

How Relevant is Targeting to the Success of an Antipoverty Program?

Martin Ravallion

The World Bank
Office of the Director
Development Research Group
November 2007



Abstract

Policy-oriented discussions often assume that “better targeting” implies larger impacts on poverty or more cost-effective interventions. The literature on the economics of targeting warns against that assumption, but evidence has been scarce. The paper begins with a critical review of the strengths and weaknesses of the targeting measures found in practice. It then exploits an unusually large micro data set for China to estimate aggregate and local-level

poverty impacts of the country’s main urban antipoverty program. Standard measures of targeting are found to be uninformative, or even deceptive, about impacts on poverty and cost-effectiveness in reducing poverty. In program design and evaluation, it would be better to focus directly on the program’s outcomes for poor people than to rely on prevailing measures of targeting.

This paper—a product of the Director’s office, Development Research Group—is part of a larger effort in the department to assess the reliability of the methods used in practice for guiding policy making. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at mravallion@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

How Relevant is Targeting to the Success of an Antipoverty Program?

Martin Ravallion¹

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

Keywords: Poverty, cash transfers, errors of targeting, China

JEL: I32, I38, O15

¹ For helpful discussions on this topic and help with the data used here the author is grateful to Shaohua Chen, Jean-Yves Duclos, Emanuela Galasso, Garance Genicot, Margaret Grosh, Pilar Garcia Martinez, Philip O'Keefe, Adam Wagstaff, Dominique van de Walle, Youjuan Wang, Xiaoqing Yu, and seminar participants at Beijing University and at the Ministry of Finance-World Bank Roundtable on Public Finance, 2006. The findings, interpretations and conclusions of this paper are those of the author, and should not be attributed to the World Bank.

Various measures of the “targeting performance” of antipoverty programs have been widely used to inform policy discussions. These measures are typically interpreted by both analysts and policy makers as indicators of a program’s performance in “...directing benefits toward poorer members of the population” (Coady, Grosh and Hoddinott, 2004a, p.81). Comparisons of such measures across different programs have informed public choices on which programs should be scaled up and which should be dropped.²

As in any situation in which measurement is used to inform policy, “the indicators need to be related to the overall policy problem, with an explicit formulation of the objective and constraints” (Atkinson, 1995, p.31). It is widely agreed that the objective of this class of public programs is to reduce poverty, subject to the relevant constraints, including those related to the information available and the behavior of relevant agents, as well as resources. Better targeting is not seen as desirable in its own right, but rather as an instrument for reducing poverty.

Do the measures of targeting used in policy discussions provide useful indicators for this policy problem? The most widely used measures quantify some aspect of how well a given program concentrates its benefits on the poor, which is essentially what “targeting” has come to mean. An example is the share of transfers going to the poor. Cornia and Stewart (1995) have been influential in arguing that measurement practices and policy discussions have put too high a weight on avoiding one type of error—the “Type 1 error” of having (ineligible) non-poor participants—relative to the “Type 2 error” of incomplete coverage of the poor.³

Cornia and Stewart did not present data linking these aspects of targeting performance to poverty outcomes (though they do point to this as an important direction for further research). However, the literature warns us against assuming that better targeting, as assessed by standard measures, will necessarily enhance a program’s total impact on poverty.⁴ A number of factors cloud the relationship between targeting performance and total impact on poverty, including

² Early empirical studies by Mateus (1983) and Grosh (1992) were influential in arguing the case for finer targeting. The meta studies of Grosh (1994) and Coady et al. (2004a,b) have provided the most comprehensive comparative data on program performance based on targeting measures.

³ The distinction between these two errors goes back to Weisbrod (1970) who called them “vertical-” and “horizontal target efficiency.” Cornia and Stewart (1995) used the terms “E-mistakes” and “F-mistakes;” Smolensky et al., (1995) called them “errors of inclusion” and “errors of exclusion.” The literature on social welfare policy in developed countries has also suggested that coverage of the poor is given little weight by standard measures of targeting. See, for example, the results of Duclos (1995) on the implications of incomplete take up of Britain’s welfare benefits for measures of targeting. However, the relationship between these problems and overall impacts on poverty has received little attention.

⁴ For an overview of the arguments and evidence see van de Walle (1998).

aspects of program design, implementation and the context in which a program operates. Incentive issues have been a theme of one strand of the literature, pointing to the possibility that fine targeting will impose high marginal tax rates on recipients, possibly creating poverty traps.⁵ The literature has also warned that fine targeting can undermine political support for an antipoverty program; concentrating gains on the poor may induce a lower overall transfer to the poor, with benefits spread too thin, or covering too few people.⁶

It is also unclear how useful these measures are as indicators of cost-effectiveness, as an input to scaling-up decisions. Here it is not the total impact on poverty that one is focusing on, but rather the impact per unit of the resources devoted to a given program. (The total impact then depends on the allocation of resources across programs, weighted by their cost-effectiveness ratios.) Intuitively, the impact on poverty will depend on both the share of transfers going to the poor and the total transfer. Plainly a large uniform transfer (received by everyone, whether poor or not) can have more impact on poverty than a small well-targeted transfer. But will the latter type of program, with low leakage to the non-poor, necessarily be more cost-effective? The answer is far from obvious on *a priori* grounds. The factors noted above that cloud the relationship between targeting performance and a program's total impact on poverty will not, in general, vanish when total impact is normalized by total spending.

For example, finer targeting typically entails administrative costs, which are debits against the total budget in determining the government's total transfer payment. Then the share of transfers going to the poor does not even identify the transfer to the poor per unit public spending. Less obviously, but no less importantly, targeting can generate "hidden" costs to participants, notably when there are conditionalities, such as work requirements, behavioral conditionalities or sources of social stigma. Given the costs of targeting, it is not difficult to imagine cases in which the better targeted program (with the higher share of transfers going to the poor) is less cost effective in reducing poverty, and the literature already contains examples.⁷ In short, avoiding leakage to the non-poor can reduce the amount actually going to the poor, with theoretically ambiguous implications for poverty and cost-effectiveness in fighting poverty.

⁵ Besley and Kanbur (1993) pointed to this problem and other issues raised by targeting. Also see Smolensky et al. (1995). Kanbur et al. (1995) study the incentive issues in fine targeting, including characterizing an optimal scheme for poverty reduction, taking account of labor supply responses.

⁶ For theoretical analyses see De Donder and Hindriks (1998) and Gelbach and Pritchett (2000).

⁷ Ravallion and Datt (1995) and Murgai and Ravallion (2005) provide examples for workfare programs in India.

Whether better targeting, as measured in practice, implies a greater impact on poverty, or a more cost-effective intervention, is ultimately an empirical question. Yet, beyond a few suggestive examples, we really know rather little about how well these popular targeting measures perform in practice.

This paper tries to help fill this gap in knowledge using a detailed case study of one program, namely China's "Minimum Livelihood Guarantee Scheme," popularly known in China as *Di Bao* (DB). This has been the government's main response to the new challenges of social protection in urban areas. A number of factors—the decentralized nature of the program, its scale and the availability of a large data set representative at local level—combine to make this an unusual opportunity to put targeting measures to the test. The program's targeting performance and impacts on poverty are estimated under standard assumptions across each of the 35 major municipalities of China. The most popular targeting measures found in policy-oriented discussions are thus tested as indicators of program performance in reducing poverty.

Measures

In principle, one can measure "targeting performance" by a program's impact on poverty relative to an explicit counterfactual, such as an un-targeted allocation of the same budget (as in Ravallion and Chao, 1989). Then the interpretation for poverty is unambiguous. That is not, however, the approach that has dominated the literature and practice. This discussion will focus on the main measures of targeting performance found in practice, and on which much of our current knowledge about "what works and what doesn't" is based. More precise definitions of the measures can be found in Table 1.⁸

Targeting measures

I focus on four main measures, the first three of which are based on the concentration curve, $C(p)$, giving the cumulative share of transfers going to the poorest $p\%$ of the population ranked by (say) household income per person (Figure 1).

The first measure is the share of transfers going to the poorest $H\%$, such as the poorest 40%. This is denoted $S=C(H)$. In the empirical work discussed later, it will be natural to identify the poorest $H\%$ as the target group, i.e., the set of people deemed to have incomes

⁸ Table 1 relates only to the measures used in this study. For a more comprehensive discussion of these and other measures, including their analytic properties, see the excellent volume by Lambert (2001).

below the municipal *Di Bao* poverty line; more precisely, we can set $H = H_0$, which is the pre-intervention headcount index of poverty—the proportion of the population living in households with pre-transfer income per person less than the poverty line. (The post-transfer headcount index is H_1 .) For much of the present discussion we can just take the poorest $H\%$ to be some reference group of poor or relatively poor people, without presuming that it is the precise target population for the program in question.

The popularity of S is evident in the fact that the meta-studies by Grosh (1994, 1995) and Coady, Grosh and Hoddinott (2004a,b) found that this was the most readily available measure in their primary sources.⁹ The measure's popularity may well stem from its ease of interpretation. Against this advantage, the measure has some obvious drawbacks. For one thing, it tells us nothing about how transfers are distributed amongst the poor; two programs can have the same share of transfers going to the poor, but in one case the gains are heavily concentrated amongst the poorest, while in the other they case they only reach those just below the poverty line. Another concern is that this measure does not directly reflect the overall size of the transfer program, which will clearly matter to impacts on poverty, as discussed in the introduction.¹⁰

The second measure is the normalized share, NS , obtained by dividing S by H (Figure 1). Coady et al. (2004a, b) preferred to use NS as their measure of targeting performance, arguing that this was more comparable than S because it measures performance relative to a "... common reference outcome...that would result from neutral (as opposed to progressive or regressive) targeting" (p.69).¹¹ By "neutral targeting" they mean that everyone gets the same transfer amount (whether poor or not), i.e., a "uniform transfer."¹² If the transfer is uniform then clearly $NS=1$. However, finding a value of NS close to unity does not imply that the allocation is "close" to being uniform. There are many ways one could get a value for NS of unity (or nearly so), with rather different interpretations. Similarly to S , the NS measure is insensitive to how transfers are

⁹ Coady et al. provide the shares going to the poorest 10%, 20% and 40% for 85 of the antipoverty programs in their study (though with missing data in some cases).

