Spatial Decentralization and Program Evaluation: Theory and an Example

Mark M. Pitt, Brown University
Nidhiya Menon, Brandeis University

Version: February 14, 2018

Abstract

This paper proposes an instrumental variable method for program evaluation that only requires a single cross-section of data on the spatial intensity of programs and outcomes. The instruments are derived from a simple theoretical model of government decision-making in which governments are responsive to the attributes of places and their populations, rather than to the attributes of individuals, in making allocation decisions across space, and have a social welfare function that is spatially weakly separable, that is, that the budgeting process behaves as if it is multi-stage with respect to administrative districts and sub-districts. The spatial instrumental variables model is then estimated and tested with a single cross-section of Indonesian census data. The results offer support to the identification strategy proposed but also highlight some critical issues affecting validity.

Keywords: Spatial Decentralization, Program Evaluation, Instrumental Variables
JEL Classification Number: C21, H44, O12, C50

Thanks to Moshe Buchinsky and Andrew Foster and to participants at the North American Summer Meeting of the Econometric Society at Duke, the Northeast Universities Development Consortium Conference at MIT, the Royal Economic Society Meetings at Cambridge University, and the 2014 AEA meetings. The usual disclaimer applies. Contact information: Mark M. Pitt, Professor Emeritus of Economics and Research Professor of Population Studies, Brown University, Box 1836, Providence, RI 02912. Tel: 401.863.2445. Email: Mark_Pitt@brown.edu. Nidhiya Menon, Associate Professor, Department of Economics & IBS, MS 021, Brandeis University, Waltham, MA 02454. Tel: 781.736.2230. Email: nmenon@brandeis.edu.

1. Introduction

Governments in developing countries earmark significant proportions of their budget towards establishing programs that seek to alter the behavior of target populations. By influencing fertility, health, and schooling outcomes, these programs are often the government's main tools for spreading economic well-being and for spurring economic growth. A fundamental problem in program evaluation is that the coverage of programs and the timing of program initiatives -- program placement -- are not likely to be random to the extent that governmental decision rules are responsive to attributes of the targeted populations that are not measured in the data. Simple measured associations between programs and program outcomes, anticipated or unanticipated, will therefore not provide correct estimates of program effects. Data on the spatial distribution of programs and population characteristics at more than one point in time can be used to identify program effects with relatively simple methods (fixed effects) when program placement depends on unmeasured time-persistent or permanent characteristics of locations but varies as a function of aggregate economy-wide trends or shocks.[1] The longitudinal data required for fixed effects estimation are not always available in developing countries, or are too closely spaced so that program change is small relative to noise, and the assumption of the time invariance of the confounding unobservable component may not always hold. This paper illustrates an application of an instrumental variable method for program evaluation that only requires a single cross-section of data on the spatial intensity of programs and outcomes. The instruments are derived from a theoretical model of government decision-making that requires that the government's social welfare function is spatially weakly separable, that is, that the budgeting process behaves as if it is multi-stage with respect to administrative districts and sub-districts. The spatial instrumental variables model is then estimated and tested with a cross-section of Indonesian census data in order to illustrate how existing spatial methods may be used to address local public finance questions.

The spatial instruments illustrated in this paper can be derived from models that optimize household and government behavior subject to resource and information constraints. The idea is that the nature of the budgetary process with spatially-defined administrative districts generates a set of exclusion restrictions that can be used as instruments for the observed allocation of public

[1] Other approaches to identification with single-cross sections of data make use of natural or quasi-experiments (e.g. Pitt and Khandker 1998), regression discontinuity designs (e.g. van der Klaauw 2002), or make use of the time history of program placement (Duflo 2001).

programs across space. Governments are, rather innocuously, assumed to be responsive to the attributes of places and their populations, rather than to the attributes of individuals, in making allocation decisions across space. If the attributes of a district rather than individual characteristics decide program placement, and the means (or higher moments) of outcomes for all "competing" districts enter into the government's social welfare function, then the district means of individual and district exogenous determinants of these outcomes in *other* districts may be used as instruments for the placement of programs in any particular district. The intuition underlying this is the same as the notion of strategic interaction between decentralized bodies of government in the public finance literature where spatial attributes of places matter in decision making (Besley and Case 1995, Brueckner 2003). The assumption of weak separability of a social welfare function having as argument the mean outcomes of every administrative unit (district) is sufficient to generate *spatially decentralized budgeting*, an allocation process that yields restrictions sufficient for identification. The separability assumption says that the marginal rate of substitution in human capital between any two locations is independent of all other locations, thus allowing for the existence of aggregator functions by area. With separability, the process can be thought of as a multi-stage budgeting process where allocation decisions are made sequentially at progressively lower levels of government. Decision-making of this type will arise as a consequence of the costliness of acquiring and processing information on the returns to and health status of every single uniquely identified household or spatial aggregation of households. If characteristics of locations (which may be defined by political boundaries or other criteria as we discuss below) enter the government's operand to influence program placement, with budget constraints, competition for resources will arise at increasingly decentralized levels of local authority. Program intensity in an area will hence depend on the characteristics of other areas, leading to exclusion restrictions for identification. The assumptions of this model, set out in detail below, are not specific to Indonesia or developing countries more generally. Although it simplifies consideration of the spatial allocation process to think of it as a process in which larger administrative districts allocate a budget to smaller districts who in turn allocate them to even smaller administrative units*, it is not required that the government actually act in this manner*. What is required is that the social planner's program allocation decision is *spatially informed* -- knowing the location of places (districts) matters in the allocation decision.

The type of model implied by the theory is known as a *spatial-x model* in the spatial econometrics literature (Anselin 1988). Related versions of the model in this literature have been used to estimate cross-sectional models that include a lagged dependent variable on the right hand side with or without an error structure that is spatially autoregressive (Kelejian and Prucha 1998). Other methods of estimating models with spatial autoregressive errors consistently include a dynamic spatial GMM estimator and an approach that combines spatial and non-spatial instruments (Baltagi *et al.* 2014). Our innovation is to translate these models from the spatial econometrics literature to demonstrate that spatial instruments may be used in the program evaluation field. We apply the intuition from existing spatial techniques to data from Indonesia in order to demonstrate the manner in which spatial instrument methods can work. In particular, we use data from the 1980 Population Census of Indonesia merged with detailed information on the presence of government programs in all of the villages of Indonesia that were collected as part of the Population Census. In order to illustrate the manner in which spatial instrument methods may be applied in program evaluation, we assume some type of spatial dependence from an exogenous predictor in other sub-districts of Indonesia to the programs available in any particular sub-district. Three sets of spatial instruments are constructed defined by both spatial contiguity (shared boundaries) and by membership in a larger administrative unit. Four outcomes (girls' and boys' school enrollment, recent fertility, and contraceptive use) and four programs (primary schools, secondary schools, family planning clinics, and public health clinics) are investigated. The spatial instrumental variables results generate significantly different estimated program effects as compared to models that assume programs are placed exogenously, thus demonstrating the manner in which spatial techniques may be used to evaluate programs when standard methods fail. Hansen-Sargan tests of overidentification and tests of instrument orthogonality and instrument redundancy are reported. Identification is threatened by *spatial network effects* that hold that exogenous predictors in other sub-districts directly affect outcomes in any particular sub-district conditional on program availability. Below we estimate models that identify program effects even allowing for spatial network effects at varying levels of spatial distance, and use the estimates to test for spatial network effects and the efficacy of the identification strategy. As always in instrumental variable models, some assumptions are untestable, but we develop some empirical criteria that at least suggest whether the identification of program effects from two-stage budgeting is reasonable in an empirical setting.

Section 2 of the paper sets out a theory of household and government decision-making that generates the required exclusion restrictions. Section 3 discusses issues of empirical implementation, and Section 4 describes the data and variable construction. In Section 5, models of program evaluation with spatial instruments are estimated. Models that permit spatial network effects across neighboring sub-districts are also estimated and compared to those that do not. Allowing for spatial network effects leave the results qualitatively unchanged but flags weaknesses of the approach in two of four cases. Section 6 describes robustness checks and examines a particular nonlinear model of spatial dependence. Section 7 summarizes the results.

2. Economic model

Model for the household

To illustrate the sources of endogeneity confounding the evaluation of spatially-sited programs and the assumptions underlying our identification strategy, we model both household and government behavior in the context of a multi-district nation. The country consists of L administrative districts each with N households, and K sub-districts per district each with $J = N/K$ households.[2] The pair of sub-district and district indices (κ, ℓ) uniquely identifies a sub-district. The household side is completely conventional. We abstract from issues of allocation within the household. As models of household demand when some goods, such as human capital H, are home produced are well known, we proceed directly to the (solved out) conditional demand equations. The linearized conditional demand equation for H_{jkl}, the human capital of household j in sub-district k of district l[3] is

(1) $H_{jkl} = \beta_0 + P_{jkl}\,\beta_1 + (W^hP)\,\beta_2 + \delta r_{kl} + \mu_{kl} + \eta_{jkl}$, $j=1,\ldots,N/K$

where H_{jkl} is a vector of KL individual-level observations containing the jth household of each sub-district, r_{kl} is an KL-vector that denotes the level of a composite public good input (e.g. public health services) provided by the government to sub-district k in district l, P_{jkl} is an KL x M matrix of M strictly exogenous variables affecting human capital, some of which vary by sub-district (attributes of local environment and the prices for goods) and some of which vary by household (parental schooling and age, religion, languages spoken), P is the KL x M (spatial) matrix of strictly exogenous variables P_{jkl} at all places[4], W^h is a KL x KL spatial weight matrix, the term μ_{kl} is the sub-

---

[2] It is not required that there actually are administrative districts and sub-districts that have control over resource allocation in order for the process that allocates programs to places to be a spatially two-stage process as modeled here. It is sufficient that the central planner consider spatially aggregated program determinants and program outcomes in allocating programs that have spatially limited effects on outcomes. The spatial aggregation of the information set will arise as a consequence of the costliness of acquiring and processing information on the returns to and health status of every single uniquely identified household, as described below.
[3] More generally, k is a place in space that is contained within a set of places l.
[4] For variables in P_{jkl} that vary across households within a sub-district, the corresponding elements of the spatial matrix P are the sub-district means.

district specific unobservable heterogeneity, and η_{jkl} is a non-systematic household specific error term representing deviations from the sub-district average of unobserved factors.[5]

Equation (1) is of the form of known as the *spatial-x model* in the spatial econometrics literature (Anselin 1988, Kelejian and Prucha 1998, Baltagi *et al.* 2014). This model assumes some type of spatial dependence from an exogenous predictor in sub-districts other than (κ,ℓ) to the program intensity r_{kl} in sub-district (κ,ℓ). The spatial weight matrix W^h is a positive and symmetric matrix whose elements w_{kl}^h represent exogenous geography such as measures of distance between spatially defined units, or more simply, nearest neighbors or contiguity, the most common measure in the spatial econometric literature. The diagonal elements of W^h are zero by convention, and rows sum to one as a normalization. This facilitates the interpretation as an averaging of neighboring values. There are clearly more unique elements in W^h than observations when the number of observations is equal to the number of sub-districts and, in accord with the spatial econometrics literature, W^h is assumed to be fixed and known. In the case of contiguity, elements of W^h for places that are not contiguous to place (κ,ℓ) have the value of zero and elements of contiguous places have the value of one (prior to normalization). Thus, W^h simply picks out the matrix of strictly exogenous variables for *relevant* neighbors from P.[6]

A standard assumption in the very large literature on program evaluation that makes use of data on outcomes in a district matched to programs in the same district (as in state- or county-level analysis in the US) is that $\beta_2=0$.[7] This assumption, typically never stated, holds that none of the exogenous variables in other places affect outcomes in place (k,l), that is, there are no spatial spillovers arising from P. Below, we show how instruments for endogenous program placement can be derived from models that optimize household and government behavior subject to resource and information constraints even when $\beta_2\neq 0$.

Model for the social planner

[5] In the empirical example, six public goods rkl are specified. Consequently, while in equation (1) the parameter δ is a scalar, it is a vector of six parameters in the application. The parameter vectors β_1 and β_2 are M x 1 vectors. Since P_{jkl} is an M-vector for each observation, one could specify a different KL x KL matrix W^h for each element of that M-vector. We do not do so here in order to avoid clutter and because we have a single W^h matrix for all of the exogenous variables in the empirical application.

