Democratization and Infrastructure Investment: Evidence from Healthcare in Indonesia

Allan Hsiao *
Massachusetts Institute of Technology

January 31, 2019
(most recent version here)

Abstract

How does democratization impact the spatial allocation of public goods, and what are the implications for social welfare? To answer these questions, I study Indonesia’s expansion of healthcare infrastructure — hospitals, clinics, and subclinics — before and after democratization. First, I use geo-coded panel data on healthcare usage to estimate spatial demand for healthcare services. The estimated demand system allows me to quantify the consumer surplus gains of any given spatial allocation of facilities. Second, I use these estimates to construct a measure of misallocation, which captures the difference between the maximum and realized surplus gains for a given construction budget. I construct this measure over time, and I show that misallocation is rising leading up to Indonesia’s democratization but levels off afterwards. Third, I use a structural model to decompose the channels driving these trends. I model the decision over facility placements as a dynamic discrete choice problem, and I estimate the model by revealed preference. I find that the reform decreased misallocation by 10 percentage points, driven by electoral concerns that prevent authorities from allocating resources according to their own preferences. Local elections mean that local governments undervalue benefits to non-constituents, but I find that the distortion arising from these uninternalized spillovers is small. The effect of elite capture is similarly small.

*Email: ajhsiao@mit.edu.
1 Introduction

The role of democracy in economic development has long been the subject of intense policy interest. Globally, recent advances suggest a positive effect on economic growth (Acemoglu et al., 2019). In developing countries in particular, Pande (2011) surveys a large body of evidence suggesting that, when voters are well informed, electoral accountability enforces better governance. Infrastructure investment in particular is a key part of governance and has important implications for economic growth. How does democratization impact infrastructure investment, and what are the implications for social welfare? To answer these questions, I study the expansion of healthcare infrastructure in Indonesia before and after democratization.

Following the fall of Suharto, Indonesia implemented a set of major reforms in 1999. At the forefront of these reforms was democratization, which brought the first free elections since 1955 and subsequently the direct election of local mayors. Local governance was further emphasized with the the transfer of significant decision-making power from Jakarta to local governments. In this paper, I study the welfare implications of this transition to local governance via its impact on infrastructure investment. On one hand, local elections are beneficial if they prevent central government authorities from commandeering resources for their own use. On the other hand, there may be welfare losses if investments have spillover effects, and local governments fail to internalize the effects on non-constituents.

Focusing specifically on healthcare infrastructure, I quantify welfare effects by estimating a model of demand for healthcare facilities over space. I draw on a panel dataset geocoded at the village level that captures the expansion of healthcare infrastructure over time, as well as corresponding increases in usage levels. Conditional on being sick, individuals choose among visiting their closest public hospital, private hospital, clinic, or subclinic, or the outside option of not seeking treatment. They have disutility from the distance and congestion of a given facility, and the expansion of healthcare infrastructure over time generates panel variation in both. I then estimate a logit demand model over these characteristics and a set of village fixed effects.

The estimated demand system allows me to construct a measure of facility misallocation over time. With the demand system, I can predict the consumer surplus gains generated by any given spatial allocation of facilities. For a given district in a given three-year interval, I calculate the gains generated both by the actual allocation and by the allocation that maximizes social welfare (including benefits to non-constituents). My measure of misallocation is one minus the ratio of the two, such that a government that places its facilities in the welfare-maximizing way has zero misallocation by this measure. Plotting the measure over time, I find that misallocation is rising leading up to Indonesia’s democratization but levels off afterwards.

To analyze the channels driving this trend, I build and estimate a model of facility placement. In choosing where to place facilities, the government considers the consumer surplus, vote response, and costs associated with a given facility placement. I also accommodate unobserved factors driving the government’s placement decision. The government’s weight on votes reflects electoral accountability, and for spillovers I allow the government to have a non-zero weight on non-constituent consumer surplus. The unobserved factors allow for some government preferences to be uncorrelated with what is best for citizens, as might be the case with elite capture. For estimation, I develop a novel application of moment-inequality techniques to estimate the government’s objective function parameters by revealed preference. In doing so, I face two key challenges
that complicate estimation: the dynamic nature of infrastructure investments, which are durable, and the accommodation of factors that are known to the agent but unobserved to the econometrician. I sidestep both issues with a matched resequencing strategy similar in spirit to matched difference-in-differences.

The estimated government preferences are directly interpretable. These parameters describe how the government trades off votes and costs against consumer surplus in dollar terms. The estimates therefore capture the government’s effective valuation of a vote, and its cost of government spending. First, I find that the government values an additional vote at $17 of consumer surplus. This magnitude is large, and it implies that patient welfare and votes receive roughly equal consideration in determining facility placement. Second, to justify an additional dollar of spending, I find that the government requires a return of $3 of consumer surplus.

With the estimated model in hand, I decompose the channels driving the change in misallocation generated by the reform. For the metropolitan area of Jakarta, I find that the reform decreases misallocation in the post-reform by 10 percentage points. Electoral concerns drive the change and in isolation would have decreased misallocation by 12 percentage points. The uninternalized spillovers associated with localized decision-making create 3 percentage points of welfare loss, but this distortion is small relative to the benefits of electoral accountability. Compared to pre-reform levels, I find a small decrease in post-reform elite capture that in isolation decreases misallocation by 2 percentage points.

A large literature on political decentralization studies the impacts of electoral accountability in local elections. Seabright (1996) emphasizes voter information in a theoretical model to argue that local elections improve accountability by allowing local issues take center stage. Empirically, Ferraz and Finan (2008) show that voters hold candidates accountable by responding to performance, and Casey (2015) finds that the shift to local elections increases political accountability by empowering voters. At the same time, decentralized decision-making may be socially suboptimal in the presence of spillovers and economies of scale (Oates, 1972). Sigman (2002), Kahn et al. (2015), and Lipscomb and Mobarak (2017) show empirically that uninternalized spillovers result in lower water quality where water flows from one jurisdiction into another. Finally, Bardhan and Mookherjee (2000) and Bardhan and Mookherjee (2006) argue that elite capture may also be worsened in local elections, although Alatas et al. (2012) find elite capture to be quantitatively small in an empirical, albeit non-electoral, setting. I contribute to the literature by providing an empirical evaluation of these trade-offs from a national planning perspective. While other empirical work has generally focused on particular mechanisms, I use a structural model to quantify welfare effects and the relative strengths of the mechanisms at play.

I also build on a growing literature in industrial organization that applies moment-inequality techniques introduced in Pakes (2010) and Pakes et al. (2015). However, unlike the typical IO application in which a firm’s objective function is well approximated by a profit function, in my setting patient welfare, votes, and costs offer a rather incomplete characterization of the government’s objective function. As such, I develop a methodology that flexibly accommodates the government’s unobserved decision factors while also allowing these factors to vary at a fine level of disaggregation. To do so, I combine techniques applied in Holmes (2011), Ho and Pakes (2014), and Ishii (2007) in order to address unobservables in my dynamic setting.

Lastly, I draw on a literature that documents sources of quasi-random policy variation in Indonesia, and I show how this variation provides identification in my structural model. Martinez-
Bravo et al. (2017) studies appointment dates of Suharto-regime district mayors, who were allowed to complete their terms after the fall of Suharto, and finds that districts with more exposure to these mayors experience worse governance outcomes. This variation identifies the government’s electoral concerns by pointing to a period of time in which some district heads were not subject to electoral accountability. Burgess et al. (2012) and Bazzi and Gudgeon (2017) analyze the timing of redistricting around two moratoria to study deforestation and ethnic conflict, respectively. Redistricting helps to pin down local governments’ internalizing of non-constituent surplus by generating variation in who is and who is not a constituent. Compared to the literature, I therefore highlight how these sources of variation can be combined with a structural model to generate further insights.

In what follows, section 2 provides background information on Indonesia’s decentralization reform and its public healthcare system. Section 3 outlines the data. Section 4 models and estimates the spatial demand system. Section 5 describes the measure of misallocation and how it varies over time. Section 6 presents the facility-placement model, and section 7 details the estimation procedure and results. Section 8 is the channel decomposition exercise, and section 9 concludes.