¹⁰ The literature has pointed to the possibility that the share going to the poor can vary with the scale of a program, though the political economy of program capture; see Lanjouw and Ravallion (1999).

¹¹ Coady et al. used $H=40\%$ when it was available, which was the case for about half the programs in their study, and the next lowest available number (20% or 10%) when the value for $H=40\%$ was not available. In the earlier comparative study of targeting performance by Grosh (1994), the value of H is set at 40% in all programs studied, in which case the first two measures will (of course) rank

¹² This is sometimes called an "un-targeted transfer" in the literature, although it is not clear the absence of any effort at targeting would yield a uniform transfer.

distributed amongst the poor. The poor can receive $H\%$ of the transfers, but different people amongst the poor receive very different amounts; for example, the money could all go to either the poorest person or the least poor person; either way $NS=1$. NS also approaches unity as H approaches 100%, no matter how the money is distributed. When the reference outcome is this ambiguous, the usefulness of the measure becomes theoretically questionable.

The third measure is the concentration index, CI , which is a widely used in studies of fiscal incidence. This can be thought of as a “generalized S ” in that, instead of focusing on one point on the concentration curve, CI measures the area between the curve and the diagonal (along which the transfer is uniform); in Figure 1, CI is just twice the area marked A.¹³ The index is bounded above by 1 (at which point the poorest person receives all payments) and below by -1 (the richest person receives all). This measure has the attraction that it reflects distribution amongst the poor, and (indeed) over the whole range of incomes. A disadvantage is that it is not as easy to interpret as S or NS . And, as with the previous measures, it tells us nothing directly about the scale of transfers.

Although these measures are all based on the concentration curve, they can give quite different results. Of course, S and NS will always be in the same ratio to each other when the same value of H is used for all programs. However, these two measures can rank programs differently when H varies, as in the case study presented later in this paper, and would presumably be the case in many applications.

To illustrate, consider a transfer scheme operating in two cities and giving all participants the same amount. In city A all the transfers go to the poorest 20% and the overall poverty rate is 50% while in city B the transfers go to the poorest 40% and the poverty rate is 10%. A far higher share of the transfers goes to the poor in A ($S=100\%$ versus 25% in B). City A also has the higher concentration index ($CI=0.8$ in A versus 0.6 in B). By contrast, it is in city B where the scheme is deemed to be better targeted according to the normalized share ($NS=2.5$ for B versus 2 for A). More generally, the concentration curve for program A could lie everywhere above that for program B and yet NS is higher for B, given its lower H .

The fourth measure is the “targeting differential,” TD , which is the difference between the participation rate for the poor—which I will call the coverage rate (CR)—and that for the

¹³ To assure that all measures go in the same direction, I multiply the usual definition of CI by -1.

non-poor (Table 1).¹⁴ Alternatively, one can normalize the targeting differential by the mean transfer over all recipients; call this TD^* . (When all recipients get the same transfer, $TD=TD^*$.) However, it turns out later that the choice between TD and TD^* makes little difference in the case study. Since TD is easier to interpret I shall focus on this measure.

To interpret the targeting differential, note that when only the poor get help from the program and all of them are covered, $TD = 1$, which is the measure's upper bound; when only the non-poor get the program and all of them do, $TD = -1$, its lower bound. (In the "two cities" example above, $TD=0.67$ for city B and 0.4 for A.) This measure is easy to interpret, and it automatically reflects both leakage to the non-poor and coverage of the poor.

How are these measures related to the incidence of Type 1 and Type 2 errors? A Type 1 error can be defined as incorrectly classifying a person as poor, while a Type 2 error is incorrectly classifying a person as not poor. A Type 1 error entails a leakage of transfers to the non-poor, while a Type 2 error implies lower coverage of the poor. Let the proportions of Type 1 and Type 2 errors in the populations of the non-poor and poor (respectively) be $T1$ and $T2$, as defined more precisely in Table 1.¹⁵ (Note that $T2=1-CR$.) Consider S . This can be written as a function of $T1$, namely $S = 1 - T1(1 - H) / P$ where P is the overall program participation rate. (Alternatively $S = 1 - T1^*$ where $T1^*$ is the proportion of participants who are Type 1 errors.) But one can equally well write S as a function of Type 2 errors, namely $S = (1 - T2)H / P$. (Or $S = (H / P) - T2^*$.) Nor is P likely to be independent of $T1$ and $T2$; for example, higher coverage of the poor (lower $T2$) may tend to come with larger programs. Thus S can be taken to depend on both $T1$ and $T2$. (The corresponding formulae are more complex for CI , and are omitted.) For the targeting differential, however, the relationship is very clear: TD automatically gives equal weight to both errors; more precisely: $TD = 1 - (T1 + T2)$.

Thus standard targeting measures depend on the incidence of both types of errors. For the measures based on the concentration curve it should not be presumed that they will be largely unaffected by Type 2 errors. Is an empirical question what weights are attached to these two "errors of targeting."

¹⁴ This measure was proposed by Ravallion (2000). Also see Galasso and Ravallion (2005) on the properties of this measure and the discussion in Stifel and Alderman (2005).

¹⁵ One might prefer to normalize by population size; similar formulae for this case are easily derived, but the essential point remains.

Poverty impacts

In testing the relevance of these targeting measures to a program's poverty impacts I use three poverty measures: the headcount index, the poverty gap index (PG), and the squared poverty gap index (SPG) (introduced by Foster et al., 1984). The measures are defined in Table 1. The pros and cons of each are well-documented; for a review see Ravallion (1994). Briefly, H is the easiest to interpret, and is the most popular measure, but is unaffected by income gains or losses to the poor unless they cross the poverty line. PG reflects mean income of the poor, but not inequality amongst the poor, which is the main advantage of SPG .

Impacts are measured by pre-transfer less post-transfer poverty measures ($H_0 - H_1$ and similarly for PG). Impacts on these measures are estimated on the same data, and under the same assumptions about how the scheme works (including behavioral responses), as used in measuring targeting. In particular, I shall assume that income in the absence of the program is observed income less payments received under DB. This assumes that there is no displacement of other income sources through behavioral responses, such as reduced work effort or lower private transfer receipts. This is the most common assumption in the literature on measuring targeting performance; indeed, it appears that virtually all of the primary studies used by Coady et al. (2004a,b) made this assumption. The assumption is questionable, however; I offer some tests that, while not conclusive, suggest that the data are at least consistent with the assumption.

In assessing cost effectiveness I will normalize the poverty impacts by the cost of the program, though a more flexible econometric method of controlling for total spending will also be used. Given the costs of targeting, it is not difficult to imagine cases in which the better targeted program by any of the above measures is less cost-effective against poverty. Consider again the example of cities A and B above in which the program in city A is better targeted according to both S and CI (but not NS or TD). Suppose that the total cost to the government is the same, but that the finer targeting of city A's program (for which it will be recalled that all of the transfers go to the poorest 20%, versus 40% in city B) entails extra costs to both the government and participants such that only 25% of participants in city A escape poverty, while in B all poor participants are able to do so. The headcount index falls by 5% points in A, but 10% points in B. B's program has higher impact on poverty and is more cost-effective.

There is a special case in which one of these measures, namely S , is a perfect indicator of cost-effectiveness for PG . That special case is when the program has no impact on H and there

are no fiscal costs besides the transfers. Then it can be readily shown that the impact on PG per unit public spending is simply S . Of course, this special case is unlikely to be of much practical interest, given that people in a neighborhood of the poverty line will presumably be transfer recipients and there will undoubtedly be other costs.

Under the same assumptions, it can be readily shown that the normalized share, NS , is a perfect indicator of cost-effectiveness in reducing the income-gap ratio (I) (Table 1). This is (implicitly) the poverty measure relevant to comparisons of program performance based on the normalized share. However, as a poverty measure, the income-gap ratio is known to have a number of undesirable properties; for example, if a poor person living above the mean for the poor escapes poverty then this measure perversely suggests higher poverty. (PG does not have this property.)

It should be noted that these measures of poverty impacts and cost-effectiveness can all be calculated from the same data required for the various measures of targeting performance described above. Of course, if one knows the impacts on poverty—which we agree to be the objective—then one does not need the targeting measures. However, since these targeting measures are widely used in assessing antipoverty programs and in comparative work, it is of interest to test their value as indicators for that policy problem.

Program and data

While economic reforms and structural changes in the Chinese economy have meant high rates of economic growth, it is believed that certain sub-groups have been adversely affected or have been unable to participate in the new economic opportunities due to their lack of skills, long-term illness or disability. The collapse of the old safety-net provided by guaranteed employment has clearly left some households vulnerable. Some of the “left behind” households started poor and some became poor, even though aggregate poverty rates have tended to fall over time. Urban areas have figured prominently in these concerns about the “new poor.”

On paper, the *Di Bao* program provides a transfer to all urban households with incomes below a DB line sufficient to bring them up to that line. The scheme became a national policy in 1999 and expanded rapidly; by 2003 participation had leveled off at 22 million people, representing 6% of urban residents. Municipal authorities have considerable power over the program, including setting the DB lines, funding (the center provides partial co-financing) and

implementation. China's cities vary in ways that could well be relevant to the outcomes of *DB*; for example, across the 35 largest urban areas studied in the paper, the highest mean household income per person (the city of Shenzhen) is over four times that of the lowest (Chongqing). The proportion living below the *DB* poverty line varies from 2% (in Fuzhou) to 19% (Haikou).

