

Are There Lasting Impacts of Aid to Poor Areas?

Shaohua Chen, Ren Mu and Martin Ravallion¹

*Development Research Group, World Bank
1818 H Street NW, Washington DC*

Summary: The paper re-visits the site of a large, World Bank-financed, rural development program in China, 10 years after it began and four years after disbursements ended. The program emphasized community participation in multi-sectoral interventions (including farming, animal husbandry, infrastructure and social services). Data were collected on 2,000 households in project and non-project areas, spanning 10 years. A double-difference estimator of the program's impact (on top of pre-existing governmental programs) reveals sizeable short-term income gains that were mostly saved. Only small and statistically insignificant gains to mean consumption emerged in the longer-term — though in rough accord with the gain to permanent income. Certain types of households gained more than others. The educated poor were under-covered by the community-based selection process — greatly reducing overall impact. The main results are robust to corrections for various sources of selection bias, including village targeting and interference due to spillover effects generated by the response of local governments to the external aid.

Keywords: Poor-areas, aid, credit, rural development, impact evaluation, China

JEL: D91, H43, I32, O22

¹ This study would not have been possible without the survey data collection effort over 10 years by the Rural Household Survey Team of China's National Bureau of Statistics (NBS). We are particularly grateful to Wang Ping Ping at NBS, who ably supervised the surveys. The authors have benefited from discussions with Guido Imbens and the comments of Kathleen Beegle, Richard Blundell, Solveig Buhl, Shubham Chaudhuri, Richard Chiburis, Alan De Brauw, Quy-Toan Do, Gershon Feder, Emanuela Galasso, Garance Genicot, Stuti Khemani, Alice Mesnard, Alan Piazza, Dominique van de Walle and seminar participants at University College London, the Overseas Development Institute, the University of Namur, the Indian Statistical Institute and the World Bank. The support of the Bank's Research Committee and the Knowledge for Change Trust Fund is gratefully acknowledged. The paper's findings, interpretations and conclusions are those of the authors and should not be attributed to the World Bank, its Executive Directors, or the countries they represent.

1. Introduction

Publicly-supported grants and loans to poor areas have long been an important vehicle for development assistance. For example, China's anti-poverty policies have emphasized such poor-area programs since the mid-1980s,² motivated by the observation that the country's success against poverty over the last 25 years has been geographically uneven, with marked disparities in living standards emerging.³ Advocates of such programs claim that credit constraints in poor areas perpetuate their poverty and that targeted aid can relieve those constraints. By this view, capital-market failures in poor areas entail that the investments made under such a program would be infeasible otherwise, implying both efficiency and equity gains.

It remains an open question how much impact can be expected. While not perfect, capital markets may still work well enough to assure that marginal products of capital come into rough parity between poor and non-poor areas in steady state. Then the problem of lagging poor areas is not so much lack of capital as low productivity of capital, such as due to poor natural conditions or poor policies. And even with credit constraints, some people are clearly more constrained than others. Then beneficiary selection will be crucial to the outcomes. If those selected are not credit constrained, their participation is voluntary, and the interest rate is no different from other credit sources, then there will be no net gain from the extra availability of credit through poor-area programs. We know very little about how well these interventions have performed in reaching the most credit-constrained households in poor areas, who need not be the poorest.

This paper provides the first rigorous assessment of the longer-term impacts at the micro (household and village) level of a large poor-area program. The program is the Southwest China Poverty Reduction Project — the Southwest Program (SWP) for short. This comprised a package of multi-sectoral interventions targeted to poor villages using community-based participant and activity selection. The overall aim was to achieve a large and sustainable reduction in poverty. The paper reports results from an intensive

² See (*inter alia*) Leading Group (1988), World Bank (1992, 1997), Jalan and Ravallion (1998) and Park et al. (2002).

³ See, for example, Knight and Song (1993), Jian et al., (1996), Khan and Riskin (1998), Ravallion and Jalan (1996, 1999), World Bank (1992,1997), Kanbur, and Zhang (1999) and Ravallion and Chen (2006).

survey data collection effort over 10 years, initiated by two of the authors and done in close collaboration with the Rural Survey Organization of China's National Bureau of Statistics.

Assessing development aid effectiveness at the project level raises a number of challenges. A long-term commitment to collecting high-quality longitudinal survey data is crucial, but it is not sufficient. Impact can only be meaningfully assessed relative to a counterfactual; our counterfactual is the absence of the SWP, which means that we assess the incremental impacts, on top of pre-existing governmental spending. As in any observational study, there are concerns about selection bias, i.e., differences in counterfactual outcomes between SWP participants and non-participants. Our data collection effort allows us to “difference out” the time-invariant component of the selection bias (arising from non-random placement). However, we argue that it is implausible that the bias would be constant over time, given that the initial village characteristics that attract the program (such as poor infrastructure) also influence the growth rate under the counterfactual. We use both propensity-score weighted regression and kernel-matching methods to balance the observable covariates between sampled SWP and non-SWP villages.

A further problem is that development projects are likely to violate the common assumption in impact evaluations (both experimental and non-experimental) of no interference with the comparison units.⁴ A plausible source of interference in this setting is through local public-spending spillover effects to non-SWP villages. We propose and implement a new test for spillover effects and we find that the bias is modest in our case.

Given that it is rare to assess impacts by repeated observation over a long period, we also use this as an opportunity to study less costly evaluation methods, based on respondent recall using subjective-qualitative questions. These methods are found to be deceptive about impact given the revealed biases in respondent recall.

The paper's principle finding is that there were sizeable income gains from the SWP during its disbursement period, but these gains did not survive four years later. The

⁴ This assumption is often implicit in impact evaluations but it was made explicit in Rubin (1980), who dubbed it the stable unit treatment value assumption (SUTVA). SUTVA is known to be implausible in certain bio-medical evaluations; Miguel and Kremer (2004) provide an interesting example in assessing the impacts of a school de-worming program in Kenya.

longer-term impact on mean income is neither large nor statistically significant. However, we do find significant gains for some sub-groups, notably those among the poor with better schooling. Our results point to substantial losses from the community-based beneficiary selection process.

The following section describes the SWP while sections 3 and 4 describe our data and methods. Section 5 presents the main results while section 6 gives our tests for bias due to spillover effects. Section 7 draws some lessons for future evaluations.

2. Background on the program

In 1986, the Government of China designated that about 15% of the 2,200 counties nationally were “poor counties,” which would receive extra assistance, mainly in the form of credit for development projects. Past research has suggested that the designated poor counties are in fact poor (by a range of defensible criteria) and that they have seen higher growth rates than one would have otherwise expected (Jalan and Ravallion, 1998; Park et al., 2002). The gains have not been sufficient to reverse the underlying tendency for growth divergence (whereby poorer counties tend to have lower growth rates) and there is evidence that the impacts on economic growth may have declined in the 1990s (Park et al, 2002). Within these designated poor counties, geographic pockets of extreme poverty have persisted to the present day, mainly in upland areas.

The SWP was introduced in 1995 with the aim of reversing the fortunes of selected poor villages in the designated poor counties of Guangxi, Guizhou and Yunan. About one-quarter of the villages were selected for the SWP (1,800 out of 7,600 villages). The aim was to choose relatively poor villages within these counties, with selection based on objective criteria, although not formulaic. The selection was done by the county government’s project office in consultation with provincial and central authorities and the World Bank.

The total outlay on the SWP was US\$464 million, which was financed by World Bank loans and counterpart funding from China’s central and provincial governments. The total investment per capita under the SWP was only slightly lower than mean annual income per capita of the project villages.

As in other World Bank projects, 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 counties and villages. Two of the authors participated in some of these missions and also revisited a number of the sampled villages over two weeks in May 2005 (including some visited 10 years earlier) and had informal discussions about the SWP with numerous ex-participants.

Within the selected villages, virtually all households were expected to benefit from the infrastructure investments, such as improved rural roads, power lines and piped water supply. Widespread benefits were also expected from the improved social services, including upgrading village schools and health clinics, and training of teachers and village health-care workers. Those with school-aged children also received tuition subsidies (essentially conditional cash transfers). Over half of the households in SWP villages also received individual loans (accounting for about 60% of disbursements). The interest rate was set at the same level as for loans from the government's poor-area programs and the Agricultural Development Bank of China, although this is a lower rate than for commercial sources of credit. The loans financed various activities including initiatives for raising farm yields, animal husbandry and tree planting. There was also a component for off-farm employment, including voluntary labor mobility to urban areas and support for village enterprises. The selection of project activities aimed to take account of local conditions and the expressed preferences of participants, although it is unclear how well this worked in practice.⁵

Household selection was a less transparent process than village selection, which could be based on data and field observations. The household selection was typically done by the pre-existing "farmers' committee" in each village and was not subject to rigorous monitoring. From our discussions in field work, it appears that credit-worthiness criteria and successful past experience with similar project activities played an important role. No doubt local level connections also played a role.

In common with other development projects, the SWP provided the capital and technical assistance, but it did not provide insurance, and many of the activities are likely to entail non-negligible risk; the income gains will depend on a number of contingencies,

⁵ Farmers' preferences were sometimes over-ruled by local cadres (World Bank, 2003).

including the vagaries of the weather, uncertain demand for the new products and risks associated with out-migration.

The *ex ante* expectation was that the SWP would virtually eliminate poverty in the selected villages over the longer term. The World Bank's Implementation Completion Report (ICR) — the final document giving the *ex post* "self-assessment" of a lending operation by the relevant operational unit — claimed that the SWP had a substantial impact on poverty, citing survey data indicating that the poverty rate had been more than halved in the project areas over 1995-2001 (World Bank, 2003).⁶ However, the attribution of these gains to the SWP is questionable. The evaluative claims in the ICR are reflexive comparisons, which only reveal the true impact under the assumption that there would have been no progress against poverty in the absence of the project. That assumption must be deemed highly implausible in this setting.

Ravallion and Chen (2005) studied the impacts of the SWP over the disbursement period, 1995-2000, using survey data for 2,000 randomly sampled households in both SWP and observationally similar non-SWP villages that had first been surveyed in 1995 (at the beginning of the project) and then annually until project completion. On comparing income changes in SWP villages with those in the matched non-SWP villages, they found an average income gain over five years of around 10% of baseline mean income, representing an average rate of return of 9-10%. The gains are not as dramatic as suggested by the reflexive comparisons in the ICR, but they are still sizeable.

However, a surprisingly large share of the income gain was saved. On comparing the final year of disbursement with the first, Ravallion and Chen found only a modest impact on mean consumption or consumption poverty. The savings rate from the project's income gains was well above the pre-intervention average savings rate..

Why was there such a high savings rate from the initial income gains? A number of explanations can be suggested, carrying rather different implications for the long-term impact of the SWP. Possibly households saved more to assure they could repay the loans. That depends on the extent to which repayment was enforced. While the World Bank's loan is made to the (central) Government of China, and repayment is virtually

⁶ This was confirmed by researchers at the Chinese Academy of Social Sciences, who also pointed to a substantial increase in primary school completion rates and a decline in the infant mortality rate which they attributed to the SWP (Guobao et al., 2004).

certain, that is not the case for the loans made at local level, where enforcement problems are common. Indeed, local repayment rates on loans for poverty reduction under the government's own program were less than 25% in the three provinces covered under SWP.⁷ However, it may be that the necessity of the center repaying the World Bank "trickled down" in the form of greater local enforcement of SWP repayments than for the loans made under the government's own poor-area programs.

Another possibility is that the high initial savings rate reflected a perception on the part of participants that the longer-term income gains from SWP would be modest or uncertain at best — raising concerns about the sustainability of the program's impacts.⁸ When interpreted in terms of the Permanent Income Hypothesis, the Ravallion-Chen findings imply that participants felt that a large share of the income gain was transient, and (hence) it was saved. While this would happen even without uncertainty about the future income gains, such uncertainty is likely, and would probably lead to precautionary saving in response to the project.⁹ In this regard, it is instructive that Ravallion and Chen found large year-to-year differences in impact, which were primarily due to variability in the annual returns to the program's investments rather than the level of investment. This variability in the returns suggests that participants would have had a hard time assessing the program's impact on permanent income.

The transient-income explanation suggests that the income impacts of SWP would diminish appreciably after disbursement. Precautionary saving would also start to fall as participants learn more about the impacts. Consumption gains should become evident in due course, consistent with the project's underlying impact on permanent income.