2 Institutional Details

Healthcare in Indonesia is an ideal setting for studying the effect of decentralization on public goods provision. The country underwent decentralization in 1999 following the fall of Suharto’s highly centralized regime. Since decentralization, local district governments have been responsible for providing healthcare services. As well, free elections have allowed citizens to reclaim political power previously amassed by Suharto and his deputies. In this section, I provide further details both on the decentralization reforms and on the Indonesian healthcare system.

2.1 Decentralization

“Decentralization” in Indonesia refers to the set of administrative and political reforms following the end of Suharto’s 31-year regime in 1998. Following Suharto’s fall, a transitional government passed legislation calling for the transfer of power to local governments.1 Within two years, local district governments received additional authority in the form of two million civil servants, 30% of government expenditures, and responsibility for the provision of a range of public services. At the same time, Indonesia held its first free elections since 1955. Previously, Suharto’s regime had suppressed opposition parties by forcing them to merge into two parties, one Islamic and one non-Islamic, controlling the opposition leadership, and implementing a recall system that enabled the removal of individual legislators. As a result, between 1973 and 1998 Suharto’s party, Golkar, won landslide victories in five legislative elections.

Furthermore, after 2001 a number of new district governments were established as existing districts split into smaller districts. I refer to this reform as “redistricting.” District governments were required to apply for central government approval to redistrict, and the central government placed a moratorium on redistricting from 2004 and 2006 and again from 2009 to 2012. Burgess et al. (2012) and Bazzi and Gudgeon (2017) argue that the timing of redistricting around first moratorium is plausibly exogenous.

2.2 Administrative entities

“Districts” are subdivisions of provinces (propinsi/provinsi) and refer collectively to both regencies (kabupaten) and cities (kota). Districts are subdivided into subdistricts (kecamatan), which are further subdivided into rural villages (desa) and urban communities (kelurahan). These rural villages and urban communities form the smallest administrative entities in Indonesia, and in this paper I will refer to both as “villages.” In sum, the administrative hierarchy is as follows: nation, province, district, subdistrict, and village.

2.3 Healthcare

Indonesia relies on a multi-layered, referrals-based system to provide care to its 260 million citizens. Below, I summarize the institutional details that are relevant for this paper.

The public healthcare system is organized in layers.

The public system consists of hospitals, clinics, and smaller facilities. Hospitals are themselves divided into classes: class A hospitals average 1,450 beds and cover a range of specialties, while class D hospitals are district-level facilities that average 70 beds and offer only general care. Below hospitals are clinics (puskesmas), which are usually staffed by a physician and focus on providing primary care. Some clinics are equipped to provide basic inpatient services. Clinics are further supported by a network of subclinics (pustu) and village facilities, including village health posts (poskesdes), village maternity posts (polindes), and neighborhood health posts (posyandu). Subclinics are staffed with one to three nurses and visited weekly to monthly by a physician. Village facilities are often staffed by local volunteers trained by health workers and may operate on borrowed premises.

Clinics provide referrals to in-district hospitals.

The system is based on referrals, with clinics acting as gatekeepers to the public hospital network. As such, clinics are the main providers of basic services and primary care, while district-level hospitals are the main providers of curative care. Hospital resources are rationed to those with high need, while clinics are mandated to treat patients suffering from minor illnesses without further referrals. In practice, however, it is possible for patients to receive hospital services without a clinic referral, either by presenting with an emergency condition or by obtaining a referral from on-site clinics that commonly operate within hospitals.

Clinics are required to provide referrals only to in-district hospitals except in exceptional cases. Such restrictions are enforceable even when the nearest out-of-district hospital is significantly closer to the patient than the nearest in-district hospital. Before the implementation of electronic data systems in recent years, however, it was difficult to enforce in-district residence requirements in practice.

Access expanded with healthcare infrastructure and insurance programs.

The Indonesian government has expanded access to healthcare services by devoting significant resources to building infrastructure. Since the origins of the clinic system in the 1970s, the government has worked toward its formal goal of one clinic per 30,000 people and one subclinic per 10,000 people. In the 1990s, the government implemented the Bidan di Desa initiative, which sought

---

2 Anecdotally, there is a list of 99 minor illnesses that are officially discouraged from referral to hospitals.

3 Clinics also have a mandate to serve populations at the subdistrict level, even for subdistricts with fewer than
Indonesian Hospitals

Figure 1: Hospitals versus population density in Java, 1990-2014

(a) Java 1990

(b) Java 2014

Orange dots are public hospitals, blue dots are private hospitals, and gray shading conveys population density. There are 390 hospitals in 1990 and 1,258 in 2014. Expansion seems strongest in areas of initial concentration and high population density – the cluster to the northwest is Jakarta. Data are from PODES.

to station a midwife in every village. At the same time, the hospital network has continued to grow. Figure 1 shows the visible expansion of hospitals in Java, Indonesia’s most populous island, over more than two decades. Today, there are about 2,500 hospitals, 10,000 clinics, and 25,000 subclinics in Indonesia.4

Insurance programs have also played an important role. The Indonesian government launched a universal healthcare program, Jaminan Kesehatan Nasional (JKN), in 2014 with plans to achieve full coverage by 2019. This program builds on the Askeskin (2004) and Jamkesmas (2008) programs, which provided coverage to the poor and near poor.

Under decentralization, district governments place new facilities.

Before decentralization, the central government funded facility construction and possessed

---

4 Despite wide-reaching coverage, poor quality of care is an ongoing concern. By this metric, Indonesia lags behind its Southeast Asian neighbors. Absenteeism among health personnel is a chronic issue, and hospital beds are scarce at 0.9 beds per 100,000 people. Health workers are thinly spread, with 0.2 physicians and 1.2 nurses per 1,000 people. There is a particular need for specialists.
broad authority over the placement of new facilities. The clinic system, for example, was originally funded by the same INPRES program that funded the large-scale construction of more than 60,000 schools in the 1970s. Later, central funding continued through the Ministry of Health.

Since decentralization in 2001, district governments have been responsible for the direct implementation of healthcare services. Funding continues to come from the central government, including through disbursements from the Special Allocation Fund (Dana Alokasi Khusus) that are earmarked for facility construction. District governments negotiate with the central government for funds, including with proposals for new facilities. But afterwards, the central government cannot enforce agreed-upon proposals, and in some cases has limited information on the completion status of funded projects. The process is therefore conceptually similar to two-stage budgeting, with district governments having full leeway to choose the placement of their allocated facilities.

The private system primarily serves the wealthy.

Private hospitals cater to the wealthy and operate outside of the public system. Growth in the number of private hospitals has largely involved the establishment of smaller, single-specialty hospitals – particularly in dentistry. Private doctor practices (praktek dokter) and polyclinics (poliklinik) are the private counterparts to public clinics and often result from public doctors who open secondary practices. These facilities also serve relatively wealthy clientele, although to a lesser extent than private hospitals do.

3 Data and Descriptive Evidence

How does decentralization affect facility placement? I draw on geocoded, village-level panel data on healthcare infrastructure in order to analyze the reduced-form relationship between decentralization and the allocation of facilities across space. I find that decentralization affects facility placement through both channels of interest – spillovers and electoral accountability.

3.1 Data

Village-level data on health infrastructure come from the Village Potential Statistics (PODES), a census of Indonesian villages conducted every few years. I use data from 1990 to 2014. The core data cover a range of facility types, including hospitals, clinics, subclinics, and village-level facilities, and record the number of facilities of each type by village. In 2011, the data also contain information on facility quality for clinics, subclinics, and a subset of village-level facilities. The PODES data also contain village-level voting results in the 1999 and 2004 legislative elections.

The PODES data do not distinguish between public and private hospitals, so I also draw on Rumah Sakit Online (RSO), an online database of hospitals maintained by the Indonesian Ministry of Health. This database lists approximately 2,500 hospitals and contains information on address, type (public or private), number of beds, number of personnel, and some measures of hospital

---

5 The Ministry of Health still plays a role in Indonesia healthcare by setting national policy and allocating public doctors. Province governments conduct training, run a limited number province-level hospitals, and supervise districts.

6 The full set of facility types covered is: hospitals, clinics, subclinics, maternity hospitals, polyclinics, doctor practices, midwife practices, village health posts, village maternity posts, and neighborhood healthcare posts.

7 The 2011 PODES Infrastructure module has data on village health posts (poskesdes), village maternity posts (polindes), and neighborhood health posts (posyandu).
quality. The SUSENAS data distinguish between public and private hospitals, but the PODES data do not. Thus, I use the RSO distinction between public and private hospitals in 2016 in an attempt to classify hospitals in the PODES data.