The analysis uses China's Urban Household Short Survey (UHSS) for 2003/04, as discussed in Chen et al. (2006). The UHSS was done by the Urban Household Survey Division of the National Bureau of Statistics (NBS). I use the UHSS sample for the 35 largest cities, giving a total sample of 76,000, varying from 450 (in Shenzhen) to 12,000 (in Beijing). For these 35 cities, the definitions of geographic areas in the UHSS coincide with those for the *DB* lines and the entire data set has been cleaned by NBS staff and made available for this research.

While the UHSS is a relatively short survey, it allows us to measure a fairly wide range of household characteristics. The survey also included a question on household income and questions were added on *DB* participation and income received from *DB*.

As noted in the last section, in measuring targeting and poverty impacts I assume that income in the absence of the program is observed income less payments received under *DB*. While this is a common assumption, it is clearly questionable. Testing the assumption is difficult without panel data (and even then there can be severe identification problems). With only a single cross-sectional survey it is hard to be confident in the results, given the likelihood of omitted variables correlated with both program placement and the behaviors of interest.

However, I can offer some observations that are at least consistent with this assumption. The design of *DB* intends that the benefits received will decrease as income rises, implying that participants face a positive marginal tax rate. Indeed, if the program works the way it is supposed to then it exactly fills the gap between current non-*DB* income and the *DB* line. Then participants will have no incentive to work (under the usual assumptions that leisure is a normal good and work yields no direct utility). Earned income net of *DB* will fall to zero. The program will have created a poverty trap, whereby participants do not face an incentive to raise their own incomes, because of the loss of benefits under *DB*.

The extent to which this is a real problem in practice is unclear. Benefits are unlikely to be withdrawn quickly. There are reports that local authorities allow *DB* benefits to continue for some period after the participant finds a job (O'Keefe, 2004). Observations from field work also indicate that a notion of "imputed income" was used in a number of provinces. This was a

notional level of income that reflected the potential income given the household labor force; this was apparently done with the aim of minimizing work disincentives.¹⁶

Figure 2 plots DB payment (per capita) against the DB gap, given by the difference between the relevant DB line and income net of DB (both per capita). If the program exactly filled these gaps (when positive) then DB payments would rise with a slope of unity, but would be zero for those with income above the DB line. We see a marked tendency for mean DB payments (conditional on the DB gap) to rise with the DB gap, though the conditional expected value (as measured by a non-parametric regression) has a slope appreciably less than unity. The regression line starts to be noticeably positive at per capita incomes that are about 2,000 Yuan above the DB line and peaks at a mean of around 300 Yuan per capita, at a DB gap of around 4,000 Yuan. (The conditional mean is, of course, positive throughout, but very close to zero below 2,000 Yuan.) Thus Figure 2 suggests that the average benefit withdrawal rate (BWR) — the amount by which mean DB payments change with an extra Yuan of income — is around -0.05; on average, a 100 Yuan increase in income entails a drop of only 5 Yuan in DB payments.

An alternative method of estimating the average benefit withdrawal rate is to regress the per capita DB payment received on income per person less DB receipts, with a complete set of dummy variables for municipalities (to capture the differences in the generosity of the program). The implied BWR is very low, at -0.0012 (t-ratio=-17.51, n=76,808). This does not allow for the censoring that is evident in Figure 2. Using a Tobit regression, the estimate is -0.004 (t=-76.23). Estimating the Tobits separately for each municipality, I obtained statistically significant BWRs in all cases, but all were very low, with none higher (in absolute value) than -0.001.

There is almost certainly attenuation bias in these estimates, due to income measurement errors. There is the usual source of measurement error in asking incomes using only one question, plus the fact that income net of DB payments will probably underestimate income in the absence of DB if there are behavioral responses. To address this concern, I tried an Instrumental Variables Estimator (IVE), in which a set of household-level characteristics (including demographics, education attainments, occupation, housing conditions) are used as instrumental variables for income in estimating the BWR; Chen et al. (2006) provide details on the variables used in the first-stage regressions. Note that this only works for the unconditional

¹⁶ This is based on a personal communication with Philip O’Keefe at the World Bank, drawing on his field-work discussions with local administrators.

regression coefficient of DB payments on pre-DB income, so the instrumental variables are automatically excluded from the main regression of interest; the conditional BWR is unidentified. The IV estimate of the unconditional BWR is -0.0021 ($t=-28.33$), again very low. I also repeated these calculations separately for each municipality, using the IVE for the full sample in each municipality. The estimates were significantly negative for all municipalities and ranged from -0.0102 to -0.0001.

While each of these tests requires an assumption that can be questioned, they all suggest that the benefit withdrawal rate for *Di Bao* is very small. It would thus appear unlikely that the program would provide any serious disincentive for earning income, thus supporting our assumption that income in the absence of DB is simply observed income minus DB payments received. However, at the same time, such a low BWR raises concerns about how well the program reaches the poorest and how well it adapts to changes in household needs. The BWR for the program is almost certainly too low; Kanbur et al. (1995) find that an optimal BWR around one half is consistent with evidence on the relevant income elasticity of labor supply.

Targeting performance and poverty impacts

On calculating all these measures on the same data set and under the same assumptions, one can test the assumption commonly made in policy discussions that better targeting allows a greater impact on poverty and/or a more cost effective antipoverty program. One can also revisit some of the findings from past research on the factors relevant to targeting success. I begin with the aggregate results and then turn to the city-level analysis.

Aggregate results

I find that 7.7% of the total population of the 35 cities had a net income (observed income minus DB receipts) below the DB line (Table 2). The program's total participation is equivalent to about half of the eligible population by this definition. About 40% of DB recipients are ineligible according to these data ($0.43=1.69/3.91$). The proportion of these Type 1 errors amongst the non-poor is clearly very low at 0.018 ($=1.69/92.29$). But there is a high proportion

of Type 2 errors, with almost three-quarters of those who are eligible not being covered by the program ($0.71=5.48/7.71$, i.e., $CR=0.29$).¹⁷

Nonetheless, targeting performance appears to be excellent by international standards, with $S=64\%$, $NS=8.3$ and $CI=0.78$.¹⁸ Coady et al. (2004a,b) provide estimates of NS for 85 programs. Argentina's *Trabajar* program has a $NS = 4.0$, making it the best performer by this measure amongst all programs surveyed by Coady et al.¹⁹ The median NS is 1.25. By this measure, *Di Bao* is a clear outlier in targeting performance internationally.

Turning to the fourth measure of targeting performance I find that while 29% of the poor receive DB, this is only true of about 2% of the non-poor. Thus I find that $TD=0.27$. The mean DB payment across all those with $Y<Z$ is 87.61 Yuan per person per year, while the corresponding mean for those with $Y \geq Z$ is 4.15. The overall mean DB payment across all recipients is 270.33, so $TD^*=0.31$.

These calculations indicate that, while the program is good at concentrating benefits on the poor, it still falls well short of perfect targeting ($TD=1$) in which all of the poor and only the poor are covered, as would be implied by the program's design (as was plain from Figure 2). Another way to see this is to calculate the total receipts for those with net income below the DB line. One then finds that only 12.1% of the aggregate DB gap is filled by the program. DB is a long way off reaching its own aim of bringing everyone up to the DB line.

The weak coverage of the program — in terms of both coverage of those living below the DB line and coverage of the DB gap — is naturally limiting its impact on poverty, despite excellent targeting in the sense of avoiding leakage to the non-poor. Table 3 gives various poverty measures before and after DB transfers. To test robustness to the location of the DB lines, Figure 3 gives the empirical cumulative distribution functions of income — with

¹⁷ Ravallion (2007) proposes a method of testing robustness of such calculations to income measurement errors and implements the method for the same program. The results indicate that, if anything, the survey-based incomes probably understate the coverage of the eligible population.

¹⁸ To calculate CI from the micro data I used the regression-based method of Jenkins (1988).

¹⁹ *Trabajar* is a combination of a workfare program and social fund, whereby participants are offered low-wage work to do things of value to poor communities; see Jalan and Ravallion (2003). Coady et al., calculate the normalized share from *Trabajar* from Jalan and Ravallion (2003) who estimate that 80% of *Trabajar* participants come from the poorest 20% of the Argentine population ranked by income net of *Trabajar* receipts (which are roughly constant across recipients). The corresponding normalized share for the poorest decile is much higher, at about 6.0, though still less than for DB.

households ranked by income normalized by the relevant DB line — with and without DB receipts for both participants and the full (35-city) sample.

The program is having a sizeable impact on poverty amongst the participants (Table 3). The proportion of the participant population falling below the DB line is 45% with DB transfers, but it would have been 57% without them. However, the impact on poverty in the population as a whole is much less. The proportion falling below the DB lines falls from 7.7% to 7.3% after DB transfers. Proportionate impacts are slightly higher for *PG* than for the headcount index (and slightly higher again for the squared poverty gap); this indicates that the program has increased the mean income of those below the DB line and reduced inequality amongst them.

Targeting and poverty across cities of China

Given the scheme's decentralized financing and implementation and the differences observed across China's cities, heterogeneity in outcomes across municipalities is to be expected. There will, of course, be differences in local resources and administrative capabilities, but there will also be (less obvious) differences in the local political economy. Here I only aim to describe the differences in DB performance across municipalities, and to use these differences to assess how well prevailing targeting measures perform in predicting impacts on poverty.

