the review of income and wealth

Review of Income and Wealth
Series 64, Number 1, March 2018
DOI: 10.1111/roiw.12273

CAN POVERTY BE ALLEVIATED IN CHINA?

BY CONG QIN

School of Finance, Renmin University of China

AND

TERENCE TAI LEUNG CHONG*

Department of Economics and Lau Chor Tak Institute of Global Economics and Finance, the Chinese University of Hong Kong, Department of International Economics and Trade, Nanjing University

The 2000s witnessed the third poverty alleviation wave in China. Compared with its predecessors, the third wave distinguished itself by new interventions and redefined standards for the National Poor Counties. This paper evaluates the effectiveness of the new program using a data set consisting of 1,411 of China's western and central counties from 2000 to 2010. It combines the propensity score matching method with the difference-in-differences approach, which helps to avoid selection bias and track the policy impact on variables of interest at each time point. It is found that the non-western local governments tended to manipulate data on income and output growth to maintain the special transfer payments disbursed exclusively to the National Poor Counties. It is also shown that the program failed to improve the infrastructure and sanitary conditions in general.

JEL Codes: H23, H71

Keywords: average treatment effect on treated, poverty alleviation, propensity score matching, receiver operating characteristic curve, National Poor Counties

1. Introduction

The 2000s experienced a new wave of poverty alleviation in China. Compared to its two predecessors—namely the first wave (1986–1993) of China's poverty reduction and the second wave (1994–2000), which was also called the 8-7 Plan—the third wave was characterized by a dramatic change in the standard of poverty and the formulation of new policies against rural poverty. The world's largest regional poverty targeting program, which resulted in phenomenal economic growth in China, proved to be a huge success in poverty alleviation in the developing world. This led to a drastic reduction in the poor population from about 125 million in 1986, the first year of China's war against rural poverty, to around 32 million in 2000, the final year of the 8-7 Plan. In the

We owe much to Shi Li, Hongliang Zhang, Kai Yuen Tsui, and Julan Du for their helpful comments. We are also very much indebted to Yingshi Chen and Mandy Cheung for their assistance in the research. Any remaining errors are ours.

*Correspondence to: Terence Tai-Leung Chong, Department of Economics, the Chinese University of Hong Kong, Shatin, N.T., Hong Kong (chong2064@cuhk.edu.hk). Web page: http://www.cuhk.edu.hk/eco/staff/tlchong/tlchong3.htm.

TABLE 1
THE DISTRIBUTION OF POVERTY POPULATION IN 2000, 2005 AND 2010

	Year	2000	2005	2010
Poverty Population Share With "Low-Income Standard" (%)	All	10.2	6.8	2.8
	Eastern	2.9	1.6	0.4
	Central	8.8	6.6	2.5
	Western	20.6	13.3	6.1

Source: Poverty Monitoring Report of Rural China 2010.

third wave, the number of people living in poverty dropped to about 26.88 million in 2010.[1]

Although the first two waves were viewed as exemplars in China's war on poverty, there existed two main limitations. First, political factors had affected the selection of the National Poor Counties, which was the major intervention in each wave (Park *et al*., 2002). As time went by, the effectiveness of the targeting program deteriorated, since lobbying efforts and political resistance prevented the government from taking National Poor County status away from counties that were no longer below the poverty line. Second, the distribution of the poor population also changed over time. At the beginning of the reform and opening up policy, most of the poverty-stricken population lived in contiguous areas. In 1993, the poverty-stricken population living in the National Poor Counties increased to about 72 percent, up from 50 percent in 1986, and two-thirds of this population were located in the eastern and central provinces. However, the coverage of the counties shrank to about 60 percent in 2001, and the poverty level in the eastern provinces greatly decreased due to their soaring economic growth.

In response to the newly apparent drawbacks of the old program, the Chinese government made two main changes in the third wave. To avoid political obstacles, the Chinese government adopted a more comprehensive method when selecting new National Poor Counties, which was known as the "631 index" method. Furthermore, to increase the coverage of the rural poor population, along with county-level policies, the Chinese government considers including poor villages in the counties. It is also worth noting that since almost all the eastern counties eliminated poverty in the first two waves, the Chinese government excluded those counties from the new program.[2] This rationalizes our exclusive focus on central and western counties in this study. Table 1 describes the distribution of the poverty-stricken population in the third wave.

This paper is closely associated with the previous literature on anti-poverty programs in China. The previous two poverty reduction waves are studied separately by Park *et al*. (2002) and Meng (2013). The former analysis exploits a

[1]During the third wave, China raised the poverty line and extended the coverage of the population living in poverty to benefit not only the previous rural poor but also the low-income population. The number here follows the new statistical criteria.

[2]Given the relatively severe poverty situation in Hebei and Hainan provinces, which are located in the eastern region, the Chinese central government exclusively remains the identity of those poverty counties.

four-period income growth model, estimates the impact of poverty investments, and uses the propensity score matching method as a robustness check. The main finding is that the National Poor Counties program increased the rural income per capita by 2.28 percent per year in the first wave. Taking advantage of the policy change in 1994, the latter paper employs a regression discontinuity approach, and uses the propensity score as an instrumental variable to estimate the impact of the second wave. The author finds that the second wave brought about a 38 percent increase in rural income for the newly designated National Poor Counties.

Though the importance of the propensity score is limited in these two papers, the usefulness of the newly introduced method is revealed. Thus, we believe that when combined with various matching methods, the propensity score can play a more important role in program evaluation. Moreover, the existing literature mainly focuses on the impact of a program on income increment, but ignores other changes caused by the policy. Our paper evaluates sanitary and infrastructure conditions, since China's central government for the first time explicitly included targets to improve health in the program and reinforced the importance of improving infrastructure conditions in poor counties.

It has always been a great challenge to estimate the causal effects of a program when a randomized controlled trial is not attainable. However, when public interventions are based on manipulable variables, causation can be found (Rubin, 1986). The manipulation may wrongly attribute the effect caused by the pretreatment characteristic difference between the treated and the control cohorts to the assigned treatment if the two cohorts are unbalanced. To reduce the selection bias resulting from the problem of imbalance, we adopt the propensity score matching method to transform the quasi-experimental studies into randomized experimental ones.

Employing a panel data set consisting of 1,411 central and western Chinese counties over 11 years, we estimate reliable propensity scores by comparing five differentiated matching methods, and evaluate the impact of the National Poor Counties program in China's third poverty alleviation wave by using the DID approach. Our main findings include two parts. First, we identify two possible mechanisms through which the program functions. Based on our estimation, the program generally had no or negative effects on the whole sample. Moreover, rather than boosting the economy, local governments outside of the western regions tended to manipulate data on income and output growth so as to maintain special transfer payments disbursed exclusively to National Poor Counties. Second, we find that the program failed to influence the western and central counties as a whole and achieve its goals, which included improving infrastructure and sanitary conditions. However, for certain subsamples, the program was found to have had some positive effects.