The National Socioeconomic Survey (SUSENAS) dataset contains annually collected, individual-level data on healthcare usage. I use data from 1993 to 2010. The data record healthcare usage by facility type (e.g., public hospital visits), but they do not track the specific facility used. The data also contain village locations, demographic information, and a limited set of health outcomes.

The village location information is crucial. First, I use the village locations to construct a panel by linking the PODES waves over time. I do the same with the SUSENAS data. Second, the village locations allow me to link the PODES and SUSENAS panels to each other at a fine geographic level. Finally, villages are sufficiently small that I can geocode the data and calculate distances between individuals and facilities with a relatively high degree of accuracy.

Finally, I use data from the Jakarta Smart City initiative, which estimated land prices in 2015 at the sub-block level for the province of Jakarta. These prices were constructed using administrative records, market transactions, and field visits for the purpose of updating the provincial land tax records. Harari and Wong (2018) use the data to study slum development, and they describe the data in detail. I aggregate these data to construct village-level land prices in Jakarta, and I am working on compiling prices for the rest of Indonesia.

Table 1 summarizes the village-level data by year. For the more than 62,000 villages for which I was able to construct a balanced panel, the number of facilities has grown over time for all facilities types. The panel data cover a tripling of hospitals from 1990 and a doubling of clinics and subclinics. Consistent with this growth, facility distance and congestion has declined. Lastly, the vast majority of these villages are rural and continue to be designated as such, although the population-weighted average has declined from approximately 0.70 to 0.55 over the period of study.

3.2 Internalizing spillover effects

To build intuition, consider facility placement in border villages. If decentralized decision makers fail to fully internalize facilities’ spatial spillovers, then these villages will be less likely to receive new facilities after decentralization as compared to before. Under decentralization, a district will avoid placing facilities near the border, where they will be easily accessible to non-constituents. Non-constituents from across the border consume facility resources but contribute neither taxes nor votes. On the other hand, under centralization the central government has no aversion to placing facilities near a district border. Individuals on both sides of the border contribute taxes and votes to the central government.

I test for this effect using variation generated by redistricting. After decentralization, a number of districts split into smaller districts, generating new border villages where the splits occurred.

---

8 The data do not contain village locations past year 2010.
9 An empirical concern is that, within a district, most hospital construction occurs in the district capital. With a dataset coded at the city level, one would therefore struggle to find any location effects if hospitals are always built in the same city. Within a city, however, there are many neighborhoods, and data geocoded at the village level are capable of detecting shifts toward certain neighborhoods over others.
10 I consider the district’s placement response because out-of-district pricing (akin to out-of-state tuition) is difficult in this context. In practice, facilities have only imperfect means of determining patients’ districts of residence, particularly in the years covered by the data.
Table 1: Summary statistics by year (PODES)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Public hospitals</td>
<td>664</td>
<td>750</td>
<td>798</td>
<td>863</td>
<td>942</td>
<td>1,084</td>
<td>1,279</td>
<td>1,526</td>
<td>1,840</td>
</tr>
<tr>
<td>Private hospitals</td>
<td>231</td>
<td>260</td>
<td>282</td>
<td>307</td>
<td>351</td>
<td>395</td>
<td>465</td>
<td>544</td>
<td>654</td>
</tr>
<tr>
<td>Clinics</td>
<td>5,202</td>
<td>6,021</td>
<td>6,435</td>
<td>6,868</td>
<td>7,199</td>
<td>7,719</td>
<td>8,533</td>
<td>9,398</td>
<td>10,788</td>
</tr>
<tr>
<td>Subclinics</td>
<td>12,412</td>
<td>15,660</td>
<td>17,140</td>
<td>19,154</td>
<td>20,196</td>
<td>21,480</td>
<td>23,217</td>
<td>24,767</td>
<td>27,744</td>
</tr>
<tr>
<td>Distance, public hospital</td>
<td>30.58</td>
<td>28.71</td>
<td>28.36</td>
<td>27.43</td>
<td>26.21</td>
<td>23.91</td>
<td>21.51</td>
<td>19.72</td>
<td>18.32</td>
</tr>
<tr>
<td>Distance, private hospital</td>
<td>66.22</td>
<td>65.78</td>
<td>64.63</td>
<td>63.17</td>
<td>61.69</td>
<td>59.36</td>
<td>56.36</td>
<td>53.60</td>
<td>50.83</td>
</tr>
<tr>
<td>Distance, clinic</td>
<td>6.95</td>
<td>6.32</td>
<td>6.07</td>
<td>5.75</td>
<td>5.49</td>
<td>5.17</td>
<td>4.68</td>
<td>4.34</td>
<td>4.07</td>
</tr>
<tr>
<td>Distance, subclinic</td>
<td>4.21</td>
<td>3.46</td>
<td>3.18</td>
<td>2.79</td>
<td>2.63</td>
<td>2.47</td>
<td>2.25</td>
<td>2.07</td>
<td>1.81</td>
</tr>
<tr>
<td>Congestion, public hospital</td>
<td>442.83</td>
<td>424.52</td>
<td>418.10</td>
<td>374.01</td>
<td>363.22</td>
<td>328.79</td>
<td>296.96</td>
<td>258.44</td>
<td>225.54</td>
</tr>
<tr>
<td>Congestion, private hospital</td>
<td>1,274.83</td>
<td>1,280.61</td>
<td>1,261.05</td>
<td>1,127.77</td>
<td>1,087.34</td>
<td>1,035.04</td>
<td>984.82</td>
<td>929.34</td>
<td>811.11</td>
</tr>
<tr>
<td>Congestion, clinic</td>
<td>38.13</td>
<td>33.90</td>
<td>32.41</td>
<td>29.67</td>
<td>29.89</td>
<td>29.11</td>
<td>28.54</td>
<td>27.04</td>
<td>25.28</td>
</tr>
<tr>
<td>Rural (dummy)</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.82</td>
<td>0.81</td>
<td>0.82</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>Population</td>
<td>2,901</td>
<td>3,026</td>
<td>3,090</td>
<td>3,022</td>
<td>3,199</td>
<td>3,325</td>
<td>3,529</td>
<td>3,682</td>
<td>3,776</td>
</tr>
</tbody>
</table>

Each observation is a village. The first four rows are totals, and all other rows are averages. Distance is to the closest facility of a given type and is measured in kilometers. Congestion is of the closest facility of a given type and is measured as the number of people (in thousands) for whom this facility is the closest of its type.

Expanding the intuition of border villages, consider a general measure of a village’s exposure to individuals from other districts,

$$
\text{exposure}_{vt} = \frac{\sum_{v'|D_t(v') \neq D_t(v)} \text{population}_{v'} \cdot \max \left\{ 1 - \frac{1}{100} \text{distance}(v, v'), 0 \right\}}{\sum_{v'} \text{population}_{v'} \cdot \max \left\{ 1 - \frac{1}{100} \text{distance}(v, v'), 0 \right\}},
$$

such that village $v$’s exposure is given by the proportion of people within 100 km of a village, downweighting linearly by distance, who reside in a district different from that of the village.\textsuperscript{11} This measure will be high for border villages and lower for non-border villages. Redistricting, which can divide neighboring villages into different districts, therefore generates variation in this exposure measure in a given village over time. Table A1, appended, shows that this exposure measure is increasing over time among villages that undergo redistricting, while it is stable among villages that do not.

I study the variation in exposure generated by redistricting in an event-study framework. I use the post-decentralization data (2003 and onward in PODES), and I restrict the sample to villages in districts that split between 2003 and 2008, during which the timing of redistricting is exogeneous, as argued in Burgess et al. (2012) and Bazzi and Gudgeon (2017).\textsuperscript{12} The following specification

---

\textsuperscript{11} I take population$_v$ to be the average population of a given village over the length of the panel. Doing so means that the over-time variation in exposure$_v$ arises solely from redistricting (i.e., $D_t(v)$) rather than the relatively small variation in a village’s population from one time period to the next.