There is considerable variation in targeting performance across municipalities. (The Appendix gives results by city.) *S* varies from 31% to 98%; *NS* varies from 2.8 to 18.8, and *CI* varies from 0.64 to 0.93, while *TD* varies from 0.06 to 0.53. (All cities except one, Kunming, have *NS* higher than the best performing program surveyed by Coady et al., 2004.)

Recall that *TD* automatically gives equal weights to the Type 1 and Type 2 error proportions. The weights for the three measures based on the concentration curve depend on the analytic properties of these measures and how the design features and setting influence the overall program participation rates (as discussed above). A simple way of summarizing this (potentially complex) relationship, is to regress each measure of targeting performance on *T1* and *T2* (as defined in Table 1). Table 4 gives the regressions. *T2* has only a small and statistically insignificant effect on *S*, *NS* or *CI*; by contrast, *T1* has a strong and significant effect for *NS* and *S*; the coefficient is also significant at the 3% level for *S* if one drops *T2*.²⁰

However, the normalizations used in defining *T1* and *T2* have bearing on the weights attached to Type 1 and Type 2 errors. If instead one normalizes by total populations of the

²⁰ The regression coefficient on Type 2 errors is then -1.361 ($t=-2.28$; $\text{prob}=0.03$).

municipality the results change noticeably; thus the regressors become H_0T1 and $(1 - H_0)T2$. As can be seen in Table 4, the three measures based on the concentration curve all attach negative weights to Type 1 errors per capita, but now we find that *NS* also puts a negative weight on Type 2 errors (as does *CI*, although it is not significant). And we find that *S* puts a positive weight on Type 2 errors, while *TD* puts a positive weight on Type 1 errors. Only for *CI* do we find that the weights attached to the two errors of targeting are robust to the normalization. The generalizations found in the literature about the relative importance of Type 1 and Type 2 errors to standard measures of targeting are questionable.

However, the most important point for the present analysis is that none of the measures based on the concentration index have strong correlations with the coverage rate of the poor. The simple correlation coefficients with *CR* are -0.28, -0.30 and -0.40 for *S*, *NS* and *CI* respectively. By contrast, *TD* has a correlation of 0.98 with the coverage rate.

There are some clear covariates of the heterogeneity in targeting performance, echoing some past findings in the literature. There is a high correlation between DB spending and targeting performance as measured by *TD* ($r=0.73$), though the correlation with targeting performance is appreciably weaker for the other measures.²¹ Figure 4 shows the relationship for *TD*. This pattern in the data is consistent with evidence for antipoverty programs in other settings indicating that *TD* tends to improve as programs expand, and to deteriorate in fiscal contractions (Ravallion, 2004). It appears that the early benefits (at a low level of spending) tend to be captured more by the non-poor while the poor benefit more when the program expands and are the first to bear the costs of contractions. As can be seen from Figure 5, the differences in program scale, as measured by participation rates (*P*), are also highly positively correlated with coverage rates for the poor ($r=0.80$). As we will see, this ‘scale effect’ will be key to understanding why measures of targeting performance based on the concentration curve perform so poorly as predictors of poverty impacts.

The impact of higher initial poverty on targeting performance has also been discussed in the literature.²² I can confirm the Coady et al. (2004a,b) finding that the normalized share is higher in richer cities. The correlation coefficient between the normalized share and the pre-DB

²¹ By contrast, the other targeting measures tend to be negatively correlated with DB spending, though only significantly so for *CI* ($r = -0.52$)

²² For a theoretical analysis (using *TD* as the measure of targeting) see Ravallion (1999).

headcount index is -0.81. However, this is entirely due to the normalization; if one uses the ordinary share, S , the correlation is positive and significant ($r=0.55$).²³ (One wonders whether the Coady et al results would be robust to using S , or NS with a uniform H across all programs.)

Impacts on poverty also vary across cities (see the Appendix for details). Subtracting the post-DB poverty rate from the pre-DB rate, the impact on the headcount index varies from 0.0% to 1.5% points. Table 5 provides the correlation coefficients between the targeting measures and the program's impacts on both the headcount index and poverty gap index. The results for SPG were very similar to PG and are omitted for brevity. Correlations are given for both the levels impact (pre-DB poverty measure *less* post-DB measure) and the proportionate impact (normalized by the pre-DB poverty measure).

Amongst the four main measures of targeting performance, the strongest indicator (by far) of the impact on poverty is the targeting differential, TD . Strikingly, I find no sign of a positive correlation between the impacts on poverty and any of the three most popular measures, S , NS and CI . For S the correlation coefficient with the impacts on the level of the headcount index is only 0.03, while for CI and NS the correlation coefficients with poverty impacts turn out to be negative. This switches for proportionate impacts, which are negatively correlated with S but virtually orthogonal to NS . Figures 6 and 7 plot the impacts on the level of the headcount index against the NS and TD measures respectively. We see clearly that municipalities with a higher normalized share going to the poor tended to have lower impacts on poverty (Figure 6). The targeting differential does not have this perverse property (Figure 7).

In Table 5, the correlations are pair-wise. Instead, Table 6 gives regressions of the poverty impacts (columns 1 and 3) on all four measures jointly. (The table also includes regressions that control for spending, which I return to.) TD remains the strongest predictor of poverty impacts. S now emerges as a positive predictor, though still not significant at the 1% level and NS is no longer a significant negative indicator, at given values of the other targeting measures. CI remains a negative predictor of poverty impacts.

Table 5 also gives the correlation coefficients for the incidence of Type 1 and Type 2 errors. Strikingly, one finds that the proportion of Type 1 errors ($T1$) is positively correlated with poverty impacts; cities that did a better job excluding the non-poor tended to do less well in

²³ Using mean income instead one obtains $r=0.57$ for NS and -0.34 for S .

reducing poverty. It is only *T2* that is negatively correlated with impacts.²⁴ I found the same pattern in the partial correlations, by regressing poverty impacts on both the proportions of Type 1 and Type 2 errors; the regression coefficients on the Type 1 (Type 2) error proportions were significantly positive (negative) at the 1% level for the impacts on both *H* and *PG*.

It is now evident that the main reason why the three targeting measures based on the concentration curve perform so badly as indicators of poverty impact is that they are not even positively correlated with the program's coverage of the poor, which is highly correlated with poverty impacts (as can be seen from the correlation coefficients for Type 2 errors in Table 5).

What about cost-effectiveness? Cities with higher DB spending tend to have higher impacts on poverty; the correlation coefficients between DB payments per capita (of the population) and the impacts of *H* and *PG* are 0.80 and 0.86 respectively. The simplest way to test how well the targeting measures predict poverty impacts at given levels of spending is by normalizing poverty impacts by DB spending, to give the cost-effectiveness ratio.²⁵ Table 5 gives the correlation coefficients.

The share going to the poor now emerges as having a positive correlation with cost-effectiveness and statistically significantly so for *PG*. Figure 8 plots the data points in this case; a positive relationship is evident, with a regression coefficient of 0.033 ($t=5.61$, based on a White standard error). Even so, $R^2=0.42$, so that the majority of the variance in cost-effectiveness in reducing *PG* is left unexplained.

This is the exception though. None of the measures show significant correlations with cost-effectiveness in reducing the headcount index. The normalized share still has the perverse negative correlation found for the total poverty impacts.²⁶ The reason why the normalized share, *NS*, emerges as a perverse indicator (even though *S* is uncorrelated with impact, and positively correlated with cost-effectiveness) is that there is a positive correlation between the (pre-

²⁴ Expressing the incidence of targeting errors on a per capita basis gives virtually identical correlations for Type 1 errors. However, the per capita incidence of Type 2 errors is uncorrelated with poverty impacts. This stems from the strong positive correlation between (pre-DB) *H* and poverty impacts (and positive correlation between *H* and Type 2 errors per capita); using a regression to control for *H* a significant negative correlation reemerges between the poverty impacts and Type 2 errors.

²⁵ Data were not available on administrative costs of the program at municipal level, so the spending variable is solely based on transfer payments. The correlation coefficients will only be unaffected if the administrative cost share is the same across different cities.

²⁶ The normalized share is significantly correlated with cost-effectiveness in reducing the income-gap ratio ($r=0.53$), although recall that this is a flawed measure of poverty.

program) headcount index and the program's impact on poverty; the correlation coefficients are 0.36 and 0.51 for H and PG , respectively. The incidence of the two types of errors shows little or no correlation with the cost-effectiveness ratios.

A more flexible way of seeing whether the targeting measures reveal poverty impacts at given spending is to use a regression of poverty impacts on the targeting measures with controls for spending. To allow for nonlinearity in a reasonably flexible way, I used a cubic function of DB spending per capita. Testing each targeting measure one by one, only S turns out to be a significant (positive) predictor of impacts on poverty at given program spending; the correlation was significant at the 4% level for H and 1% level for PG . The results using the four measures together are given in Table 6 (columns 2 and 4; for PG the higher order terms can be dropped, giving column 5). Again, one finds that only S is a significant predictor of impacts on poverty at given program spending, though now this holds for for H at the 2% level as well as PG .

Conclusions

The three most popular measures of targeting performance found in practice are the share of transfers going to the poorest $H\%$, the share normalized by H , and the concentration index. Each has its strengths and weaknesses. However, I find that none of these measures reveal much about the success of China's *Di Bao* program in achieving its objective of eliminating extreme urban poverty. Only one of the measures studied here, namely the targeting differential, has a statistically significant positive correlation with the program's poverty impacts. The cities of China that are better at targeting this program are generally not the ones where the scheme came closest to attaining its objective.

Possibly more surprising, I find that these measures are not very informative about the program's cost-effectiveness, i.e., poverty impact at given program spending. The one exception is that the share going to the poor is a statistically significant predictor of cost effectiveness in reducing the poverty gap index. But, even then, about 60% of the variance in the cost-effectiveness ratio is left unexplained. All other measures perform poorly, or even perversely.