The remainder of this paper proceeds as follows. Section 2 comprehensively reviews the history of China's poverty alleviation waves, including the successes achieved in each wave. Section 3 validates and describes the selected variables to estimate the propensity scores. Section 4 discusses the empirical strategies. Results are reported in Section 5. Finally, Section 6 presents the conclusion.

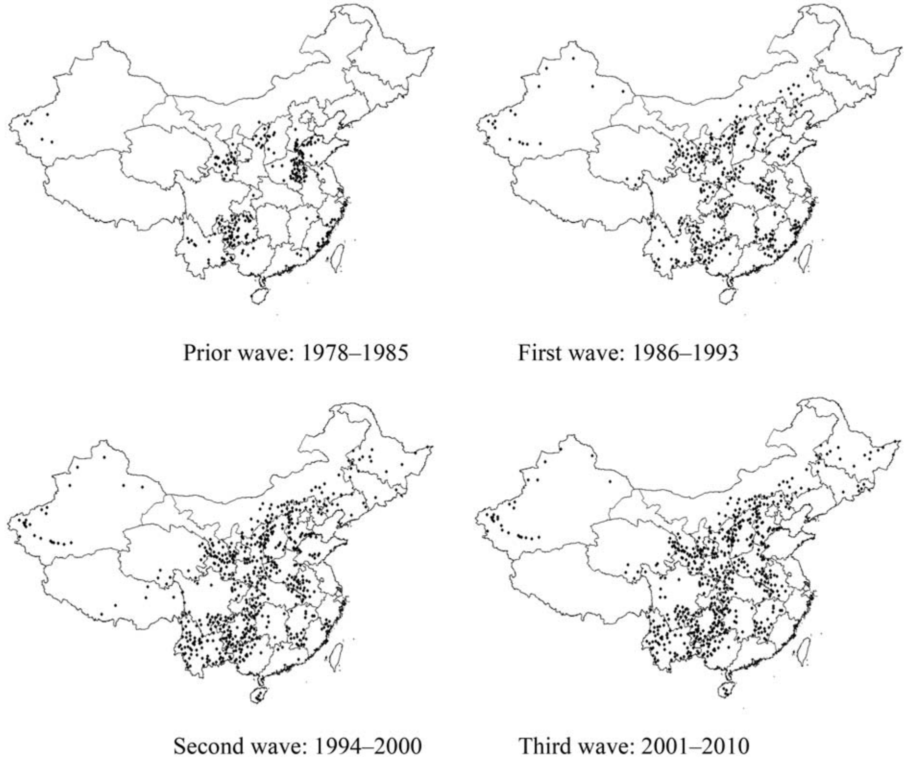

Figure 1. The Distribution of National Poor Counties

Note: Since the National Poor Counties program was not yet established in the prior wave, we use "the poor counties list 1977–1979" instead.

2. Background

2.1. *Poverty in China*

There is no doubt that poverty is mainly a rural phenomenon in China (World Bank, 2001; Meng, 2013). This seems to be an inevitable result of the existence of the dual economic structure between China's urban and rural areas. For a long time, the Chinese government had exclusively provided urban citizens with various social welfare services, to which rural residents had no access. The rigorous household registration system further exacerbated the welfare disparity between urban and rural households. Soon after reform and opening up, a difference in incidence of rural poverty began to exist between opulent coastal and less developed inland areas. Since then, China's poverty has switched from a country-wide phenomenon to one that is concentrated in its western and central regions. Moreover, residents in former revolutionary areas, minority autonomous areas, and certain remote areas are more prone to poverty than the rest of the rural population.

Three regional poverty alleviation programs have been identified since 1986: the first wave was implemented from 1986 to 1993, the second wave from 1994 to

2000, and the third wave from 2001 to 2010. Because of its huge success in poverty reduction, we include the period from 1978 to 1985 as the precursor wave as well, even though it lacked any explicit poverty reduction policy. Figure 1 shows the geographic distribution of the National Poor Counties in each wave. Points shaded in black stand for the National Poor Counties. We can see that as time went by, the government spent more efforts on poverty reduction in inland areas.

2.2. *The Prior Wave (1978–1985)*

Despite the absence of explicit poverty-targeting programs and coordination at the central government level prior to 1986, poverty alleviation has been a significant priority for Chinese leaders since reform and opening up. Given that more than 90 percent of the population living in poverty resided in rural areas, poverty in China was mainly a rural phenomenon. According to a report by China's National Bureau of Statistics,[3] about 250 million Chinese people, accounting for 30.7 percent of the total rural population, were identified as poor. Members of this population had an annual income below 100 yuan. Based on a variety of surveys, three main causes of widespread poverty during this period proved to be systemic. The rural land system, which restricted rural productivity due to its lack of incentives, was identified as a source of rural poverty. In addition, to accumulate funds for industry, the Chinese market system adopted a system of unified purchase and sale, and utilized the scissors gap between the prices of agricultural products and industrial products, which aggravated rural poverty. Meanwhile, the employment and household registration systems constrained the flow of the surplus rural labor force, further exacerbating poverty.

System changes were needed to address these issues. In order to enhance incentives for peasants, the household contract responsibility system replaced the highly collective people's commune system. From then on, rural households were able to independently cultivate their farmlands. Meanwhile, subsidized primary product sales (Park *et al*., 1994), which aimed to reduce the scissors gap between the prices of agricultural and industrial products, were announced. In 1979, the Chinese government increased the prices of 10 types of agricultural products, including grain, cotton, and oil-bearing crops. From 1978 to 1985, the total income generated by the policy was 125.74 billion yuan, accounting for a 15.5 percent increase in rural household income. Another policy was to encourage the growth of township enterprises. Starting in 1983, township enterprises began to flourish. Within three years, the number of township enterprises increased from about 1.34 million in 1983 to 12.22 million in 1985, with an increase in total output value from 101.68 billion yuan to 272.84 billion yuan. Respectively, the above three reforms enhanced land productivity, increased agricultural income for farmers, and opened doors for farmers to embark on non-agricultural tasks. In the meantime, a series of policies was implemented to boost the rural economy, relieve rural poverty, and ameliorate the industrial structure.

The achievement of this wave was outstanding. By the end of 1985, the population below the poverty line had reduced from 250 million to 125 million, and

[3]Data source: *Poverty Monitoring Report of Rural China 2000*.

the poverty incidence decreased from 30.7 percent to 14.8 percent. Meanwhile, average annual income increased from 134 yuan to 397 yuan.[4]

2.3. *The First Wave (1986–1993)*

Benefiting from the nationwide economic growth caused by the previous wave's institutional reform, an overwhelming majority of the rural population, who had suffered in poverty due to the lack of economic opportunities, were able to take advantage of their superiority in geography and resources. This uneven development caused rural poverty to become a regional problem rather than a national phenomenon, as it had been over the past 30 years. The rural poor were mainly concentrated in the old revolutionary regions (*lao qu*), minority autonomous areas (*minzu zizhiqu*), and certain inland parts, which in total formed 18 large, contiguous areas.