\textsuperscript{12} These papers argue that the timing of redistricting is plausibly exogeneous from 2001 to 2003 and from 2007 to 2008 given the national moratoria placed on redistricting from 2004 to 2006 and from 2009 to 2012. I use the
Table 2: Facility construction by exposure to other districts

<table>
<thead>
<tr>
<th></th>
<th>Public Hospitals</th>
<th>Clinics</th>
<th>Subclinics</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>0.000299</td>
<td>-0.00867**</td>
<td>-0.0248***</td>
<td>89,740</td>
</tr>
<tr>
<td></td>
<td>(0.00132)</td>
<td>(0.00350)</td>
<td>(0.00456)</td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>0.00131</td>
<td>0.00138</td>
<td>0.000939</td>
<td>57,644</td>
</tr>
<tr>
<td></td>
<td>(0.000846)</td>
<td>(0.00328)</td>
<td>(0.00677)</td>
<td></td>
</tr>
</tbody>
</table>

Each cell is a single event-study regression with village and year fixed effects. In each, I restrict the sample to villages in districts that undergo redistricting. The first row examines the impact of exposure to other districts on facility construction in the post-decentralization period, where the change in exposure is generated by redistricting. As a placebo experiment, the second row shows the impact of redistricting in the pre-decentralization period. Exposure to other districts is calculated as the proportion of people within 100 km of a village, downweighting linearly by distance, who reside in a district different from that of the village. The outcome is the number of facilities in a village, where the outcome facility is denoted by the column labels. Controls include include population, ruralness, and the number of facilities for non-outcome facility types, including private hospitals. The unit of observation is a village-year, and the number of observations applies for all regressions in the corresponding row. Standard errors are clustered by village. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

compares village outcomes before and after the first split that occurs from 2003 to 2008.

\[ \text{facilities}_{vt} = \beta^e (\text{exposure}_{vt} \times \text{post-split}_{vt}) + x_{vt}\beta + \delta_v + \delta_t + \varepsilon_{vt} \] (1)

The outcome variables are the number of public hospitals, clinics, and subclinics in a village. I also control for a village’s other facilities, including private hospitals, its population and ruralness, and village and time fixed effects.

Table 2 presents the results. The first panel shows that, after redistricting, increases in exposure to other districts results in less clinic and subclinic construction. Figure A1, appended, presents the event study graphically.\(^\text{13}\) Pre-trends are largely absent, although the drop in subclinics seems to start one time period before a split. Pre-trends are somewhat difficult to assess in this context, however, as the short nature of the panel limits the number of pre-event periods in the data.\(^\text{14}\)

To this end, the second panel of table 2 presents a placebo experiment in which I perform the same exercise around redistricting in the pre-decentralization data.\(^\text{15}\) Before decentralization, the central government makes placement decisions regardless of how districts are drawn. Redistricting should therefore have no effect on facility construction, and the results indeed support this hypothesis. The statistical insignificance of the estimates is not only because the sample sizes are smaller

\(^\text{13}\) I plot the exposure-over-time coefficients \(\beta^e_s\) of the following specification, where \(s = -1\) is the base case.

\[ \text{facilities}_{vt} = \sum_s \beta^e_s (\text{exposure}_{vt} \times \text{periods to split} = s)_{vt} + x_{vt}\beta + \delta_v + \delta_t + \varepsilon_{vt} \]

\(^\text{14}\) I cannot use the pre-decentralization data to assess pre-trends because during this period Indonesia was centralized, such that the effect of exposure to other districts is entirely different in these waves from what it is in the post-decentralization data.

\(^\text{15}\) The sample is villages in districts that redistrict between 1990 and 2000. The villages in this sample are different from the villages in the post-decentralization sample, but the results are robust to focusing on the intersection of the two samples, namely the subset of villages in the post-decentralization sample that experience redistricting in the pre-decentralization period.
Table 3: Effect of electoral competitiveness on facility construction

<table>
<thead>
<tr>
<th></th>
<th>Public Hospitals</th>
<th>Clinics</th>
<th>Subclinics</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>0.000235</td>
<td>0.00568***</td>
<td>0.000822</td>
<td>559,746</td>
</tr>
<tr>
<td></td>
<td>(0.000660)</td>
<td>(0.00135)</td>
<td>(0.00229)</td>
<td></td>
</tr>
<tr>
<td>Rural villages</td>
<td>0.00120**</td>
<td>-0.00855***</td>
<td>-0.0229***</td>
<td>279,873</td>
</tr>
<tr>
<td></td>
<td>(0.000517)</td>
<td>(0.00207)</td>
<td>(0.00377)</td>
<td></td>
</tr>
<tr>
<td>Urban villages</td>
<td>0.000958</td>
<td>0.0173***</td>
<td>0.00925***</td>
<td>279,873</td>
</tr>
<tr>
<td></td>
<td>(0.00121)</td>
<td>(0.00180)</td>
<td>(0.00258)</td>
<td></td>
</tr>
</tbody>
</table>

Each cell is a single difference-in-difference regression with village and year fixed effects, and the unit of observation is a village-year. I report the coefficient on the interaction term competitive_v × post-decentralization_t. Competitive_v records whether a village is in a district that had a large number of effective political parties in the 1999 legislative elections. “Large” is defined as being above the median. The first row presents results for the full sample, the second for urban villages, and the third for rural villages. Urban villages are those with above-median population density in 1996. The outcome is the number of facilities in a village, where the outcome facility is denoted by the column labels. Controls include include population, ruralness, and the number of facilities for non-outcome facility types, including private hospitals. The unit of observation is a village-year, and the number of observations applies for all regressions in the corresponding row. Standard errors are clustered by village. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

than those of the first panel – the point estimates themselves are also much closer to zero. Figure A2, appended, presents the analysis graphically and shows the absence of pre-trends.

3.3 Responding to votes

In addition to the shifting of administrative authority, the fall of Suharto brought free elections to Indonesia for the first time since 1955. In this section, I ask if this electoral accountability affected facility placement. I adopt a difference-in-differences strategy to compare villages in electorally competitive and non-competitive districts (treatment and control, respectively), before and after the re-introduction of free elections.

\[
\text{facilities}_{vt} = \beta (\text{competitive}_v \times \text{post-decentralization}_t) + x_{vt}\beta + \delta_v + \delta_t + \epsilon_{vt}
\]  

(2)

Competitive_v records whether a village is in a competitive district. To determine electoral competitiveness, I calculate the number of effective political parties using district-level vote share data from the 1999 legislative elections. The number of effective political parties, proposed by Laakso and Taagepera (1979), is simply the inverse of the Herfindahl index calculated from vote shares. I define competitive districts as those with an above-median number of effective political parties. The outcome variables are the number of public hospitals, clinics, and subclinics in a village, and I control for a village’s other facilities, including private hospitals, its population and ruralness, and village and time fixed effects.

Table 3 presents the results, which show that villages in electorally competitive districts receive more clinics with the introduction of free elections. To test whether electoral accountability impacts facility placement within a district, I split the sample into urban and rural villages – defined as having above- or below-median population density in 1996 – and run the same specification. These
Figure 2: Effect of electoral competitiveness on facility construction, rural vs. urban

Each figure plots the graphical analogue of the difference-in-difference regressions presented in table 3. The unit of observation is a village-year. Standard errors are clustered by village, and the standard error bars are at the 95% significance level.

results, also presented in table 3, show that the full-sample analysis masks important heterogeneity. Within electorally competitive districts, clinics and subclinics shift from rural to urban villages.\textsuperscript{16} Figure 2, which presents the analysis graphically, illustrates the divergence between rural and urban villages.\textsuperscript{17}

4 Model

I present a general model of facility placement that captures the channels of interest, and I describe how it applies to both the centralized and decentralized regimes. A government chooses where to place new facilities over time in order to maximize patient welfare and votes net of costs. I define patient welfare as the consumer surplus resulting from facility usage. I also allow for unobserved factors that rationalize placements unexplained by the welfare, votes, and cost functions. Lastly, I detail the assumptions I use for estimating the model.

\textsuperscript{16} To ensure that the rural-urban results are not driven by differences in ruralness across districts (as opposed to within-district differences in ruralness), I repeat the analysis dropping the 25% most-rural and the 25% most-urban districts. The results are similar.