These findings echo some of the warnings in the literature against relying on standard measures of targeting performance for informing policy choices concerning antipoverty programs. The paper's findings also cast doubt on the generalizations found in the literature about what type of program "works best," and so should be scaled up, based on cross-program

comparisons of targeting measures. The external validity of these programmatic comparisons is clearly questionable if the targeting measures have such a poor fit with poverty impacts. It is also unlikely that past findings on the socio-economic factors influencing targeting performance at country level are robust to seemingly arbitrary differences in the measures used.

One question is left begging: Why have the literature's warnings carried so little weight in practice? Possibly the more "theoretical" objections to these targeting measures have fallen on deaf ears for lack of clear evidence on how the measures perform in practice. The results of this case study will then help. One can also conjecture that the preference for targeting measures that put a high weight on avoiding leakage to the non-poor stems from fiscal pressures, given that reducing leakage helps cut public spending, while expanding coverage does the opposite. While I do not doubt that such thinking has had influence at times, it is surely misguided. For if the problem was to minimize public spending (unconditionally) then why would governments bother with such programs in the first place? Evidently there is a demand for these policies, as part of a comprehensive antipoverty strategy. A more credible characterization of the policy problem would then give positive weight to both avoiding leakage and expanding coverage of the poor.

From that perspective, measures of targeting performance that penalize both errors of targeting make more sense—again echoing recommendations found in the literature. However, that conclusion would still miss the point. Even the targeting measure studied here that is found to be the best predictor of poverty impacts is a long way from being a perfect indicator. If there is a single message from this study it is that analysts and policy makers would be better advised to focus on the estimable outcome measures most directly relevant to their policy problem. In the present context, impacts on poverty can be assessed with the same data and under the same assumptions as required by prevailing measures of targeting performance.

References

- Atkinson, Anthony B., 1995, "On Targeting and Social Security: Theory and Western Experience with Family Benefits," in Dominique van de Walle and Kimberly Nead (eds) *Public Spending and the Poor*, Johns Hopkins University Press for the World Bank.
- Besley, Timothy and Ravi Kanbur, 1993, "Principles of Targeting," in Lipton, Michael and Jacques van der Gaag (eds.) *Including the Poor*. Washington D.C.: World Bank.
- Chen, Shaohua, Martin Ravallion and Youjuan Wang, 2006, "Di Bao: A Guaranteed Minimum Income in China's Cities?" Policy Research Working Paper, WPS 3805, Washington DC, World Bank.
- Coady, David, Margaret Grosh and John Hoddinott, 2004a, "Targeting Outcomes Redux," *World Bank Research Observer* 19(1): 61-86.
- _____, _____ and _____, 2004b, *Targeting Transfers in Developing Countries: Review of Lessons and Experience*, Washington DC: World Bank.
- Cornia, Giovanni and Frances Stewart, 1995, "Two Errors of Targeting," in Dominique van de Walle and Kimberly Nead (eds) *Public Spending and the Poor*, Johns Hopkins University Press for the World Bank.
- De Donder, Philippe, and Jean Hindriks, 1998, "The Political Economy of Targeting," *Public Choice* 95: 177-200.
- Duclos, Jean-Yves, 1995, "On the Equity Aspects of Imperfect Income Redistribution," *Review of Income and Wealth* 41(2): 177-190.
- Foster, James, J. Greer, and Erik Thorbecke, 1984, "A Class of Decomposable Poverty Measures," *Econometrica* 52: 761-765.
- Galasso, Emanuela and Martin Ravallion, 2005, "Decentralized Targeting of an Antipoverty Program," *Journal of Public Economics* 85: 705-727.
- Gelbach, Jonah and Lant Pritchett, 2000. "Indicator Targeting in a Political Economy: Leakier can be Better. *Journal of Policy Reform* 4: 113-45.
- Grosh, Margaret, 1992, "The Jamaican Food Stamps Program: A Case Study in Targeting," *Food Policy* 17(1): 23-40.
- _____, 1994, *Administering Targeted Social Programs in Latin America: From Platitudes to Practice*. Washington DC: World Bank.
- _____, 1995, "Toward Quantifying the Trade-Off: Administrative Costs and Incidence

- in Targeted Programs in Latin America,” in Dominique van de Walle and Kimberly Nead (eds) *Public Spending and the Poor*, Johns Hopkins University Press for the World Bank.
- Jalan, Jyotsna and Martin Ravallion, 2003, “Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching,” *Journal of Business and Economic Statistics*. 21(1): 19-30.
- Jenkins, Stephen, 1988, “Calculating Income Distribution Indices from Micro-Data,” *National Tax Journal* 41(1): 139-42.
- Kanbur, Ravi , Michael Keen and Matti Tuomala, 1995, “Labor Supply and Targeting in Poverty-Alleviation Programs,” in Dominique van de Walle and Kimberly Nead (eds) *Public Spending and the Poor*, Johns Hopkins University Press for the World Bank.
- Lambert, Peter, 2001, *The Distribution and Redistribution of Income, Third Edition*, Manchester: Manchester University Press.
- Lanjouw, Peter and Martin Ravallion, 1999, “Benefit Incidence and the Timing of Program Capture” *World Bank Economic Review*, 13(2):.257-274.
- Mateus, Abel, 1983, *Targeting Food Subsidies for the Needy: The Use of Cost-Benefit Analysis and Institutional Design*. World Bank Staff Working Paper 617, Washington DC.
- Murgai, Rinku and Martin Ravallion, 2005, “Is a Guaranteed Living Wage a Good Antipoverty Policy?” Policy Research Working Paper 3640, World Bank, Washington DC.
- O’Keefe, Philip, 2004, “Social Assistance in China: An Evolving System,” mimeo, World Bank, Washington DC.
- Ravallion, Martin, 1994, *Poverty Comparisons*, Harwood Academic Press, Chur: Switzerland.
- _____, 1999, “Are Poorer States Worse at Targeting their Poor?” *Economics Letters*, 65: 373-377.
- _____, 2000, “Monitoring Targeting Performance when Decentralized Allocations to the Poor are Unobserved,” *World Bank Economic Review* 14(2): 331-45.
- _____, 2004, “Who is Protected from Budget Cuts?” *Journal of Policy Reform* 7(2): 109-22.
- _____, 2007, “Miss-Targeting or Miss-Measurement?” *Economics Letters*, forthcoming.
- Ravallion, Martin and Calvin Chao, 1989, “Targeted Policies for Poverty Alleviation Under

- Imperfect Information: Algorithms and Applications,” *Journal of Policy Modeling* 11(2): 213-224.
- Ravallion, Martin and Gaurav Datt, 1995, “Is Targeting Through a Work Requirement Efficient? Some Evidence for Rural India,” in Dominique van de Walle and Kimberly Nead (eds) *Public Spending and the Poor: Theory and Evidence*, Baltimore: Johns Hopkins University Press.
- Stifel, David and Harold Alderman, 2005, “Targeting at the Margin: The ‘Glass of Milk’ Subsidy Programme in Peru,” *Journal of Development Studies* 41(5): 839-864.
- Smolensky, Eugene, Siobhand Reilly and Erik Evenhouse, 1995, “Should Public Assistance be Targeted?” *Journal of Post Keynesian Economics* 18: 3-28.
- van de Walle, Dominique, 1998, “Targeting Revisited,” *World Bank Research Observer*, 13(2): 231-48.
- Weisbrod, Burton, 1970, “Collective Action and the Distribution of Income: A Conceptual Approach,” in R. H. Haverman and J. Margolis (eds.) *Public Expenditure and Policy Analysis*, Chicago, Ill.: Markham. (Reprinted in Nicholas Barr (ed.) *Economic Theory and the Welfare State, Volume 2*, Edward Elgar, Cheltenham, UK, 2001.)

Table 1: Measures of targeting and poverty

Measure	Definition	Formula
Targeting measures		
Concentration curve	Share of total transfers going to the poorest $p\%$ of the population ranked by household income per person.	$C(p) = \frac{1}{\bar{t}} \int_0^p t(x) dx$ where $t(x)$ is the transfer to quantile x ranked by income per person and \bar{t} is the mean transfer.
Share going to the poor (S)	Share of transfers going to those who are initially deemed poor (or other reference group based on income).	$S = C(H_0)$ where H_0 is the pre-program headcount index of poverty (see below).
Normalized share (NS)	Share going to the poor divided by proportion who are poor.	$NS = C(H_0) / H_0$
Concentration index (CI)	Area between the concentration curve and the diagonal (along which everyone receives the same amount).	$CI = 2 \int_0^1 C(p) dp - 1$
Coverage rate (CR)	Program participation rate for the poor.	$CR = N(D=1, Y < Z) / N(Y < Z)$
Targeting differential (TD)	Difference between the coverage rate and the participation rate for the non-poor.	$TD = \frac{N(D=1, Y < Z)}{N(Y < Z)} - \frac{N(D=1, Y \geq Z)}{N(Y \geq Z)}$ where $N(D=1, Y < Z)$ is the number of people who are both poor and receiving the transfers (other variables defined similarly).
Proportion of Type 1 errors ($T1$)	Proportion of (ineligible) non-poor who are assigned the program.	$T1 = N(D=1, Y \geq Z) / N(Y \geq Z) = (1-S)P / (1-H_0)$; $P = N(D=1) / N$
Proportion of Type 2 errors ($T2$)	Proportion of the poor who fail to receive the program.	$T2 \equiv N(D=0, Y < Z) / N(Y < Z) = 1 - CR$
Poverty measures		
Headcount index (H)	The proportion of the population living in households with income per person less than the poverty line.	$H = F(z)$ where F is the cumulative distribution function (CDF) for incomes and z is the poverty line.
Poverty gap index (PG)	Mean distance below the poverty line as a proportion of the line where the mean is taken over the whole population, counting the non-poor as having zero gap.	$PG = \int_0^H (1 - y(p) / z) dp$ where $y(p)$ is the quantile function (inverse of CDF) giving income of the p th quantile.
Income gap ratio (I)	Mean distance below the poverty line as a proportion of the line, amongst the poor alone.	$I = 1 - PG / H$
Squared poverty gap index (SPG)	As for PG except that the proportionate poverty gaps are weighted by themselves.	$SPG = \int_0^H (1 - y(p) / z)^2 dp$
Poverty impact	Pre-intervention poverty measure minus post-intervention measure	$H_0 - H_1$ and similarly for PG and SPG

Note: The table only refers to the measures used in this paper. The text contains further discussion of the pros and cons of each measure.