There is another explanation for the high savings rate from the short-term income gains. This postulates that the SWP systematically alters the returns-to-saving in the participating villages. By this view, the project provided local public goods that increased the marginal product of private capital, and so stimulated higher savings to

⁷ The repayment rates on loans for poverty reduction in 1997 ranged from 8% in Yunnan to 23% in Guizhou. Repayment rates were somewhat higher for other types of loans but the overall average was still only 30% (Government of China, 1998).

⁸ The ICR rated "sustainability" as "highly likely." The Bank's internal evaluation of SWP by its Operations Evaluation Department pointed to the need for further evidence on the longer-term sustainability and impact of SWP.

⁹ There is evidence of precautionary savings in response to uninsured risk in the same region of rural China; see Jalan and Ravallion (2001).

support the desired private investment, which would yield longer-term income gains beyond the life of the project.¹⁰ This assumes that there are capital-market imperfections, which entail that investment depends on own-savings and that the marginal products of private capital are not equalized across locations. With the poor facing severe constraints on access to credit and yet having higher marginal products of capital in their own (farm and non-farm) enterprises (given low capital stocks and concave production functions) one might expect to see a sizeable (and pro-poor) investment response. Clearly, this explanation offers a more positive view of the prospects of a sustained impact on poverty from the SWP, in that it suggests that income gains will persist well beyond the disbursement period (as the returns to investment start to be realized) and that sustainable consumption gains would emerge.

By re-surveying in 2004/05 the same sample studied by Ravallion and Chen (2005) we hope to throw light on which of these explanations is most plausible.

3. Data

The original plan for the impact evaluation of SWP was to do a baseline survey in 1995 and to only do follow-up surveys during the Bank's disbursement period, up to 2000. However, we decided to re-survey the original sampled households in 2004/05, to try to resolve the issues about longer-term impact discussed in the previous section.

All surveys were implemented by the Rural Household Survey (RHS) team of the government's National Bureau of Statistics (NBS). The surveys covered 2,000 randomly-sampled households in 200 villages, with roughly half not participating in the SWP. All villages were in counties covered under the government's poor-area program, to assure that we will identify the impact of the SWP, on top of the government's program. There are 112 SWP villages and 86 non-SWP villages.¹¹ The SWP villages were a random sample from all project villages, while the non-SWP villages were a

¹⁰ Jalan and Ravallion (2002) provide a micro model of growth with imperfect capital markets that is consistent with this property, and find supportive evidence in the same region of rural China.

¹¹ In the 2004/05 survey, two villages (one SWP and one not) were inadvertently replaced by two different villages in the same township.

random sample from all other villages in the designated poor counties. Ten randomly sampled households were interviewed in each village.

The 1996-2000 and 2004/05 surveys included community, household and individual questionnaires. The community schedule collected data on natural conditions, infrastructure and access to services. The household survey collected data on (*inter alia*) incomes, consumptions and assets. The individual questionnaires covered gender, age, education and occupation. A data set was collected from 1997 to 2001 on development project activities (both SWP and under other existing government programs). There are 34 project activities identified in these data, in seven categories (farming, animal husbandry, forestry, infrastructure, education, health and labor migration).¹²

We follow Ravallion and Chen (2005) in using 1996 as the baseline. There are serious comparability problems between the 1995 survey and later surveys.¹³ As a baseline, the 1996 data are not free of contamination; 17% of the program's total disbursement on household projects had been made by the end of 1996. We check robustness to using 1995 as the baseline.

Relative to other household surveys, unusual effort went into obtaining accurate estimates of consumption and income from the 1996-2000 and 2004/05 surveys. While the community, individual and project activity surveys used conventional one-time interviews, the household survey was quite different. The household surveys from 1996 onwards were closely modeled on NBS's Rural Household Survey (RHS) (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. Similarly to the RHS, sampled households maintain a daily record on all transactions plus log books on production. Local interviewing assistants visited each household at two-three weekly intervals, at which time inconsistencies found at the local (county-level) NBS office are checked. Other trained interviewers also visited at regular

¹² A project activity survey in 1998 also gathered information about the scale and the starting year of each SWP sub-project at village and household level, as well as data on other funding these villages and households received from the government and other sources.

¹³ Because of delays in NBS being told the locations of SWP villages, the first survey in December 1995 had to use a one-time interview method, asking recall over the full year. The use of this long recall period is likely to lead to underestimation of income and consumption (though this is of less concern for village-level characteristics). The subsequent surveys used the daily-diary method over the full year, allowing more accurate income and consumption data.

intervals to collect additional data. This intensive interviewing method is a marked contrast to most surveys in which the respondent is visited only once or twice.

The consumption aggregate (termed “living expenditures” by NBS) is built up from very detailed data on cash spending on all commodities and imputed values of in-kind spending, which is mainly consumption from household production, valued at local selling prices. Living expenditures exclude spending on production inputs (which are accounted for in net income from own-production activities). They also exclude transfer payments, 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. Income is measured net of all production costs, including interest on debt (including loans from the SWP). The migrant workers were not tracked, although the income aggregate includes remittances received from family members who migrated, including those supported by the SWP. Remittances are expected to be the main means by which the out-migration component reduced poverty in the short run.

Given the unusual effort that went into data collecting and checking the consumption and income data, we expect that subtracting consumption from income will give reasonably accurate estimates of savings. We also look into what forms the savings took. There are many forms of saving in this setting, including money balances and investment in own-production activities. The survey was not designed to allow a complete independent accounting of all forms of saving. Some data were collected on assets and liabilities, although the reliability of the reported values is questionable. We also study impacts on holdings of specific assets.

For the 2004/05 follow-up survey we used exactly the same survey instrument as for the 1996-2000 surveys, augmented with a module to elicit perceptions of both welfare and the project’s impacts. The module asked respondents to assess whether various aspects of their lives had improved over the preceding 10 years. (The questions in this module were asked in 2005.) These involved a long list of aspects of well-being and in each case the respondent was asked whether this item had improved or not over the last

10 years, on a 10 point scale (from “extremely worse off” to “extremely better off”).¹⁴ (The sample was restricted to adults who were at least 28 years of age at the time of the interview.) Our idea here is to see whether such a rapid appraisal tool — which does not require any prior surveys, including a baseline — gives similar results to our more costly longitudinal survey-based method.

Over 1996-2005, the attrition rate was 12% (6% over 2000-05). Using a probit model for attrition over 1996-2005 we found a number of significant predictors, including age of head, share of children, landholding and some geographic variables.¹⁵ (Being an SWP village was not a significant predictor of attrition.) NBS survey teams were instructed to find replacement households as similar as possible to those that dropped out. We also tested how well this replacement worked, using a regression for the probability of being a replacement household estimated on the pooled sample of replacement and “drop-out” households.¹⁶ Among the same set of covariates for attrition, no regressor was a significant predictor for replacement and the regression has very low overall explanatory power. It appears that the sample with replacements can be considered representative of the population.

We checked the robustness of our results to several potential data problems. One problem concerns the aggregation of total living expenditures. It appears that in processing the 2004/05 survey data, living expenditure in one county may have failed to include in-kind consumption.¹⁷ The data for three other households whose in-kind income was more than six times larger than their total living expenditure seem to have a similar problem. We re-estimated the impacts on consumption and income, dropping this one county and the three households. The results reported below were robust to this change (details available from the authors).

¹⁴ The Chinese and local language versions of the module were refined over time on the basis of field tests in poor villages in a number of locations.

¹⁵ A statistical addendum is available from the authors giving full details.

¹⁶ Note that we have baseline data for the “drop-outs” and the current year’s data for the replacements. To deal with the time difference we did a pro-rata adjustment of the data on drop-outs to 2004 values according to the ratios of the means over time for each variable, based on the balanced panel. In calculating the ratios, we also weighted by the attrition probability.

¹⁷ We suspect there is a problem because the total living expenditure of 68% the sample in that county is equal to cash expenditure, whereas net in-kind income is about half of overall total.

Another potential data problem is related to the coding of SWP projects. We find in the village-level project data base that all ten villages in one county claim to have an SWP funded project, even though six of them were officially designated as non-SWP villages.¹⁸ It may well be that there was significant SWP participation by villages that had not been selected for the project in this particular county (although we cannot rule out coding errors). On deleting this county we found that our main results were robust (details are available from the authors).

4. Estimation methods and sources of bias

Our aim is to estimate average treatment effects on the treated. The double-difference method identifies a project’s impact if the changes in outcomes for the non-participants reveal the counterfactual changes for participants. The key assumption is that the selection bias (the counterfactual difference in outcomes) is constant over time and additive. In the present context, we point to two sources of time-varying selection bias: (i) outcome changes are correlated with initial differences between the participating and non-participating areas, and (ii) spillover effects, whereby the project itself alters the subsequent path of outcomes for the non-participants.

4.1 Biases due to targeting

Let us begin with the classic evaluation problem. We have data on an outcome measure Y_{it} for the i ’th unit observed at dates, $t=0,1$. Each unit is observed to be either a participant ($T_{it} = 1$) or non-participant ($T_{it} = 0$). We can write the outcome measure as:

$$Y_{it} = Y_{it}^C + T_{it} G_{it} \quad (t=0,1; i=1,\dots,N) \quad (1)$$

where $G_{it} = Y_{it}^T - Y_{it}^C$ is the gain (“impact”) due to the program, Y_{it}^T is the outcome under treatment and Y_{it}^C is the counterfactual outcome. G_{it} is not directly observable for any i (or in expectation) since we do not know Y_{it}^T for $T_{it} = 0$ and Y_{it}^C for $T_{it} = 1$. The selection bias is the mean difference in counterfactual outcomes (dropping the i subscripts):

¹⁸ There are scattered minor reports of SWP activity in non-SWP villages elsewhere, but these appear to be random, and are probably coding errors.

$$B_t = E(Y_t^C | T_1 = 1) - E(Y_t^C | T_1 = 0) \quad (2)$$

We call this unconditional bias, given that we have not yet allowed for any other control variables. Given the purposive targeting of the SWP it must be presumed that $B_t \neq 0$.

The standard double-difference estimator assumes that $B_1 = B_0$, implying that the change in mean gains for period 1 participants is consistently estimated by:

$$DD = E[(Y_1^T - Y_0^T) | T_1 = 1] - E[(Y_1^C - Y_0^C) | T_1 = 0] = E[G_1 - G_0 | T_1 = 1] \quad (3)$$

If period 0 is a true baseline, with $T_{0i} = 0$ for all i (by definition), then $Y_{0i} = Y_{0i}^C$ for all i , and so $DD = E(G_1 | T_1 = 1)$, i.e., mean impact on the treated units.

However, time-invariant unconditional bias ($B_1 = B_0$) is implausible for poor-area development programs. The targeted poor areas typically lack infrastructure and other initial endowments, which could (in turn) affect the subsequent growth rates. DD will then be a biased estimator, since the subsequent outcome changes are a function of initial conditions that also influenced the assignment of the sample between the two groups. In other words, the selection bias will not be constant over time.¹⁹

The direction of bias in DD depends on whether the underlying growth process is convergent or divergent. For the government's poor-area programs in southwest China in the 1980s, Jalan and Ravallion (1998) found that failure to control for the initial heterogeneity between the targeted counties and non-participating counties yields a downward bias in a DD estimator, consistent with growth divergence.²⁰ However, it is unclear whether this also holds across villages within the same (poor) counties; indeed, the results of Jalan and Ravallion (2002) (also for southwest China) suggest that inter-county divergence can occur side-by-side with intra-county convergence.

We address this issue by balancing treatment and comparison units in terms of the initial conditions that may have influenced program placement. These variables are represented by the vector X . Our key identifying assumption is that the selection bias is time-invariant conditional on X , i.e., that:

$$E(Y_1^C | T_1 = 1, X) - E(Y_1^C | T_1 = 0, X) = E(Y_0^C | T_1 = 1, X) - E(Y_0^C | T_1 = 0, X) \quad (4)$$

¹⁹ This echoes more general concerns about the importance of correcting for selection bias based on observables (Rosenbaum and Rubin, 1983; Heckman et al., 1998).

²⁰ Also see Jalan and Ravallion (2002) who find evidence of divergence at the county level.