[6] Following Elhorst (2014) spatial weights matrices commonly used in applied research are: (i) p-order binary contiguity matrices where if p = 1 only first-order neighbors are included; (ii) inverse distance matrices which may have set weights to zero if distance exceeds some maximum; (iii) q-nearest neighbor matrices (where q is a positive integer); (iv) block diagonal matrices where each block (or "district") represents a group of spatial units that interact with each other ("sub-districts") but not with observations in other blocks.

[7] Or, alternatively, that W^h is the zero matrix.

The central government has a social welfare function that may include many outcomes in addition to H_{jkl}, the levels of human capital enjoyed by its citizens. From some unspecified process that determines the full public expenditure of the country, the "Ministry of Human Capital" (social planner) receives a fixed lump-sum (V) to allocate towards the production of human capital H_{jkl}.[8] The most general form of the sub-utility function for human capital of the social planner contains the individual human capital outcomes of each household (person) as arguments:

(2) $\boldsymbol{\mathcal{W}}=(H_{111},H_{211},\ldots,H_{n11},H_{121},\ldots,H_{NKL})$.

In principle, the social welfare function can be written with sub-district level outcomes, such as the sub-district means or higher moments, in place of individual specific outcomes as arguments. This will arise as a consequence of the costliness of acquiring and processing information on the health status of every single uniquely identified household or small aggregations of households.[9] For example, decision makers may only have information from sample surveys on the human capital stock of its citizens, and may only be aware of the first few moments of the sampled distribution in sub-districts, and have just these statistics in allocating resources. The weakly separable social welfare function with mean sub-district outcomes as arguments is:

(3) $\boldsymbol{\mathcal{W}}=\boldsymbol{\mathcal{W}}(\boldsymbol{w}_1[\boldsymbol{w}_{11}(H_{11}), \boldsymbol{w}_{21}(H_{21}),\ldots, \boldsymbol{w}_{K1}(H_{K1})],\ldots, \boldsymbol{w}_L[\boldsymbol{w}_{1L}(H_{1L}), \boldsymbol{w}_{2L}(H_{2L}),\ldots, \boldsymbol{w}_{KL}(H_{KL})]$

where $\boldsymbol{w}_l$ is a utility function for district *l* defined over the second-level sub-utilities $\boldsymbol{w}_{kl}$ of all of the sub-districts in *l*. The implications of weak separability are discussed below.

Maximizing the social welfare function (3) subject to the fixed budget V and solving for r_{kl} yields the reduced-form equations for program intensity

(4) $r_{kl} = r(P_{11},.., P_{KL}, \mu_{11},\ldots, \mu_{KL}, V)$

There are two implications of equation (4) worthy of attention. First, the sub-district level error μ_{kl} is a determinant of both H_{jkl}, the human capital of individual *j* in sub-district *k* in district *l*, and of the program intensity in that sub-district, r_{kl}. Consequently, consistent estimation of the effect of program intensity on human capital must deal with this confounding unobserved variable. Second, P_{kl}, the exogenous determinants of human capital in districts and sub-districts *other* than

[8] In accord with most of the program evaluation literature, we abstract from taxation and other public finance issues that may affect the determination of V.

[9] In addition, the social planner's allocation decisions only affect the sub-district availability of publicly provided goods such as schools and clinics and not person-specific availabilities.

sub-district (κ,ℓ) (that is, when $(k,l) \neq (\kappa,\ell)$), affect program intensity in sub-district (κ,ℓ), $r_{\kappa\ell}$, but not necessarily human capital in (κ,ℓ).[10]

This model can be linearized and compactly represented as:

(5) $r_{\kappa\ell} = P_{\kappa\ell}\,\beta + WP\theta + \varepsilon_{\kappa,\ell}$

where $r_{\kappa\ell}$ is the composite program spatially allocated to place indexed by (κ,ℓ), W (like W^h) is the KL x KL spatial weight matrix associated with program placement, β and θ are M x 1 parameter vectors to be estimated, and ε is an error term that contains (i) μ_{kl} (which include the $P_{\kappa\ell}$ that are unobserved by us but presumed known by the social planner)[11] and (ii) any other sources of randomness in the data.

Differences between W and W^h give rise to the exclusion restrictions that we propose for instrumental variables estimation of the determinants of the spatial distribution of public programs. Aside from the usual issues concerning the number of instruments and their strength found in standard instrumental variables modeling, three restrictions on the spatial weight matrix W and on WP in equation (5) are necessary for identification:

(A) (trivially) at least one off-diagonal element must be non-zero;

(B) all of the off-diagonal elements on W cannot have the same value.

If the w_{kl} all have the same value then the product WP creates the means of P across all districts except district (κ,ℓ), and the model is not identified since then P and WP are perfectly negatively correlated.[12] We refer to this case as one in which the social planner's program allocation decision is *spatially uninformed* in that knowing the location of places (districts) does not matter in her allocation decision.

(C) $\theta \neq 0$, that is, the social planner is *spatially informed* when making allocation decisions across sub-districts. Weakly separable preferences (as in (3)) is sufficient for this to hold.

Given (A) – (C), there are then two cases that generate the exclusion restrictions required for identification of δ, the effect of program r_{kl} on H_{jkl}, in equation (1):

Case I. No network effects on H_{jkl} conditional on r_{kl} in equation (1) ($\beta_2=0$)

---

[10] To be clear, the "prices" of human capital inputs that are included in P_{kl} may represent the costs of acquiring information about existence of the public program r_{kl} and the services it provides, as well as transport and other direct or indirect costs involved in taking advantage of its services.

[11] We allow for spatial correlations in μ_{kl} (and hence in $\varepsilon_{\kappa,\ell}$, the error term in equation (5)) in our estimations and do not assume that they are independently distributed. All standard errors reported in the paper allow for unrestricted correlation between sub-districts in a district in order to account for this non-independence.

[12] To see this, note that sum of an exogenous variable across districts is a constant, so that WP is a constant minus P_{kl}, and thus *corr*(constant – P_{kl}, P_{kl}) = -1.

Under Case I all of the terms WP in equation (5) are available as instrumental variables for the estimation of the determinants of the spatial distribution of public programs.

Note that if household's behavior is only influenced by the attributes of their neighbors who reside in their sub-district, then $\beta_2=0$ is not contradicted. The larger are sub-districts, the more credible are network effects arising wholly within a sub-district likely to be true. The Indonesian administrative district structure used in this analysis are, in order: 1. province, 2. district (*kabupaten*), and 3. sub-district (*kecamatan*). There are roughly 3000 sub-districts with a current average population of 87,000 people, so this is a sizable "neighborhood."

Furthermore, the general empirical finding that human capital outcomes (H_{jkl}), exogenous attributes of places (P_{kl}), and public programs (r_{kl}) are spatially correlated is not sufficient to invalidate $\beta_2=0$ or invalidate the identification strategy set out and applied in this paper.[13] However, there may indeed be attributes of the structure of preferences, biology, and information that may generate structural cross-sub-district effects. We consider this possibility in what follows below.

Case II. There are network effects on H_{jkl} conditional on r_{kl} in equation (1) ($\beta_2\neq0$)
In this case, a sufficient condition for identification is that some of the nonzero off-diagonal elements of W correspond to zero elements of W^h (and $\Theta\neq0$). More precisely, if the elements of W and W^h are either zero or one, with the value of one picking out those sub-districts associated with cross-sub-district effects, the matrix $[W - W^h]$ must have at least one positive element in the upper diagonal. This means that in equation (5) there are exogenous variables for some sub-districts other than (k,l) that determine program intensity in district (k,l), r_{kl}, but not human capital H_{jkl} in (k,l).[14] This latter assumption is scrutinized at length below.

---

[13] The P_{kl} attributes of places are almost surely spatially correlated in all countries of size. In Indonesia, the sub-districts of the island of volcanic Bali share attributes of a physical environment that differ from the common environment shared by sub-districts in lowland West Borneo. Sub-districts along coasts share attributes different from those in high plateaus. Any spatial correlation in P_{kl} will necessarily feedback into correlations in program intensity, r_{kl}, and human capital outcomes, H_{jkl}, without the need to postulate models of strategic interaction among governmental units or spatial network effects. The data generating processes described below are sufficient to generate correlations among all of the variables considered.

[14] The proof that this condition is sufficient for identification is straightforward. In the common case where the elements of the (not-normalized) spatial weight matrix are either zero or one, the exogenous variables WP and W^hP are the means of the variables P for the *cross* sub-districts picked out by W and W^h. If W has the value of one for at least one element and that element has the value zero in W^h, then WP and W^hP are not perfectly collinear. If there are M variables in P, this generates M identifying restrictions. Consequently, the predicted value of the program (r_{kl}) in the sample, conditional on the exogenous variables in equation (1) which include the instrument WP, is not collinear with the exogenous variables in equation (1), which include W^hP, and is thus identified. In terms of a spatial lag structure, if WP includes a lag not found in W^hP (as in the application reported below), then there are KL positive elements in WP

While these are sufficient conditions, they are not necessary for a simple generalization of the model. Consider partitioning W into two submatrices W^1 and W^2 with associated parameters Θ^1 and Θ^2 in equation (5). With the two matrices W^1 and W^2 and two parameter vectors Θ^1 and Θ^2, the effects of sub-district exogenous variables on program placement are no longer all proportional as they are when specifying a single full matrix W. A single identifying restriction is created since there is one parameter β_2 associated with W^hP in equation (1), but two parameters Θ^1 and Θ^2 associated with the spatial term WP when W is partitioned into W^1 and W^2. This is true even if the W and W^h only contain zeros and ones, and exactly overlap so that $[W - W^h]$ is the zero matrix. We will examine a case of this form (given in equation (6), below), and its justification.

The question is: do theories of public finance meets the requirements above for identification? For example, when is a nonzero W matrix sensible? Since having non-zero elements in W is necessary to identify our model, this is crucial. And, if there are *network effects* determining the outcome H_{jkl} (Case II, above), are there then restrictions on $[W - W^h]$ that still identify the model? Can empirical tests suggesting the validity of these restrictions be constructed and used to guide an application?

Spatially decentralized budgeting as a source of parameter identification

The assumption of weak separability of the social welfare function with respect to the spatially defined good H_{kl} is sufficient to generate *spatially decentralized budgeting*, an allocation process that yields restrictions sufficient for identification. The separability restriction says that the marginal rates of substitution between the human capital of any two sub-districts are independent of the human capital of all other sub-districts in other districts, hence aggregator functions exist for sub-districts. As in the case of consumer demand theory, the spatial allocation process with separability can be conceptualized as a two-stage budgeting process in which budget allocations are made to districts and then, conditional only on the budget available to the district to which a sub-district belongs, allocations are made to each sub-district. This process looks like (spatially) political decentralization, in which decision-making is devolved to local levels of government from larger units of government. Local units of government are likely to be better informed about the distribution of human capital outcomes and the health environment, and so devolving decision-making to them may be efficient when conveying this information is expensive or fraught with strategic misrepresentation. Whether or not decision-making is actually devolved to lower-levels

---

that are zero in W^hP. Identification via specifying an extra spatial lag, implying KL additional positive elements, generates the same number of identifying restrictions, M, as a smaller set of positive off-diagonal elements in WP.

of government does not matter for the estimation strategy, only that the social welfare function is weakly separable. More generally one can think of the social planner as examining regions that do not necessarily conform to administrative district borders in deciding where to place programs. Regions to the social planner may be defined by ethnicity or religion (e.g. Kurdish regions in Turkey and the Middle East), areas of poverty (e.g. *Appalachia* and the *Rust Belt* in the US, the West Midlands in the UK, and western areas of China), areas with security concerns, or areas defined by electoral politics.[15] Simply put, what is required is that the social planner's program allocation decision is *spatially informed* -- knowing the location of places (districts) matters in the allocation decision, not just the (non-locational) attributes P_{kl}.