Table 2: Leakage and coverage of the *Di Bao* program

% of population	Net income below DB line		Total
	Yes	No	
Receiving DB	2.22	1.69	3.91
Not receiving DB	5.48	90.60	96.09
Total	7.71	92.29	100.00

Note: n=76,443 (for the 35 municipalities).

Table 3: Impacts on aggregate poverty measures for urban China

	Poverty measures (%)	
	Before <i>Di Bao</i> (income net of DB receipts)	After <i>Di Bao</i> (income including DB receipts)
<i>(a) Population (participants + non-participants)</i>		
Headcount index (%)	7.71	7.26
Poverty gap index (%)	2.28	2.06
Squared poverty gap index (x100)	1.02	0.88
<i>(b) Participants only</i>		
Headcount index (%)	56.85	45.49
Poverty gap index (%)	19.92	14.23
Squared poverty gap index (x100)	10.21	6.44

Table 4: Targeting measures regressed on Type 1 and Type 2 errors

	Share (S)		Normalized share (NS)		Concentration index (CI)		Targeting differential (TD)	
Constant	0.560 (2.43; 0.02)	0.542 (10.97; 0.00)	9.689 (1.89; 0.07)	16.355 (13.54; 0.00)	0.902 (11.79; 0.00)	0.871 (27.71; 0.00)	1	0.288 (8.65; 0.00)
Prop. Type 1 errors	-0.938 (-1.04; 0.30)	n.a.	-0.390 (-2.21; 0.03)	n.a.	-1.354 (-4.59; 0.00)	n.a.	-1	n.a.
Prop. Type 2 errors	0.138 (0.42; 0.67)	n.a.	0.015 (0.21; 0.84)	n.a.	-0.052 (-0.49; 0.63)	n.a.	-1	n.a.
Type 1 errors per capita	n.a.	-1.044 (-1.81 0.08)	n.a.	-0.672 (-6.88; 0.00)	n.a.	-1.382 (-6.77; 0.00)	n.a.	2.165 (3.34; 0.00)
Type 2 errors per capita	n.a.	2.728 (4.76; 0.00)	n.a.	-0.903 (-5.16; 0.00)	n.a.	-0.153 (-0.35; 0.73)	n.a.	-1.000 (-2.72; 0.01)
R ²	0.097	0.412	0.141	0.639	0.369	0.365	1	0.449

Note: t-ratios and prob. values in parentheses, based on White standard errors; n=35.

Table 5: Is targeting performance correlated with poverty impacts?

Correlation coefficients	Impact on poverty measure		Proportionate impact on poverty (normalized by pre-transfer value)		Cost-effectiveness ratio (impact on poverty per unit spending)	
	Headcount index	Poverty gap index	Headcount index	Poverty gap index	Headcount index	Poverty gap index
<i>(a) Measures of targeting performance</i>						
Share of spending going to the poor (<i>S</i>)	0.03	0.07	-0.26	-0.33	0.32	0.65*
Share going to the poor normalized by headcount index (<i>NS</i>)	-0.44*	-0.53*	0.04	0.12	-0.21	-0.42*
Concentration index (<i>CI</i>)	-0.40	-0.40	-0.19	-0.17	0.04	0.26
Targeting differential (<i>TD</i>)	0.61*	0.65*	0.63*	0.74*	0.13	-0.01
<i>(b) Type 1 and Type 2 errors</i>						
Prop. of Type 1 errors (<i>T1</i>)	0.63*	0.72*	0.44*	0.51*	-0.06	-0.11
Prop. of Type 2 errors (<i>T2</i>)	-0.66*	-0.71*	-0.63*	-0.75*	-0.09	0.03

Note: n=35; * indicates significant at 1% level.

Table 6: Which targeting measure best predicts poverty impacts and cost-effectiveness?

	Impact on headcount index		Impact on poverty gap index		
	(1)	(2)	(3)	(4)	(5)
Constant	1.092 (1.59; 0.12)	0.401 (0.86; 0.40)	0.540 (1.78; 0.09)	0.017 (0.10; 0.92)	0.001 (0.00; 0.99)
Share of spending going to the poor (<i>S</i>)	0.722 (2.53; 0.02)	0.608 (2.56; 0.02)	0.413 (2.08; 0.05)	0.343 (2.49; 0.02)	0.338 (2.65; 0.01)
Normalized share going to the poor (<i>NS</i>)	-1.056 (-0.92; 0.37)	-0.306 (-0.46; 0.65)	-1.033 (-2.13; 0.04)	-0.706 (-1.77; 0.08)	-0.692 (-1.85; 0.07)
Concentration index (<i>CI</i>)	-1.800 (-2.02; 0.05)	-1.181 (-1.87; 0.07)	-0.881 (-1.80; 0.08)	-0.188 (-0.62; 0.54)	-0.183 (-0.67; 0.51)
Targeting differential (<i>TD</i>)	1.613 (4.61; 0.00)	-0.072 (-0.15; 0.88)	0.932 (5.92; 0.00)	0.177 (0.97; 0.34)	0.176 (1.04; 0.31)
Program cost	n.a.	1.099 (3.11; 0.00)	n.a.	0.125 (0.51; 0.62)	0.179 (4.65; 0.00)
Program cost ²	n.a.	-0.441 (-2.44; 0.02)	n.a.	0.044 (0.24; 0.81)	n.a.
Program cost ³	n.a.	0.064 (2.51; 0.02)	n.a.	-0.007 (-0.25; 0.80)	n.a.
R ²	0.540	0.764	0.643	0.849	0.847

Note: t-ratios and prob. values in parentheses, based on White standard errors; n=35.

Figure 1: Targeting measures based on the concentration curve

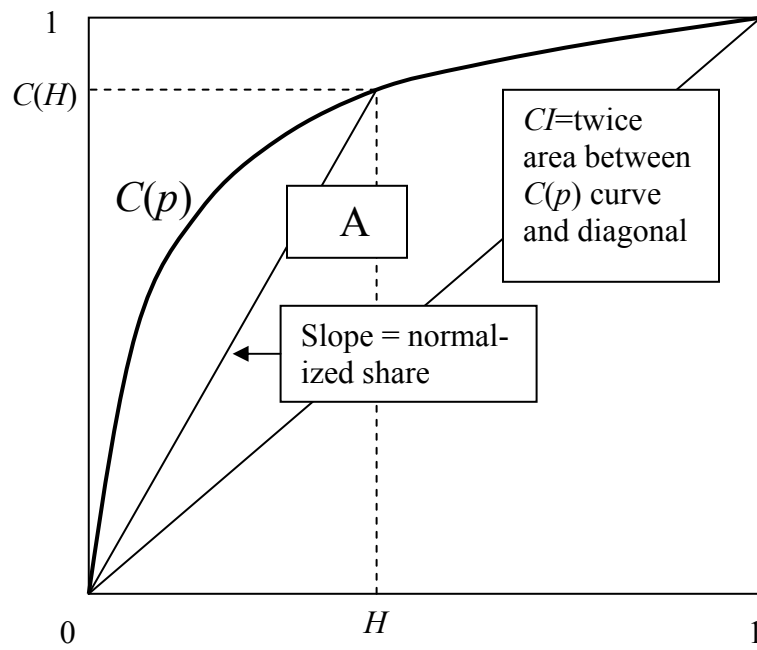
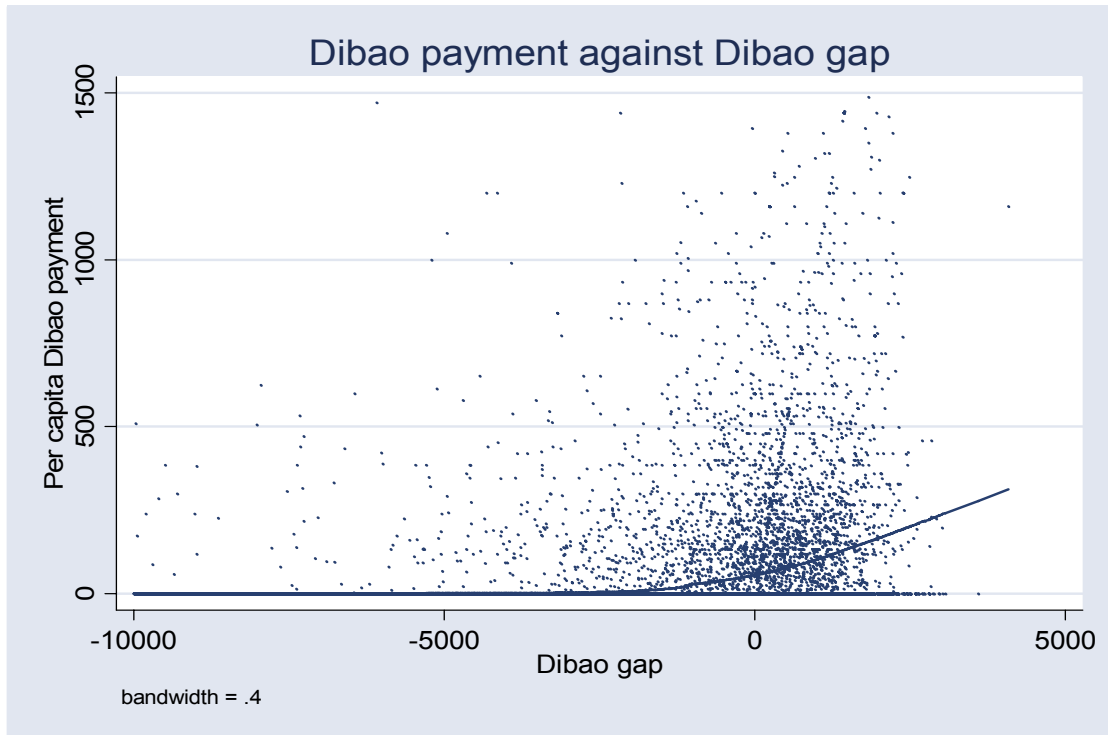


Figure 2: Di Bao payment received plotted against Di Bao gap



Note: Extreme values trimmed (after estimation).