On applying a result due to Rosenbaum and Rubin (1983), if outcome changes are independent of participation given X , then they are also independent of participation given the propensity score: $P(X_i) = \Pr(T_i = 1|X_i)$, ($0 < P(X_i) < 1$). This justifies balancing on $P(X)$ to remove selection bias based on X . Note that this only addresses time-varying selection bias based on observables; a bias will remain if there are any latent (time-varying) factors correlated with the changes in counterfactual outcomes. As discussed later, a remaining bias due to unobservables appears to be more likely for household selection than village selection.

We use various methods for assuring covariate balance on $P(X)$. One method is to limit comparisons to a trimmed sub-sample with sufficient overlap in propensity scores. For our data, the region of common support (minimum score for treated, maximum score for untreated) is (0.11-0.95). For our “trimmed sample” we chose a slightly tighter interval (0.1, 0.9), which are also the efficiency bounds recommended by Crump et al. (2006) for estimating average treatment effects with minimum variance.²¹

We also use the weighted-regression method proposed by Hirano, Imbens and Ridder (2003). Thus we estimate the DD from the following regression:

$$Y_{it} = \alpha + DD.T_{it}t + \beta T_{it} + \delta_t + \varepsilon_{it} \quad (5)$$

where $E(\varepsilon|T) = 0$. This is estimated with weights of unity for treated units and $\hat{P}(X)/(1 - \hat{P}(X))$ for controls, where $\hat{P}(X)$ is a consistent estimate of $P(X)$ and $0 < \hat{P}(X) < 1$. Hirano et al. show that weighting the controls this way yields an efficient estimator.²² We estimate (5) on both the pooled sample (for $t=0,1$ and including replacement households) and for both the total sample and trimmed sample.

²¹ Using the formula in Crump et al. (2006), the exact bounds are 0.0997 and 0.9003. For estimating the average treatment effect on the treated Crump et al. also recommend dropping treatment units with scores less than about 0.8 (for our data), but keeping all untreated units. We did not follow this recommendation. For one thing, we felt that this entailed the loss of too many treatment villages, raising concerns about inference for the population of treated villages. Secondly, our balancing tests performed better when we also deleted the low-score untreated villages, which are clearly poor comparison units.

²² If we wish to estimate the average treatment effect for the population, the weights are $1/\hat{P}(X)$ for the treated units and $1/(1 - \hat{P}(X))$ for the controls (Hirano and Imbens, 2002).

To interpret (5) note that, in a balanced panel, we could instead estimate the equivalent regression in the more familiar “fixed-effects” form:

$$Y_{it} = \alpha^* + DD.T_{it} + \delta_t + \eta_i + v_{it} \quad (6)$$

Here the fixed effect is $\eta_i = \eta_i^T T_{i1} + \eta_i^C (1 - T_{i1}) = \beta T_{i1} + \bar{\eta}^C + \mu_i$ ($E(\eta_i | T_{i1}) \neq 0$) where $\beta = \bar{\eta}^T - \bar{\eta}^C$, $\mu_i = (\eta_i^T - \bar{\eta}^T) T_{i1} + (\eta_i^C - \bar{\eta}^C)(1 - T_{i1})$, $E(\mu_i) = 0$, $\varepsilon_{it} = v_{it} + \mu_i$ and $\alpha = \alpha^* + \bar{\eta}^C$. Thus, βT_{i1} in (5) picks up differences in the mean of the latent individual effects, such as would arise from initial selection into the program. The advantage of (5) is that it does not require a balanced panel, and hence it gives estimates that are robust to selective attrition (recalling that the replacements appear to have preserved the sample’s ability to represent the population).

As a robustness check, we compare these estimates with matching on the propensity score. Note first that the sample estimate of mean impact can be written as:

$$\left(\sum_{i=1}^{N_T} (Y_{i1}^T - Y_{i0}^T) - \sum_{j=1}^{N_C} W_{ij} (Y_{j1}^C - Y_{j0}^C) \right) / N_T$$

where N_T is the number of SWP participants, N_C is the number of control observations, and W_{ij} is the propensity score-based weight given to the j ’th non-participants in making a comparison with the i ’th participant. How many non-participants to include in the control group and how to assign weights to each non-participants are practical questions in implementing PSM. One option is to use the popular method of nearest-neighbor matching. However, because of the non-smoothness of nearest neighbor matching, the conventional bootstrapping method is inappropriate for estimating the standard errors (Abadie and Imbens, 2006). In order to assure valid bootstrapped standard errors, we choose to apply nonparametric kernel matching in which all the non-participants are used as controls and weights are assigned according to a kernel function of the predicted propensity score (following Heckman et al., 1997, 1998). The weights can be written as $W_{ij} = K_{ij} / \sum_k K_{ik}$, where $K_{ij} = K((\hat{P}_j(X) - \hat{P}_i(X)) / a_n)$, in which $K(\cdot)$ is a kernel function and a_n a bandwidth parameter. We use the normal density function as the kernel and the odds ratio (rather than propensity score) because SWP villages are over-sampled relative to their frequency in the population eligible for the project.

The conditional independence assumption motivates a specification test of whether there are differences in observables between the project and non-project villages after conditioning on $P(X)$ through matching or re-weighting. Following Rosenbaum and Rubin (1985) and Abadie and Imbens (2006) we test for covariate balancing using differences in standardized means between the SWP villages and matched or re-weighted non-SWP villages. To achieve a better balance of covariates and to allow for a more flexible estimate of propensity scores, we also include polynomial terms for the initial income levels (see, for example, Smith and Todd, 2005). We will show that the matching and re-weighting procedures produce a satisfactory balancing of the observables between SWP and comparison villages.

4.2 *Biases due to spillover effects*

All the methods described above assume that an observationally similar comparison group pre-intervention reveals the counterfactual of what would have happened over time to mean outcomes for the treatment group in the absence of the treatment. This will clearly not be the case if there are any spillover effects, whereby the intervention changes outcomes for non-participants. This is a distinct source of bias to the problem of non-random assignment.

Spillover effects due to residential mobility between villages are unlikely in this setting given the village-level administrative land allocation. Under China's rural land laws, a migrating household would have little prospect of getting a share of the land available (and almost certainly cultivated) at the destination and would also risk losing their land at the origin.

Another source of spillover effects is inter-village trade (possibly via urban hubs). To the extent that the project has an impact on local incomes and prices, trade-induced general equilibrium effects will entail spillover effects to the non-SWP villages used to infer the counterfactual. We will test for impacts on prices as well as incomes, distinguishing cash-incomes (as derived from inter-village trade) and incomes-in-kind.

Local public spending responses to project aid can also be confounding. Recall that there were other development activities supported by the local (county and provincial) governments, side-by-side with the aid-financed SWP. As has long been recognized in the public finance literature, whether the resources transferred to

participants actually financed the identified project is unclear given that, to some degree, external aid is fungible. In the present context, non-SWP villages could be affected by a re-allocation of public spending by local authorities. With roughly half the loans not repaid by farmers or local agencies (while the loans from the World Bank are repaid by the central government), the incidence of the implicit grant component would presumably entail a reduction in other public spending on the SWP villages. In principle, the induced gains to non-SWP areas could entail widespread spillover effects, possibly beyond the provincial borders. However, a geographic “flypaper effect” is plausible, whereby the displacement entails gains to villages in the same county (though possibly with some spillover to other counties in the same province).²³ The implication for our evaluation is that comparing outcome changes over time between SWP and (matched) non-SWP villages in the same counties will under-estimate the project’s true impact.

There is a counter-argument, whereby the local public response to SWP could entail negative displacement. By this argument, the SWP’s counterpart funding requirements, and the desire of local cadres to be seen to have succeeded with this (high-profile) external aid project, would attract local resources from non-SWP villages and counties.²⁴ This would impart the opposite bias, entailing an over-estimation of impact.

We test for spillover effects. The presence of non-SWP development projects in the SWP villages provides the clue. We use the same evaluation methods described above, but the “outcome variable” becomes the extent of non-SWP project activity in the SWP villages.

5. Estimated impacts

Table 1 gives mean income and consumption and the poverty rates for 1996, 2000 and 2004/05. The poverty line of 808 yuan per person per year in 1995 prices (corresponding to the \$1 a day line used by Ravallion and Chen, 2005, at 1993 purchasing power parity) as well as poverty lines above and below this figure. We see

²³ On the flypaper effect see Zampelli (1986). It is sometimes argued on theoretical grounds that such effects should not be observed under standard assumptions, although see Roemer and Silvestre (2002) for a counter-argument. Evidence (for Vietnam) of a sectoral flypaper effect in a development project can be found in van de Walle and Mu (2007).

²⁴ Pack and Pack (1990) found evidence of such an effect of earmarked aid to Indonesia.

that the income gains in SWP villages between 1996 and 2000 were larger than among non-SWP villages, but that this reverses between 2000 and 2004/05. Ten years after its commencement, the SWP does not appear to have allowed the selected poor villages to catch up with the rest of these (poor) counties.

Table 1 suggests that SWP had little or no impact on income and consumption. However, before accepting that conclusion we need to probe more deeply into some potential sources of bias.

We begin with selection bias due to non-random placement of the SWP. Table 2 gives probits for whether a village was selected for SWP, as used to estimate the propensity scores. The variables were chosen to reflect the selection criteria used by the project staff (based on our interviews at the time).

We find that project villages tend to be in more hilly/mountainous 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.²⁵ The SWP villages also tend to have larger populations, with lower mean income in 1995 (from the village-level data), lower mean consumption in 1995 (from the household survey) and more land per capita. The latter characteristic probably reflects lower population density and lower land quality in the project villages. In most respects, the results of Table 2 suggest that the SWP villages tend to be poorer than other villages within the project counties, consistently with Table 1.

Using the propensity scores based on Table 2 to re-weight the data we were able to obtain a close balancing of the characteristics of the two samples (including in the means of the initial outcome variables), particularly after trimming the samples, as discussed in the previous section. The Appendix provides details on the balancing tests, which pass comfortably; this was also the case for a full set of covariates in Table 2, for which the balancing tests are reported in the Addendum available from the authors.

5.1 Double-difference estimates of average impacts

In assessing impacts on mean consumption and income, we begin with the simple *DD* estimates of the mean impacts for income, consumption and saving, as given in Table

²⁵ Remote villages are more likely to have a very basic health clinic, to compensate for the inaccessibility to more comprehensive township facilities.

3. We give estimates for both 2000 (at the end of disbursements) and 2004/05 and for both the levels and the logs; the latter gives higher weight to the gains to poorer households. The baseline is 1996 in both cases.

Focusing first on the disbursement period, we see that there was a sizeable and statistically significant impact on mean income, but not consumption; the bulk of the income gain up to 2000 was saved. (The same pattern was found using 1995 as the baseline.) On decomposing income (as wage income, farming, animal husbandry, fishery, forestry, non-farm enterprises, transfers and asset income), the only component that showed a statistically significant impact was animal husbandry, for which the simple *DD* estimate of the impact on net income was 90.85 yuan ($t=2.92$), which rose to 117.26 ($t=3.37$) and 136.15 ($t=3.55$) using weighting and matching (respectively) to correct for selection bias (Table 4).²⁶

Another way of disaggregating income is into cash or kind (which will be relevant when we consider trade spillovers in section 6). We found that the bulk of the short-term income impact was income in-kind from animal husbandry, as is evident from Table 4. This is puzzling, as a sizeable share of income-in-kind from husbandry in a rural economy is also consumed directly, and should then show up in consumption. However, the income in-kind that is being affected by the project appears to be small non-productive animals and new litters of productive animals, which are counted as income in kind but are held over for consumption or sale at a later date rather than consumed.²⁷ We will return to this point when we discuss the longer-term impacts.

We can also disaggregate consumption expenditure. On separating food staples (rice, wheat etc) from non-staples and other foods we found significant impacts in 2000 for non-staple foods (meat, vegetables etc); the simple *DD* for this category was 26.26 yuan ($t=1.68$) though rising to 40.64 yuan ($t=2.69$) and 42.58 yuan ($t=2.70$) for the PS weighted and kernel matched estimators respectively. This is likely to entail nutritional gains through higher protein and more micro-nutrients.

²⁶ We only report the results for husbandry, and summarize those for other components; a statistical addendum is available with full details.

²⁷ We do not have data on this, but the practice of counting such animals as income in-kind is discussed in the manual for enumerators provided by NBS.