Consider multi-stage budgeting arising from separable preferences in which the central planner allocates resources for the program to districts, and the *district social planner* allocates resources to sub-districts. The allocation of program intensity to sub-district (κ,ℓ) depends on two different own and cross-price influences by which sub-districts "compete" for resources from the central planner and the district planner (as in Brueckner 2003).[16] First, at the topmost level of budgeting, the attributes of district ℓ are compared to all other districts $l \neq \ell$. Second, at the second-level of budgeting, the attributes of sub-district κ in district ℓ are compared to that of all other sub-districts in district ℓ.[17]

To simplify the exposition, we re-write the spatial weight matrix associated with sub-district (κ, ℓ) in terms of district-level sub-matrices W_ℓ that contains elements $w_{k\ell}$ of sub-districts located within district ℓ, so that $W^\ell = [0 \quad \cdots \quad W_\ell \quad \cdots \quad 0]$, where for all sub-districts not in district ℓ

[15] Alternatively, the indices (k,l) can be the latitude and longitude of points (or centroids of grids) on a map, and the social planner compares the social benefit of locating a program at (k,l) in comparison to points in the environs of (k,l), giving lesser weight to the attributes of points distant from (k,l).

[16] Brueckner considers the case where taxes are chosen in strategic fashion by smaller units of governments taking into account the mobility of factors of production to differences in tax rates across space. This strategic process, although not relevant to the Indonesian case, similarly generates a dependence of one districts public programs, funded by local taxes, on other districts attributes. In countries in which even small administrative units set tax rates and hence public program availability through a process of strategic interaction, the instrumental variable methods proposed here will require a different theoretical structure to be applicable.

[17] The social welfare function used above to motivate the exclusion restrictions arising from multi-stage budgeting posits an apparently benevolent social planner who derives welfare from the well-being of citizens. This assumption is not necessary to generate the exclusion restrictions. The social planner can be motivated by electoral or other political considerations such as regional harmony. All that is required is that the informational complexity of weighting the political characteristics of all individuals requires some level of spatial aggregation consistent with multi-stage budgeting.

the spatial weight elements are zero, and at least one element in W_ℓ in nonzero.[18] The linearized allocation equation with two-stage budgeting is

(6) $r_{\kappa\ell} = P_{\kappa\ell}\,\beta + W^{\ell}P\theta^{\ell} + WP\theta + \varepsilon_{\kappa,\ell}$

where the term $WP\theta$ corresponds to the first stage of budgeting that allocates the program to districts all over the country, and the term $W^{\ell}P\theta^{\ell}$ corresponds to the second stage of budgeting that allocates the program to sub-districts within a district. Specification (6) is a generalization of specification (5) and was first introduced in the discussion of Case II identification, above, in that it allows the spatial weights to be partly determined by the data in the regressions.[19] In particular, if $\theta^{\ell}=0$ then (6) collapses to equation (5). In addition, it does not require strong assumptions about W in order to conform to restriction B above that requires that weights vary across space. In equation (6), weights need not vary either in W or in W^{ℓ}, reflecting an agnostic view of spatial competition for public resources within the two-stage budgeting framework. More precisely, both the national and district-level social planner's decisions can be spatially uninformed so that WP is equal to the (unweighted) mean of district attributes P, and $W^{\ell}P$ is equal to the (unweighted) mean of all attributes P in district ℓ, that is, all weights are equal.

To be clear, equation (6) generalizes the first-stage equation (5) only by allowing more than two stages of budgeting (e.g. allocations to provinces, then allocations to districts within a province, then allocations to sub-districts of a district, etc.). Its structure here only arises from the assumption of a weakly separable social welfare function and not from any type of model of strategic interaction among government units such as those surveyed in Brueckner (2003). Below we will also consider models of the type that Brueckner (2003) has considered in which the attributes of other places affect the outcomes H_{jkl} and/or the placement of programs $r_{\kappa\ell}$ through strategic interaction among units or benefit "spillovers." It is those models that pose a direct threat to identification.[20]

---

[18] The spatial weights for any sub-district in district ℓ in spatial weight matrix W^{ℓ} need not be the same as the weights for that same sub-district in spatial weight matrix W.

[19] Letting the data determine the weight is akin to restricting the lag length so that it is less than the time dimension of the data in time-series models and then estimating the lag parameters freely or with some smoothing restrictions. Several selection criteria have been used to determine time lag length the most common of which are the Akaike Information Criterion (AIC) and the Schwarz' Bayesian Information Criterion. As in time-series estimation, it is clear that all of the elements of W and W^h cannot be treated as parameters to be estimated (along with β and θ) in the spatial-x model given by equations (1) and (5) because the number of unique elements is so large. In the Indonesian case described below there are approximately 3000 administrative districts; that means there are almost 4.5 million unique off-diagonal elements in the symmetric W matrix.

[20] In most of the program evaluation literature it is assumed that treatments received by one household does not affect outcomes for other households. In the statistics literature this assumption is referred to as the Stable-Unit-Treatment-Value-Assumption (SUTVA, Rubin, 1978). We do not make that assumption here as the "treatment" received by anyone

Under fairly general conditions, consistent estimates of program effects can be obtained even if the spatial weights are not those actually used by social planners. This is because consistent estimation of linear two-stage least squares models only requires that the first-stage be a linear combination of the exogenous variables. If the strictly exogenous variables $P_{k'l}$ ($k' \neq k$) are uncorrelated with ($\mu_{kl} + \eta_{jkl}$), the errors of the second-stage equation (1), then any linear combination of the strictly exogenous variables is still uncorrelated. Choosing a sub-optimal W or W^{ℓ} matrix only affects the efficiency of the estimation, not its consistency, as long as the linear combination is not made "weak" by the weighting. Consequently, the two-stage budgeting assumption does not require that the investigator be further informed *a priori* about the spatial preferences of the planners to consistently identify program effect parameters.

It is not even required that that the two-stage budgeting assumption be true in order to obtain consistent estimation of the model given by equations (1) and (6). If there is a single social planner who allocates programs to every spatial level throughout the country in a single-stage, and if that social planners decision-making process were *spatially uninformed* so that requirement B does not hold, consistent estimates of program effects may obtain by arbitrarily choosing the elements of W so that they vary, which is sufficient for requirement B, and that the weighted-average instrument that is obtained in not weak. As in over-identified models more generally, the arbitrarily weighted average of valid instruments is inefficient relative to the optimal two-stage least square or GMM weighted linear combination.

Consider the limiting case were the number of instruments equals the number of observations which number KL. Then the first-stage equation (5) predicts the program placement variable perfectly, and instrumental variables estimation accomplishes nothing as the second-stage equation (1) in this case is the same as OLS. As the number of instruments approaches the sample

in the sub-district is allowed to affect all others in the sub-district. Except for the Hausman-Taylor model that we test and reject below, we allow for unlimited interactions to occur among households within the broader group contained within the sub-district (average size of 87,000 individuals). We then estimate reduced-form models of spatial spillover by including spatial lags in the sub-district attributes P_{kl}. However, we do not include spatial lags in the treatment effects (r_{kl}0 themselves. Thus what we estimate is what is often referred to as "intention-to-treat effects", one of the most common parameter estimated in the program evaluation literature. Notice that in many settings, the most relevant policy decision is whether or not to provide individuals with access to treatment. In those cases, intention-to-treat effects have direct policy. Moreover, spatial lags in treatment effects are smaller the more community-based are the program evaluated and the larger is the size of the spatial unit of observation. In Indonesia, few individuals would seek care from community based health centers or family planning clinics outside of their sub-district, although that would be the case for hospitals or specialized care facilities, programs that we do not study. Similarly for primary and secondary schools as compared to universities or technical institutes. We have now added text to the paper that makes these points as well as the interpretation of our estimated treatment effect parameters (δ) as local "intention to treat" effects.

size, the 2SLS estimator tends towards the OLS estimator, which is expected to be biased, so that the 2SLS estimator may also have a serious finite sample bias if too many instruments are used relative to the sample size - even though all the instruments are valid.[21] There is also the distinct problem of weak instruments where finite sample bias arises even if there are few instruments but those instruments provide very little (additional) information about endogenous programs. Using many instruments may improve the efficiency of estimators asymptotically, but the bias might be large in finite samples, making inference inaccurate. Our spatial case has the feature that the number of instruments grows at a faster rate than the number of observations.[22] The spatially weakly separable social welfare function provides a theoretical structure for a *spatially informed* W matrix, the source of parameter identification, that is less likely to suffer from the weak instruments bias problem associated with arbitrarily setting elements of W to zero. This is demonstrated in the empirical work reported below that compares spatially uninformed estimates with those informed by spatially weak separability.

Spatial network effects: theoretical threat to instrument validity

The validity of using the exogenous attributes of other locales as instruments for a sub-district under Case I identification requires that they not directly affect human capital outcomes in that sub-district. However, it is possible that human capital in a sub district, H_{jkl}, is determined not only by the characteristics of that sub district, P_{kl}, but by the characteristics of the neighboring sub-districts arising from spatial spillovers of some kind. This generates *spatial network effects* of the form given by (1) with $\beta_2 \neq 0$. In this case the instrumental variable model is not identified if W^h and W exactly coincide (*Case II* identification).[23] However, the model given by (1) and (6) based upon more than two stages of budgeting is sufficient to identify this model. Furthermore, the more restrictive model (1) and (5) can be identified if some elements of W are nonzero when the same elements of W^h are zero. There is a theoretical and empirical argument why this may be so. A household's cost of acquiring information is increasing with distance from the source. In a variety of empirical applications, demand-side network effects have been found to be stronger for closer neighbors than those residing further away, presumably reflecting a cost of acquiring information

[21] In addition, if the non-zero elements of W are the same, corresponding to simple averages of other places, then as the number of non-zero elements becomes large ("almost spatially uninformed planner") then WP and $P_{\kappa,\ell}$ approach perfectly negative correlation by construction, as footnote 12 points out.

[22] The number of unique elements in W is $(KL^2 - KL)/2$ and thus increases at a quadratic rate.

[23] In addition, the error terms μ_{kl} and η_{jkl} need to be re-defined to include the unobserved attributes of other places because if the observed exogenous variables of other places effect outcomes, then the unobserved exogenous variables will as well. New notation for the errors are not introduced here in order to avoid clutter. Spatially correlated errors are specifically allowed for in the empirical application.

that is increasing with distance. However, for a social planner basing decisions on district-level data provided by ministries or the statistical bureau, there is no obvious reason why distance from the national capital or provincial capitals alters access to information in the same manner. All of which suggests that W^h contains nonzero elements for nearby locations and that W contains more nonzero elements than W^h. It is on this basis that models with spatial network effect are identified below.

In the empirical example reported below, we allow for a non-zero W^h matrix that allows for *spatial network effects* between (i) contiguous sub-district neighbors and (ii) sub-districts located in a neighboring district and are thus one step further spatially distant. Correspondingly, we allow for the W matrix to also contain non-zero elements for varying levels of spatial distance and use the estimates to test for *spatial network effects* and the efficacy of our identification strategy. As always in instrumental variable models, some assumptions are untestable. Below we develop some empirical criteria that at least suggest whether the identification of program effects from two-stage budgeting is reasonable in an empirical setting.[24]

Finally, we have specified linear models above since doing so requires us to be explicit about the restrictions on W and W^h (and θ and β_2) required to identify program effects. If, however, $WP\theta$ were replaced by $g(WP)\theta$ where g() is a nonlinear function, then the model could be identified solely on the basis of this nonlinearity even if none of the restrictions on W and W^h required in a linear model were satisfied. Pörtner *et al.* (2014) in an extension and application to an earlier version of this paper suggests that a reasonable nonlinear transformation of WP is ordinal rankings, claiming that "ranks as instruments are intuitive in that they mimic expectations about the underlying resource allocation process" because they "likely reflect what policymakers care about when distributing family planning programs, but are not directly related to fertility."(p. 4). By using ranks within a country, one can dispense with a spatial weighting matrix W altogether when the human capital demand equation is linear. Pörtner *et al.* (2014) provide no evidence for their claim that ranks of P reflect the social planner's actual allocation rule, and it seems a very strong claim indeed. It implies, among other things, that program allocations are unchanged if the underlying P rises for the highest rank sub-district and falls for the lowest rank sub-district.

[24] We explicitly test for spatial network effects using distant non-neighbors across all the outcomes we consider, and address sorting (based on migration) directly. Further, we present several tests for instrument validity including a model under-identification test, C or GMM distance test, Kleibergen-Paap, and Hansen's *J*-test for over-identification. The results of these tests indicate that the terms in WP in equation (5) provide valid exclusion restrictions, although, as we discuss, validity of our instruments does not work as well in some contexts.

However, transformation into ranks seems a much more reasonable nonlinear transformation than any other that comes to mind. Below, we provide some evidence that instruments based on rank fit the data less well than our approach but yield essentially the same results qualitatively.