Figure 3: Impacts of the program on poverty

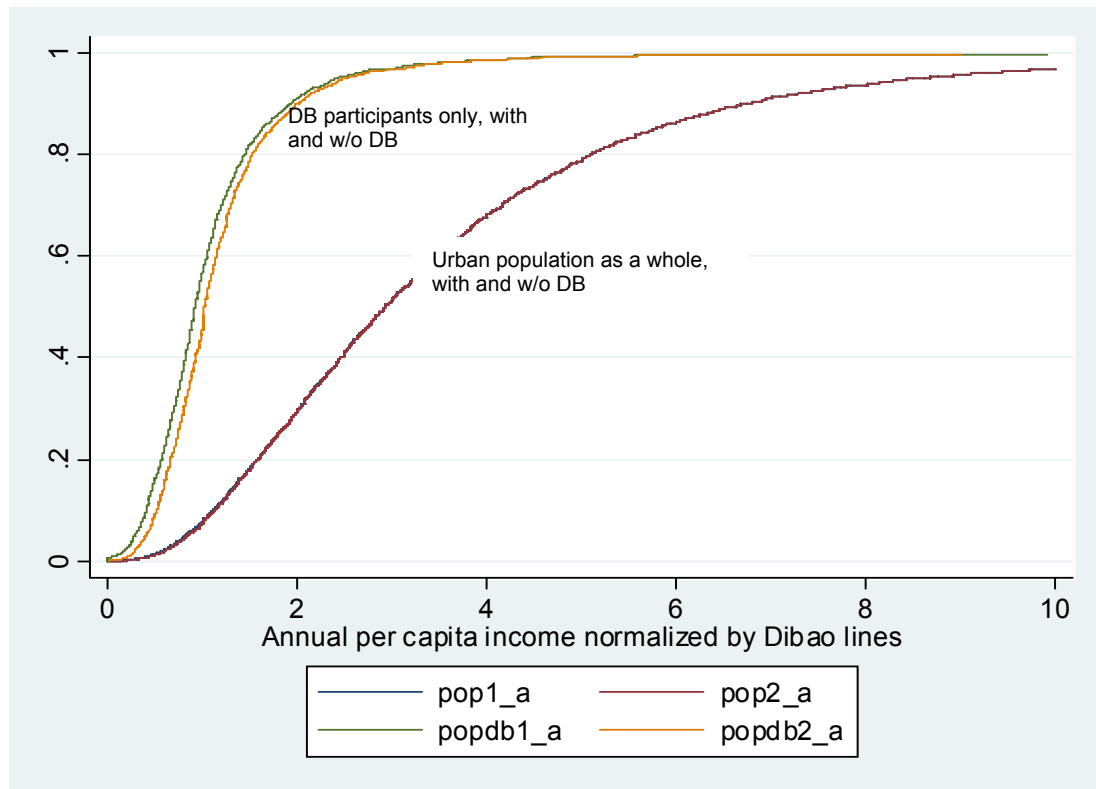


Figure 4: Targeting differential plotted against Di Bao spending per capita across cities

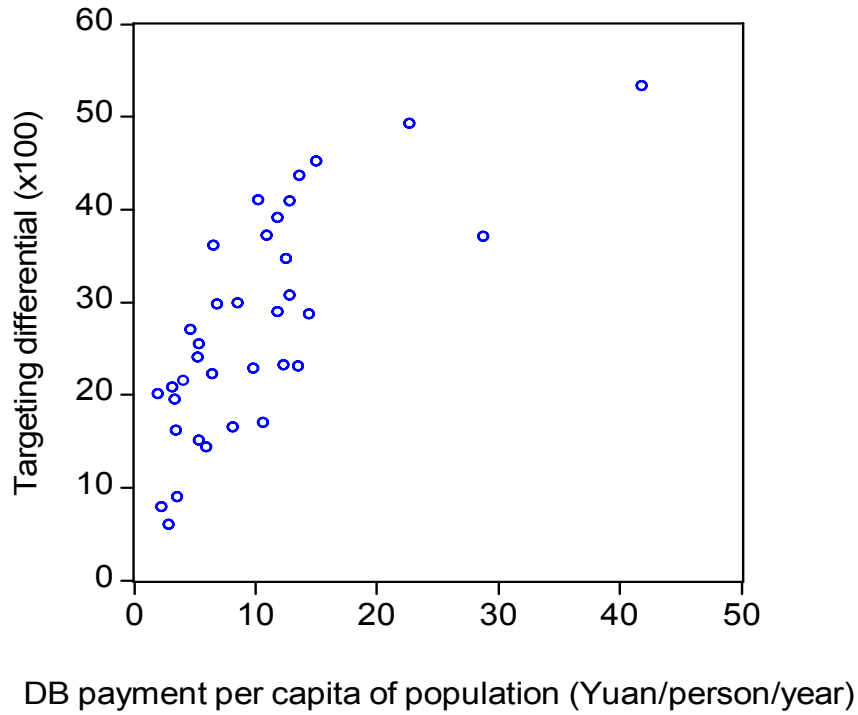


Figure 5: Coverage of the poor against coverage of the population

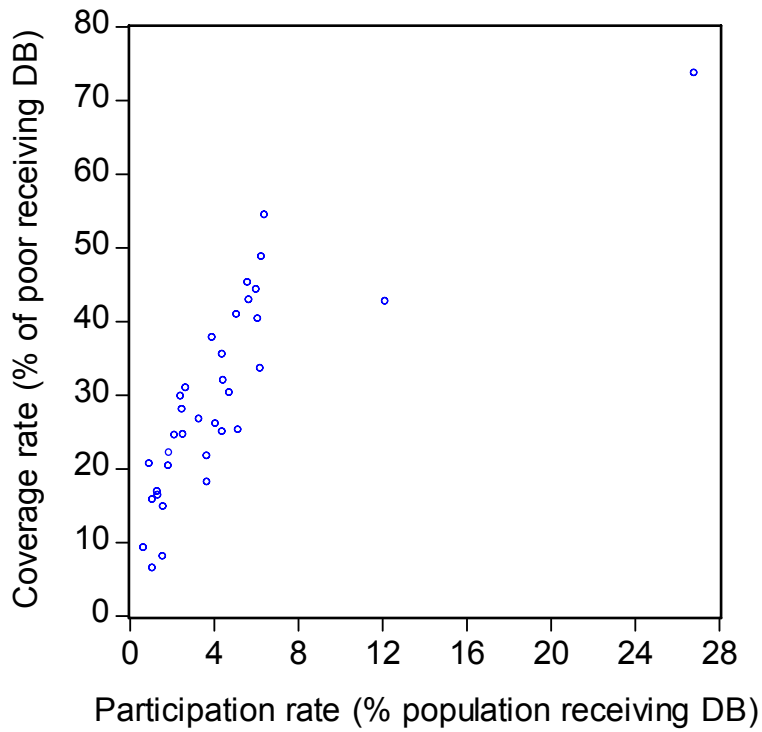


Figure 6: Impact of *Di Bao* on poverty plotted against the *SHARE measure of targeting performance across cities**

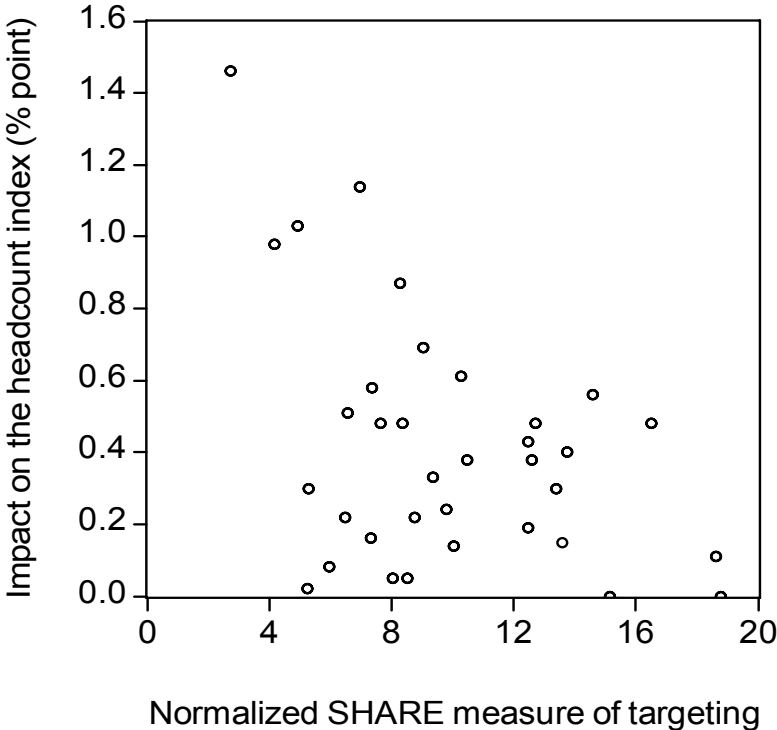


Figure 7: Impact of *Di Bao* on poverty plotted against the targeting differential