To address this newly emergent regionality of poverty, the Chinese government launched the largest regionally targeted anti-poverty program in the developing world. In 1986, the State Council set up the Leading Group for Economic Development in Poor Areas (hereafter Leading Group), a specialized inter-ministerial, anti-poverty institution consisting of all ministers whose duties were associated with poverty alleviation, in order to administer and coordinate the new poverty alleviation program. As a major targeting device, the Leading Group enacted the National Poor Counties policy soon after its establishment. For national and political considerations, the Leading Group adopted a mixed set of standards to identify the National Poor Counties. The basic poverty line for selecting the National Poor Counties was a rural net income per capita of below 150 yuan in 1985. However, for counties located in old revolutionary regions or minority autonomous areas, the poverty line was raised to 200 yuan. The standard was further relaxed to 300 yuan for counties in very important revolutionary regions and minority autonomous areas in Inner Mongolia, Xinjiang, and Qinghai.[5] According to these standards, 258 counties were initially designated as National Poor Counties in 1986. In the following two years, another 70 counties were selected as National Poor Counties as well. By 1988, provinces identified an additional 370 counties as provincial poor counties. Compared to the National Poor Counties, the provincial ones usually needed to meet more rigorous standards, and received fewer benefits.

The regional targeted poverty alleviation program proved to be a huge success. The population living in poverty continued to decrease in this period, from 125 million in 1986 to 80 million in 1993. Correspondingly, the poverty incidence dropped from 14.8 percent following the last wave to 8.7 percent at the end of 1993. Meanwhile, rural net income per capita in the National Poor Counties increased from 206 yuan to 484 yuan.[6]

[4]Data source: *Poverty Monitoring Report of Rural China 2000*.

[5]In 1988, another standard was set up to include a few pastoral and semi-pastoral counties, based on the rural net income data from 1984 to 1986. Pastoral counties with an average net income per capita below 300 yuan and semi-pastoral counties with an average net income per capita below 200 yuan were identified as the new National Poor Counties.

[6]Data source: *Poverty Monitoring Report of Rural China 2000*.

2.4. *The Second Wave (1994–2000)*

Although the regional targeted program in the first stage covered a sizeable amount of the rural poor, Park *et al*. (2002) claim that the heavy political compromise during the selection process undermined the program's efficiency. Certain qualified counties had to give up their eligibility to politically favored counties that had a rural net income higher than the poverty line (World Bank, 2001). In addition, some researchers cast doubt on the validity of the poverty line used to select the National Poor Counties (Meng, 2013). In response to previous criticisms and the change in geographic distribution of the poor, the Chinese government renewed the program in 1994. Known as the 8-7 Plan, the government promised to lift the majority of the remaining 80 million rural poor from poverty by 2000 (within seven years).

The entire program was still overseen by the Leading Group, which changed the poverty line and hence the list of the National Poor Counties in 1994. The revised list initially included only 326 counties with a rural net income per capita of below 400 yuan in 1992. However, facing political pressure from the National Poor Counties selected in the last wave, the Chinese government raised the poverty line for those counties to 700 yuan. Finally, the 8-7 Plan covered 592 counties, which accounted for about 28 percent of all county-level administrative units in China. By 2000, the majority of the goals of the 8-7 Plan had been achieved. Rural net income per capita in National Poor Counties increased from 648 yuan in 1993 to 1,337 yuan in 2000. The population living below the poverty line continued to decline, from 80 million in 1993 to 32 million in 2000.[7]

2.5. *The Third Wave (2001–2010)*

The huge success in poverty alleviation of the 8-7 Plan shrank the majority of the remaining population living in poverty down to 14 large areas, most of which were located in the western and central parts of China. Meanwhile, there existed a number of isolated villages distributed in other parts. In spite of the unprecedented achievement, the second wave was still criticized for its compromise with political interference that might have led to improper selection of the National Poor Counties. Moreover, the program only targeted the absolutely poverty-stricken population, whose living conditions were lower than the international standard. In response to the previous criticisms and the existence of both concentrated and dispersed populations of the rural poor, the central government launched another anti-poverty program in 2001, aiming to relieve the remaining poverty-stricken people and to enhance infrastructure, education, and health conditions in the targeted regions.

To improve the targeting accuracy, the Leading Group renewed the list and the poverty line again in 2001. The new standard was called the "631 index," which took into consideration the poverty incidence (weighted at 60 percent), rural net income per capita (weighted at 30 percent), and annual GDP, as well as local government revenue per capita (weighted at 10 percent). The basic poverty

[7]Data sources: *Poverty Monitoring Report of Rural China 2000, Poverty Monitoring Report of Rural China 2001, and Poverty Monitoring Report of Rural China 2010*.

TABLE 2
THE RURAL POVERTY POPULATION AND THE INCIDENCE OF POVERTY

Year	Poverty Line Income Level (Yuan)	Rural Poverty Population (million)		The Incidence of Poverty (%)	
		Whole Country	National Poor Counties	Whole Country	National Poor Counties
2000	865	94.22	–	10.2	–
2001	872	90.3	–	9.8	–
2002	869	86.45	48.28	9.2	24.3
2003	882	85.17	47.09	9.1	23.7
2004	924	75.87	41.93	8.1	21
2005	944	64.32	36.11	6.8	18
2006	958	56.98	31.1	6	15.4
2007	1067	43.2	26.2	4.6	13
2008	1196	40.07	24.21	4.2	11.9
2009	1196	35.97	21.75	3.8	10.7
2010	1274	26.88	16.93	2.8	8.3

Source: Poverty Monitoring Report of Rural China 2010.

line was 1,300 yuan for rural net income per capita, 2,700 yuan for GDP per capita, and 120 yuan for government revenue per capita. However, the rural net income per capita standard rose to 2,700 yuan for counties with large minority populations and old revolutionary areas. According to the new standard, the Leading Group designated 592 National Poor Counties, from which all the counties in the eastern coastal provinces were eliminated. In addition, in 2007 the Chinese government expanded the coverage of the anti-poverty program, and included not only the absolute poverty-stricken population, but also the low-income population.

To support the program, the Leading Group disbursed three main funds, including the Food-for-Work program (*yigong daizhen*) supervised by the State Planning Commission, the budgetary grant program (*fazhan zijin*) overseen by the Ministry of Finance, and the subsidized loan program (*tiexi daikuan*) managed by the Agricultural Development Bank and the Poor Area Development Office (under the Leading Group). Rather than merely financial support, the central government proposed another three interventions, namely the Whole-Village Advance (*Zhengcun Tuijin*), Labor Force Transfer Training (*Laodongli Zhuanyi Peixun*), and Agricultural Industrialization Poverty Alleviation (*Nongye Chanyehua Fupin*).