Identification using own sub-district attributes only: the Hausman-Taylor model

In the models introduced above, the attributes of a subset other sub-districts are the exclusion restrictions that identify the impacts of non-randomly allocated public programs. Another related set of spatially-defined exclusion restrictions arise if data on program outcomes and household- and individual-level exogenous attributes that affect those outcomes are available. The assumption is that it is only the sub-district *means* of the household-level attributes (parental schooling and age, religion, languages spoken) that form a sub-matrix of P_{kl} that affect the placement of programs. Quite sensibly, this assumes that program placement is not attentive to the individual household values of these variables. This *Hausman-Taylor* (1981) assumption (which we present for the sake of completeness despite our belief that it is restrictive) generates an extra set of exclusion restrictions. Using our *spatial-x* model notation, consider the determinants of programs available to household *i* in sub-district (κ, ℓ) as a linear function of all of the household-specific exogenous variables $P_{i\kappa\ell}$, and all of the exogenous variables of all other *households* in the country P^i.[25] The reduced-form linear program allocation equation is

$$(7)\quad r_{i\kappa\ell} = r_{\kappa\ell} I_{\kappa\ell} = P_{i\kappa\ell}\,\beta + W^i P^i \theta + \varepsilon_{i,\kappa,\ell}$$

where $I_{\kappa\ell}$ is a vector of ones with length equal to N, the number of households in district (κ,ℓ). W^i is a vector of household-specific spatial weights and takes the form $W^i = [0 \quad \cdots \quad W^i_{\kappa,\ell} \quad \cdots \quad 0]$, that is, the only non-zero elements in W^i correspond to households that reside in sub-district (κ,ℓ). If, as assumed, the social planner is only attentive to the sub-district means (or higher moments) of the household-specific exogenous variables rather than to those of specific households, then $\beta=0$. If all the non-zero elements $w_{\kappa\ell}$ take the same value, we have the Hausman-Taylor (1981) model of identification with panel data in which the sub-district means of exogenous variables that vary within a sub-district are instruments for endogenous variables that do not vary within a district.[26] Implicit in the Hausman-Taylor model in this context is a social planner who allocates program resources to a place (sub-district) based only on the attributes of that place without regard to the

[25] Not all the exogenous variables affecting human capital outcomes are household-specific. All that is required is that some of them (such as parental age and education, exogenous components of wealth (land), and religion) are and they are in sufficient number to satisfy the rank and order conditions for identification. We do not distinguish between columns of $P_{i\kappa\ell}$ that are household-invariant from those that are household-varying simply to reduce notational clutter.

[26] The instruments need not be the sub-district means of $P_{i\kappa\ell}$. If the spatial weights $W^i_{\kappa,\ell}$ vary across households, then the sub-district weighted average of $P_{i\kappa\ell}$ are the instruments.

attributes of other places, that is, there are no cross-sub-district effects. This is consistent with a planner without a budget constraint who allocates resources until the point at which a social rate of return criterion is met. Hypothetically, the government has no liquidity constraints – it can borrow for all program investments that yield at least some social rate of return.[27] These may seem like strong restrictions (again, we believe that these are - the model is presented for the sake of completeness) but alternatively, Hausman-Taylor spatial weights do not require multi-stage budgeting assumptions on the allocation process of the social planner.

In the case of Hausman-Taylor (*H-T*) identification, the assumption required is that of independence of individual-level human capital levels and the attributes of one's neighbors in the sub-district. For example, it requires that the age and schooling of a child's parents influence that child's human capital, but not the human capital of other children in the same sub-district conditional on public programs. In the case of schooling, if labor markets are spatially separated and spatial mobility is costly, then the characteristics of parents affect wages and the return to those investments in the sub-district, violating the assumption above. There are other scenarios under which the spatial independence assumption is not valid. If there are neighborhood externalities to health such as in the case of de-worming (Miguel and Kremer, 2004), then spatial independence may not hold. If there are peer group effects or other types of non-independent preferences within spatially defined areas, then spatial independence will also not hold. We test and reject the validity of the Hausman-Taylor estimator as applied to our data below.

Sources of identification summarized

To summarize, a model of the placement of spatially-sited public programs affecting human capital in which the social planner's welfare function is defined over the human capital outcomes of all spatially defined districts, yields identifying instrumental variables (exclusion restrictions) for estimating the effects of public program on human capital outcomes of the following types:

A. *Spatially informed program allocation*. A spatially weakly separable social welfare function (two-stage budgeting) is sufficient but not necessary for *spatially informed* program allocation in which sub-districts "compete" with each other for resources. In this case,

[27] This is very much like Becker's *efficient schooling hypothesis* which holds that liquidity unconstrained parents who care about the earnings of their children when they are adults will invest in the human capital of each child based solely on a rate of return criterion without respect to the different abilities of their multiple children who otherwise "compete" for parental resources. With efficient schooling there are no cross-child effects – the innate intellectual ability of one sibling has no effect on the parental schooling investment in the other. The Hausman-Taylor assumptions correspond to program allocation model characterized by "efficient program allocation" as there are no cross-district effects. It also implies that the social welfare function (3) is additive in the sub-utility functions. The "efficient program allocation model" like the efficient schooling model has testable implications (no cross-effects) that we examine below.

strictly exogenous variables that vary across spatially defined sub-districts, such as within-sub-district means (or higher moments) of household-level variables (such as parental age and schooling, religion, and ethnicity), as well as sub-district level attributes (climate, topography, geology, prices) of *other* sub-districts are potentially available as instruments for programs in any particular sub-district.

B. *Hausman-Taylor identification*. Program allocation can be *spatially uninformed* but requires household-level data for estimation. Sub-district means of the household-specific determinants of human capital (such as parental schooling and age) that vary within sub-districts can serve as instruments for program availability in the same sub-district under the assumption that the mean of these characteristics are uncorrelated with the unobserved sub-district-level effect.

3. Issues in empirical implementation

The spatially decentralized budgeting model that we propose means that, to paraphrase the terminology of Conley and Ligon (2002), *political economic distance* matters. What characterizes the economic distance relevant to program placement is not obvious. The two-stage budgeting example discretizes economic distance by characterizing it as whether administrative sub-districts (κ,ℓ) and (κ',ℓ') are both contained within the same district ($\ell = \ell'$). This notion is more appropriate to formally decentralized program allocation in which decision-making is devolved to sequentially lower units of government than to a single social planner. The use of membership of a place (sub-district) in a district as the measure of economic distance has the implication that two places that are spatially contiguous but in different districts (they straddle the district border) are as economically distant as places at the opposite ends of a country. Other general metrics include physical distance (distance between centroids of sub-districts) or even travel time distance. In the empirical example using data from Indonesia presented below, we lack data on physical distance. Instead, we construct three measures of economic distance, one based on spatial proximity, one based on shared district status, and one based on contiguity to a shared district. In the *neighbors* measure, sub-districts that are contiguous are considered "spatially proximate" irrespective of their district. In the second measure labeled *non-neighbors*, membership in a common district is what matters. In the case of neighboring districts (contiguity), some but not all of the other sub-districts in district ℓ are likely to be contiguous with (κ,ℓ), and sub-districts outside of district ℓ may be contiguous with (κ,ℓ). The idea is that competition for public programs with neighboring sub-districts (*kecamatan*) with which a sub-district shares program effects across a common border may differ from competition with non-neighboring sub-districts with which, by dint of membership in a larger political unit, it shares an allocation for a program to be spatially distributed by the district (*kabupaten*) social planner. To avoid overlap and for other reasons discussed below, the *non-neighbors* designation refers to sub-districts in the same district as (κ,ℓ) that are not neighbors with it. Thus the union of *neighbors* and *non-neighbors* is all other sub-districts in district κ except (κ,ℓ), plus any sub-districts contiguous to (κ,ℓ) that are not in district ℓ. For the third measure, labeled *distant non-neighbors*, what matters is membership in a district (rather than sub-district) that is contiguous to the district ℓ in which sub district (κ,ℓ) belongs.

An additional consideration arises if the benefits of spatially-sited public programs can spill over the boundaries of spatially defined sub-districts. In principle, programs can serve clients from a catchment area larger than a single sub-district. Given an existing spatial distribution of hospitals or secondary schools, for example, the social planner may be less likely to invest in a new facility in a particular district if that district is already served by similar facilities in nearby districts. If population was distributed uniformly across a homogenous plane and administrative boundaries did not affect access to program services, an efficient planner would locate new facilities depending only on distance to existing facilities. If a facility like a hospital or secondary school can serve a group of spatially proximate sub-districts, its location depends on the attributes of those spatially proximate districts differently than it does on less proximate districts. Consequently, the attributes of neighboring districts may have an especially large effect on whether a program is sited in a particular district. More importantly, it renders the exogenous characteristics of *neighbors* invalid as exclusion restrictions in the estimation of the human capital equation (1) conditional on own-sub-district programs. This is the *spatial network effects* issue discussed earlier. We use the exogenous characteristics of *non-neighbors* and *distant non-neighbors* as additional instruments sets to (i) test for the orthogonality of *neighbors* and *non-neighbors* with respect to the unobserved components of equation (1), and (ii) estimate and test models with spatial network effects.

A common approach to estimating a policy response function such as (1) is to condition on a single policy, for example increasing the spatial coverage of a single program r_{kl} from the set of programs R_{kl}. This single-program policy response function provides the effect of increasing the single program's "intensity" (say schools) on the outcome of interest (school enrollment). However, it is clear that in general, the spatially defined instruments (including those of the Hausman-Taylor type) may affect the social planner's allocation of the full set of programs R_{kl}, and that the spatial allocation of programs may be correlated.[28] If these other programs also affect the outcome of interest, for example fertility control and public health programs may affect schooling of children in a model of child quantity and quality, then the orthogonality conditions underlying the instruments will not hold. Consequently, our approach may only be valid when estimating a policy function (1) that conditions on the complete set of programs. This is testable with the data as long as a sufficiently large set of instruments are available, and we report these tests below.

In most of the program evaluation literature it is assumed that treatments received by one household does not affect outcomes for other households. In the statistics literature this assumption is referred to as the Stable-Unit-Treatment-Value-Assumption (SUTVA, Rubin, 1978). We do not make that assumption here as the "treatment" received by anyone in the sub-district is allowed to affect all others in the sub-district. Except for the Hausman-Taylor model that we test and reject below, we allow for unlimited interactions to occur among households within the broader group contained within the sub-district (average size of 87,000 individuals). We also allow *spatial spillovers* by including spatial lags in the sub-district attributes P_{kl}. However, we do not include spatial lags in the treatment effects (r_{kl}) themselves. Thus, the estimated treatment parameter δ can be interpreted as a local "intention-to-treat effect", one of the most common parameters estimated in the program evaluation literature. Notice that in many settings, the most relevant policy decision is whether or not to provide individuals with access to treatment. In those cases, intention-to-treat effects have direct policy. Moreover, spatial lags in treatment effects are smaller the more community-based are the program evaluated and the larger is the size of the spatial unit of observation. In Indonesia, few individuals would seek care from community based health centers or family planning clinics outside of their sub-district, although that would be the case for hospitals or specialized care facilities, programs that we do not study. Similarly for primary and secondary schools as compared to universities or technical institutes.

Finally, the spatial decision-making process that we propose implies that unobserved attributes that affect human capital outcomes and spatially-sited public programs are likely to be correlated across space (errors are correlated across space); our standard errors account for this non-independence by estimating random effects (also called variance components) models with unrestricted correlation between sub-districts in a district.

[28] Pair-wise correlation coefficients for the four government programs in our data are consistently large, positive and significant at the 5% level, suggesting that there are likely to be complementarities in the spatial placement of programs and in their effects on outcomes.