\textsuperscript{17} I plot the competitive-over-time coefficients $\beta^c_s$ of the following specification, where $s = -1$ is the base case.

\[
\text{facilities}_{vt} = \sum_s \beta^c_s (\text{competitive}_v \times \mathbb{1}(\text{periods to decentralization} = s)_t) + x_{vt}\beta + \delta_v + \delta_t + \varepsilon_{vt}
\]
4.1 The facility allocation problem

A government faces a facility location problem over time. At time $t = 0$, the government chooses a policy $a$ that specifies where to build facilities in every period. Let $n_v(a^t)$ be the number of facilities in village $v$ resulting from placement $a$ as of time $t$, and let $n(a^t)$ be the collection of these terms for all villages. The government chooses a placement policy to maximize its objective function, which is a discounted sum of payoffs over time, including patient welfare, votes, costs, and unobservables.

$$\pi(a) = \sum_{t=1}^{\infty} \beta^{t-1} \left( W(n(a^t)) + V(n(a^t)) - C(n(a^t)) + \xi(n(a^t)) \right)$$  \hfill (3)

The welfare and votes terms sum over villages and represent the consumer surplus and votes response generated by a given placement. I predict these values by estimating demand using observed data on usage and vote choices and obtaining demand parameters $\omega$. Costs, which I calculate for each facility type $f \in \{\text{hosp}, \text{clin}, \text{sub}\}$, also sum over villages. These values are directly observed in the data.

$$W(a^t) = \sum_{v \in V} \text{surplus}_{vt}(a^t, X_{vt}; \omega)$$  \hfill (4)

$$V(a^t) = \sum_{v \in V} \tau^V \left( \text{votes}_{vt}(a^t, X_{vt}; \omega) \right)$$  \hfill (5)

$$C(a^t) = \sum_{v \in V} \sum_{f} \tau^L \left( n_vf(a^t) \cdot \text{landprice}_{vt} \cdot \text{plotsize}_f \right) + \sum_{v \in V} \sum_{f} \tau^P \left( n_vf(a^t) \cdot \text{popden}_{vt} \right)$$  \hfill (6)

I assume homogeneity across older and newer facilities – it is only the total number of facilities at a given point in time that enters the model. To keep notation light, in what follows I write $n(a^t)$ as simply $a^t$ when it appears as a function argument.

4.2 Patient welfare $W(a^t)$

I assume consumer demand for facilities is described by a logit demand system. An individual chooses among the closest facility of each facility type $f \in F$, namely set $\mathcal{F}$ containing the individual’s closest public hospital, closest private hospital, closest clinic, and closest subclinic. The utility of facility type $f$ for individuals living in village $v$ at time $t$ is

$$\text{utility}_{vtf} = x_{vtf}\beta_f + p_{vtf}\alpha + \xi_{tf} + \delta_v + \delta_t + \varepsilon_{vtf},$$  \hfill (7)

where individuals consider facility characteristics $x_{vtf} = [\text{distance}_{vtf}, \text{congestion}_{vtf}]$ and facility prices $p_{vtf}$.\(^{18}\) For added flexibility, I allow the preferences over facility characteristics to vary

\(^{18}\) I construct price indices using household-level data on health spending. Consider data from a given year. For each village, I can run the following regression using the household data

$$\text{spending}_h = \mu_0 (\text{household members who visit no facility})_h + \mu_{F} \sum_{f \in F} (\text{household members who visit } f)_h$$
freely by facility type. Individuals also consider facility quality $\xi_{tf}$, which is unobserved by the econometrician and allowed to vary over time. Finally, $\delta_v$ and $\delta_t$ are village and time fixed effects, and $\varepsilon_{vtf}$ are logit errors. I use the free normalization to set the utility of the outside option to zero.

Consumer surplus is a function of the compensating variation associated with a given facility placement. As in McFadden (1981), I use the estimated price elasticity $\hat{\alpha}$ to calculate the change in prices needed to compensate for some change in facility characteristics.

$$CV_{vt} = \frac{1}{\hat{\alpha}} \left[ \ln \left( \sum_{f \in \mathcal{F}} \exp(V'_{vtf}) \right) - \ln \left( \sum_{f \in \mathcal{F}} \exp(V_{vtf}) \right) \right]$$

(8)

The change in consumer surplus arising from a change in facility placement is therefore

$$\Delta \text{surplus}_{vt}(a^t, X_{vt}; \omega) = \text{population}_{vt} \cdot CV_{vt}.$$  

(9)

As is generally the case with logit demand, this quantity is only identified in changes, not levels.

### 4.3 Votes $V(a^t)$

The incumbent party’s chance of winning an election in a given village is a function of the facilities within the village’s choice set. I consider the change in the distance and congestion variables from the previous time period in order to isolate the incumbent party’s own contribution to facility placement. The binary logit specification is

$$\logit \left[ \Pr(\text{incumbent\_party\_won}_{vt}) \right] = \sum_{f \in \mathcal{F}} \Delta x_{vtf} \beta_f + \delta_v + \delta_t,$$

(10)

which includes a village and time fixed effect as before. The total vote response to a change in facility placement is therefore

$$\Delta \text{votes}_{vt}(a^t, X_{vt}; \omega) = \text{population}_{vt} \cdot \Delta \Pr(\text{incumbent\_party\_won}_{vt}).$$

(11)

Note that I use the free normalization in this discrete-choice framework to set consumer surplus as the numeraire. Preferences over other factors are therefore denominated in dollars of consumer surplus. Parameter $\tau^V$ can therefore be interpreted as the amount of consumer surplus, measured in dollars, that the government is willing to trade off for one additional vote. It quantifies the value of a vote.

---

19 For example, I allow for the distance elasticity for public hospitals to differ flexibly from that of clinics.
20 A more flexible specification would allow for facility-time heterogeneity instead of just facility type-time heterogeneity. The trade-off is that estimation would involve a large matrix of fixed effects, although the grouped fixed effects literature could help in making the implementation more tractable.
Table 4: The model under decentralization vs. centralization

<table>
<thead>
<tr>
<th></th>
<th>Centralized</th>
<th>Decentralized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Before 2001</td>
<td>After 2001</td>
</tr>
<tr>
<td>Decision maker</td>
<td>Central gov.</td>
<td>Individual dist.</td>
</tr>
<tr>
<td>Villages considered</td>
<td>$V_{all}$</td>
<td>$V_{in}$</td>
</tr>
<tr>
<td>Time-variant parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Votes</td>
<td>$\tau^V$</td>
<td>0</td>
</tr>
<tr>
<td>Unobservables</td>
<td>${\lambda_v^{\text{pre}}}$</td>
<td>${\lambda_v^{\text{post}}}$</td>
</tr>
<tr>
<td>Time-invariant parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand</td>
<td>$\omega$</td>
<td></td>
</tr>
<tr>
<td>Land costs</td>
<td>$\tau^L$</td>
<td></td>
</tr>
<tr>
<td>Population density</td>
<td>$\tau^P$</td>
<td></td>
</tr>
</tbody>
</table>

4.4 Costs $C(a^t)$

On the cost side, $\text{landprice}_{vt}$ captures differences in facility costs arising from differences in land values. In conversations with government officials, land prices are mentioned as the key factor in deciding where to construct new facilities. I multiply the unit cost a facility’s plot size to obtain the total land cost in dollar terms. Parameter $\tau^L$ is the rate at which the government trades off consumer surplus with an additional dollar of government expenditures.

To capture additional cost factors, I also model facility costs as a function of $\text{popden}_{vt}$, the population density of the village in which a facility is located. This term captures rural-urban differences, such as in labor costs, while also broadly proxying for a village’s infrastructure, wealth, and education levels. I omit the fixed costs $\tau_{hosp}^1, \tau_{clin}^1, \tau_{sub}^1$ of each facility type without loss of generality because these costs are fixed across locations.

4.5 Unobservables $\xi(a^t)$

Unobservables capture additional choice factors that impact the allocation of facilities to villages but that are not observed in the data, such as disease outbreaks, unobserved costs, and underlying heterogeneity in preferences over villages. I consider a decomposition of village, time, and village-time factors, and I assume that these factors enter the unobservables term linearly.

$$\xi(a^t) = \sum_{v \in V} \sum_f (\lambda_v f + \psi_v f + \varepsilon_{v f}) n_{v f}(a^t)$$ (12)

where, for a facility type $f \in \{\text{hosp}, \text{clin}, \text{sub}\}$, $n_{v f}$ is the number of facilities in village $v$. The linear functional form implies that unobservables $\xi$ increase by $(\lambda_v f + \psi_v f + \varepsilon_{v f})$ for each facility of type $f$ built in village $v$. 