First, recognizing the decentralization of the population living in poverty, along with the new National Poor Counties, the Chinese government identified 148,000 poor villages, encompassing about 80 percent of the total rural poor. The Whole-Village Advance was a community-based program. Each targeted village committee could decide on its own development plan, through a democratic process with the full participation of its village members. Since the central government believed that the improvement of living conditions and amelioration of productivity would increase household income, the plan focused on improving infrastructure and social welfare services. By 2009, about 108,400 villages had started their Whole-Village Advance plan. Among those villages, about 38,400 were located in old revolutionary areas, minority autonomous areas, and inland regions.

Second, the Labor Force Transfer Training program was a short-term job--training plan. The plan focused on training in work skills and agricultural techniques. After obtaining new skills within a relatively short period, the more qualified rural labor force could be transferred to towns and cities to obtain job opportunities with higher wages. From 2004 to 2009, the central government disbursed 3 billion yuan for the program, benefiting about 400,000 farmers. A survey showed that a worker who was involved in the program usually had a 300 to 400 yuan higher monthly salary than those who were not (PMRRC, 2011).

Third, the Agricultural Industrialization Poverty Alleviation program emphasized the development of large corporations in the industrialized agricultural industry. By subsidizing those companies, the Chinese government aimed to promote the regional economy and in turn indirectly increase rural household income. The Agricultural Industrialization Poverty Alleviation program was market-oriented, aggregating local pillar industries to form a complete industry chain. In 2004, the Leading Group identified 260 large corporations based on the recommendations of each province. Since then, 8,000 targeted villages had been included in the program. By 2009, the program had received 1.2 billion yuan from the central government and lifted more than 4 million households out of poverty.

The success of this wave was also astonishing. Rural net income reached 5919 yuan in 2010, about 1.6 times higher than in 2000. Table 2 depicts changes in the number of rural poverty-stricken people and the incidence of poverty from 2000 to 2010. Through the steadily increasing poverty line and the Leading Group's adoption of a more extensive standard to identify the poor during the decade, the rural population living in poverty in National Poor Counties decreased by more than 60% from almost 50 million in 2002 to about 17 million in 2010. Moreover, poverty incidence in the designated counties also decreased sharply after 2002, from 24.3% to less than 10 percent in 2010.

3. Data

In this paper, we use a county-level data set from the year 2000 to 2010,[8] covering 1,549 counties in the central and western regions of China. Because of the missing information on some key variables, especially in pretreatment years, our final sample contains 1,411 counties. The majority of the data are obtained from the University of Michigan's China Data Online database, which contains various important social and economic variables at the county level used to designate National Poor counties, such as GDP and local government revenue[9] per capita, as well as other social welfare variables we are interested in.

To establish a more comprehensive database, we supplemented the current data source with a threefold effort. First, since most of the rural household net

[8]To fulfill the common trend requirement, we include the information of six outcome variables in 1998 and 1999, namely two and three years before the implementation of the National Poor Counties program.

[9]Local government revenue in China exclusively consists of various tax revenues. Transfer payments are included in the index named local government total revenue.

income data are missing in China Data Online, we collected the data for 2000 to 2010 in the China National Knowledge Infrastructure (CNKI) database supported by China Statistics Press. Second, to explore the relationship between fiscal support and poverty alleviation, we added some fiscal data from the *Statistics on Public Finance of the Districts, Cities and Counties (Quanguo Di Shi Xian Caizheng Tongji Ziliao)*. Third, lists of old revolutionary areas and minority autonomous areas were obtained from *An Outline of Chinese Rural Economic Statistics by County 1988* (*Zhongguo Fenxian Nongcun Jingji Tongji Gaiyao 1988)*. Moreover, we referred to the CNKI database for lists of land frontier counties and mountainous counties. In addition, all the variables associated with money are measured in yuan, taking the form of a logarithm. They are deflated to the 2000 price level using price deflators calculated by annual CPI, which we also collected from the CNKI database.

Considering the successively appearing statistical covariate selection methods, the causal relationships among potential outcomes, treatment assignment, and covariates should initially depend on theory and previously estimable analyses (Sianesi, 2004; Smith and Todd, 2005). Since the propensity score matching method only deals with the overt bias, to obtain a precise estimation of propensity scores and avoid omitted variable bias, variables should not be excluded from our estimation unless the causal relationship fails both theoretically and statistically. In theory, only variables that can simultaneously affect the potential outcomes and treatment assignment should be included in our model. In practice, however, we face a trade-off between bias reduction and increasing variance. Moreover, Cuong (2013) demonstrates with Monte Carlo simulations that when estimating the average treatment effect on treated (ATT), greater efficiency is achieved if all the determinants of the outcome variable are included in the matching process. Therefore, besides the variables related to the potential outcomes and the treatment assignment, we also include determinants associated with the outcome variables.

Based on the principles mentioned above, we categorize matching covariates into four categories. Since the program was implemented in 2001, all the matching covariates are lagged one year (in 2000) to avoid contemporaneous endogeneity. The first category includes variables that would affect the potential outcomes and treatment assignment simultaneously. These are six variables announced explicitly by the Leading Group to construct the "631 index," including pre-trend GDP per capita, rural household net income, local revenue per capita, and three dummy indicators of revolutionary areas, minority autonomous areas, and land frontier counties. In this paper, the first three are also variables of interest. We use GDP per capita to measure the degree of economic growth. There is little doubt that economic growth contributes significantly to poverty alleviation (Ravallion and Datt, 2002; Ravallion and Chen, 2007), especially in China. Rural household net income is the most direct assessment of a welfare program, and has been widely used to evaluate the previous poverty alleviation waves (Park and Wang, 2010; Meng, 2013). Finally, to explore the revenue-generating ability of local governments, we measure local revenue per capita.

The second category consists of three variables of interest in the pre-trend period. As mentioned in the plan of the program, the improvement of infrastructure was one of the main goals. Following Fan *et al.* (2004), we use agricultural machinery power (AMP) per capita to measure infrastructure conditions. Moreover,

TABLE 3
Descriptive Table for Treatment and Control Cohorts in 2000

Variables	Combined (N=1441)	Control (N=892)	Treatment (N=549)	Difference
	Mean			
Log(GDP)	8.152	8.403	7.746	0.657***
Log(income)	7.433	7.638	7.101	0.536***
Log(revenue)	4.991	5.197	4.656	0.541***
revolutionary	0.198	0.164	0.253	−0.090***
minority	0.335	0.266	0.448	−0.182***
land	0.066	0.062	0.073	−0.011
density	256.539	309.431	170.601	138.830***
Log(special)	4.057	3.856	4.383	−0.528***
Log(expenditure)	5.857	5.853	5.864	−0.011
AMP	0.485	0.570	0.347	0.223***
beds	1880.772	2063.155	1584.443	478.712***
education	6.953	7.270	6.439	0.831***
first	36.404	33.428	41.239	−7.811***
second	32.597	35.567	27.772	7.794***

Note: ***, ** and * denote statistical significance at the 1, 5 and 10% levels, respectively.