4. Data and variable construction

We use data from two sources in this study: the 1980 *Potensi Desa* (Village Potential) survey of Indonesia (*PODES*) and the 1980 *Sensus Penduduk* (Population Census) of Indonesia. The 1980 *PODES* data provides information at the village level on the government programs that are studied: health clinics (*PUSKESMAS*[29]), family planning clinics, and grade and secondary schools. Information on area specific geographical characteristics such as the occurrences of natural shocks (droughts, floods, earthquakes, and other shocks) in the last five years, and other information such as distance from the coast and proportion of households in urban areas, is also reported. It is important to note that these natural shocks are typically not catastrophic events. For example, Indonesia experiences many earthquakes annually that result in little more than collapsed bridges and ruptured irrigation dikes. Records on earthquakes from 1917 to 2010 with an intensity of 6.0 (medium intensity) or higher show that there were 10 such earthquakes in 2005 and 2009 with a median of 0 fatalities in either year.[30] Approximately 62,000 villages (*desa*), almost all of the villages of Indonesia, are covered by the 1980 PODES data which was carried out in conjunction with the 1980 Population Census.[31]

The 1980 long-form Census data sample provide detailed individual level information on the dependent variable outcomes: current school enrollment for girls and boys, and women's birth histories and contraceptive use. Data are also provided on other individual and household characteristics such as age and schooling attainment of household heads and spouses, as well as area of land owned by the household, indicators for whether the household owns its own home, religion of the household, and language of the household head. The sub-district level data were constructed by merging the information on programs and other geo-physical variables from the

---

[29] PUSat KESehatan MASyarakat – literally, Peoples Health Centers.

[30] This information is available from the U.S. Geological Survey at http://earthquake.usgs.gov/earthquakes/world/historical_country.php. In terms of significant earthquakes, where earthquakes are classified as being significant if they caused moderate damage (about $1 million or more), 10 or more fatalities, had an intensity of 7.5 or higher, a modified Mercalli Intensity of X or higher, or the earthquake generated a tsunami, data for Indonesia from the National Geophysical Data Center of the National Oceanic and Atmospheric Administration shows that many of these instances occurred in fairly narrow geographic areas on land, and in retrospective data from 1629 to 2010, the median number of recorded total deaths is only 30 (many occur out at sea, and many do not create tsunamis large enough to cause fatalities). These data are available at National Geophysical Data Center/ World Data Center (NGDC/WDC) Significant Earthquake Database, Boulder, CO, USA. (Available at http://www.ngdc.noaa.gov/nndc/struts/form?t=101650&s=1&d=1). The earthquake and tsunami of December 26, 2004 is an outlier, since it is the only instance when deaths from a natural event in Indonesia exceeded the approximate 10,000 mark, a level that was previously set for the country in 1815.

[31] It is not possible to link village data from the *PODES* dataset to household-level data from the Population Census at the *desa* level. Although *desa* translates to village in common usage, every part of Indonesia is part of a *desa* in these data; *desa* are units of governance with well-defined boundaries so that the largest cities contain hundreds of them and even smaller towns will have more than one.

1980 *PODES* data with the 1980 Indonesian census data aggregated to the *kecamatan* (sub-district) level. For each outcome, we have a random sub-sample of between 80,000-95,000 individuals who are "at risk" for the outcome. In particular, our sample consists of 82,891 girls aged 10-18, 82,889 boys aged 10-18, 87,655 women of ages 21-30 with fertility histories, and 95,372 women of ages 21-30 with recent contraception data.

Table 1 presents the means and standard deviations for each of the four outcomes analyzed in this study. In particular, the outcomes we study are current school enrollment rates of girls and boys ages 10-18 years, whether last child's year of birth lies in the previous two years (between 1978-1980) for women between the ages of 21 and 30, and whether any contraceptives are currently being used by women ages 21-30. Table 1 shows that current school enrollment rates range from 59% for girls to 66% for boys. Fertility, as measured by the incidence of births in the previous two years for women ages 21-30 years, is relatively high at 69%. In keeping with this, contraceptive prevalence among women in this age group is relatively low at 28%. The samples that are used to study the four outcomes we consider include data from approximately 3000 sub-districts in Indonesia. Table 1 also presents summary statistics for four government programs. Approximately 77% of households reside in villages (*desa*) in which there is a grade school, and about 40% live in village with a junior or secondary school. Coverage of *PUSKESMAS* health clinics is about 25%, and the coverage of family planning clinics is approximately 49%.

The strictly exogenous variables are classified as follows:

1. *Variables describing sub-district (κ,ℓ) that vary across but not within sub-districts (these are the P_{kl} from the theory)*. These describe the physical and economic environment of a sub-district (*kecamatan*). These include measures of the recent (five-year) occurrence of a natural shock (droughts, floods, earthquakes, and other natural shocks), whether the sub-district is located on an ocean coast, and the proportion of households in villages of the sub-districts that are urban. These sub-district variables are determinants of individual human capital outcomes for residents of that sub-district. The exclusion restrictions based upon multi-stage budgeting are that the sub-district means of these variables in competing sub-districts influence program placement in a particular sub-district but not human capital outcomes in that sub-district conditional on program placement. There are 6 such variables.
2. *Variables that vary within sub-districts (these are the P_{jkl} from the theory)*. These are the household and individual variables that are determinants of the individual human capital

outcomes, including parental age and schooling, area of land owned, religion, and the language of the household head, plus interactions of some of these variables with the proportion of the sub-district population that is urban. These variables are part of P_{kl} and are included in all regressions, although their estimated parameters are not reported in the tables. Interactions of all the (level) exogenous variables and the urban variable are also included. Interactions are not used as instruments. There are 21 such variables.

3. *Hausman-Taylor identifying instruments*. These are the sub-district means of the individual exogenous variables of #2 above. The underlying individual-level variables are presumed to affect individual human capital outcomes but the sub-district means of these household-level variables are assumed to influence only program placement in a sub-district (not individual human capital outcomes). There are 21 such variables.

4. *Neighbor instruments*. These are the means of the variables of #1 and #3 taken over the sub-districts that are spatially contiguous to sub-district (κ,ℓ), excluding the urban interactions. For each sub-district, we determine neighbors that share a geographical boundary using detailed province-level maps of Indonesia. There are at most 14 neighboring sub-districts in the data, although most sub-districts in the data have many fewer neighbors. There are 12 such variables.

5. *Non-neighbor instruments*. These are the means of the variables of #1 and #3 taken over the sub-districts that are in the same district (district ℓ) as sub-district (κ,ℓ) but are not spatially contiguous to sub-district (κ,ℓ), excluding the urban interactions. There are approximately 300 *kabupaten* (districts) in the data. There are 12 such variables.

6. *Distant non-neighbor instruments*. These are the means of the variables of #1 and #3 taken over the sub-districts that are in all districts contiguous to district ℓ, excluding the urban interactions. Instrument sets #4, #5, and #6 arise from the weak separability of the social welfare function. There are 12 such variables.

Table 2 reports descriptive statistics for the strictly exogenous individual and household characteristics, as well as the interactions of these characteristics with proportion of households in the sub-district located in an urban zone as defined by the Population Census, plus the standard deviations for all variables that make up the *neighbor* and *non-neighbor* instrument set. In principle, higher moments of the distribution of sub-district characteristics, or other measures of central tendency such as medians, could be used in the spatial aggregations. Means are what are reported in the statistical reports of the Indonesian Central Bureau of Statistics and of government ministries and thus are particularly relevant. There seems to be plenty of variation in spatial program placement attributable to the instruments derived from sub-district means, and adding higher moments may only add weak instruments. Table 2 also presents descriptive statistics for environmental variables that are measured at the sub-district (*kecamatan*) level. From this table, about 13% of households are in urban areas, and the proportion of households that have experienced natural shocks such as droughts, floods, or other events, ranges between 9% - 26% in these data. All of the own sub-district variables, including means of individual-level characteristics listed in Table 2, are included in the second stage of all of the spatial instrumental variable models reported.

5. Results

We start by estimating models that treat program placement as exogenous (OLS) using the individual data. If maximizing returns on a per-program basis underlies the geographic distribution of public funds, OLS will underestimate impacts. This is evident from the discussion and results in Pitt *et al.* (1993) who studied the same programs using fixed effects methods, and can be clearly seen on comparing the size of the OLS coefficients in Table 3 with the magnitude of the preferred estimates in Tables 4 and 5.

Hausman-Taylor regression models using the individual-level data for all four outcomes are presented in Online Appendix Table 1. The Hansen *J*-test statistics imply that the tests of overidentifying restrictions reject their null (at the 0.05 level) in every case and give us no confidence that the Hausman-Taylor instrument set is appropriate.[32] To summarize, the results suggest that additional secondary schools increase school enrollment, additional public health clinics (*PUSKESMAS*) reduce school enrollment, additional primary schools reduce fertility and increase contraceptive use, and additional family planning clinics reduce fertility and reduce the use of contraception. This pattern of program effects is qualitatively similar to the OLS estimates except for opposite signs for the effect of family planning clinics on contraceptive use.

---

[32] In addition to the strong assumptions it imposes, the failure of the Hausman-Taylor instruments may reflect a Tiebout (1956) type of world in which households sort themselves spatially among communities offering different mixes of public services and other attributes. Household that highly value quality schools may themselves be highly educated and have births at later ages, and sort themselves into communities with households having similar preferences and personal attributes. In this example, the age and schooling of parents in neighboring and non-neighboring districts are not likely to be as strongly affected by Tiebout-like spatial sorting of households than they would be within districts. Robustness checks along this line are reported below.

Tables 4 and 5 present the district-level random effects results of models estimated with sub-district level data using the instrument sets derived from the multi-stage budgeting model of program allocation. [33] The Kleibergen-Paap rk LM statistic for model under-identification is reported in Online Appendix Table 2. In all cases, the *p*-values strongly reject the null that the first stage matrix of reduced form coefficients is under-identified.[34] Column (1) of Table 4 uses the *neighbor* sub-district instruments and conditions only on school availability in the determination of girl's school enrollment. Column (2) adds family planning clinics and *PUSKESMAS* to the specification. Hansen's *J*-test fails to reject the null hypothesis in either specification at the 0.10 level (p=.41 and p=.25, respectively), providing some confidence in the validity of this instrument set. The failure to reject the model that conditions on only school availability means that our estimation strategy permits, in principle, the estimation of the usual sort of policy function – the effect of schools on schooling – without having to include other public programs in order not to reject the test of overidentification. As this is a linear probability model, a 1 percent point increase in secondary school availability is estimated to increase school enrollment rate of girls ages 10-18 by 0.39 percentage points in the model with all four programs, which is four times larger than the exogenous program estimate reported in Table 3.[35]

---

[33] The parameters on programs and their standard errors, as well as the Hansen-Sargan *J*-tests, are invariant to whether individual-level data or sub-district level data are used because the programs vary at the sub-district level. In order to estimate program effects there is no reason to use the individual-level data unless using Hausman-Taylor methods, in which case they are required. Test statistics from GMM with standard errors that are clustered at the district level are reported in Tables 4 and 5. These "cluster-robust" standard errors originally proposed by White (1984, p.134-142) do not require specification of a model for within-cluster error correlation, but do require the additional assumption that the number of clusters, rather than just the number of observations, goes to infinity. As Cameron and Miller (2015) point out this method extends naturally to the IV estimator and are closely related to spatial-robust variance matrix estimates.

[34] In addition to demonstrating validity, these tests on the strength of the first stage are important for another reason. Since the government has to spend a predetermined budget on schools and health programs in a year, the theoretical model developed above is static. However, the PODES 80 data has information on all existing programs, some of which may be been built long before and others that are more recent. Lacking details on year of initiation of programs, we use all the data available. The results of the diagnostic tests for the first stage in Online Appendix Table 3 are reassuring in that they show that the strength of the first stage is little affected by this. Further, following the theoretical structure underlying dynamic panel data models with lagged dependent variables outlined in Holtz-Eakin *et al*. (1988) and Arellano and Bond (1991), the IV framework is still appropriate when data on all existing programs is used, assuming there is no correlation in errors (which we take into account).

[35] The parameter associated with secondary schools is 18 percent larger in the model that conditions on all four programs than the model that conditions only on schools, but it corresponds to a different policy experiment. In the schools only model, the secondary school parameter measures the marginal effect of secondary school provision conditional on fixed primary schools with the other two programs variables. In the model with all four programs, the secondary school parameter measures the marginal effect of secondary school provision conditional on all three other programs fixed. The spatial covariation of programs in conjunction with the impact of non-school programs on school enrollment can account for the different parameters in these two different conditional demand equations.