The results change dramatically when we track the impacts through to 2004/05, as is evident when we return to Table 3. We find no significant impacts on mean income or consumption over the longer observation period.²⁸ (This also was also true for staples and non-staples separately.) Table 3 also gives the *DD* estimates for mean income using the propensity scores to balance project and non-project villages; we give results using both weighting and matching, for both end dates, and for both the trimmed sample and total sample. The basic pattern in the simple *DD* estimates is still evident. The results are robust to using kernel matching instead of the re-weighted regression method.²⁹

While there is clearly some sensitivity to the choice of estimation method, the pattern is still reasonably robust, indicating significant and sizeable income gains during the disbursement period but much less in the longer term. The estimated income gains in 2000 tend to be larger when we correct for purposive selection of SWP villages; this is consistent with a divergent growth process between villages. However, no such pattern is evident for the 2004/05 impacts.

We did find significant longer-term impacts on income in-kind. On breaking up income in-kind by source, we found that both farming and husbandry accounted for almost all these long-run impacts, though only husbandry was significant (Table 4). The simple *DD* estimate of impact in 2004/05 on income in-kind was 130.30 yuan ($t=2.11$), though this fell somewhat when we corrected for selection; with weighting we obtained $DD=111.90$ ($t=1.89$) while with matching, $DD=96.98$ ($t=1.78$). We found no other significant impacts in the long run amongst cash income components.

In contrast to the period up to 2000, we find consumption gains in the post-disbursement period. The impact on total consumption in 2004/05 is not statistically significant (Table 3). However, when we break this up according to cash or kind, we do find signs of larger impacts on consumption in kind. The simple *DD* estimate for consumption in kind in 2004/05 is 118.40 yuan ($t=2.54$), although this drops appreciably when we correct for selection bias; using PS weighting the impact is 74.46 ($t=1.50$). The

²⁸ The same pattern was evident using 1995 as the baseline, although impacts were somewhat lower.

²⁹ The results were also robust to deleting the troublesome county and the observations with problematic data (section 3).

longer-term impacts on consumption in kind probably include consumption of the income in-kind from animal husbandry that we observed in the SWP disbursement period.

For either the simple *DD* or the score-weighted *DD*, the consumption gains exceed what one could reasonably expect under the permanent income hypothesis (PIH) if the income gains from SWP were purely transient. For then the consumption gains in the four-year period following SWP would simply be the rate of interest times the permanent-income equivalent of the transient income gain. For the simple and score-weighted *DD*, plausible rates of interest would imply lower consumption gains than we see in Table 3, although this is not true for the kernel-matched *DD* for which the post-disbursement consumption gains equal the increment to permanent income at a rate of interest of about 10%. Statistically, however, we cannot reject the null hypothesis that the post-disbursement consumption gain equals the increment to permanent income (at reasonable interest rates) treating the SWP income gain as transient.

The PIH interpretation begs the question as to why we saw no consumption gains in the disbursement period. If SWP participants knew at the outset that the project would entail only a transient income gain then consumption would have immediately reflected the implied gain to permanent income. However, from what we know about the SWP, it is unlikely that participants could have formed a reliable estimate of the gain to permanent income due to SWP until at least project completion. As noted in section 2, there was considerable uncertainty about the income gains, and high initial savings may have been a short-term precautionary response.

We tested for impacts on a number of other variables.³⁰ We found no evidence of impacts on interest payments on loans or the proportion of households paying interest or paying back loans, for either 2000 or 2004/05. So we find no support for the idea that either the high savings from the short-term gains or the lower longer-term impacts on incomes stem from greater enforcement of interest or repayment requirements under the SWP, compared to other credit sources.

With weak enforcement of the SWP loan repayments, it might be conjectured that taxes on SWP areas would increase, to help local authorities pay back the SWP loans to higher levels of government. However, we did not find any evidence of impacts on taxes

³⁰ Again we only summarize the results here; the addendum gives full details.

or fees paid per capita, in either 2000 or 2004. It appears that higher levels of government treated the SWP as, in large part, a transfer payment to lower levels.

In testing for impacts on agricultural productivity, we used total farm income per unit area.³¹ We found no evidence of impacts.

Nor did we find much evidence of impacts on holdings of productive assets and wealth (including housing). This was true for both the disbursement period and the longer-term. An exception is that the village data base revealed a significant impact on livestock holdings, notably cows and goats.³²

There is some sign of a demographic impact. Household size fell in both SWP and non-SWP villages over 1996-2000, but more so in the former. The simple *DD* for household size is -0.13 persons ($t=-1.75$) and it is slightly larger with the corrections for selection bias (the PS-weighted estimate is -0.16, $t=-1.64$, and it was similar for kernel matching). The demographic effect was associated with slightly fewer children. However, the demographic impact was not evident in 2004.

Nor did we find any evidence of impacts on remittances received from family members migrating out, or on the probability of a family member migrating.³³

We did find significant impacts on school enrolment rates during the disbursement period; our PS-weighted *DD* estimate was 0.074 (with a t-ratio of 2.20), i.e., a 7.4% point increase in the school enrollment rate of children aged 6-14 by the year 2000 is attributed to SWP.³⁴ However, this impact had dropped substantially by 2004/05; the corresponding *DD* estimate fell to 0.032 ($t=1.00$). The transient schooling impact probably reflects the fact that the tuition subsidies ended with other SWP disbursements. Of course, even though the non-SWP village caught up substantially with the SWP

³¹ Ideally we would use physical output for a given crop per unit area under its cultivation. However, only total land area under cultivation was collected. Instead we used an overall farm productivity measure, obtained by dividing total net income from farming by total cultivated area; this can be interpreted as a mean crop-specific yields weighted by both prices and shares of land.

³² The simple *DD* for cows per person in 2000 was 0.05 ($t\text{-ratio}=2.47$); with score-weighting it rose to 0.07 ($t=3.54$) and it was the same with kernel matching ($t=4.33$). By 2004 the impacts were slightly higher and equally significant statistically; the simple *DD* estimate was 0.07 ($t=3.69$) while the score-weighting the impact was 0.09 ($t=4.05$) and with kernel matching it was 0.10 ($t=3.92$). Significant impacts were also evident for sheep, although with lower t-ratios.

³³ Out migration in the previous year is only measured for those present in the village at the time of the interview, although NBS made an effort to ask the individual questions at times of the year when migrants are more likely to be present. Remittances may well be the better indicator.

³⁴ The uncorrected *DD* was 0.046 ($t=1.41$) and the kernel matched *DD* was 0.072 ($t=2.40$).

villages in schooling by 2004/05, that still means that there were children in SWP villages who entered school earlier than without the SWP.

There was almost no sign of impacts on the prices of agricultural outputs and purchase prices for inputs for 13 items.³⁵

We found positive impacts during the disbursement period for a number of types of infrastructure, although they are generally not statistically significant. We found little sign of impacts in the 2004/05 data. The exception was TV reception, which showed significant impacts in the longer-term as well as during the disbursement period.

We turn now to the impacts on income and consumption poverty. Table 5(a) gives the estimated impacts on the incidence of income poverty for various poverty lines; Table 5(b) gives the corresponding results for consumption poverty. Again we give estimates using the poverty line of 808 yuan per person per year as well as selected poverty lines above and below this figure.³⁶ The poverty impacts in the SWP disbursement period are broadly consistent with our findings for the impacts on the means in Table 3.³⁷ In Figure 1 we also give the results graphically, by plotting the *DD* estimate of the impact on the headcount index of poverty (for income and consumption poverty in panels (a) and (b) respectively) against the poverty line, which we vary over virtually the whole distribution. Impacts on the income poverty rate are largest just below the 808 poverty line, for both end dates. The impacts on consumption poverty echo our results for mean consumption around the middle of the range of poverty lines, where 2004/05 consumption-poverty impacts exceed those for 2000; the results imply a sizeable 10 or more percentage point drop in the consumption poverty rate at poverty lines around 600 yuan. However, this is not true at lower and higher lines, where impacts over the two time periods agree closely.

³⁵ The only exceptions were that diesel oil had a significantly higher price in the SWP villages by 2004/05 and edible oil crop had a slightly lower price. Details are available from the authors.

³⁶ The table only gives results for the trimmed sample, which is better balanced. However, although the precise estimates differ between the two samples, the basic pattern was the same, and our main conclusions do not depend on this choice.

³⁷ The results were also robust to deleting the county in which some SWP activity was recorded in non-SWP villages. We found marginally significant impact on consumption poverty reduction at low poverty line in 2004 after deleting the consumption outliers (The weighted D-D at 500 consumption poverty line is -8.06 with t-ratio of -1.72; the weighted D-D at 600 is -9.20 with t-ratio of -1.67.)

For all of the above impact estimates, the counterfactual is the absence of the SWP. There is an alternative counterfactual of interest, namely the absence of direct participation in any anti-poverty program, including the government's programs. For identifying this counterfactual we can use those households in non-SWP villages who did not participate in any other program; this applied to 69% of the households in non-SWP villages. So we repeated the above calculations dropping those who recorded any direct participation in other programs. (The balancing tests passed comfortably.) The impacts for 2000 were similar to those above. However, the long-run impacts on mean income and consumption were larger. For example, the simple *DD* estimate of the impact on mean income in 2004 rose to 125 yuan per person (as compared to 45 yuan in Table 5) although this fell to 99 yuan when we corrected for selection bias using PS weighting. Nonetheless, the impacts relative to this alternative counterfactual were still not significantly different from zero; for example, the t-ratio on the simple *DD* for mean income was 1.47, which dropped to 1.13 with PS weighting.

5.2 *Heterogeneity in impacts*

We focus on two dimensions of possible heterogeneity, namely whether the household itself received a loan under the SWP, and the initial characteristics of the household. With regard to the first, we can think of participation in SWP as being either “direct,” for those households who receive a SWP loan, or “indirect,” when the household only benefits from the local public goods supported by SWP. 56% of sampled households in SWP villages received SWP loans; a further 12% received loans under the government's own poor-area program.

A simple *DD* estimate indicates that the income gains in the disbursement period were substantially larger for the direct participants; the *DD* estimate was 263.07 yuan ($t=2.64$) for loan recipients versus 126.73 ($t=1.70$) for indirect participants.³⁸ No such differences were evident in the post-disbursement period. However, this calculation ignores the possibility of time-varying selection bias due to purposive targeting of the direct participants. Given that the household-level loans under the SWP were targeted to the poor, a convergent (divergent) process within villages — whereby initially poorer

³⁸ On using PS weights to correct for village selection the *DD* estimate for direct participants was 270.88 ($t=3.05$) versus 55.89 ($t=0.72$).

households experienced higher (lower) subsequent income growth — would generate an upward (downward) bias in the simple *DD* estimator.

To address this concern, we estimated a household-level probit for the selection of household loan recipients within SWP villages as a function of initial income as well as a number of other baseline attributes including household demographics, education attainments, land owned, dwelling value, ethnicity, village characteristics and county effects. This indicated that the beneficiary selection process within project villages favored certain types of households, notably those who were better educated households and with employment links to TVEs or state-owned enterprises. We found evidence of targeting in favor of those with fewer productive assets in 1996, but there was no sign of targeting in favor of those with low initial income. On using this regression to estimate a score-weighted *DD* for the differential impact of direct participation we could not reject the null that the income gains were the same for direct as indirect participants within SWP villages in both 2000 and 2004. We also tested robustness to the possibility that loans recorded as being received from the government's programs in SWP villages were in fact SWP loans; the results were very similar.

These findings do not support the view that income gains from SWP were any larger for direct than indirect participants. However, we caution that non-ignorable latent factors in the household selection process are of greater concern here than for village selection. While (for household-level data) the probit for direct participation had satisfactory explanatory power (pseudo $R^2=0.364$), there are potentially confounding omitted variables, such as related to the household's connections within the village.

With regard to the second dimension of heterogeneity, we tested for differences in impacts according to the initial values of income, education and ethnicity.³⁹ The score-weighted triple-difference was not significant for any of our outcome variables when we stratified by education or ethnicity. However, we found a notable difference when stratified by initial income (above or below the median), with significant longer term gains for the low-income group. When we interacted income with education we found that the longer-term gains were strongest for the relatively well educated (at least junior

³⁹ We distinguish Han Chinese from all other ethnic minorities. The ICR points to concerns about how well ethnic minorities were reached by the SWP (World Bank, 2003).

high school) amongst the low-income households, as can be seen in Table 6. We also find evidence of significant longer-term impacts on assets and housing for this group, but not on farm productivity. However, we did find some signs of productivity gains for those above the median income.