Loayza and Raddatz (2010) argue that medical care influences the capacity of poverty reduction programs. Thus, we employ the number of beds per million people in hospitals and sanitation agencies to measure medical care. To support the newly emphasized social welfare aspects, the special transfer payments, which are allocated from higher levels of government to develop the targeted infrastructure, education, and sanitary conditions, are considered to be the main fund resource.

The third category includes the remaining covariates in 2000. We use the population density of counties as basic variable. Local government expenditure per capita reflects the size of the governments. Average years of education measures local educational conditions, while primary and secondary industry shares of are used to estimate the industrial structure.

To meet the common trend assumption required by the DID estimator, the last category involves outcome variables in 1999 and 1998 for each specific logit regression. We summarize the variables for the year 2000 in Table 3. Before matching, a significant difference exists between the control group and the treated group, except for the inland frontier area dummy and local government expenditure per capita. Table 3 shows that designated counties are generally those with lower economic development and social welfare conditions. They also tend to be agricultural counties. This suggests that effects may be attributable to pretreatment imbalances in the National Poor Counties. Therefore, we need to rebalance the data, rather than perform a simple comparison.

4. Methodology

Since the pretreatment data set is unbalanced, we have to find certain methods to remove the potential selection bias. In this paper, we utilize the propensity score matching method to balance the two groups. In response to the existence of potential bureaucratic selection bias, which might lead to a possible endogeneity problem that is not easily addressed by the matching method, two points need to be explained. First, as already mentioned in Section 2, in recognition of the political interference on the program selection, the Chinese central government adopted a more comprehensive standard—the "631 index"—to select National Poor Counties. The new standard relieves the political intervention in the selection procedure and consequently controls for endogeneity. Second, since the matching method only deals with the overt bias, we remedy the weakness by including more covariates (Baser, 2006; Guo and Fraser, 2014). In addressing the bias, rather than arbitrarily using one matching method, we use several methods and then compare their performances in bias reduction.

4.1. *Propensity Score Estimation*

The propensity score is the predicted probability that a unit gets treated, usually obtained from a logistic regression. Combined with matching methods, it enables us to transform multi-dimensional matching into one-dimensional matching, which increases matching efficiency dramatically. We can also remove the overt bias with the propensity score matching process. Second, compared with regular linear regression methods, we do not need a valid instrumental variable and can make use of the logistic function as follows:

$$P(W_i|X_i=x_i)=E(W_i)=\frac{1}{1+e^{-x_i\beta_i}} \tag{1}$$

where W_i stands for the binary treatment status of observation i ($W_i = 1$, if unit i is treated; $W_i = 0$, otherwise), X_i is a vector of conditional variables, and β_i is the coefficient of X_i.

In practice, we usually make the logarithmic transformation after the estimation. Since the logistic regression may underestimate the probability of rare events (Tomz *et al.*, 2003), the rule of thumb is to choose a control group data set with a sample size that is no more than nine times the treatment group (Baser, 2006). In our sample, the control data set and treatment group ratio is about 2:1, and it is possible to apply logistic regression in our model.

4.2. *Matching Algorithms*

In this part, we introduce the three most commonly used matching methods, from which our five methods extend. The purpose of applying matching algorithms is to achieve balance between the control group and the treated group. Due to the lack of randomly assigned treatments, the control units and the treated ones are usually unbalanced before the treatment. Hence, different outcomes may be attributed to those pretreatment imbalances rather than treatment effects. The matching algorithm with the highest quality is the one that can eliminate the difference between the treated group and the control group in the data set we analyzed.

Let I_1 and I_0 stand for sets of counties in the treated group and control group, respectively. S_p is the region of common support. The ATT for the National Poor Counties program is defined as follows:

$$(2) \qquad \tau_{ATT}^{PSM} = \frac{1}{n_1} \sum\nolimits_{i \in I_1 \cap S_p} \left\{ Y_{1i} - \sum_{j \in I_0 \cap S_p} W(i,j)\, Y_{0i} \right\}$$

where n_1 denotes the number of units in the set of $I_i \cap S_p$, and W(i, j) are the matching weights, which will be allocated to control counties to form a reliable counterfactual. Depending on the choice of the functional form of W(i, j), a variety of matching methods has been proposed, which includes nearest-neighbor matching (NNM), radius matching (RM), and kernel matching (KM).

NNM compares the distance in terms of propensity score between non-participants and a participant; the closest one in the control group is chosen to be a matching partner for the treated individual (Caliendo and Kopeinig, 2008). RM is a method combining NNM with a caliper, which is a sort of tolerance level associated with the maximum distance in terms of propensity score. As a non-parametric matching algorithm, KM treated members are matched with a weighted average of a subgroup of control members, depending on the bandwidth we choose. The first NNM variant is NNM without replacement, which means that individuals in the control group can be chosen as a matching partner no more than once. The second variation is n-to-1 NNM. The main difference between this variant and NNM is that we can use the closest *n* control individuals in distance as matching partners for a treated individual.

To further eliminate unobservables that may affect the National Poor Counties' assignment and the outcome variable, drawn on the panel data, we adopt the DID matching strategy (Heckman *et al.*, 1997; Gebel and Voßemer, 2014). Rather than directly focusing on the outcome variable, this variation matches the before-and-after-treatment differences in the outcome between the treated and the matched controls. In this way, besides the overt bias, we are able to further eliminate constant hidden differences and additive selection biases, and significantly improve the quality of our estimation results. The DID propensity score matching (PSM-DID) estimator is defined in the following way:

$$(3) \qquad \tau_{ATT}^{DID-PSM} = \frac{1}{n_1} \sum_{i \in I_1 \cap S_p} \left\{ (Y_{1ti} - Y_{1t'i}) - \sum_{j \in I_0 \cap S_p} W(i,j)(Y_{0ti} - Y_{0t'i}) \right\}$$

where Y_{1ti} (Y_{0ti}) and $Y_{1t'i}$ ($Y_{0t'i}$) are the outcome variables of interest for the treated (control) counties in time t and t', which are after and before the implementation of the National Poor Counties program.