In column (3) of Table 4, the *neighbor* instrument set is augmented with the *non-neighbor* instrument set. A *C* or *GMM distance test* (Baum and Schaffer, 2007) is used to test the validity of the *neighbor* orthogonality conditions that underlie the random effects IV estimator of columns (1) and (2). Denote J as the value of the GMM objective function for the efficient GMM estimator that uses the neighbor plus the non-neighbor augmented instrumental variable set orthogonality conditions, and J_N as the value of the efficient GMM estimator that uses only the *non-neighbor* orthogonality conditions, then under the null that the *neighbor* orthogonality conditions are actually satisfied, the test statistic $(J - J_N) \sim \chi^2$ with degrees of freedom equal to the number of variables in the *neighbor* instrument set. The null hypothesis that the orthogonality conditions associated with the neighbor instrument set is satisfied is not rejected in the model with all four programs in column (3) (p=0.11). In addition, we estimate an LM version of the Kleibergen–Paap (2006) rk test (Kleibergen and Schaffer, 2007) of the redundancy of the *non-neighbor* instrument set by testing the rank of the matrix E(X'Z). The null hypothesis that the non-neighbor instrument set is redundant when added to the neighbor instrument set is clearly rejected in column (3) (p=.00). The *J*-test of overidentification cannot reject the null hypothesis with the augmented instrument set of column (3) as was the case with *neighbors* as the only source of instruments.[36]

Having a set of instruments for more distant sub-districts than *neighbors* allows for not only the orthogonality and redundancy tests of column (3) but also for *spatial network effects* in which the attributes of *neighbor* sub-districts may directly affects outcomes. Estimates from this spatial network model is presented in column (4) in which the exogenous attributes of *neighbors* are now included in the second-stage equation. A test that the set of neighbor attributes are jointly zero cannot be rejected (p=0.09) suggesting that *neighbor* spatial network effects are not important in the determination of girls schooling. This test, along with the orthogonality and redundancy tests, support the use of neighbor attributes as instruments and the credibility of the estimates of column (3).

Column (5) of Table 4 uses the *neighbor* sub-district instruments and conditions on only school availability in the determination of boy's school enrollment. As for girls schooling, column

[36] The robustness of results in Table 4 and Table 5 below were tested with a reduced instrument set that excluded parental age and natural shock variables (recent history of droughts, floods, earthquakes and other disasters). Dropping these instruments ameliorates concerns that (i) parents who care more about their children's welfare may postpone having children until they are older and have more resources (parental age is endogenous), and (ii) shocks in other sub-districts may affect human capital in a particular neighboring sub-district through temporary migration (refugees) or variations in prices. Re-estimation with the reduced instrument (available on request) leaves the original results in Tables 4 and 5 essentially unaltered.

(6) adds family planning clinics and *PUSKESMAS* to the specification. Once more, Hansen's *J*-test fails to reject the null hypothesis in either specification at the 0.10 level (with p=0.27 and p=.15, respectively) providing some confidence in the validity of this instrument set. As with girl's schooling, augmenting the instrument set with *non-neighbors* (column (7)) permits the test of redundancy of the *non-neighbors* instrument set (rejected) and orthogonality of the neighbor instrument (not rejected) in a model for which the overall *J*-test is not rejected. A 1 percent point increase in secondary school availability is estimated to increase the school enrollment rate of boys ages 10-18 by 0.26 percentage points in column (6).

The test for spatial network effects reported in column (8) of Table 4 cannot reject the null hypothesis that the attributes of neighboring sub-districts affect boys schooling (p=.00), thus calling into doubt the use of neighboring sub-districts as instruments.[37] Note, however, that the parameter estimates of column (8) are only little different from those reported in column (7) in which neighboring sub-districts are identifying instruments. The empirical issue now is whether the *non-neighbor* instruments of column (8) also directly affect boys schooling and are thus also questionable. The estimates of column (9) use *distant non-neighbors* as instruments and control for both *neighbor* and *non-neighbor* attributes in the second-stage equation. Here the null hypothesis that the attributes of non-neighboring sub-districts affect boys schooling cannot be rejected (p=.00), casting into doubt non-neighbors as valid instruments. The parameter estimates are qualitatively the same as for the previous model (column (8)) in which *non-neighbor* attributes are the instruments, and virtually the same as those of column (6) in which *neighbor* attributes are instruments and no spatial effects are allowed. The model under-identification test (not shown in table) rejects weak identification at better than the 0.01 level in spite of the use of only *distant non-neighbors* as instruments. This suggests that there may be room to find instruments that are even further distant than our *distant non-neighbors*. The boys schooling estimates illustrate an empirical strategy of using increasing spatial lags as instruments while including less distant lags in the second-stage equation to allow for spatial network effects at varying distances. For boys schooling, network effects are found even two spatial lags away. In comparison, for girls schooling, network effects are not found even for *neighbors*.

[37] This test is in seeming contradiction to the orthogonality test. The orthogonality test, which is based on the difference in the Hansen *J*-test statistics ("C-test"), fails to reject the null hypothesis and yet when these instruments are included in the second stage, which also constitutes a test of orthogonality, we reject that their parameters are jointly zero. These tests can differ in finite samples due to differences in the underlying covariance estimators. Of course, the validity of this test requires that the identifying subset of instruments are, in fact, valid.

Table 5 presents instrumental variable estimates of recent fertility and contraceptive use. Columns (1) and (6) estimates the effect of family planning programs on these two outcomes without conditioning on other programs, as in the usual policy function. The Hansen *J*-test rejects the null hypothesis (at the .05 level) for both fertility (p=.01) and contraceptive use (p=.00). As noted above, the "failure" of these instruments when there is only a single program may result from incorrectly omitting the other programs (schools and PUSKESMAS health clinics) in the regression when they are part of the social planner's allocation problem, thus invalidating the instruments. The spatial instruments affect the allocation of all programs, and these omitted programs affect fertility and contraceptive choice. Adding the other three programs is sufficient to lead to non-rejection for fertility (p=.10) but not for contraceptive use (p=.00) in columns (2) and (7) of Table 5. The rejection of Hansen's *J*-test of overidentification for contraceptive use apparently reflects other issues, including spatial network effects, which are examined below.

In the case of recent fertility, adding *non-neighbors* to the instrument list in column (3) yields non-rejection of the orthogonality of *neighbors* (p=.47) and the rejection of redundancy of *non-neighbors* (p=.00), however, the null hypothesis of the absence of spatial network effects (the parameters on *neighbors* in the second stage regression) in column (4) is easily rejected (p=.00). The rejection calls into question the estimates of columns (2) and (3) in the same way as for boys schooling. The *spatial network* estimates of column (4), which do not use neighbors as instruments, may be consistent if the *spatial network* effect does not extend beyond neighbors. To explore that possibility, we make use of the *distant non-neighbor* set of instruments to test whether the *non-neighbor's* variables are statistically unimportant in the second-stage equation. However, the model firmly rejects this null hypothesis (p=.00) suggesting that even *non-neighbor* attributes induce spatial network effects in fertility. One could, of course, continue on this line of inquiry by using as instruments regions even further distant from any sub-district and then stop once no further additional spatial network effects are found. The parameter estimates are qualitatively the same as for the previous model (column (4)) in which *non-neighbor* attributes are the instruments, and virtually the same as those of column (3) in which *neighbor* attributes are instruments and no spatial effects are allowed.

For the determinants of contraceptive use, the Hansen *J*-test is rejected at the .05 level in the first four models presented in Table 5 (columns 6-9). However, we cannot reject the absence of spatial network effects from neighboring sub-districts (p=.26) in column (9). We extend the instrument set to include *distant non-neighbor* attributes in column (10) and cannot reject that (i) spatial network effects include *non-neighbors* (p=.00), and (ii) the model is over-identified (p=.09). In short, the spatial IV model seems to fail in the case of contraception as in none of the models estimated do statistical tests not reject that the instruments have the required properties.[38] The parameter estimates are qualitatively the same as all the models that include the full set of four programs.

To summarize, the estimates imply that grade school availability has a statistically significant and negative effect on fertility and a positive effect on contraceptive use, and that the IV estimates of these effects are larger in absolute value than the OLS estimates in Table 3. Family planning clinics are estimated to reduce recent fertility but have insignificant positive effects on contraceptive use, while PUSKESMAS availability increases fertility and reduces contraceptive use.

6. Robustness checks

We present falsification tests to rule out spurious correlation and demonstrate the robustness of results discussed above, plus compare our model in which the instruments are the *levels* of attributes of other sub-districts to one in which the instruments are the *ranks* of other sub-district attributes among the set of sub-districts belonging to a large administrative district. The falsification tests consider the impact of our school and health programs on alternate outcomes that include household religion and male old-age mortality. The religion indicator denotes whether the household religion is Islam, and there are few compelling reasons for why school and health programs should impact this variable. The *p*-values in the first two columns of Online Appendix Table 3 confirm this. These two specifications correspond to the second and third columns of each of four dependent variables in Tables 4 and 5. Next, using data from SUPAS 85 which provides information on cause of death, an indicator was created for whether an individual over sixty years who died was male.[39] Conditional on having lived to this age, the gender of the person who died is plausibly random. Further, this age cohort of individuals is unlikely to have been affected by programs established as of 1980 (their year of birth is 1925 or earlier). The *p*-value in the last column of Online Appendix Table 3 is consistent with this – we cannot reject the null hypothesis that the school and health programs are jointly zero at the conventional .05 level.

[38] Lacking more distant instruments at the sub-district level than *distant sub-districts*, we are unable to test if the *distant non-neighbors* instruments directly affect contraceptive behavior.

[39] There is no *kecamatan* identifier in the SUPAS 85, so these data were matched to the 1980 Census data at the *kabupaten* level. The estimations are at the *kecamatan* mean level, as before.

We also look for the presence of spatial sorting effects which, if found, may invalidate our instruments. If individuals sort into areas based on characteristics such as amenities, services or public goods that are unobserved from our perspective and thus not controlled for, then the estimates may be biased (Tiebout 1956). To address selective migration, we re-estimated the girls' schooling and recent fertility models including endogenous interactions of the school and health programs with two separate indicators of migration. The first measures whether duration of residence in the current province is one year or less and the second considers whether the current province of residence is different from the province lived in five years ago. Both these variables are present in the 1980 Census data. Online Appendix Table 4 reports results from these interaction terms for girls schooling and recent fertility. The hypothesis that all of the interaction between migration status and program intensity are jointly zero cannot be rejected in all cases.

In Section 2, we noted that if $WP\theta$ were replaced by $g(WP)\theta$ in equation (5), where g() is a nonlinear function, then the model could be identified solely on the basis of this nonlinearity even if none of the restrictions on W and W^h required in a linear model were satisfied. Pörtner *et al*. (2014) in an extension and application to an earlier version of this paper, suggests that a reasonable nonlinear transformation of WP is ordinal rankings. Below, we briefly summarize differences between the "levels" model presented above and a model where the instruments are the ranks of sub-districts within variously sized administrative districts – districts and provinces. The estimates of the ranks model are provided in Online Appendix Tables 5 and 6. In summary:

(i) The ranks model fit the data less well and the instruments are weak. In particular, when rank with a kabupaten and rank within a province are the instruments, they are statistically significant at the .01 level in the determination of only one program (secondary schools), and are not statistically significant at even the .10 level for two programs (grade schools and family planning clinics). In contrast, when the identifying instruments are level attributes of *neighbors* and *non-neighbors*, the identifying instruments are, as a set, statistically significant determinants of each of the four public programs at better than the .00 level.

(ii) The rank model fails to pass the Hansen *J*-test for overidentification in three of four cases at the .05 level. The levels model only fails this test for contraceptive use.

(iii) Qualitatively, the estimates based on levels and ranks are very similar, demonstrating the same pattern of signs and statistically important parameters.

7. Summary and conclusion

This paper proposes an instrumental variable method for program evaluation that requires a single cross-section of data on the spatial intensity of programs and outcomes. The instruments are derived from a simple theoretical model of government decision-making that requires that the government's social welfare function is spatially weakly separable, that is, that the budgeting process is multi-stage with respect to administrative districts and sub-districts. If the attributes of a district rather than individual characteristics decide program placement, and the means (or higher moments) of outcomes for all "competing" districts enter into the government social welfare function, then the district means of individual and district exogenous determinants of these outcomes in *other* districts may be used as instruments.