The heterogeneity in returns suggests that a different assignment of the loans would have increased overall impact. The household participation rate was slightly higher for the group of relatively poor but well educated households; 61.1% of this group in SWP villages participated, as compared to 58.8% of those with above median income and higher education, 50.0% of those with high income but low schooling, and 47.8% of those with both low income and schooling. (The program slightly favored better educated households both above and below median income.) Suppose that beneficiary selection had focused solely on the relatively well-educated poor, and saturated this group, with no change to mean impacts, which were zero for other groups (consistently with Table 6). Then the impact of the program as a whole would have risen substantially, from a mean impact of about 40 yuan per person to about 150 yuan.⁴⁰ To achieve this outcome, the program would have had to over-ride the community-based selection process, which evidently put too little weight on reaching the educated poor, even though this group was already favored in the selection process.

While we found no impacts on average remittances and out-migration, significant positive impacts were evident when we stratified by initial income and education; the impacts were significant for those who were initially above median income and (among those with above-median income) were larger for those with more schooling.

5.3 *The rapid-appraisal method*

Table 7 summarizes our findings from asking in 2005 whether various aspects of well-being had improved over the previous 10 years. Note that since the question already embodies the change over time, the single difference can be interpreted as a double-difference estimate of the impact on the underlying level of that variable. We give the estimated impact for all line items identified in the module added for implementing the rapid appraisal method. We give results with and without PS weighting and trimming.

⁴⁰ This is based on an impact of about 200 Yuan for this group (Table 5), scaled down by 25% to reflect the number of households in this group, which would then represent 75% of the total number of SWP participants.

Since the results without weighting or trimming only require the 2004/05 data, only this case would be feasible in a single survey round. We give the single difference between SWP villages and non-SWP villages in the proportion of the population saying that the item in question had “obviously improved” or better.⁴¹

The subjective assessments by SWP participants of whether their living standards had improved since the project began are not significantly different to those found for the non-SWP villages.⁴² For example, 36% of those in the SWP villages reported that their overall standard of living had “obviously improved” over the last 10 years. But this was also true of 36% of those in the non-SWP villages, implying zero impact of the project. The lack of evident impact also extends to human development and social impacts; for example, despite the aim of the SWP to promote “local community participation” (World Bank, 2003), there is no sign that knowledge of village affairs, participation in village decision making, or local democracy improved any more in the SWP villages than in the comparison villages (Table 7). Adding the corrections for differences in baseline characteristics between the treatment and comparison villages does not change the main conclusion from this exercise, although the impacts on household and village living standards are now significant at roughly the 10% level (Table 7). Nor did we find any impacts among the subset of households who initially had income per person below the median and were relatively better educated; the rapid appraisal method is unable to detect the longer-term gains found for this group using the quantitative longitudinal surveys.

However, we question how reliable this method is for identifying impacts. There is little correlation between the perceived changes in standard of living and the changes in log consumption per person between 1996 and 2004/05; the correlation coefficient is 0.09 for SWP villages, which is only significant at the 8% level ($t=1.78$); in the non-SWP villages, the correlation is even lower at 0.01. Table 8 gives regressions of the perceived change in overall standard of living on the change in log consumption per person, with controls for respondents’ gender and age. With the controls, we do find a significant partial correlation between perceived gains in living standards and the growth rate in

⁴¹ We also tested sensitivity to using both a lower and higher cut-off; in neither case did we find any significant difference between SWP villages and the comparison villages.

⁴² One item, “ecology”, becomes close to significant (though negative); with this number of items identified, one can expect at least one significant difference purely by chance.

consumption in the SWP villages (but not in the non-project villages). Note, however, that the coefficients on the change in log consumption and 1996 log consumption are very similar; indeed, we cannot reject the null that they are the same (though the rejection is stronger for SWP villages than non-SWP villages). Under this null, it is current consumption that is driving perceptions of the welfare gains.

Subjective recall over such a long period is evidently not reflecting well the changes in living standards as measured by consumption. The recall responses appear to put too high a weight on current level of living (particularly in the project areas) and were clearly also affected by many idiosyncratic factors not accountable to consumption (as is evident from the low R^2 in Table 8). By not adequately reflecting the baseline outcomes, the recall method has a hard time identifying any longer-term impacts.

The subjective module included questions on perceived current living conditions (for the same items in Table 7). We examined the relationship between the answers and consumption per person in the 2004/05 survey data, to see if there are any signs that the relationship is different between SWP and non-SWP villages, as might arise from impacts of “non-income” dimensions of welfare captured in the subjective assessments. Our test entailed regressing each subjective measure on log consumption per capita in 2004/05, a dummy variable for SWP villages and the interaction effect between these two variables. Only for roads (“are you satisfied with village road conditions?”) was there any significant difference between SWP and non-SWP villages; households with higher consumption in the SWP villages tended to rate road quality higher, but there was no such gradient in non-SWP villages. One might take this to suggest that the SWP enhanced perceived road quality for better-off households, although we are more inclined to the view that finding one significant result in 30 tests is purely by chance.

The questions on perceptions of credit availability offer a clue to whether the SWP helped relieve liquidity constraints. There is no sign in Table 7 that credit availability improved more in SWP villages. However, direct participants (who received a SWP loan) gave a significantly higher response on both current access to credit and the improvement over the last 10 years; on the 10-point scale, the mean response was 0.61 ($t=2.34$) points higher for the level and 0.50 ($t=1.85$) higher for the change in credit availability. These impacts were attenuated somewhat when we used our probit for direct

participation to correct for selection bias; the impact for the current level of credit availability dropped to 0.39 ($t=1.43$) and 0.49 ($t=1.91$) for the change. So there is some sign in the data that access to SWP loans did enhance perceived credit availability.

6. Are we underestimating the impacts due to spillover effects?

Biases in long-term impact estimates can arise from interference due to spillover effects, as discussed in section 4. Our results do not offer much support to the idea of trade-induced spillover effects. We have seen that there were no significant impacts on prices, although it might be argued that arbitrage eliminated any price differentials. More damaging to the notion that there were significant trade-spillovers across villages is the fact that we did not find significant impacts on cash income, even during the disbursement period; the short-term income gains were in kind, and mainly from animal husbandry. Since much inter-village trade is likely to involve cash, there must be a presumption that such trade was affected rather little by SWP.

What about bias due to the responses of the local political economy? From the data on project activities, we counted the number of non-SWP projects of each type that started between 1996 and 2001 (inclusive). In the cases when loans were made to households, the project data also give counts of the total number of beneficiary households. There are two main concerns about these data. Firstly: the enumerators and respondents may have had a hard time identifying which activities were not SWP in SWP villages. (This is of less concern at village level since the list of project villages was provided by the government.) Secondly: we cannot tell what happened in the post-disbursement period since the project data we use for these calculations were only collected during the SWP disbursement period.

Table 9 gives the results for various project activities.⁴³ Large displacement effects are evident for virtually all non-SWP activities.⁴⁴ For most categories, the mean in SWP villages is half or less that in non-SWP villages, implying that 40% or more of

⁴³ The main activities excluded are minor infrastructure projects none of which showed any significant displacement. When there is no response from a village for a specific activity we treat it as a zero; this is plausible, although we test robustness to treating it as a missing value.

⁴⁴ We repeated these tests using the total samples and treating all cases in which no entry was made as missing values. The results in Table 9 were reasonably robust. (The effects tended to be stronger under the alternative treatment of “no response” entries.)

the non-SWP spending allocation to SWP villages was cut, and re-allocated to non-SWP villages.⁴⁵ Such large displacement effects would imply that the benefits of the SWP are likely to have spilled over to our comparison villages, leading us to under-estimate the impacts of SWP.

How large is the bias in our estimates of the impact on income due to these spillover effects? Let $DD^* = DD + SPILL$ be the true impact where $SPILL$ is the income gain to the comparison group due to the spillover effect. Let I_{SW} denote the level of investment per capita in the SWP villages and let I_{NSW} be that for non-SWP projects (primarily the government's programs). We can write $I_{NSW} = wI_{NSW}^T + (1-w)I_{NSW}^C$ where w is the population share of the SWP villages and I_{NSW}^T, I_{NSW}^C are the mean investments on non-SWP projects in the SWP and non-SWP villages respectively. Let $k \equiv I_{NSW}^T / I_{NSW}^C$ and let r_{SW} and r_{NSW} denote the rates of return to investment in the SWP and the government's program respectively. We assume that $I_{NSW}^T = I_{NSW}^C$ under the counterfactual. Then $SPILL = r_{NSW}(I_{NSW}^C - I_{NSW})$. We can then derive the following formula for the proportionate bias:

$$\frac{DD^*}{DD} = 1 + \delta \frac{r_{NSW} I_{NSW}}{r_{SW} I_{SW}} \quad \text{where} \quad \delta = \frac{w(1-k)}{1-w(1-k)} \quad (7)$$

The rates of return to the government's poor area programs may well be somewhat lower than for SWP, at least during the disbursement period, but for the present purpose it is probably reasonable to assume that $r_{NSW} = r_{SW}$. One-quarter of villages in the poor countries participated in SWP, so $w=0.25$. Based on Table 9 we can take $k=1/3$ to be a reasonable lower-bound (noting that DD^* / DD is strictly decreasing in k).⁴⁶ The level of investment per capita under the non-SWP projects appears is about half

⁴⁵ Recall that about one quarter of villages in SWP counties received the aid project, so that a non-SWP village will receive, on average, one third of the displaced spending.

⁴⁶ Using the project data base to comparing average loan amounts for non-SWP in SWP villages with those in non-SWP villages gives $k=0.58$.

of than under SWP.⁴⁷ Inserting these numbers into equation (7) we obtain

$DD^* / DD = 1.10$ — only a 10% overestimation.

It is possible that some of the non-SWP activity in SWP villages recorded in the project activity data should in fact have been classified as (counterpart) SWP activities. To test sensitivity to such miss-classification, we can calculate an upper bound to DD^* / DD by making the extreme assumption that $I_{NSW}^T = 0$ ($k=0$). Then it is evident from (7) that $DD^* / DD = 1.17$. Even at this upper bound, the bias due to spillover effects is modest.

Recall that the tests for displacement in Table 9 do not cover the post-disbursement period. It might be expected that the local spending balance between the treatment and comparison villages would be restored once the external aid ceased. However, while the displacement effect is presumably greater in the disbursement period, it cannot be ruled out post-disbursement. If there are in fact longer-term gains from the SWP and this is known locally then continuing positive displacement will be expected, making it harder to identify those gains.

Although the data used in Table 9 are not available for 2004/05, we can test for long-term impacts on new loan activity from non-SWP sources, as an indication of whether the SWP displaced other sources of finance in the post-disbursement period. (In 1995 we know who had received SWP loans so we can net this out of total loans received. In 2004/05 there were no new SWP loans.) By these calculations, we found no significant impacts on non-SWP loans in 2004/05. This does not suggest there was long-term displacement of other sources of finance.

7. Conclusions

The longer-term impacts of aid to poor areas depend crucially on why these areas are poor in the first place. If persistently poor areas arise from generalized capital-market failures then external aid can relieve the credit constraints and so enhance long-run growth. If instead the credit market failures are specific to certain (liquidity-constrained) subgroups of the population then the aid will need to be targeted to those groups.

⁴⁷ According to the project data, mean lending per capita under non-SWP projects (whether in SWP or non-SWP villages) represents 53% of the corresponding mean loan under the SWP.

However, persistently poor areas can arise from other causes, such as governance failures or (possibly policy-induced) distortions in other markets (including labor, such as due to restrictions on migration). Then the benefits from extra aid to poor areas may be modest or even zero. Unfortunately, the absence of rigorous studies of the long-term impacts of aid to poor areas has left a gap in our knowledge about both the causes of geographically concentrated poverty and aid effectiveness.