5. Results

5.1. *Propensity Score Estimation Results*

In this paper, we have six variables of interest, each of which has distinct determinants. Consequently, we independently apply logistic regressions for each of the six outcome variable. As displayed in Table A1, most of the covariates

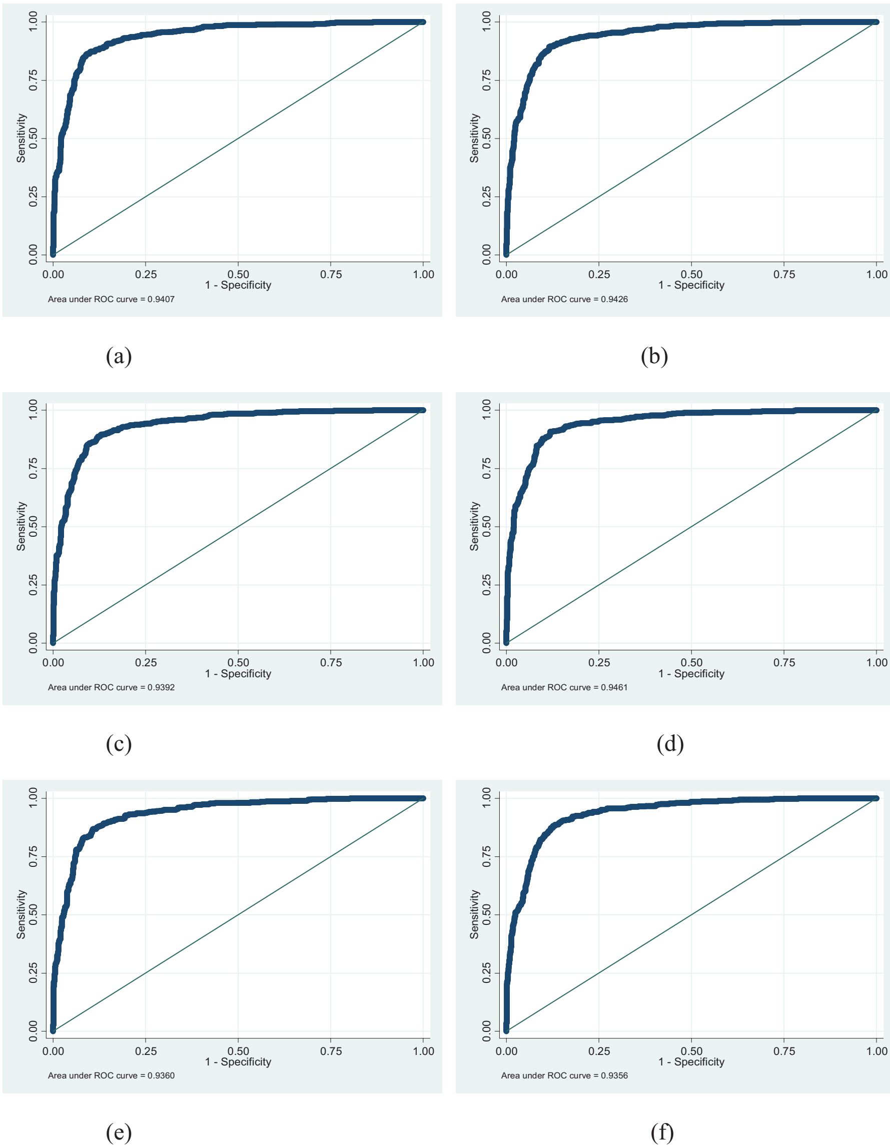

Figure 2. Receiver Operating Characteristic Curve and C-Statistics [Colour figure can be viewed at wileyonlinelibrary.com]

Note: From (a) to (f), the outcome variables are ΔLog(GDP), ΔLog(income), ΔLog(revenue), ΔLog(special), Δbeds and ΔAMP in order.

included in the logistic regression are significant and exhibit reasonable signs as expected. No matter what outcome variables we choose, the GDP per capita, rural household net income, local government expenditure, and sanitation conditions are negatively related to the probability of a county being chosen as a National Poor County. However, a county that is identified as a revolutionary county and with a higher share of secondary industry and special transfer payment per capita is more likely to be selected as a National Poor County. Given that the six pseudo-R^2 are all above 0.500, our equations explain more than 50 percent of the variation in the choice.

Moreover, to see the goodness of fit of our regression, we calculate the area under the receiver operating characteristic (ROC) curve. The greater the predictive power of our estimation, the more bowed the ROC curve, and the larger the area created under the curve. Therefore, the area under the curve (C-statistic) can be used to measure the predictive power of our logistic regression. To achieve a valid classification, the C-statistic should be greater than 0.80, which is met in our case, as shown in Figure 2.

5.2. *Matching Results*

Low quality matching with small bias reduction (and therefore imbalances) between the treated and the control groups can lead to biased estimation of the average treatment effect on the treated. To identify the most suitable type of matching, we implement five matching algorithms, including NNM with replacement, NNM without replacement, 3-to-1 NNM, RM, and KM. Since a caliper can significantly improve the matching quality, according to the rule of thumb, we impose 0.25*SD (standard deviation) calipers on each matching algorithm. By doing so, we rely on the fact that the performance of different matching methods varies case by case, mainly according to the data set we use.

As exhibited in Table A2, we test the group balance before and after matching based on three criteria.[10] First, we apply a t-test to calculate the t-statistics for the treated group and the control group after matching. Except for NNM without replacement, the other four algorithms perform equally badly, leaving around 10 unbalanced variables between the participant and non-participant counties. As for the NNM without replacement, it induces no significant difference between two cohorts (all the p-values are greater than 0.100).

Secondly, we check the standardized percentage bias, which is the percentage difference of the sample means in the treatment and matching subsamples, as a percentage of the square root of the average of the variances in the treated and the control cohorts (Rosenbaum and Rubin, 1985). Even without an explicit standard under which we can treat a standardized percentage bias as a success, 5 percent, 8 percent, and 20 percent are commonly used as sufficient thresholds (Girma and Görg, 2007; Caliendo and Kopeinig, 2008). It is clear that the performance of NNM without replacement is also the best under this criterion. Figure A1 depicts the standardized percentage bias across covariates after NNM without replacement.

Lastly, we calculate the percentage bias reduction in the means of the independent variables before and after matching. The results are almost the same as those we obtained from the previous two criteria. As for the NNM without replacement, the percentage bias reductions of all the explanatory variables exceeds 70 percent (except for the land frontier), which are much higher than the

[10]Because of the limited space, all balancing test results reported in this subsection are based on the PSM method when the outcome variable is the ΔLog(GDP). When the outcome variable varies, the balancing tests share similar results.

results from other algorithms. Taking the three criteria into consideration, the best matching type for our data set is the NNM without replacement. In addition, we also performed the sensitivity analysis with more rigid calipers, and the results varied little. Hence, we stick to NNM with 0.25*SD calipers for the following analysis.

5.3. *Impact of the Third Wave*

Although the National Poor Counties program was designed to promote economic growth and enrich the poor, the designated counties might well deviate from the original intention of the policy for their own interests. There were several distinct potential responses by local governments to the program, leading to two opposite mechanisms through which the program functioned. Since economic growth would increase a local government's revenue by expanding the tax base, local governments had an incentive to develop the economy, which in turn benefited them. This turned out to be a virtuous circle. Meanwhile, economic growth was supposed to enrich the poor through the trickle-down effect. In the case of China, National Poor Counties exclusively received various transfer payments from both central and provincial governments and enjoyed preferential policies; however, these would be removed if a county developed and surpassed the poverty line. Therefore, there existed two ways of development for local governments. First, taking advantage of the various benefits brought by the program, they could spare more effort on local development to pass the poverty line and continue their economic and welfare growth without the previous transfer payments and preferential policies. Second, being afraid of losing the previous financial and political support and even slipping into retrogression, the designated counties might attempt to maintain their benefits from their National Poor County identities by lowering their growth rates to avoid passing the poverty line.