The identification strategy proposed potentially has broad applicability to the evaluation of public programs when the data consists of spatial variation in program intensity coupled with measures of outcomes for individuals who are matched to places. The validity of using the strictly exogenous attributes of other locales as instruments for a sub-district requires that they not directly affect human capital outcomes in that sub-district. However, it is possible that human capital in a sub-district is determined not only by the characteristics of that sub-district but also by the characteristics of the neighboring sub-districts. This generates *spatial network effects* which confound the instrumental variable approach but do not necessarily rule it out. What is required is that the distance for which *spatial network effects* are relevant is less than the distance for which the attributes of places matter in making program allocation decisions. Arguably, spatial network effects are relevant only close to the area in which behavior is observed. Spatial sorting in response to program placement is another challenge to the validity of the method.

To illustrate the method, the spatial instrumental variables model is estimated and tested with a single cross-section of Indonesian census data. A variety of tests of instrument validity are performed. In particular, sub-sets of instruments from areas at increasing distance from the area in which behavior is observed are used to test for the validity of instruments derived from less distant places. In the case of contraceptive use, these tests are unable to validate the use of the proposed instrumental variable method. In the case of girls schooling, the tests are supportive of our approach in that they fail to reject the requirements of the instrumental variables estimation method. Reassuringly, in none of the four cases does allowing for spatial network effects qualitatively affect the estimated program effects. Nonetheless, the lesson implied is that the validity of the instruments used in this program evaluation strategy proposed must be subjected to empirical test and may not be suitable to every setting or behavior.

References

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58: 277-297.

Baltagi, Badi, Fingleton, Bernard, and Alain Pirotte. 2014. "Estimating and Forecasting with a Dynamic Spatial Panel Data Model." *Oxford Bulletin of Economics and Statistics* 76(1): 112-138.

Baum, Christopher, and Mark E Schaffer. 2007. "Enhanced Routines for Instrumental Variables/GMM Estimation" Boston College, Department of Economics, WP No. 667, September.

Besley, Timothy and Anne Case. 1995. "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition." *American Economic Review* 85(1): 25-45.

Brueckner, Jan K. 2003. "Strategic Interaction Among Governments: An Overview of Empirical Studies." *International Regional Science Review* 26(2): 175-188.

Conley, Timothy, and Ethan Ligon. 2002. "Economic Distance and Cross-Country Spillovers." *Journal of Economic Growth* 7(2): 157-187.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Experiment." *American Economic Review* 91: 795-813.

Elhorst, Paul. 2014. *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer: Berlin New York Dordrecht London.

Hausman, Jerry A., and William E Taylor. 1981. "Panel Data and Unobservable Individual Effects." *Econometrica* 49(6): 1377-1398.

Holtz-Eakin, Douglas, Newey, Whitney, and Harvey S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56(6): 1371-1395.

Kelejian, Harry, and Ingmar Prucha. 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." *Journal of Real Estate Finance and Economics* 17(1): 99-121.

Kleibergen, Frank, and Mark E. Schaffer. 2007. "RANKTEST: Stata Module to Testing the Rank of a Matrix with the Kleibergen-Paap rk Statistic." At: http://ideas.repec.org/c/boc/bocode/s456865.html

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72(1): 159-217.

Pitt, Mark M., Rosenzweig, Mark R., and Donna M Gibbons. 1993. “The Determinants and Consequences of the Placement of Government Programs in Indonesia.” *World Bank Economic Review* 7(3): 319-348.

Pörtner, Claus, Beegle, Kathleen, and Luc Christiaensen. 2014. “Does Family Planning Reduce Fertility? Evidence from Rural Ethiopia.” Unpublished manuscript.

Rubin, Donald B. 1978. “Bayesian inference for causal effects: The Role of Randomization.” *Annals of Statistics* 6: 34–58.

Tiebout, Charles M. 1956. “A Pure Theory of Local Expenditures.” *Journal of Political Economy* 64(5): 416-424.

van der Klaauw, Wilbert. 2002. “Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach.” *International Economic Review*, 43(4): 1249–87

Table 1: Means and standard deviations for the endogenous variables

Variables	
Outcomes	
Current school enrollment for girls ages 10-18 years	0.593 (0.196) N=2921
Current school enrollment for boys ages 10-18 years	0.659 (0.178) N=2919
Whether last child's year of birth lies between 1978-1980 for women ages 21-30 years	0.689 (0.163) N=2914
Whether any contraceptives are currently being used by women ages 21-30 years	0.280 (0.244) N=3033
Programs	
Proportion of households in villages	
with grade schools	0.774 (0.279) N=2921
with PUSKESMAS clinics	0.245 (0.196) N=2921
with family planning clinics	0.486 (0.335) N=2921
with junior or secondary schools	0.394 (0.388) N=2921

Standard deviations in parentheses. "N" denotes the number of *sub-district (kecamatan)* observations.

Table 2: Means and standard deviations for the exogenous variables

Variable	Sub-district Mean (1)	Sub-district SD (2)	*Neighboring* Sub-districts SD (3)	*Non-neighboring* Sub-districts SD (4)
Environmental variables				
Proportion of households in villages				
with urban status	0.129	0.262	0.197	0.180
with drought, flood, earthquake or other shocks in the last five years	0.410	0.344	0.238	0.188
with a coastal environment	0.169	0.283	0.227	0.194
Individual and household attributes				
Dummy for household religion is Islam	0.826	0.325	0.177	0.302
Dummy for household religion is Christianity	0.131	0.288	0.049	0.260
Land owned by household (acres)	0.648	0.718	8.514	5.115
Dummy for household owns its own home	0.921	0.124	0.300	0.184
Dummy for household head's language is Indonesian	0.074	0.193	0.263	0.150
Mother's age (years)	40.308	2.722	8.341	3.398
Household head's age (years)	46.068	3.273	13.191	13.010
Mother's schooling (years)	2.441	1.611	15.651	3.466
Household head's schooling (years)	3.422	1.733	1.158	1.158
Proportion of households in villages with urban status				
interacted with land owned by household	2.680	8.515	4.682	3.612
interacted with dummy for household owns home	0.101	0.194	0.138	0.123
interacted with mother's schooling	0.538	1.377	1.045	0.998
interacted with household head's schooling	0.701	1.729	1.331	1.260
interacted with dummy for head's lang. is Indonesian	0.032	0.130	0.109	0.126
interacted with dummy for religion is Christianity	0.013	0.058	0.039	0.053
interacted with dummy for religion is Islam	0.109	0.221	0.165	0.215
interacted with mother's age	0.522	1.064	0.801	1.027
interacted with father's age	0.597	1.210	0.909	1.168
interacted with drought, flood, earthquake or other shocks in the last five years	0.051	0.155	0.109	0.142

"SD" denotes standard deviation. Since the means for *neighboring* sub-districts (*kecamatans*) and the means for *non-neighboring* sub-districts are approximately the same as the sub-district means, the table presents only SDs in columns (3) and (4). As non-neighboring sub-districts and distant non-neighbors are both means within districts (*kabupatans*), standard deviations for distant non-neighbors are not separately reported as these will be close to the standard deviations of non-neighboring sub-districts.

Table 3: Models with exogenous programs

	Current school enrollment for girls ages 10-18	Current school enrollment for boys ages 10-18	Whether last child's year of birth lies between 1978-1980 for women ages 21-30	Whether any contraceptives are currently being used by women ages 21-30
Proportion of households in villages				
with grade schools	0.042***	0.031**	-0.047***	0.085***
	(0.016)	(0.015)	(0.016)	(0.023)
with junior or secondary schools	0.096***	0.103***	0.010	-0.022
	(0.025)	(0.024)	(0.022)	(0.032)
with PUSKESMAS clinics	-0.022	-0.027	-0.038**	-0.017
	(0.018)	(0.017)	(0.016)	(0.023)
with family planning clinics	0.005	-0.012	-0.017	0.090***
	(0.012)	(0.011)	(0.011)	(0.016)
Observations (individuals)	82,891	82,889	87,655	95,372
Number of sub-districts	2,921	2,919	2,914	3,033

Random effects estimates. *p*-values in square brackets. *** Denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level.

Table 4: Current school enrollment for girls and boys ages 10-18

	Girls enrollment				Boys enrollment				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Proportion of households in villages									
with grade schools	0.049	0.080	0.103**	0.172***	0.047	0.083*	0.139***	0.171***	0.089**
	(0.041)	(0.049)	(0.044)	(0.055)	(0.036)	(0.045)	(0.041)	(0.045)	(0.037)
with junior or secondary schools	0.328***	0.389***	0.315***	0.247**	0.192**	0.259***	0.174**	0.167**	0.241***
	(0.085)	(0.088)	(0.078)	(0.096)	(0.078)	(0.086)	(0.073)	(0.080)	(0.079)
with PUSKESMAS clinics		-0.095	-0.043	-0.069		-0.062	-0.015	-0.029	-0.193***
		(0.074)	(0.063)	(0.083)		(0.069)	(0.060)	(0.071)	(0.068)
with family planning clinics		-0.030	-0.055	-0.037		-0.054	-0.091**	-0.147***	-0.166***
		(0.042)	(0.037)	(0.045)		(0.041)	(0.038)	(0.042)	(0.034)
Neighboring sub-districts	IV	IV	IV	Ind	IV	IV	IV	Ind	Ind
Non-neighboring sub-districts	No	No	IV	IV	No	No	IV	IV	Ind
Distant non-neighbors	No	No	No	No	No	No	No	No	IV
Hansen's *J*-test χ^2	10.379(10)	10.170(8)	27.683(20)	9.579(8)	12.243(10)	12.138(8)	29.870(20)	12.451(8)	9.319(8)
	[0.408]	[0.253]	[0.117]	[0.296]	[0.269]	[0.145]	[0.072]	[0.132]	[0.316]
Orthogonality test χ^2			18.372(12)[a]				19.458(12)[a]		
			[0.105]				[0.078]		
Redundancy test χ^2			114.544(48)[b]				107.123(48)[b]		120.600(48)[c]
			[0.000]				[0.000]		[0.000]
Spatial network test χ^2				18.860(12)[a]				46.480(12)[a]	40.000(12)[b]
				[0.092]				[0.000]	[0.000]
Observations (sub-districts)	2,921	2,921	2,921	2,921	2,919	2,919	2,919	2,919	2,919

IV-random effects parameter and standard error estimates. Test statistics from GMM with standard errors that are clustered at the district level. Degrees of freedom for the χ^2 tests reported in brackets after the test-statistic. *p*-values in square brackets. *** Denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level. IV indicates included in the first stage and Ind indicates included in the second stage. [a] Denotes for neighboring sub-districts, [b] denotes for non-neighboring sub-districts and [c] denotes for distant non-neighbors.

Table 5: Fertility and contraception

	Whether last child's year of birth is between 1978-1980 for women ages 21-30					Whether any contraceptives are currently being used by women ages 21-30				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Proportion of households in villages										
with grade schools		-0.049	-0.139**	-0.175***	-0.180***		0.541***	0.512***	0.578***	0.406***
		(0.077)	(0.056)	(0.066)	(0.053)		(0.096)	(0.087)	(0.137)	(0.080)
with junior or secondary schools		-0.055	0.010	0.056	-0.082		-0.079	-0.071	-0.058	0.203
		(0.134)	(0.096)	(0.114)	(0.107)		(0.176)	(0.148)	(0.221)	(0.149)
with PUSKESMAS clinics		0.238**	0.201***	0.160*	0.083		-0.435***	-0.447***	-0.478***	-0.267**
		(0.101)	(0.074)	(0.088)	(0.088)		(0.113)	(0.102)	(0.153)	(0.118)
with family planning clinics	-0.304***	-0.326***	-0.225***	-0.164***	-0.083*	0.323***	0.088	0.089	0.088	0.008
	(0.041)	(0.071)	(0.049)	(0.055)	(0.044)	(0.053)	(0.089)	(0.079)	(0.132)	(0.075)
Neighboring sub-districts	IV	IV	IV	Ind	Ind	IV	IV	IV	Ind	Ind
Non-neighboring sub-dist.	No	No	IV	IV	Ind	No	No	IV	IV	Ind
Distant non-neighbors	No	No	No	No	IV	No	No	No	No	IV
Hansen's *J*-test χ^2	25.650(11)	13.438(8)	28.264(20)	17.303(8)	10.956(8)	47.048(11)	31.046(8)	44.031(20)	17.147(8)	13.782(8)
	[0.007]	[0.098]	[0.103]	[0.027]	[0.204]	[0.000]	[0.000]	[0.002]	[0.029]	[0.088]
Orthogonality test χ^2			11.670(12)[a]					19.516(12)[a]		
			[0.473]					[0.077]		
Redundancy test χ^2			119.378(48)[b]		120.026(48)[c]			113.632(48)[b]		
			[0.000]		[0.000]			[0.000]		
Spatial network test χ^2				33.130(12)[a]	50.620(12)[b]				14.680(12)[a]	49.020(12)[b]
				[0.001]	[0.000]				[0.259]	[0.000]
Observations (sub-districts)	2,914	2,914	2,914	2,914	2,914	3,033	3,033	3,033	3,033	3,033

IV-random effects parameter and standard error estimates. Test statistics from GMM with standard errors that are clustered at the district level. Degrees of freedom for the χ^2 tests reported in brackets after the test-statistic. *p*-values in square brackets. *** Denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level. IV indicates included in the first stage and Ind indicates included in the second stage. [a] Denotes for neighboring sub-districts, [b] denotes for non-neighboring sub-districts and [c] denotes for distant non-neighbors.