To help fill this gap, we have used a specially designed set of high-quality surveys collected over a 10 year period to study the impacts of a World Bank-financed poor-area development program in southwest China. We find a sizeable and statistically significant impact on mean household income in the participating villages during the disbursement period. However, there was a much smaller impact on consumption during that period; the short-term income gains were largely saved (although with some improvements in diet quality). Four years after disbursements had ended, both project and non-project villages had seen sizeable economic gains, with only modest net gain to mean income attributed to the project. Indeed, we cannot reject the null hypothesis that the longer-term average impact was in fact zero, although we do find evidence of longer-term impacts on income in-kind from animal husbandry.

The most plausible interpretation of our findings appears to be as follows. The high savings rate from the initial income gains reflected uncertainty about the future impacts — no doubt compounded by the uncertainty about the project's loan repayment and interest obligations, given uncertain contract enforcement at local level. Farm animals were clearly an important form of saving as well as being the main source of the short-term income gains. No doubt the relevant uncertainties were resolved in the longer term. Productivity gains turned out to be small. The initial income gains proved to be transient for most households, although there was some persistence in the income gains from animal husbandry. The mean consumption gains over the longer-time period are in rough accord with what one would expect from the (modest) increment to permanent income attributable to the project.

We find that there were significant and lasting income gains among the subset of households who were initially poor and relatively well educated. Presumably these households had more productive investment options, which could not be financed

otherwise given the liquidity constraints facing the poorest. The program's community-based selection process favored the better educated, but expanded coverage of those who were also poor could have greatly enhanced the program's overall impact.

Given the heterogeneity in returns, the implied (*ex-post*) deficiencies of the community-based selection process help explain the program's disappointing overall impact. While the program performed well in selecting poor villages, overall impacts were greatly attenuated by inadequate coverage of the (educated) poor within poor villages.

Some lessons emerge for assessing the impacts of development projects. The importance of investing in longer-term survey-data collection is plain. We have shown that rapid appraisal methods based on a single post-intervention survey using subjective-qualitative questions are vulnerable to severe recall-error biases, stemming from the fact that respondents' perceptions of how their living conditions have changed give far too high a weight to current circumstances.

Our results also point to the importance of taking account of the participants' inter-temporal behavior in response to the uninsured risks often associated with a development project. Those responses can cloud impacts in both experimental and non-experimental evaluations. An evaluation that focused solely on the income or consumption gains during the disbursement period (as is commonly the case) can give a deceptive picture of the true impacts.

Our findings illustrate how the responses of local development agents can cloud identification of the long-term impacts of geographically-placed projects (whether randomly placed or targeted). We found evidence of positive spillover effects on the comparison villages through the displacement of other development spending. Such interference suggests that the classic impact evaluation methods will systematically underestimate the long-term impacts. In our case, it is unlikely that these effects are imparting a large bias on our impact estimates, under plausible assumptions on the relevant parameters. But this may well be a bigger problem in other settings.

References

- Alberto, Abadie and Guido Imbens, 2006, "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74(1):235-267.
- _____, 2006, "On the Failure of the Bootstrap for Matching Estimators", mimeo, University of California, Berkeley.
- Chen, Shaohua and Martin Ravallion, 1996, "Data in Transition: Assessing Rural Living Standards in Southern China," *China Economic Review*, 7, 23-56.
- Crump, R., J. Hotz, G. Imbens, and O. Mitnik, 2006, "Moving the Goalposts: Addressing Limited Overlap in Estimation of Average Treatment Effects by Changing the Estimand," National Bureau of Economic Research, Technical Paper 330, Cambridge, Mass.
- Government of China, 1998, *Yearbook of China Agricultural Development Bank*. Beijing: China Statistics Press.
- Guobao, Wu, Qiulin Yang and Chengwei Huang, 2004, "The China Southwest Poverty Reduction Project," Paper presented at the conference, *Scaling Up Poverty Reduction*, Shanghai, China. Shortened version published in *Reducing Poverty on a Global Scale*, Edited by Blanca Moreno-Dodson, World Bank, 2005, CD-ROM, pp. 255-258.
- Heckman, James and Petra Todd, 1995, "Adapting Propensity Score Matching and Selection Model to Choice-Based Samples," Working Paper. Department of Economics, University of Chicago.
- 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.
- Hirano, Keisuke and Guido Imbens, 2002, "Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization," *Health Services and Outcomes Research Methodology*, 2:259-278.
- Hirano, Keisuke, Guido Imbens, and Geert Ridder, 2003, "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica* Vol. 71(4): 1161-1189.
- Imbens, Guido, 2004, "Nonparametric Estimation of Average Treatment Effects under

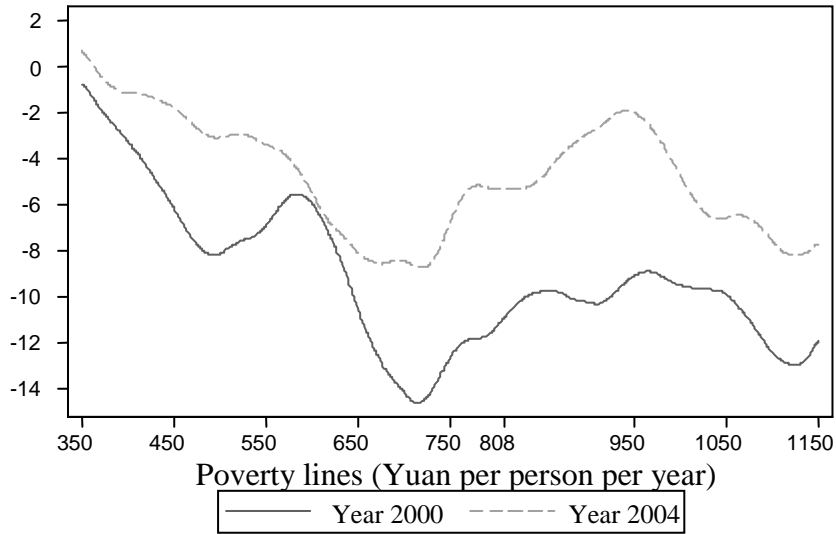
- Exogeneity: A Review,” *Review of Economics and Statistics*, 86(1): 4-29.
- Jian, Tianlun, Jeffrey Sachs and Andrew Warner, 1996, “Trends in Regional Inequality in China,” *China Economic Review*, 7(1), 1-21.
- Jalan, Jyotsna and Martin Ravallion, 1998, “Are There Dynamic Gains from a Poor-Area Development Program?” *Journal of Public Economics*, 67, 65-85.
- _____ and _____, 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.
- Kanbur, Ravi and Xiaobo Zhang, 1999. “Which Regional Inequality? The Evolution of Rural-Urban and Inland-Coastal Inequality in China from 1983 to 1995,” *Journal of Comparative Economics* 27: 686-701.
- Khan, Azizur Rahman and Carl Riskin, 1998, “Income Inequality in China: Composition, Distribution and Growth of Household Income, 1988 to 1995,” *China Quarterly*, 154, 221-253.
- Knight, John and Lina Song, 1993, “The Spatial Contribution to Income Inequality in Rural China,” *Cambridge Journal of Economics* 17: 195-213.
- 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.
- Miguel, Edward and Michael Kremer, 2004, “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,” *Econometrica* 72(1): 159-217.
- National Bureau of Statistics (NBS), 2000, *The Poverty Monitoring Report of Rural China 2000*, Beijing: China Statistics Press.
- Pack, Howard and Janet Pack, 1990, “Is Foreign Aid Fungible? The Case of Indonesia,” *Economic Journal* 100, 188-194.
- Park, Albert, Sangui Wang and Guobao Wu, 2002, “Regional Poverty Targeting in China,” *Journal of Public Economics*, 86(1), 123-153.
- Ravallion, Martin and Shaohua Chen, 2005, “Hidden Impact: Household Saving in Response to a Poor-Area Development Project,” *Journal of Public Economics*, 89: 2183-2204.

- Ravallion, Martin and Jyotsna Jalan, 1999, "China's Lagging Poor Areas," *American Economic Review, Papers and Proceedings* 89(2): 301-305.
- Roemer. John and Joaquim Silvestre, 2002, "The Flypaper Effect is Not an Anomaly," *Journal of Public Economic Theory* 4(1), 1-17.
- 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.
- _____ and _____, 1985, "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *American Statistician*, 39(1): 33-38.
- Rubin, Donald B., 1980, "Discussion of the Paper by D. Basu," *Journal of the American Statistical Association* 75: 591-593.
- Smith, A. Jeffrey and Petra E. Todd, 2005, "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 125: 305-353.
- van de walle, Dominique and Ren Mu, 2007 , "Fungibility and the Flypaper Effect of Project Aid: Micro-Evidence for Vietnam," *Journal of Development Economics* (in press)
- World Bank, 1992, *China: Strategies for Reducing Poverty*, Washington DC: World Bank.
- _____, 1997, *China 2020: Sharing Rising Incomes*, Washington DC: World Bank.
- _____, 2003, *Implementation Completion Report on the Southwest Poverty Reduction Project*, Report No. 26132, Washington DC: World Bank.
- Zampelli, Ernest M.. 1986, "Resource Fungibility, the Flypaper Effect and the Expenditure Impact of Grants-in-Aid," *Review of Economics and Statistics* 68(1), 33-40.

Figure 1: Impacts on poverty (trimmed sample)

(a) Income poverty

DD poverty impact (% points)



(b) Consumption poverty

DD poverty impact (% points)

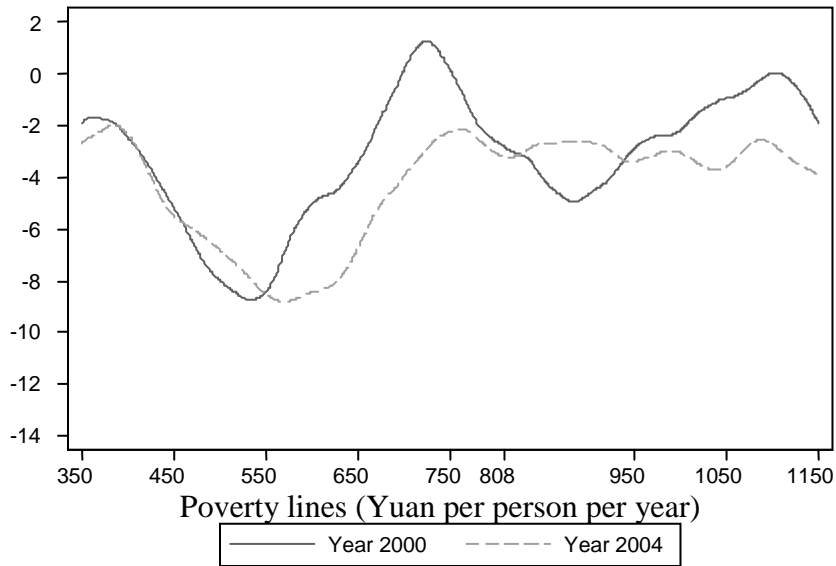


Table 1: Summary statistics on outcome indicators

	1996		2000		2004/05	
	SWP villages	Non-SWP villages	SWP villages	Non-SWP villages	SWP villages	Non-SWP villages
Mean income	996.061 (715.402)	1158.319 (604.914)	1263.412 (910.036)	1223.698 (669.843)	1390.766 (902.030)	1518.963 (930.867)
Mean consumption	843.559 (469.555)	945.201 (445.787)	943.550 (566.183)	1023.352 (698.428)	1130.588 (794.167)	1211.973 (795.499)
Income poverty rate						
Poverty line=600 yuan	0.222 (0.416)	0.127 (0.332)	0.138 (0.345)	0.112 (0.316)	0.123 (0.329)	0.095 (0.294)
808 yuan	0.453 (0.498)	0.306 (0.461)	0.290 (0.454)	0.262 (0.440)	0.242 (0.429)	0.182 (0.386)
1000 yuan	0.614 (0.487)	0.456 (0.498)	0.449 (0.497)	0.415 (0.493)	0.369 (0.483)	0.290 (0.454)
Consumption poverty rate						
Poverty line=600 yuan	0.290 (0.454)	0.183 (0.387)	0.276 (0.447)	0.219 (0.414)	0.179 (0.384)	0.135 (0.342)
808 yuan	0.576 (0.494)	0.454 (0.498)	0.509 (0.500)	0.441 (0.497)	0.385 (0.487)	0.317 (0.465)
1000 yuan	0.757 (0.429)	0.648 (0.478)	0.675 (0.468)	0.627 (0.484)	0.537 (0.499)	0.468 (0.499)

Notes: Standard deviations are in parenthesis. Income, consumption and poverty measures are weighted by household size. There are 112 project villages and 86 comparison villages. The mean of income/expenditure is Yuan per capita per year at 1995 prices.