As mentioned in Sections 2 and 3, the third wave was more ambitious than its predecessors. Besides regional economic growth and rural household income growth as the main goals of the first two waves, the explicit target of the third poverty alleviation wave consisted of three parts, namely infrastructure promotion, educational improvement, and improvement of sanitary conditions. Since the first wave, the three main interventions reflect that the promotion of infrastructure was ranked as the first priority in poverty alleviation attempts. The third wave was the first time that improving health and education as poverty alleviation measures were explicitly included. However, whether these goals have been achieved remains unknown. Therefore, in this subsection, we attempt to answer two questions. First, what was the response of local governments to the program? Second, were these poverty alleviation goals achieved?

5.3.1. What Was the Main Response of Local Governments to the Program?

If the first mechanism functions, we would expect significantly positive coefficients for GDP and revenue per capita. However, if the second mechanism dominates, we would observe negative coefficients for GDP per capita at the end of the program, which capture the attempts of local governments to limit their economic

TABLE 4
The Impact of National Poor Counties on the Growth of GDP, Income, and Local Government Revenue

	Δlog(GDP)	Δlog(income)	Δlog(revenue)
2001	−0.022**	−0.082***	−0.054**
	(0.011)	(0.025)	(0.022)
2002	−0.040***	−0.090***	−0.038
	(0.014)	(0.027)	(0.027)
2003	−0.039*	−0.110***	−0.050
	(0.021)	(0.030)	(0.034)
2004	−0.043	−0.135***	−0.071
	(0.027)	(0.031)	(0.049)
2005	−0.047	−0.166***	−0.090
	(0.034)	(0.033)	(0.064)
2006	−0.050	−0.183***	−0.069
	(0.038)	(0.033)	(0.075)
2007	−0.057	−0.164	−0.094
	(0.042)	(0.112)	(0.075)
2008	−0.067	−0.177***	−0.081
	(0.048)	(0.038)	(0.077)
2009	−0.068	−0.179***	−0.061
	(0.048)	(0.038)	(0.076)
2010	−0.073	−0.203***	−0.035
	(0.050)	(0.041)	(0.076)

Note: Standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10% levels.

growth to maintain their status in the next wave. We would also expect significantly negative coefficients for revenue per capita, given the lower economic growth and, consequently, the shrunken tax base. In addition, since we assume that local governments are inclined to sacrifice their GDP growth rates for higher transfer payments, some significantly positive coefficients for special transfer payments per capita are expected.

As shown in Table 4, DID estimators reflect the ineffectiveness of the National Poor Counties program for the whole sample. GDP growth rates are significantly negative at the beginning of the program and become insignificant in 2004. Local government revenue almost follows the same pattern. The second column in Table 4 reveals that the income disparity between participant and non-participant counties became larger after the program was implemented. In response to the bad economic performance, the central government allocated more transfer payments to support the designated counties in 2006, as the estimates become positively significant in the first column in Table 5.

We provide two possible explanations for the ineffectiveness of the program. First, after 20 years of unwavering efforts against poverty, the remaining poor that were still unable to share the benefits brought by national economic growth or the regional targeted poverty alleviation program were either residents in the most remote areas, usually with a bad environment, or physically handicapped. It was thus difficult to lift the rest of the poor population out of poverty. Second, the validity of the program in the third wave is questionable. After the success of the second wave, the distribution of the rural poverty population became evenly scattered. A large proportion of residents within certain National Poor Counties

TABLE 5

The Impact of National Poor Counties on Special Transfer Payments Growth, Sanitary Conditions, and Infrastructure Conditions

	Δlog(special)	Δbeds	ΔAMP
2001	0.156	32.760	0.008
	(0.322)	(61.302)	(0.008)
2002	0.178	−97.631*	−0.001
	(0.313)	(57.513)	(0.010)
2003	0.182	−96.508	0.010
	(0.308)	(63.539)	(0.012)
2004	0.170	−48.427	0.017
	(0.317)	(100.371)	(0.015)
2005	0.172	−221.573***	0.008
	(0.316)	(64.907)	(0.018)
2006	0.067	−215.116***	0.006
	(0.057)	(79.023)	(0.029)
2007	0.115*	−257.892***	0.015
	(0.059)	(85.900)	(0.029)
2008	0.111*	−31.028	−0.017
	(0.066)	(78.824)	(0.029)
2009	0.145**	16.327	−0.002
	(0.065)	(88.677)	(0.036)
2010		−18.913	0.029
		(95.764)	(0.053)

Note: Standard errors are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10% levels.

were not poor any more, while there indeed existed a non-negligible portion of the rural poor living outside National Poor Counties (Park and Wang, 2010). The preciseness and effectiveness of the program, which targeted counties, was therefore decreased. On the other hand, since the main interventions in the third wave were basically the same as the second wave, and the effectiveness of the corresponding polices had deteriorated at the end of the second wave (Meng, 2013), the validity of polices in the third was cast into doubt. Either potential reason would retard economic growth, as well as the trickle-down effect that was supposed to benefit the poor.

From the estimation for the whole sample, we know little about the program's mechanisms, except for its ineffectiveness. We therefore separate the sample to analyze the heterogeneity in subsamples in order to identify which mechanism functioned. We first identify the counties involved in the "Western Development Strategy," launched in 2000, as the western counties and the rest as the non-western counties. Table A3 and A4 revealed the same pattern in the western counties as in the whole sample. However, the special transfer payments received by western counties are insignificant in the sampled period. As for the non-western counties, the second mechanism functions. In the second column in Table A4, the GDP growth rate turns significantly negative in 2008 (−13.4 percent in 2008, −13.4 percent in 2009, and −17.9 percent in 2010), which means that for certain reasons the program hampered the economic growth of the non-western counties. It reflects that at the end of the third wave, the designated counties were worried about re-selection in the next wave and tried to limit their growth rate to maintain their identity in the following program. In addition, the coefficients of special transfer payments per capita are significantly positive in most years, which showed that the non-western counties are more likely to gain access to the special transfer payments after the implementation of the program. The aforementioned features demonstrate that the second mechanism was the main mechanism through which the National Poor Counties program affected local governments.

We further analyze the heterogeneity for counties located in flat areas and mountainous areas. While the program had different effects on the two types of counties compared to the whole sample, we find almost no difference between the two types. In the 2000s, the program generally had a negative effect on those counties in terms of GDP growth, income level, and local government revenue. Unfortunately, those counties did not obtain substantial transfer payments to compensate for their lower economic development.