14
4.6 Decentralization vs. centralization

Table 4 summarizes how the model specializes to the decentralized and centralized allocation problems. There are several differences. Under decentralization, the agent is a single decentralized district government. It chooses placements over villages $V_{\text{in}}$ within the district, and it only considers the welfare gains for these villages. Under centralization, the agent is the central government. It chooses over the full set of villages $V_{\text{all}}$ and internalizes all spillover effects. Votes are part of the government objective function under decentralization, but not under centralization. Finally, the set of unobserved village preferences $\{\lambda_v\}$ are allowed to differ between decentralization and centralization, but the remaining parameters of the model are assumed to be time-invariant.

The counterfactual exercises involve two key changes. As an example, consider the counterfactual of continued centralization absent the decentralization reform. First, I capture the spillover channel by changing the scope of villages considered from $V_{\text{in}}$ to $V_{\text{all}}$, which forces the government to internalize the spatial externalities of its choices. Second, I simulate the elimination of electoral accountability by setting the government’s weight $\tau^V$ on votes to zero, shifting all consideration onto patient welfare, costs, and unobservables.

4.7 Additional assumptions for estimation

**Parametric assumptions**

First, I assume that village-time shocks $\varepsilon_{vtf}$ are mean-zero for each facility type $f$

$$E(\varepsilon_{vtf}) = 0 \quad \forall f, \quad 22$$

but I make no such assumptions about the distributions of time-invariant village preferences $\lambda_{vf}$ or village-invariant time effects $\psi_{tf}$.

Second, I allow for welfare and votes to be imperfectly predicted, as the estimated demand parameters $\hat{\omega}$ may differ from the true $\omega$. The errors are

$$\epsilon^W_{vt} = \text{surplus}_{vt}(a^t, X_{vt}; \omega) - \text{surplus}_{vt}(a^t, X_{vt}; \hat{\omega}),$$
$$\epsilon^V_{vt} = \text{votes}_{vt}(a^t, X_{vt}; \omega) - \text{votes}_{vt}(a^t, X_{vt}; \hat{\omega}). \quad 14$$

I assume that these sources of prediction error are orthogonal to the government’s information set $\mathcal{J}_0$ at the time of making its placement decision.

$$E(\epsilon^W_{vt} | \mathcal{J}_0) = E(\epsilon^V_{vt} | \mathcal{J}_0) = 0 \quad 15$$

Third, I assume that the parameters $\tau^P_f$ and $\lambda_{vf}$ for each facility type $f$ are scalar transformations of those of the other facility types.

$$\begin{align*}
\tau^P_{\text{hosp}} &= \tau^P, & \lambda_{v,\text{hosp}} &= \lambda_v \\
\tau^P_{\text{clin}} &= \gamma_{\text{clin}} \cdot \tau^P_{\text{hosp}}, & \lambda_{v,\text{clin}} &= \gamma_{\text{clin}} \cdot \lambda_{v,\text{hosp}} \\
\tau^P_{\text{sub}} &= \gamma_{\text{sub}} \cdot \tau^P_{\text{hosp}}, & \lambda_{v,\text{sub}} &= \gamma_{\text{sub}} \cdot \lambda_{v,\text{hosp}}
\end{align*} \quad 16$$

21 If facilities built in rural areas tend to be larger, then I can adjust for this difference by distinguishing landsize$^{\text{rural}}$ and landsize$^{\text{urban}}$ within each facility type.

22 Note that the expectation is from the perspective of the econometrician. The government does not form an expectation because it knows the value of $\varepsilon_{vtf}$ at the time of its decision.
Doing so avoids the need to estimate the full set of parameters separately for each facility type. In equation 6, I do not include a facility-type-specific parameter for land costs because differences across facility types are directly captured by differences in plot sizes. For simplicity, I set the parameters $\gamma_{\text{clin}}$ and $\gamma_{\text{sub}}$ to be equal to the plot size ratios.

$$
\gamma_{\text{clin}} = \frac{\text{plotsize}_{\text{clin}}}{\text{plotsize}_{\text{hosp}}}, \quad \gamma_{\text{sub}} = \frac{\text{plotsize}_{\text{sub}}}{\text{plotsize}_{\text{hosp}}}
$$

A more flexible approach would estimate these parameters from the data, but by calibrating them I can solve the model as a linear program.

**Behavioral assumptions**

First, I make the behavioral assumption that the government spends its facility budget within each period. The choice space of policies is therefore set by the number of facilities observed in the data in each period.

$$
A = \left\{ a \left| \sum_{v \in V} n_{vf}(a_t) = \sum_{v \in V} n_{vf}({\ast}a_t), \forall f, t \right. \right\}.
$$

This assumption is reasonable if the government is credit-constrained, such that it cannot spend more than its budget, and if the government fears budget cuts in response to underspending, such that it does not spend less than its budget.

Second, the government has perfect foresight over the patient welfare $W(\cdot)$, votes $V(\cdot)$, costs $C(\cdot)$, and unobservables $\xi(\cdot)$ associated with any given policy. There is no expectational error, so there is no learning over time (because there is nothing to learn).

Third, the government faces exogenously given facility budgets in each period. If budget allocations tomorrow depend on placement choices today, then the payoff associated with a given placement must account for these future budget effects above and beyond the welfare, vote, cost, and unobservables functions. New facilities in Indonesia are allocated using population-based rules, and so this assumption amounts to an assumption of exogenous population growth.\(^{23, 24}\)

Fourth, I assume away strategic responses to the government’s placement decision. In reality, for a centralized national government, the private sector may respond to the placement of public facilities. Perhaps assuming away private-sector responses is without great consequence, as private hospitals serve a different market than public hospitals do (i.e., the very wealthy). For a decentralized district government, the responses of other districts may also be important.

**5 Estimation**

Estimation proceeds in two stages. In the first stage of estimation, I specify spatial models of demand for facility usage and votes. Using panel variation generated under the actual policy $\hat{a}$, I

\(^{23}\) A decentralized district government receives an allocation from the central government based on its population.

\(^{24}\) A centralized national government also has its budget determined by population – the Ministry of Health receives its share of the national budget based on the total population across districts.

Note that the government does not necessarily need to know its budget in every future period, as its chosen policy can specify a contingent plan over all possible budgets.
Indonesian Hospitals

Allan Hsiao (MIT)

estimate demand parameters $\omega$. I use the estimated demand systems to predict consumer welfare and votes under any given counterfactual policy.\(^{25}\)

The second stage of estimation takes demand parameters $\omega$ as given. In a revealed-preference approach, I estimate government preference parameters $\tau$ by finding the values that rationalize the observed placement choices. I develop an estimator that is analogous to matched difference-in-differences in order to account for unobserved choice factors, which would otherwise give rise to selection bias. Compared to applications related to firm behavior, which may be well approximated by profit-maximization, unobserved choice factors are particularly salient in this setting in which the welfare, votes, and cost functions only offer an incomplete characterization of the government’s objective function. I also show how the estimation procedure accommodates the dynamics of the model, as well as error in the predicted values of patient welfare and votes.

5.1 Demand

Welfare

In order to predict the usage response to any given placement of facilities, I estimate a system of demand for facility visits. To do so, I combine the village-level panel data on health infrastructure from PODES with individual-level data on facility visits and health spending from SUSENAS. I focus on the SUSENAS data from 1993 to 2002, which distinguish between clinic and subclinic visits. The primary constraint is that the visits data distinguish among facility types, but not among specific facilities within a type.\(^{26}\) In estimation, I restrict the sample to the set of individuals who report at least one health concern, such that the outside option consists of sick individuals who do not seek care in a hospital, clinic, or subclinic.