Table 2: Probit regression of village participation in the SWP using baseline covariates

	Coeff.	z-value
Village on the plains	<i>Reference category</i>	
Hills	4.876	(4.02)
Mountainous	2.771	(3.05)
Whether village has electricity	-0.672	(-1.82)
...telephones	-0.070	(-0.2)
...road passing through it	0.215	(0.59)
...radio transmitters	0.352	(1.09)
Whether village can receive TV transmission	0.237	(0.82)
Located <5km from the nearest market	0.028	(0.05)
...5-10 km from the nearest market	-0.494	(-0.94)
...10-20 km from the nearest market	0.740	(0.95)
...>20km	<i>Reference category</i>	
# of days in a cycle during which the market assembles	-0.115	(-0.76)
County town within 5 km	<i>Reference category</i>	
Distance from village to county town is 5-10km	1.373	(1.95)
...10-20km	-0.530	(-0.85)
>20km	-0.448	(-0.83)
Township=village	<i>Reference category</i>	
Distance from village to township is within 5km	0.137	(0.19)
...5-10km	0.229	(0.34)
...10-20km	-1.628	(-2.55)
Main mode of transportation used by the villager: bicycle	-0.296	(-0.4)
...bus	-0.305	(-0.9)
...other automobile	0.913	(1.71)
...walking	<i>Reference category</i>	
Nearest train station is within 5 km	-0.586	(-0.62)
...5-10km	0.999	(1.39)
...10-20km	1.111	(1.52)
>20km	<i>Reference category</i>	
Nearest bus station is within 5 km	0.021	(0.07)
...5-10km	0.265	(0.64)
...10-20km	0.469	(1)
...>20km	<i>Reference category</i>	
Whether village has a day-care center	0.724	(1.38)
Elementary school is in village	<i>Reference category</i>	
Nearest elementary school is within 5km	0.055	(0.16)
...5-10km	0.737	(1.6)
Middle school is in village	<i>Reference category</i>	
Nearest middle school is within 5 km	1.026	(2.09)
...5-10km	0.142	(0.21)
...10-20km	1.551	(1.63)
...>20km	0.882	(1.13)
Medical clinic in village	<i>Reference category</i>	
Nearest medical clinic is within 5 km	-1.026	(-2.79)
...5-10km	-0.420	(-1.11)
...10-20km	-0.820	(-1.24)
...>20km	-0.997	(-1.46)

Total population of the village	0.000	(1.99)
Irrigated land (mu)	-0.001	(-2.8)
Forest land (mu)	0.000	(-0.87)
# of people work in TVE over # of labor	0.139	(1.99)
Whether village has TVE	-0.798	(-1.35)
Output of grain per capita (kg/person)	0.001	(1.52)
Net income per capita	0.020	(2)
Net income per capita squared	0.000	(-1.97)
Net income per capita cube	0.000	(1.66)
(End of year) # of pigs per person	0.972	(1.75)
(End of year) # of cows per person	0.840	(0.7)
(End of year) # of sheep, goat per person	0.531	(1.12)
(End of year) # of poultry per person	0.419	(2.54)
(End of year) # of hone been per person	-5.412	(-2.27)
Workforce per capita	0.036	(1.4)
Average household size	-0.042	(-1)
Share of workforce female	-0.082	(-1.68)
Cultivated land per capita (mu)	1.438	(3.19)
Grassland per capita (mu)	1.887	(1.43)
Village mean of consumption (log)	-0.493	(0.198)
Village mean of school enrollment (age 6-14)	-2.029	(-2.84)
Guangxi	1.394	(2.73)
Guizhou	0.659	(0.92)
Yunnan	<i>Reference category</i>	
Intercept	-2.522	(-0.88)
Pseudo-R2	0.360	

Note: The village is the unit of observation (n=200) and all explanatory variables are pre-intervention (1995). Standard errors are adjusted for cluster at county level.

Table 3: Impact of SWP on household income and consumption using propensity-score weighting or matching

	1996 mean in SWP villages	Gain in SWP project	Gain in non- SWP villages	Simple		PS		Kernel	
				DD	t-ratio	weighted DD	t-ratio	matched DD	t-ratio
Trimmed sample									
2000	981.906	196.322	66.012	130.31	1.826	182.655	2.541	169.150	2.392
income	841.729	67.092	70.480	-3.388	-0.067	-17.662	-0.313	-45.762	-0.751
consumption (C)	140.223	129.185	-4.525	133.711	2.107	200.333	2.723	214.93	2.685
saving (S)	981.906	432.325	387.399	44.926	0.500	42.975	0.455	42.234	0.549
2004/05	841.729	345.947	287.687	58.26	0.870	58.535	0.786	18.312	0.223
income	140.223	86.333	99.655	-13.322	-0.159	-15.544	-0.18	23.941	0.289
consumption									
saving									
2000	6.747	0.18	0.051	0.128	2.046	0.161	2.395	0.133	2.251
log income	6.629	0.058	0.019	0.040	0.755	0.031	0.537	0.001	0.003
log consumption	0.117	0.120	0.031	0.089	1.79	0.131	2.467	0.133	2.617
log(1+S/C)	6.747	0.345	0.264	0.081	1.171	0.062	0.823	0.038	0.522
2004/05	6.629	0.299	0.210	0.090	1.707	0.067	1.130	0.025	0.474
log income	0.117	0.046	0.055	-0.009	-0.148	-0.005	-0.078	0.014	0.263
log consumption									
log(1+S/C)									
Total sample									
2000	989.45	273.962	65.379	208.583	3.346	213.605	3.287	192.731	2.985
income	843.559	99.991	78.151	21.84	0.510	-151.054	-1.180	-189.569	-1.427
consumption (C)	145.934	173.928	-12.828	186.755	3.141	364.696	3.371	382.342	3.612
saving (S)	989.45	401.316	360.644	40.673	0.537	-47.159	-0.423	-45.246	-0.344
2004/05	843.559	287.029	266.772	20.258	0.371	36.752	0.633	25.893	0.439
income	145.934	114.244	93.816	20.427	0.303	-83.874	-0.705	-71.097	-0.52
consumption									
saving									
2000	6.752	0.230	0.050	0.180	3.448	0.180	3.337	0.160	3.221
log income	6.631	0.087	0.028	0.059	1.374	-0.046	-0.577	-0.081	-0.945
log consumption	0.121	0.143	0.021	0.122	2.727	0.227	3.568	0.241	3.615
log(1+S/C)	6.752	0.310	0.231	0.078	1.314	-0.005	-0.064	-0.01	-0.112
2004/05	6.631	0.223	0.188	0.035	0.682	0.021	0.388	0.007	0.115
log income	0.121	0.087	0.043	0.044	0.915	-0.026	-0.307	-0.017	-0.185
log consumption									
log(1+S/C)									

Notes: All the calculations are weighted by household size. T-ratio of kernel matching is obtained from bootstrapping (100 repetitions). Standard errors of weighted D-D estimations are robust to heteroskedasticity and serial correlation of households within each village. In the total sample, there are 112 project villages and 86 comparison villages. In the trimmed sample, there are 71 project villages and 66 comparison villages.

Table 4: Impacts on income from animal husbandry

	Revenue or costs from animal husbandry (AH)	1996 mean in SW/PR	Gain in		Gain in non-		PS weighted		Kernel matched	
			SW/PR project	SW/PR	SW/PR	DD	D-D	D-D	D-D	D-D
2000	total revenue	326.983	107.6	33.894	73.706	2.302	100.03	2.703	118.883	2.498
	total cost of production	190.901	-15.516	1.623	-17.139	-0.801	-17.229	-0.779	-17.271	-0.717
	net income from AH	136.082	123.117	32.271	90.845	2.924	117.26	3.373	136.154	3.551
	cash income (net)	142.204	12.411	1.587	10.824	0.663	14.684	0.858	-2.356	-0.685
	in-kind income (net)	-6.123	110.705	30.684	80.021	2.79	102.575	3.099	125.853	2.895
2004	total revenue	326.983	196.889	225.753	-28.864	-0.507	-1.357	-0.023	12.356	0.246
	total cost of production	190.901	80.772	121.847	-41.075	-1.196	-36.175	-1.031	-42.896	-1.285
	net income	136.082	116.118	103.906	12.212	0.282	34.818	0.785	55.252	1.344
	cash income (net)	142.204	103.745	150.839	-47.093	-1.025	-30.646	-0.578	3.219	0.641
	in-kind income (net)	-6.123	12.372	-46.932	59.305	2.033	65.464	1.805	74.179	1.705

Notes: All the calculations are weighted by household size. T-ratio of kernel matching is obtained from bootstrapping (100 repetitions). Standard errors of weighted D-D estimations are robust to heteroskedasticity and serial correlation of households within each village. The trimmed sample is used, for which there are 71 project villages and 66 comparison villages.

Table 5: Propensity score weighted estimates of impacts on poverty

Poverty line	Poverty incidence (1996) in project villages	(1) Change in H in project villages	(2) Change in H in comparison villages	(1)-(2) Double difference	t-ratio
(a) Income poverty					
2000					
500	14.584	-6.747	0.957	-7.704	-2.138
600	22.762	-7.331	-1.672	-5.659	-1.247
700	35.116	-13.093	1.490	-14.582	-2.824
808	46.697	-15.713	-4.599	-11.114	-1.515
900	55.047	-15.193	-4.771	-10.422	-1.581
1000	62.025	-12.906	-3.606	-9.300	-1.395
1100	68.973	-10.802	1.642	-12.444	-2.195
1150	72.405	-9.981	2.484	-12.465	-2.256
2004/05					
500	14.584	-8.053	-5.021	-3.032	-0.809
600	22.762	-12.250	-6.779	-5.470	-0.857
700	35.116	-19.410	-11.533	-7.877	-1.046
808	46.697	-24.907	-19.276	-5.630	-0.693
900	55.047	-26.344	-22.915	-3.429	-0.444
1000	62.025	-28.097	-23.816	-4.281	-0.530
1100	68.973	-27.623	-19.537	-8.086	-1.352
1150	72.405	-28.378	-20.347	-8.031	-1.424
(b) Consumption poverty					
2000					
500	18.673	-2.695	6.111	-8.806	-1.691
600	29.053	0.078	5.298	-5.221	-0.841
700	40.749	1.140	1.088	0.052	0.006
808	57.392	-5.266	-1.902	-3.364	-0.386
900	67.000	-5.761	-0.715	-5.046	-0.734
1000	75.665	-6.102	-4.570	-1.532	-0.248
1100	80.898	-4.987	-5.782	0.796	0.164
1150	83.586	-5.184	-3.569	-1.615	-0.347
2004/05					
500	18.673	-11.537	-4.081	-7.456	-1.500
600	29.053	-16.661	-7.918	-8.743	-1.536
700	40.749	-18.226	-13.352	-4.874	-0.747
808	57.392	-23.241	-19.095	-4.146	-0.584
900	67.000	-24.439	-22.567	-1.872	-0.267
1000	75.665	-25.936	-23.121	-2.815	-0.520
1100	80.898	-24.192	-21.455	-2.737	-0.511
1150	83.586	-22.006	-17.962	-4.044	-0.789

Notes: All the calculations are weighted by household size. Standard errors are robust to heteroskedasticity and serial correlation of households within each village. The trimmed sample is used with 71 project villages and 66 comparison villages.