In addition, these estimation results explain the engine of success in the third wave. While the anti-poverty achievement is huge for the National Poor Counties, Table 2 reveals that the success in poverty reduction is also substantial for non-designated ones. Therefore, despite the implementation of the program, the success of the third wave could also be attributed to the country's overall economic growth. Given the negatively significant and insignificant ATTs shown in Table 4, the program seems to have been ineffective in poverty alleviation. Thus, rather than attributing successes in anti-poverty efforts, as discussed in Section 2, to the National Poor Counties program, it is more likely due to broad-based economic growth. That helps to explain the puzzle of successful poverty reduction in the third wave, and the ineffectiveness of the program.

5.3.2. Were Poverty Alleviation Goals Achieved?

Restricted by the limited data, in this paper we are only able to detect the impact of the program on infrastructure conditions and sanitary conditions. Table 5 shows that the program had an insignificant impact on sanitary conditions at the beginning and end of the 2000s, and a significantly negative impact in the middle period. This is because after special transfer payments started increasing in 2005, the negative effect on sanitary conditions faded. The western counties and mountainous counties more or less followed the same scenario as the whole sample, as shown in Table 5. However, the program had no effect on the non-western counties in terms of sanitary conditions, for all coefficients in the sample period are insignificant. As for flat areas, there was a positive effect soon after the implementation of the National Poor Counties program (104.537, significant in 2001), though it lasted for only a year.

As mentioned in Section 3, we use the AMP per capita to investigate the rural infrastructure level. In the third column in Table 5, the insignificant coefficients of the whole sample reveal that the program's attempt to improve infrastructure was in vain. Following the same decomposition method as before, we find that the results of the subsamples of western, non-western, and flat counties coincide with those of the whole sample. As for mountainous areas, the program basically had no effect except for in 2008, when the AMP per capita decreased by 9.5 percent, which is significant at the 5 percent confidence level. Combined with results with respect to the special transfer payments, we find that although the transfer payments increased for the whole sample and non-western counties in certain years, the program failed to improve the infrastructure and sanitary conditions in the designated counties. Thus, the social welfare goals of the program were not achieved.

6. Conclusion

In this paper, we estimate the impact of the National Poor Counties program on newly-designated poor counties. To eliminate the overt bias to the largest extent, we compare five differentiated types of propensity score matching algorithms and find that the NNM method is most suitable for our data set. We then combine NNM with the DID estimator to further control for the constant hidden differences and additive selection bias.

We have two main findings. First, we distinguish two potential responses of local governments to the program. According to our results, non-western local governments were inclined to manipulate their economic growth data so as to maintain access to transfer payments disbursed exclusively to designated poor counties. Second, we find that the targeted program had a negative effect on infrastructure in some years and failed to affect sanitary conditions in general. However, if we further decompose the whole sample into flat and mountainous counties, the National Poor Counties program would have different impacts on different subsamples. Our results show that the program was not totally in vain, as it worked in certain counties.

References

AOCRESC., *An Outline of Chinese Rural Economic Statistics by County 1988*, Beijing, China Statistical Publishing House, 1989.

Baser, O., "Too Much Ado about Propensity Score Models? Comparing Methods of Propensity Score Matching," *Value in Health*, 9, 377–85, 2006.

Caliendo, M. and S. Kopeinig, "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys*, 22, 31–72, 2008.

CPFDCC., *Statistics on Public Finance of the Districts, Cities and Counties*, Beijing, China Financial and Economic Publishing House, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010.

Cuong, N. V., "Which Covariates Should Be Controlled in Propensity Score Matching? Evidence from a Simulation Study," *Statistica Neerlandica*, 67, 169–80, 2013.

Fan, S., L. Zhang, and X. Zhang, "Reforms, Investment, and Poverty in Rural China," *Economic Development and Cultural Change*, 52, 395–421, 2004.

Gebel, M. and J. Voßemer, "The Impact of Employment Transitions on Health in Germany. A Difference-in-Differences Propensity Score Matching Approach," *Social Science & Medicine*, 108, 128–36, 2014.

Girma, S., and H. Görg, "Evaluating the Foreign Ownership Wage Premium Using a Difference-in-Differences Matching Approach," *Journal of International Economics*, 72, 97–112, 2007.

Guo, S. and M. W. Fraser, *Propensity Score Analysis: Statistical Methods and Applications* (2nd ed.), Los Angeles, CA: SAGE, 2014.

Heckman, J. J., H. Ichimura, H., and P. E. Todd, "Matching As an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64, 605–54, 1997.

Loayza, N. V. and C. Raddatz, "The Composition of Growth Matters for Poverty Alleviation," *Journal of Development Economics*, 93, 137–51, 2010.

Meng, L., "Evaluating China's Poverty Alleviation Program: A Regression Discontinuity Approach," *Journal of Public Economics*, 101, 1–11, 2013.
Park, A., S. Rozelle, and F. Cai, "China's Grain Policy Reforms: Implications for Equity, Stabilization, and Efficiency," *China Economic Review*, 5, 15–33, 1994.
Park, A., S. Wang, S., and G. Wu, "Regional Poverty Targeting in China," *Journal of Public Economics*, 86, 123–53, 2002.
Park A. and S. Wang, "Community-Based Development and Poverty Alleviation: An Evaluation of China's Poor Village Investment Program," *Journal of Public Economics*, 94, 790–99, 2010.
PMRRC., *Poverty Monitoring Report of Rural China*, Beijing, China Statistical Publishing House, 2000, 2001, 2010, 2011.
Ravallion, M. and G. Datt, "Why Has Economic Growth Been More Pro-poor in Some States of India than Others?," *Journal of Development Economics*, 68, 381–400, 2002.
Ravallion, M. and S. Chen, "China's (Uneven) Progress Against Poverty," *Journal of Development Economics*, 82, 1–42, 2007.
Rosenbaum, P. R. and D. B. Rubin, "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score," *The American Statistician*, 39, 33–8, 1985.
Rubin, D. B., "Comment: Which Ifs Have Causal Answers?," *Journal of the American Statistical Association*, 81, 961–62, 1986.
Sianesi, B., "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s," *The Review of Economics and Statistics*, 86, 133–55, 2004.
Smith, J. A. and P. E. Todd, "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?," *Journal of Econometrics*, 125, 305–53, 2005.
Tomz, M., G. King, and L. Zeng, "ReLogit: Rare Events Logistic Regression," *Journal of Statistical Software*, 8, 137–63, 2003.
World Bank., *China: Overcoming Rural Poverty*, Washington, D.C., World Bank, 2001.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table A1: Estimation of Propensity Score with Logit

Table A2: Balancing Comparison (Outcome Variable is ΔLog(GDP))

Table A3: The Heterogeneity Impact of National Poor Counties on Gdp Growth and Income Growth

Table A4: The Heterogeneity Impact of National Poor Counties on Local Government Revenue Growth and Special Transfer Payments Growth

Table A5: The Heterogeneity Impact of National Poor Counties on Sanitary Conditions and Infrastructure Conditions

Figure A1: Standardized Percentage Bias across Covariates