Distance is Euclidean distance between location points.\(^{27}\) Location points are restricted by the coarseness of the data, which record individual and facility locations at the village level. I calculate distances from village centroid to village centroid. Congestion is a function of how many individuals use a given facility. A simple way of proxying for congestion is to count to the number of individuals for whom a new facility is their closest facility. In the language of Donaldson and Hornbeck (2016), distance_{vtf} captures the direct effects of facility construction, while congestion_{vtf} captures the indirect effects.\(^{28}\)

For estimation, the utility described by equation 7 yields the specification

$$\ln(s_{vtf}) - \ln(s_{0vt}) = x_{vtf}\beta_F + p_{vtf}\alpha + \xi_{tF} + \delta_v + \delta_t + \epsilon_{vtf};$$  \hspace{1cm} (19)

where $s_{vtf}$ is the share of individuals who visit facility $f$ in village $v$ at time $t$.\(^{29}\) I estimate this specification on village-level data by OLS. The logit inversion used to generate this linear expression

\(^{25}\) The welfare function, which depends on both new placement and the existing stock of facilities, provides the structure necessary for making intertemporal comparisons given that the government is walking down the demand curve over time. Its estimated curvature matters: the benefit of an additional facility in the future, when the stock of facilities is high, will differ from the benefit of an additional facility today, when the stock of facilities is low, and it is the curvature of the welfare function that determines the magnitude of this difference.

\(^{26}\) That is, I observe whether an individual visits a public hospital as opposed to a clinic, but I do not observe which particular public hospital he or she visited.

\(^{27}\) I can improve this measure by using travel times, which account for terrain, instead of Euclidean distance.

\(^{28}\) A new facility directly increases usage in nearby villages by decreasing the closest-facility distance for these villages. But it also indirectly increases usage in faraway villages as patients move toward the new facility and thereby de congest other facilities. Further details are provided in the appendix.

\(^{29}\) I calculate village shares from individual-level data on the number of visits to each facility type in the last month.
is infeasible when empirical shares are equal to zero or one, so in practice I smooth the empirical choice probabilities spatially as in Scott (2013). I also consider a specification that includes a population-density interaction term \((x_{vtf} \cdot \text{popden}_{vt})\beta_dF\), such that preferences are allowed to vary between rural and urban villages. In calculating the compensating variation described by equation 8, \(V_{vtf}'\) is simply the fitted values of specification 19, and \(V_{vtf}'\) can be calculated as

\[
V_{vtf}' = V_{vtf} + (x_{vtf}' - x_{vtf}) \beta_F, \tag{20}
\]

assuming that a change in facility placement does not impact pricing.

**Votes**

I perform a similar exercise with voting behavior as the outcome variable. In this case, I use village vote-rank data from PODES to determine whether the incumbent party won in a village in the 1999 and 2004 legislative elections. The incumbent party was Golkar in 1999, and PDI-P in 2004. The explanatory variables include the full set of facility distances and congestions across all facility types. As before, I include village and time fixed effects, and I also consider a specification that includes the interacted term \((\Delta x_{vtf} \cdot \text{popden}_{vt})\beta_dF\), which allows preferences to vary between rural and urban villages.

**5.2 Moment inequalities**

Having estimated demand, I estimate government preference parameters \(\tau\) with a revealed-preference approach as described in Pakes (2010) and Pakes et al. (2015). Let \(\hat{\alpha}\) denote the actual placement policy observed in the data. Taking the true demand parameters \(\omega^0\) as known, for all policies \(a \in \mathcal{A}\), if \(\tau = \tau^0\) then

\[
\pi(\hat{\alpha}; \tau, \omega^0) \geq \pi(a; \tau, \omega^0) \tag{21}
\]

by definition of chosen policy \(\hat{\alpha}\), where \(\tau^0\) is the government’s true preference parameters. That is, at the true parameter values, the chosen option at least weakly dominates all other options. I

---

For each sick individual, which I define as those reporting at least one health concern, I classify the individual as having visited either a private hospital, public hospital, clinic, or subclinic. I do not distinguish between a single visit and multiple visits to a given facility type. For individuals who visited multiple facility types, I code them based on the most expensive facility type they visited (in the order private hospital, public hospital, clinic, and subclinic).

30 The merged dataset includes yearly usage data from 1993 to 2002 from SUSENAS and data on infrastructure in 1993, 1996, and 2000 from PODES. First, I calculate average village choice probabilities in the SUSENAS data in the periods 1993-1995, 1996-1999, and 2000-2002. Second, for each period, I smooth the choice probabilities within districts (as designated by the 1996 district boundaries). For each village, the smoothed choice probabilities are a weighted average of choice probabilities within the district. The weights are given by \((1 + \text{distance}(v,v'))^{-2}\), such that they place the highest weight on the village’s own data and the least weight on faraway villages’ data.

31 The interacted specification also allows the facility fixed effects \(\xi_F\) to vary freely between high-density and low-density villages, where the cutoff is the mean density. That is, I include \(\xi_F^d = \xi_F \cdot I\{\text{popden}_{vt} > \text{popden}\}\).

32 To obtain values for \(V_{vtf}\), I extrapolate from the usage sample, which covers a subset of villages from 1993 to 2002, to other villages and other years. As such, calculating the fitted values requires some imputation of village fixed effects \(\delta_v\) and year fixed effects \(\delta_t\). For out-of-sample villages in districts with at least 10 in-sample villages, I impute the village fixed effect as the distance-weighted average of the same-district, in-sample fixed effects. I use the same weighting scheme as when I smooth the choice probabilities, namely \((1 + \text{distance}(v,v'))^{-2}\). For the small proportion – about 0.5% – of out-of-sample villages in districts without at least 10 in-sample villages, I calculate the distance-weighted average of all in-sample fixed effects. For year fixed effects, I use the year 1993 fixed effect for pre-1993 years and the year 2000 fixed effect for post-2000 years.
Indonesian Hospitals

Allan Hsiao (MIT)

operationalize this insight by choosing a set of alternatives \(a\), constructing the associated revealed-preference inequality for each alternative, and ruling out candidate values of parameters \(\tau\) that violate one or more of these inequalities. This approach achieves dimension reduction by evaluating only a subset of the possible alternatives, although more alternatives does mean more inequalities and therefore a smaller identified set.\(^{33}\)

Expanding inequality 21 and rearranging to show the difference in payoffs between actual policy \(\hat{a}\) and alternative policy \(a\), the revealed-preference inequality becomes

\[
\sum_{t=1}^{\infty} \beta^{t-1} \left( \Delta WVC(\hat{a}^t, a^t; \tau, \omega^0) + \Delta \xi(\hat{a}^t, a^t; \lambda, \psi, \varepsilon) \right) \geq 0,
\]

where I define

\[
\Delta WVC(\hat{a}^t, a^t; \tau, \omega^0) \equiv \Delta W(\hat{a}^t, a^t; \omega^0) + \Delta V(\hat{a}^t, a^t; \tau, \omega^0) - \Delta C(\hat{a}^t, a^t; \tau)
\]

and

\[
\Delta W(\hat{a}^t, a^t; \omega^0) \equiv \sum_{v \in V} W(\hat{a}^t, X_{vt}; \omega^0) - \sum_{v \in V} W(a^t, X_{vt}; \omega^0),
\]

with \(\Delta V\), \(\Delta C\), and \(\Delta \xi\) defined similarly. The challenge is in constructing a sample analogue to inequality 22 given (1) the infinite-horizon summation in the objective function, (2) the lack of information on unobservables \(\lambda_v\), \(\psi_t\), and \(\varepsilon_{vt}\), and (3) prediction error generated by error in estimation of demand parameters \(\omega\). To address these issues, I draw on methods described in moment-inequalities applications Holmes (2011), Ho and Pakes (2014), and Ishii (2007).

For simplicity, in discussing identification I assume only one facility type and suppress all \(f\) subscripts. Estimation uses information from all facilities types – hospitals, clinics, and subclinics – jointly, and the extension is straightforward.

5.3 Infinite-horizon summation (dynamics)

Generating a sample analogue to inequality 22 is difficult because the infinite-horizon summation requires constructing sample analogues for per-period payoffs over many, many periods. This summation captures the dynamic effects of a given policy: a new facility in period \(t\) impacts patient welfare, votes, and costs not only in period \(t\) but also in all future periods.