Table 6: Estimated impacts stratified by initial income and education

	Lower education group		Higher education group		Weighted triple difference (1)-(2)	t-ratio		
	1996 mean in SW villages	Weighted <i>DD</i> for lower education group (1)	1996 mean in SW villages	Weighted <i>DD</i> for higher education group (2)				
2000	income	643.538	81.686	645.831	207.958	2.525	-126.271	-1.491
	consumption	664.573	-43.809	674.167	55.069	0.604	-98.878	-1.246
	saving	-20.989	125.518	-28.290	152.875	1.460	-27.357	-0.291
	productive assets	413.096	-58.508	311.452	86.098	1.424	-144.606	-1.737
	housing value	501.121	-39.476	611.993	173.552	0.959	-213.028	-0.947
2004/05	income	643.538	43.687	645.831	197.933	2.026	-154.246	-1.079
	consumption	664.573	97.623	674.167	219.517	2.370	-121.894	-1.105
	saving	-20.989	-53.914	-28.290	-21.598	-0.247	-32.316	-0.277
	productive assets	413.096	80.478	311.452	134.206	1.985	-53.728	-0.446
	housing value	501.121	216.285	611.993	815.739	2.481	-599.454	-2.022
Number of households			312 (173+139)		299 (169+130)			
2000	income	1465.163	305.638	1476.474	174.261	1.194	131.376	0.587
	consumption	1061.494	-237.040	1170.625	-8.934	-0.071	-228.105	-0.979
	saving	403.747	542.693	305.881	183.215	1.375	359.479	1.086
	productive assets	600.292	-160.040	609.010	-34.391	-0.374	-125.649	-1.054
	housing value	842.872	343.787	1109.570	60.008	0.303	283.780	1.118
2004/05	income	1465.163	-27.644	1476.474	-54.414	-0.348	26.770	0.107
	consumption	1061.494	-24.752	1170.625	-136.847	-0.913	112.095	0.637
	saving	403.747	-2.876	305.881	82.452	0.493	-85.328	-0.389
	productive assets	600.292	120.572	609.010	-201.500	-1.258	322.072	1.816
	housing value	842.872	432.315	1109.570	-697.603	-0.910	1129.918	1.331
Number of households			204 (97+107)		363 (170+193)			

Notes: The numbers parentheses are the number of observations in SWP villages and non-SWP villages respectively. Estimation is made on a balanced panel of 1178 households on the trimmed sample. Lower education is defined as household head education level being lower than junior high school (illiterate or primary school). Higher education is defined as household head education level being at least junior high school. Standard errors of weighted D-D estimations are robust to heteroskedasticity and serial correlation of households within each village. Balanced panel in trimmed sample is used with 67 project villages and 62 comparison villages.

Table 7: Impacts on self-assessed satisfaction with life compared to 10 years ago

	Single difference on total sample			PS weighted and trimmed sample		
	Mean in project villages	Diff. (project-comparison)	t-ratio	Mean in project villages	Diff. (project-comparison)	t-ratio
Overall standard of living of h'hold	0.357	0.001	0.018	0.343	0.108	1.635
Income	0.328	-0.005	-0.094	0.324	0.026	0.326
Food	0.377	0.017	0.334	0.356	0.073	1.050
Clothing	0.363	0.028	0.55	0.345	0.094	1.364
Housing	0.313	-0.045	-0.952	0.292	0.006	0.089
Electricity	0.464	0.029	0.515	0.426	0.083	1.020
Hygiene	0.184	-0.058	-1.457	0.186	0.005	0.089
Household appliances	0.275	-0.013	-0.298	0.244	0.010	0.152
Asset accumulation	0.173	-0.009	-0.228	0.151	-0.079	-0.974
Agriculture skill	0.101	-0.026	-0.935	0.087	0.009	0.285
Non-agricultural skill	0.152	-0.057	-1.412	0.146	0.016	0.341
Marketing of agriculture products	0.219	-0.028	-0.627	0.239	0.067	1.132
Credit availability	0.190	0.011	0.251	0.190	0.035	0.589
Affordability of primary/mid. school	0.22	0.007	0.163	0.209	0.067	1.265
Health	0.302	-0.035	-0.71	0.285	0.092	1.550
School infrastructure	0.392	-0.053	-0.928	0.382	0.039	0.497
School quality	0.306	-0.024	-0.462	0.304	0.047	0.682
Health infrastructure	0.240	-0.059	-1.217	0.219	0.006	0.090
Road conditions	0.377	0.009	0.148	0.376	0.023	0.261
Transportation	0.426	-0.050	-0.846	0.411	-0.061	-0.686
Environment	0.132	-0.030	-0.875	0.129	0.005	0.114
Ecology	0.145	-0.072	-1.852	0.114	-0.024	-0.559
Safety	0.226	0.008	0.156	0.212	0.045	0.687
Knowledge of village affairs	0.170	-0.010	-0.227	0.161	0.048	0.992
Participation in decision-making	0.174	-0.017	-0.413	0.157	0.026	0.536
Democracy	0.232	0.015	0.321	0.216	0.075	1.353
Overall village standard of living	0.345	0.034	0.626	0.325	0.116	1.761
Service to village by county govt.	0.200	0.017	0.369	0.157	0.009	0.133
Service to h'hold by county govt.	0.180	0.034	0.783	0.167	0.041	0.637

Notes: Comparison (with 10 years ago) is based on a scale of 10, 1 being "much worse off, and 10 being "totally improved". We redefine those outcomes as dummy variables, equal to 1 if the answer is "obviously improved (8)" or above, 0 if "improved (7)" or below. All the respondents were 28 years or older at the time of interview. Single double difference estimation is made on the total sample of 104 project villages and 79 comparison villages. Weighted double difference estimation is made on the trimmed sample of 66 project villages and 60 comparison villages.

Table 8: Regressions for perceived change in overall standard of living

	(1)		(2)		Difference	
	Coefficient	t-ratio	coefficient	t-ratio	(1)-(2)	t-ratio
Intercept	1.987	1.281	1.905	0.798	0.081	0.029
Change of log consumption between 1996 and 2004/05 (β_0)	0.365	2.522	0.315	1.811	0.05	0.222
Log consumption in 1996 (β_1)	0.321	1.668	0.731	3.025	-0.41	-1.333
Gender of respondent	0.217	0.936	0.400	1.811	-0.183	-0.574
age of respondent	0.083	2.166	-0.001	-0.017	0.084	1.331
age ²	-0.001	-1.685	0.000	-0.189	-0.001	-0.837
R ²	0.036		0.043			
Test $H_0 : \beta_0 = \beta_1$ (p-value)	0.820		0.094			

Notes: Estimation is made on balanced panel with 913 households in 100 project villages and 681 households in 75 non-project villages. Standard errors are robust to heteroskedasticity and serial correlation of households within each village.

Table 9: Testing for displacement of non-SWP development projects in SWP villages

	Mean in SWP villages	Mean in non-SWP villages	Difference	t-ratio	PS weighted diff.	t-ratio	Kernel matched diff.	t-ratio
Farming								
Number of projects	0.79	2.11	-1.32	-2.45	-1.68	-2.09	-2.04	-2.03
Number of households	147.63	399.44	-251.80	-2.48	-182.99	-2.43	-205.03	-2.05
Animal husbandry								
Number of projects	1.51	3.03	-1.52	-2.08	-2.21	-1.98	-2.38	-1.78
Number of households	135.09	324.87	-189.78	-1.17	-94.99	-1.18	-62.14	-1.00
Forestry								
Number of projects	0.54	1.34	-0.79	-2.50	-1.50	-1.84	-2.28	-1.66
Number of households	131.63	296.63	-165.00	-1.41	-120.06	-1.97	-117.65	-3.15
Infrastructure								
Terracing	0.12	0.65	-0.53	-2.08	-0.94	-1.58	-1.35	-1.46
Drinking water	0.31	0.90	-0.59	-3.04	-0.86	-2.58	-1.04	-2.54
Irrigation	0.24	0.60	-0.36	-1.80	-0.30	-1.42	-0.27	-1.31
Electricity	0.28	0.58	-0.30	-2.21	-0.49	-2.01	-0.61	-1.52
Roads	0.19	0.39	-0.20	-1.89	-0.24	-1.39	-0.25	-1.53
Student subsidies: No.	0.82	2.35	-1.53	-3.03	-1.74	-2.79	-1.75	-2.83
New schools: No.	0.35	0.79	-0.44	-2.10	-0.55	-1.96	-0.84	-2.06
Teacher training: No.	0.07	0.37	-0.30	-1.87	-0.39	-2.25	-0.37	-2.26
Health insurance	0.16	0.31	-0.14	-2.05	-0.12	-1.26	-0.05	-0.69
New clinic	0.06	0.24	-0.18	-3.84	-0.12	-1.85	-0.09	-1.46
Doctor training	0.07	0.26	-0.18	-1.62	-0.18	-1.56	-0.12	-1.23
Total no. projects	6.07	14.81	-8.73	-3.25	-11.72	-2.45	-13.68	-2.14
Total no. households	415.38	1026.10	-610.71	-1.74	-399.02	-2.22	-386.20	-2.98

Notes: Trimmed sample, treating "no response" as "no project". T-ratio of kernel matching is obtained from bootstrapping (100 repetitions). Standard errors of D-D and weighted D-D estimations are robust to heteroskedasticity and serial correlation of villages within each county. Trimmed sample is used with 71 project villages and 65 comparison villages...

Appendix: Balancing tests for village characteristics and household outcomes with and without weighting and trimming

		Difference in standardized means											
		Standardized means*				PS kernel- matched for total sample				PS kernel- matched for trimmed sample			
		Un-weighted		PS weighted for total sample		PS-weighted for trimmed sample		PS kernel- matched for total sample		PS-weighted for trimmed sample		PS kernel- matched for trimmed sample	
		mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
		SWP villages		Non-SWP villages									
Village characteristics (1995)													
Total population	0.009	-0.012	0.021	0.143	0.013	0.137	0.180	0.134	0.076	0.186	0.187	0.115	0.187
Electricity	-0.151	0.196	-0.347	0.141	-0.229	0.138	-0.028	0.157	0.104	0.164	0.147	-0.268	0.147
Phone	0.053	-0.069	0.122	0.143	0.109	0.141	0.373	0.132	0.155	0.168	0.171	0.072	0.171
Road	-0.061	0.079	-0.139	0.143	-0.094	0.141	0.090	0.155	0.211	0.164	0.134	-0.130	0.134
Radio	0.044	-0.058	0.102	0.143	0.075	0.135	0.241	0.126	0.271	0.155	0.170	0.193	0.170
TV	-0.084	0.109	-0.193	0.142	-0.131	0.143	0.056	0.152	0.117	0.175	0.163	-0.136	0.163
Nearest market <5km	-0.036	0.047	-0.083	0.143	-0.068	0.148	0.100	0.152	0.078	0.187	0.206	0.417	0.206
Elementary school in village	-0.009	0.011	-0.02	0.143	-0.031	0.143	0.102	0.129	-0.075	0.182	0.18	-0.005	0.18
Clinic in village	0.021	-0.028	0.049	0.143	0.051	0.141	0.258	0.129	0.043	0.170	0.172	0.073	0.172
Net income per capita	-0.162	0.211	-0.373	0.141	-0.241	0.142	0.133	0.124	0.073	0.164	0.171	0.094	0.171
Cultivated land per capita	0.134	-0.173	0.307	0.141	0.238	0.135	0.251	0.122	0.299	0.151	0.144	-0.159	0.144
Household outcomes (1996)													
Consumption per capita	-0.156	0.203	-0.36	0.141	-0.217	0.190	-0.195	0.206	-0.069	0.181	0.181	-0.007	0.181
Income per capita	-0.168	0.219	-0.39	0.141	-0.23	0.139	-0.238	0.137	-0.182	0.181	0.185	-0.153	0.185
Headcount poverty index													
___ 600 yuan (income)	0.175	-0.227	0.402	0.141	0.384	0.169	0.442	0.18	0.248	0.201	0.234	0.244	0.234
___ 808 yuan (income)	0.196	-0.256	0.452	0.14	0.345	0.172	0.375	0.196	0.194	0.192	0.194	0.108	0.194
___ 1000 yuan (income)	0.212	-0.277	0.489	0.139	0.41	0.202	0.457	0.256	0.165	0.201	0.216	0.062	0.216
___ 600 yuan (consumption)	0.161	-0.21	0.371	0.141	0.455	0.186	0.534	0.18	0.259	0.198	0.214	0.325	0.214
___ 808 yuan (consumption)	0.155	-0.202	0.357	0.141	0.253	0.194	0.242	0.213	0.07	0.194	0.218	0.008	0.218
___ 1000 yuan (consumption)	0.171	-0.222	0.393	0.141	0.319	0.268	0.291	0.308	0.006	0.182	0.203	-0.130	0.203

Notes: * (sub-group mean *minus* mean for full sample)/standard deviation for full sample. In total sample, there are 112 project villages' 86 comparison villages. In the trimmed sample, there are 71 project villages and 66 comparison villages. Household income, consumption and poverty measures are weighted by household size. The Addendum provides further balancing tests.