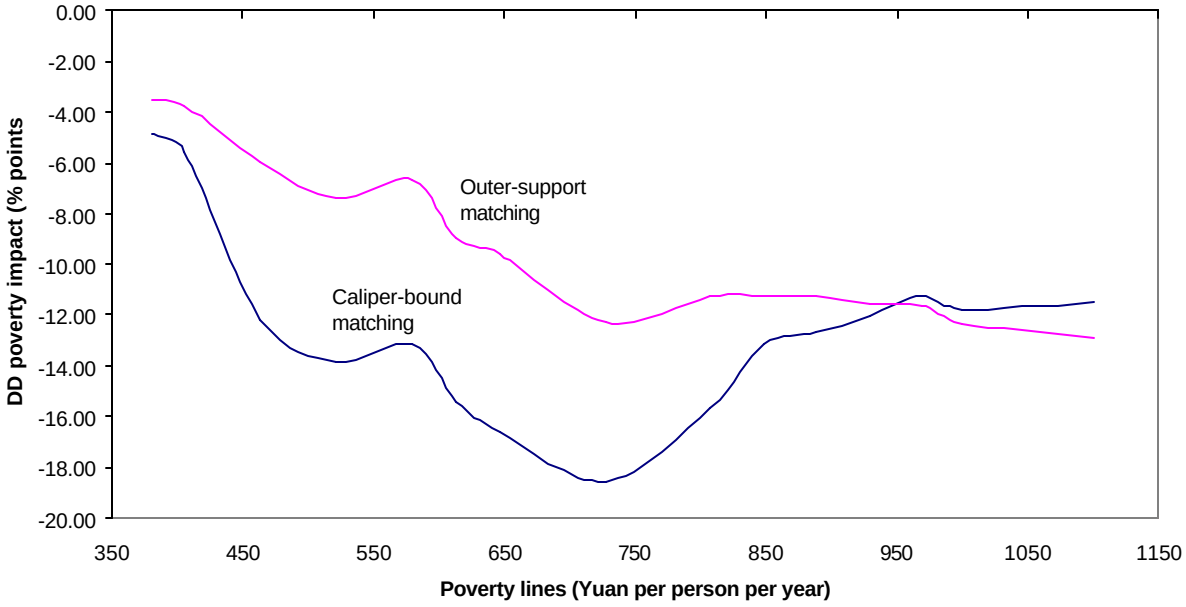


Figure 3: Impacts on income poverty



References

Besley, Timothy, 1995, "Savings, credit and insurance", in Jere Behrman and T.N. Srinivasan (eds) *Handbook of Development Economics Volume 3,* Amsterdam: North-Holland.

Chen, Jian, and Belton M. Fleisher, 1996, "Regional Income Inequality and Economic Growth in China." *Journal of Comparative Economics*, 22, 141-164.

Chen, Shaohua and Martin Ravallion, 1996, "Data in Transition: Assessing Rural Living Standards in Southern China," *China Economic Review*, 7, 23-56.

_____ and _____, 2001, "How Did the World's Poor fare in the 1990s?", *Review of Income and Wealth*, 47(3), 283-300.

Deaton, Angus, 1992, *Understanding Consumption*, Oxford: Oxford University Press.

Friedman, Milton, 1957, *A Theory of the Consumption Function*, Princeton N.J., Princeton University Press.

Gersovitz, Mark, 1988, "Savings and development," in H. Chenery and T.N. Srinivasan (eds) *Handbook of Development Economics Volume 1*, Amsterdam: North Holland.

Heckman, J., H. Ichimura, and P. Todd, 1997, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies* 64(4): 605-654.

Heckman, J., H. Ichimura, J. Smith, and P. Todd, 1998, "Characterizing Selection Bias using Experimental Data," *Econometrica*, 66: 1017-1099.

Jian, Tianlun, Jeffrey Sachs and Andrew Warner, 1996, "Trends in Regional Inequality in China," *China Economic Review*, 7(1), 1-21.

Jalan, Jyotsna and Martin Ravallion, 1998, "Are There Dynamic Gains from a Poor-Area Development Program?" *Journal of Public Economics*, 67, 65-85.

_____ and _____, 1999, "Are the Poor Less Well Insured? Evidence on Vulnerability to Income Risk in Rural China," *Journal of Development Economics*, 58(1), 61-82.

_____ and _____, 2001, "Behavioral Responses to Risk in Rural China," *Journal of Development Economics*, 66, 23-49.

_____ and _____, 2002, "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China," *Journal of Applied Econometrics*, 7(4), 329-346.

Khan, Azizur Rahman and Carl Riskin, 1998, "Income Inequality in China: Composition, Distribution and Growth of Household Income, 1988 to 1995." *The China Quarterly*, 154, 221-253.

Leading Group, 1988, *Outlines of Economic Development in China's Poor Areas*, Office of the Leading Group of Economic Development in Poor Areas Under the State Council, Agricultural Publishing House, Beijing.

National Bureau of Statistics (NBS), 2000, *The Poverty Monitoring Report of Rural China 2000*, Beijing: China Statistics Press.

Park, Albert, Sangui Wang and Guobao Wu, 2002, "Regional Poverty Targeting in China," *Journal of Public Economics*, 86(1), 123-153.

Ravallion, Martin, 1998, "Poor Areas," in David Giles and Aman Ullah, eds., *The Handbook of Applied Economic Statistics*. New York: Marcel Dekkar.

Rosenbaum, Paul R., and Donald B. Rubin, 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

World Bank, 1992, *China: Strategies for Reducing Poverty*, Washington DC.: World Bank.

_____, 1997, *China 2020: Sharing Rising Incomes*, Washington DC.: World Bank.