ONLINE APPENDIX

Online Appendix Table 1: Models with Hausman-Taylor instruments

	Current school enrollment for girls ages 10-18	Current school enrollment for boys ages 10-18	Whether last child's year of birth lies between 1978-1980 for women ages 21-30	Whether any contraceptives are currently being used by women ages 21-30
Proportion of households in villages				
with grade schools	-0.065	0.170***	-0.306***	0.855***
	(0.074)	(0.065)	(0.062)	(0.149)
with junior or secondary schools	0.244***	0.212***	0.038	0.068
	(0.060)	(0.044)	(0.049)	(0.093)
with PUSKESMAS clinics	-0.382***	-0.269***	0.190**	-0.310
	(0.104)	(0.074)	(0.081)	(0.191)
with family planning clinics	0.113	-0.125	-0.096	-0.437**
	(0.084)	(0.078)	(0.072)	(0.178)
Hansen's *J*-test χ^2	26.391(14)	34.412(14)	25.000(14)	28.335(14)
	[0.023]	[0.002]	[0.035]	[0.013]
Observations (individuals)	82,891	82,889	87,655	95,372
Number of sub-districts	2,921	2,919	2,914	3,033

Random effects estimates. Degrees of freedom for the χ^2 tests reported in brackets after the test-statistic. *p*-values in square brackets. *** Denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level.

Online Appendix Table 2: Tests for model under-identification

	(1)	(2)	(3)	(4)
Proportion of households in villages				
with grade schools	0.049	0.080	0.103**	0.172***
	(0.041)	(0.049)	(0.044)	(0.055)
with junior or secondary schools	0.328***	0.389***	0.315***	0.247**
	(0.085)	(0.088)	(0.078)	(0.096)
with PUSKESMAS clinics		-0.095	-0.043	-0.069
		(0.074)	(0.063)	(0.083)
with family planning clinics		-0.030	-0.055	-0.037
		(0.042)	(0.037)	(0.045)
Neighboring sub-districts	IV	IV	IV	Ind
Non-neighboring sub-districts	No	No	IV	IV
Neighboring districts	No	No	No	No
Kleibergen-Paap rk LM statistic χ^2	21.260(11)	22.770(9)	38.350(21)	8.810(9)
	[0.031]	[0.007]	[0.012]	[0.455]
Observations (sub-districts)	2,921	2,921	2,921	2,921

Degrees of freedom for the χ^2 tests reported in brackets after the test-statistic. *p*-values in square brackets. . *** Denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level. The null hypothesis is that the first stage model is under-identified. Tests based on the set of sub-districts used in girls' schooling models. The set of sub-districts vary slightly across outcomes. IV indicates included in the first stage and Ind indicates included in the second stage.

Online Appendix Table 3: Falsification tests: Impact of programs on alternate outcomes

	Whether household religion is Islam		Whether gender of the elderly person who died was male	
	(1)	(2)	(3)	(4)
Proportion of households in villages				
with grade schools	-0.125	-0.007	0.137	0.086
	(0.228)	(0.163)	(0.110)	(0.077)
with junior or secondary schools	-0.282	-0.274	0.531**	0.293*
	(0.392)	(0.301)	(0.249)	(0.159)
with PUSKESMAS clinics	-0.365	-0.293	-0.429**	-0.106
	(0.319)	(0.204)	(0.172)	(0.118)
with family planning clinics	0.465**	0.193	0.116	0.033
	(0.237)	(0.150)	(0.097)	(0.065)
Joint test of significance of programs χ^2	6.520(4)	4.780(4)	13.280(4)	8.580(4)
	[0.163]	[0.311]	[0.010]	[0.073]
Neighboring sub-districts	IV	IV	IV	IV
Non-neighboring sub-districts	No	IV	No	IV
Observations (sub-districts)	2921	2921	2921	2921

IV-random effects estimates. Standard errors in parentheses. Degrees of freedom for the χ^2 tests reported in brackets after the test-statistic. *p*-values in square brackets. *** Denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level. Instruments are attributes of neighboring sub-districts.

Online Appendix Table 4: Tests for effects of migration on girls schooling and recent fertility

	Migration indicator: Duration of residence in current province is less than one year		Migration indicator: Current province is different from province lived in five years ago	
	Girls schooling	Recent fertility	Girls schooling	Recent fertility
	(1)	(2)	(3)	(4)
Proportion of households in villages				
with grade schools	0.084	-0.080	0.089	-0.104
	(0.071)	(0.110)	(0.072)	(0.120)
with junior or secondary schools	0.336***	-0.081	0.329***	-0.077
	(0.128)	(0.209)	(0.125)	(0.204)
with PUSKESMAS clinics	-0.081	0.465***	-0.052	0.386**
	(0.133)	(0.166)	(0.125)	(0.185)
with family planning clinics	-0.028	-0.388***	-0.036	-0.364***
	(0.067)	(0.111)	(0.078)	(0.114)
Migration indicator	0.791	-2.633	0.510	-1.329
	(2.606)	(3.525)	(1.095)	(1.062)
Interaction of migration indicator and proportion of households in villages				
with grade schools	-2.583	7.470	-0.826	3.203**
	(3.827)	(5.308)	(1.680)	(1.566)
with junior or secondary schools	0.131	-0.809	0.376	-0.316
	(2.347)	(3.275)	(0.932)	(0.986)
with PUSKESMAS clinics	0.068	-10.729	-0.899	-2.775
	(5.787)	(7.163)	(2.028)	(2.152)
with family planning clinics	1.342	3.022	0.360	0.686
	(3.552)	(4.093)	(1.563)	(1.240)
Joint test of significance of migration interaction variables χ^2	0.520(4) [0.971]	7.350(4) [0.119]	0.670(4) [0.956]	7.330(4) [0.119]
Observations (sub-districts)	2921	2914	2921	2914

IV-random effects estimates. Degrees of freedom for the χ^2 tests reported in brackets after the test-statistic. *p*-values in square brackets. *** Denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level. Instruments are attributes of neighboring sub-districts.

Online Appendix Table 5: Current school enrollment for girls and boys ages 10-18: Ranks as IVs

	Girls enrollment				Boys enrollment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proportion of households in villages								
with grade schools	0.070	0.068	0.067*	0.122***	0.026	0.045	0.032	0.031
	(0.044)	(0.042)	(0.041)	(0.039)	(0.043)	(0.037)	(0.034)	(0.033)
with junior or secondary schools	0.245***	0.205**	0.252***	0.322***	0.320***	0.255***	0.359***	0.243***
	(0.093)	(0.090)	(0.096)	(0.083)	(0.093)	(0.081)	(0.082)	(0.075)
with PUSKESMAS clinics	-0.191**	-0.148**	-0.214***	-0.165**	-0.147**	-0.112**	-0.181***	-0.139**
	(0.079)	(0.062)	(0.064)	(0.068)	(0.068)	(0.055)	(0.053)	(0.059)
with family planning clinics	-0.016	0.017	0.013	0.015	-0.026	-0.043	-0.039	-0.035
	(0.039)	(0.035)	(0.034)	(0.032)	(0.039)	(0.032)	(0.030)	(0.029)
Neighboring sub-districts	No	No	No	Ind	No	No	No	Ind
Rank of sub-district in district	IV	IV	Ind	IV	IV	IV	Ind	IV
Rank of sub-district in province	No	IV	IV	IV	No	IV	IV	IV
Hansen's *J*-test χ^2	11.336(8)	26.896(20)	4.125(8)	25.827(20)	20.155(8)	33.414(20)	10.656(8)	26.192(20)
	[0.183]	[0.138]	[0.846]	[0.172]	[0.010]	[0.030]	[0.222]	[0.160]
Orthogonality test χ^2		14.963(12)[a]		14.221(12)[a]		13.260(12)[a]		10.052(12)[a]
		[0.243]		[0.287]		[0.351]		[0.611]
Redundancy test χ^2		94.128(48)[b]		87.970(48)[b]		90.541(48)[b]		87.332(48)[b]
		[0.000]		[0.000]		[0.000]		[0.000]
Spatial network test χ^2			17.340(12)[a]	19.860(12)[c]			28.170(12)[a]	37.520(12)[c]
			[0.137]	[0.070]			[0.005]	[0.000]
Observations (sub-districts)	2,921	2,921	2,921	2,921	2,919	2,919	2,919	2,919

IV-random effects estimates. Standard errors in parentheses. Degrees of freedom for the χ^2 tests reported in brackets after the test-statistic. *p*-values in square brackets. *** Denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level. IV indicates included in the first stage and Ind indicates included in the second stage. [a] Denotes for rank of sub-district in district, [b] denotes for rank of sub-district in province and [c] denotes for neighboring sub-districts.

Online Appendix Table 6: Fertility and contraception: Ranks as IVs

	Whether last child's year of birth lies between 1978-1980 for women ages 21-30				Whether any contraceptives are currently being used by women ages 21-30			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proportion of households in villages								
with grade schools	-0.242***	-0.179***	-0.177***	-0.125***	0.638***	0.535***	0.542***	0.465***
	(0.064)	(0.047)	(0.051)	(0.044)	(0.090)	(0.077)	(0.086)	(0.072)
with junior or secondary schools	0.113	0.013	-0.040	0.013	-0.252	-0.089	0.149	-0.046
	(0.118)	(0.093)	(0.113)	(0.083)	(0.177)	(0.149)	(0.191)	(0.141)
with PUSKESMAS clinics	0.095	0.101*	0.187**	0.068	-0.327***	-0.348***	-0.599***	-0.375***
	(0.083)	(0.059)	(0.074)	(0.061)	(0.115)	(0.094)	(0.126)	(0.098)
with family planning clinics	-0.174***	-0.145***	-0.165***	-0.090**	-0.016	0.043	0.030	0.061
	(0.057)	(0.042)	(0.047)	(0.039)	(0.078)	(0.069)	(0.080)	(0.064)
Neighboring sub-districts	No	No	No	Ind	No	No	No	Ind
Rank of sub-district in district	IV	IV	Ind	IV	IV	IV	Ind	IV
Rank of sub-district in province	No	IV	IV	IV	No	IV	IV	IV
Hansen's *J*-test χ^2	12.587(8)	50.091(20)	14.578(8)	53.127(20)	24.726(8)	59.765(20)	17.817(8)	58.711(20)
	[0.127]	[0.000]	[0.068]	[0.000]	[0.002]	[0.000]	[0.023]	[0.000]
Orthogonality test χ^2		15.660(12)[a]		13.463(12)[a]		13.747(12)[a]		11.902(12)[a]
		[0.207]		[0.336]		[0.317]		[0.454]
Redundancy test χ^2		88.712(48)[b]		91.747(48)[b]		82.051(48)[b]		84.274(48)[b]
		[0.000]		[0.000]		[0.002]		[0.001]
Spatial network test χ^2			33.320(12)[a]	47.020(12)[c]			22.870(12)[a]	25.050(12)[c]
			[0.001]	[0.000]			[0.029]	[0.015]
Observations (sub-districts)	2,914	2,914	2,914	2,914	3,033	3,033	3,033	3,033

IV-random effects estimates. Standard errors in parentheses. Degrees of freedom for the χ^2 tests reported in brackets after the test-statistic. *p*-values in square brackets. *** Denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level. IV indicates included in the first stage and Ind indicates included in the second stage. [a] Denotes for rank of sub-district in district, [b] denotes for rank of sub-district in province and [c] denotes for neighboring sub-districts.