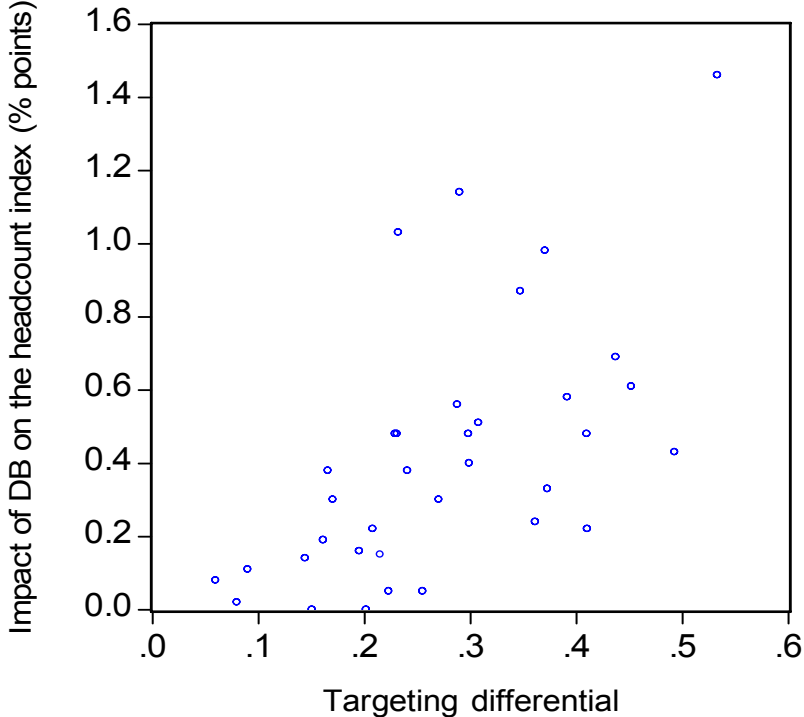
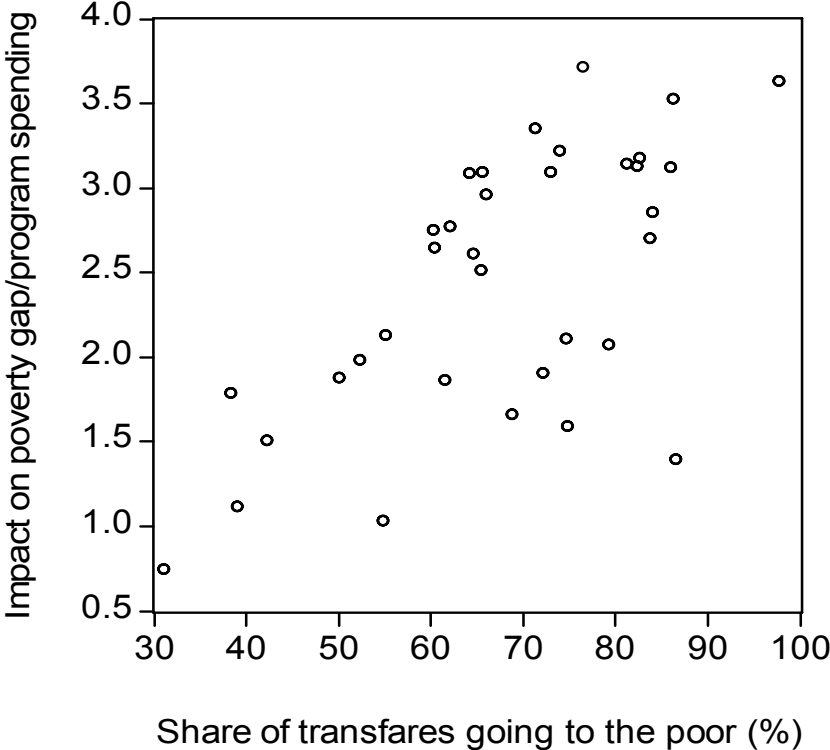


Figure 8: Cost-effectiveness for poverty gap plotted against share going to the poor



Appendix: Targeting and impacts on poverty by municipal areas of China

City	Participation rate (% of pop. with DB)	Targeting measures				Poverty impacts			Squared poverty gap (<i>SPG</i>)			
		Prop. of poor participating (CR)	Share going to the poor (<i>S</i>)	Concentration index (<i>CI</i>) (x100)	Normalized share (<i>NS</i>)	Targeting differential (<i>TD</i>) (x100)	Headcount index (<i>H</i>)	Poverty gap (<i>PG</i>)	Pre-DB	Post-DB	Pre-DB	Post-DB
Beijing	2.53	41.96	54.78	86.10	12.71	23.11	4.31	3.83	1.09	0.95	0.44	0.38
Tianjin	6.26	46.57	61.52	82.40	10.29	45.21	5.98	5.37	1.39	1.11	0.54	0.37
Shijiazhuang	3.29	65.24	64.55	81.52	8.05	25.52	8.02	7.97	2.34	2.20	1.04	0.93
Taiyuan	2.49	62.40	73.98	89.99	13.38	27.07	5.53	5.23	1.30	1.15	0.62	0.52
Huhehaote	1.08	55.88	55.03	73.12	5.96	6.00	9.24	9.16	3.19	3.13	1.55	1.50
Shenyang	4.74	75.49	82.28	85.12	6.97	29.00	11.81	10.67	3.24	2.87	1.41	1.16
Dalian	3.67	74.32	79.24	79.49	5.28	17.06	15.02	14.72	4.62	4.40	2.09	1.91
Chuangchun	4.40	43.69	65.53	87.49	8.51	22.31	7.70	7.65	2.15	1.95	1.04	0.84
Harbin	5.15	65.24	65.46	73.67	4.92	23.21	13.30	12.27	4.03	3.72	1.74	1.56
Shanghai	6.41	21.17	31.09	76.88	12.49	49.28	2.49	2.06	0.58	0.41	0.22	0.13
Nanjing	2.66	63.29	74.61	85.66	13.74	29.93	5.43	5.03	1.35	1.17	0.51	0.41
Hangzhou	0.65	66.67	86.53	89.36	18.61	9.05	4.65	4.54	1.18	1.13	0.54	0.51
Ningbo	2.42	57.97	68.82	91.75	14.61	28.78	4.71	4.15	1.27	1.03	0.56	0.44
Hefei	5.66	70.20	81.25	88.04	8.77	41.05	9.26	9.04	2.56	2.24	1.03	0.83
Fuzhou	0.93	50.00	42.25	68.70	18.78	20.19	2.25	2.25	0.52	0.49	0.18	0.17
Xiamen	2.13	79.41	72.08	87.22	10.48	24.08	6.88	6.50	1.91	1.81	0.79	0.74
Nanchang	4.44	55.26	64.10	82.66	8.36	29.83	7.67	7.19	2.10	1.89	0.88	0.73
Jinan	4.39	84.05	85.97	86.59	8.27	34.75	10.39	9.52	2.95	2.56	1.28	1.03
Qingdao	1.59	77.65	83.75	88.82	10.05	14.44	8.33	8.19	2.40	2.24	1.03	0.90
Zhengzhou	1.33	81.61	82.56	84.79	12.47	16.14	6.62	6.43	2.07	1.96	0.94	0.87
Wuhan	5.59	75.28	84.07	89.32	9.05	43.73	9.29	8.60	2.86	2.47	1.29	1.03
Changsha	6.02	48.21	50.07	76.95	7.64	40.99	6.55	6.07	2.09	1.85	0.92	0.78
Guangzhou	1.31	76.71	74.73	85.41	12.60	16.59	5.93	5.55	1.78	1.65	0.83	0.73
Shenzhen	1.08	37.50	39.08	92.74	15.15	15.09	2.58	2.58	0.92	0.86	0.42	0.40
Nanning	3.66	78.76	86.24	83.73	6.50	20.81	13.26	13.04	5.06	4.95	2.57	2.48
Haikou	1.58	95.65	97.70	93.19	5.24	8.00	18.64	18.62	6.18	6.10	2.95	2.87
Chongqing	12.13	61.61	72.98	74.59	4.17	37.07	17.49	16.51	5.74	4.85	2.66	2.04
Chengdu	1.84	57.89	38.34	81.59	7.34	19.55	5.22	5.06	1.63	1.57	0.78	0.74

Guiyang	6.20	58.98	71.23	85.57	6.55	30.78	10.87	10.36	3.75	3.32	1.83	1.50
Kunming	26.81	33.18	52.30	63.71	2.75	53.31	12.07	10.61	4.03	3.20	2.10	1.54
Xian	4.08	25.53	60.37	84.77	16.52	22.92	3.99	3.51	1.02	0.76	0.41	0.25
Lanzhou	5.08	68.04	65.92	84.43	7.36	39.16	8.44	7.86	2.36	2.01	1.08	0.84
Xining	3.92	59.17	62.12	86.33	9.82	36.12	6.14	5.90	1.91	1.73	0.92	0.79
Yinchuan	6.09	54.02	60.29	80.05	9.37	37.28	8.16	7.83	2.62	2.32	1.34	1.10
Wulumuqi	1.87	67.01	76.48	92.51	13.58	21.53	5.63	5.48	1.91	1.76	1.01	0.85
Mean	3.91	56.85	63.82			27.04	7.71	7.26	2.28	2.06	1.02	0.88