As in Holmes (2011), to address this issue I select alternatives by “pairwise resequencing,” such that alternatives swap the construction order of two facilities in the actual policy. Dynamic effects are eliminated because the alternative and actual policies are identical outside of the swap periods. For example, for actual policy \(\hat{a} = (v_1, \{v_2, v_3\}, v_4, \ldots)\), a pairwise resequenced alternative is \(a = (v_2, \{v_1, v_3\}, v_4, \ldots)\). After the second period, these policies result in the same number of facilities in every village. Formally, the set \(\mathcal{S} \subset \mathcal{A}\) of “swapped” alternatives to actual policy \(\hat{a}\) are such that

\[
n_v(\hat{a}^t) - n_v(a^t) = \begin{cases} 
0 & \text{for } t \notin s(a), \ v \in \mathcal{V} \\
-1 & \text{for } t \in s(a), \ v = w_1(a) \\
1 & \text{for } t \in s(a), \ v = w_2(a) \\
0 & \text{for } t \in s(a), \ v \in \mathcal{V} \setminus \{w_1(a), w_2(a)\}
\end{cases}
\]

\(^{33}\) By contrast, a nested fixed-point approach picks a candidate value for parameter \(\tau\) and evaluates every policy to find the optimal policy given \(\tau\). It then chooses the value of \(\tau\) that produces the predicted optimal policy most similar to the observed policy.
where \( s(a) = \{s_1(a), s_1(a) + 1, \ldots, s_2(a) - 1\} \), \( s_1(a) \) is the earlier period involved in the swap, \( s_2(a) \) is the later period, \( w_1(a) \) is the village that receives its facility earlier in the swap, and \( w_2(a) \) is the village that receives it later.\(^{34}\)

For swapped alternatives \( a \in S \), applying the first line of condition 23 to inequality 22 gives

\[
\sum_{t = s_1(a)}^{s_2(a) - 1} \beta^{t-1} \left( \Delta WVC(\hat{a}^t, a^t; \tau, \omega) + \Delta \xi(\hat{a}^t, a^t; \lambda, \psi, \varepsilon) \right) \geq 0, \quad (24)
\]

since \( n_v(\hat{a}^t) = n_v(a^t) \) for all villages \( v \) outside of the swap-relevant periods \( s(a) \). That is, the per-period payoffs of actual policy \( \hat{a} \) and swapped alternative \( a \) are identical in the periods before and after the swap, thereby eliminating the need to sum over an infinite horizon.\(^{35}\)

### 5.4 Unobservables (selection bias)

Constructing a sample analogue to inequality 24 is still difficult because \( \lambda_v \), \( \psi_t \), and \( \varepsilon_{vt} \) are unobserved. Ignoring these terms leads to selection bias – given that policy \( \hat{a} \) was actually chosen, the unobserved payoff of \( \hat{a} \) relative to unchosen policy \( a \) is unlikely to be zero in expectation. To proceed, I apply the functional-form assumption 12 on the unobservables to get

\[
\Delta \xi(\hat{a}^t, a^t; \lambda, \psi, \varepsilon) = \sum_{v \in V} \left( (\lambda_v + \varepsilon_{vt})(n_v(\hat{a}^t) - n_v(a^t)) \right) + \psi_t \left( \sum_{v \in V} n_v(\hat{a}^t) - \sum_{v \in V} n_v(a^t) \right).
\]

Substituting this expression, applying the rest of condition 23, and applying the budget-spending assumption of equation 18, inequality 24 simplifies to

\[
\sum_{t = s_1(a)}^{s_2(a) - 1} \beta^{t-1} \left( \Delta WVC(\hat{a}^t, a^t; \tau, \omega) + \lambda_{w_2(a)} - \lambda_{w_1(a)} + \varepsilon_{w_2(a),t} - \varepsilon_{w_1(a),t} \right) \geq 0. \quad (25)
\]

As in Ishii (2007), the assumption of mean-zero village-time preference shocks, \( \mathbb{E}(\varepsilon_{vt}) = 0 \), means the unobserved \( \varepsilon_{vt} \) terms can be averaged out. The selection issue is that these shocks, which are unconditionally mean-zero, may not be mean-zero after conditioning on the chosen policy \( \hat{a} \). In this case, however, the assumed linearity of the unobservable component terms delivers an inequality that is additive in the \( \lambda_v \) and \( \varepsilon_{vt} \) terms no matter the chosen policy \( \hat{a} \). That is, regardless of where \( \hat{a} \) places facilities, the alternative policy involving swap villages \( w_1(a) \) and \( w_2(a) \) yields an inequality containing the same \( \varepsilon_{w_2(a),t} \) and \( \varepsilon_{w_1(a),t} \) terms.\(^{36}\) Thus, the inequality need not condition on chosen policy \( \hat{a} \), and the unconditional average is sufficient for addressing the \( \varepsilon_{vt} \) terms.

---

\(^{34}\) If facilities are constructed at the beginning of the period, then in period \( s_2 \) both of the facilities involved in the swap have been built. As such, the equality holds in period \( s_2 \). If facilities are instead constructed at the end of the period, then the equality holds at all \( t \not\in \{s_1(a) - 1, s_1(a), \ldots, s_2(a)\} \).

\(^{35}\) If the decision maker chooses over a horizon that is limited by some lookahead depth, then swapped alternatives must swap within the lookahead depth to be valid. Even without the infinite-horizon issue, reducing the number of per-period payoffs in the inequality by choosing swapped alternatives is useful. The following section shows how swapped alternatives are helpful in addressing the unobservables issue as well.

\(^{36}\) Linearity matters a great deal here. Applying the unconditional average of \( \varepsilon_{vt} \), requires the difference between actual policy \( \hat{a} \) and alternative policy \( a \) to contain the same \( \varepsilon_{w_2(a),t} \) and \( \varepsilon_{w_1(a),t} \) terms whether villages \( w_1(a) \) and \( w_2(a) \) contain zero, one, or one thousand facilities. But similar to difference-in-differences, the hope with this identification strategy is that the \( \varepsilon_{vt} \) terms are small after conditioning on the village and time terms \( \lambda_v \) and \( \psi_t \).
The strategy involves higher-order swaps. For example, consider actual policy \(\hat{a} = (v_1, v_2, v_1, \ldots)\). A valid pair would be \(a = (v_2, v_1, v_1, \ldots)\) and \(b = (v_1, v_1, v_2, \ldots)\). Swaps need not be in adjacent or equally spaced periods – for example, a valid double swap can be formed from actual policy \(\hat{a} = (v_1, v_2, \emptyset, v_1, \ldots)\).

For pair \((a, b) \in \mathcal{D}\), I form inequality 25 for each alternative in the pair. Normalizing the discount factors and summing these two inequalities gives

\[
\frac{1}{R_a} \sum_{t=s_1(a)}^{s_2(a)-1} \beta^{t-1} \left( \Delta WVC(\hat{a}, a^t; \tau, \omega^0) + \varepsilon_{w_2(a),t} - \varepsilon_{w_1(a),t} \right) + \frac{1}{R_b} \sum_{t=s_1(b)}^{s_2(b)-1} \beta^{t-1} \left( \Delta WVC(\hat{b}, b^t; \tau, \omega^0) + \varepsilon_{w_2(b),t} - \varepsilon_{w_1(b),t} \right) \geq 0,
\]

where the \(\lambda_v\) terms are eliminated because, by condition 26,

\[
\frac{1}{R_a} \sum_{t=s_1(a)}^{s_2(a)-1} \beta^{t-1} \left( \lambda_{w_2(a)} - \lambda_{w_1(a)} \right) + \frac{1}{R_b} \sum_{t=s_1(b)}^{s_2(b)-1} \beta^{t-1} \left( \lambda_{w_2(b)} - \lambda_{w_1(b)} \right) = 0,
\]

given the normalizations

\[
R_a = \sum_{t=s_1(a)}^{s_2(a)-1} \beta^{t-1}, \quad R_b = \sum_{t=s_1(b)}^{s_2(b)-1} \beta^{t-1}.
\]

In estimation, I set discount factor \(\beta\) equal to 0.95.

The linear functional-form assumption is helpful in achieving the simple form of equation 28, which must hold in order for a candidate double swap to be valid.\(^{38,39}\) It therefore helps in expanding the possible set of double swaps to be used for estimation. The assumption is not innocuous, but in practice the number of facilities in any given village is small, such that linearity is never used to make highly non-local comparisons. Furthermore, more complex swap structures can accommodate relaxations of the linear function-form assumption if the form of the non-linearity is taken as known.\(^{40}\)

\(^{37}\) Assuming that the \(\lambda_v\) are mean-zero would imply that, on average, the patient welfare, votes, and cost functions specified in the model correctly describe the government’s objective function. Such an assumption seems strong.

\(^{38}\) The linear functional form simplifies the calculation because the \(\lambda_v\) terms are constant over time and can therefore be factored out of the discounted sum. Under a known form of non-linearity, the degree of curvature simply affects the normalization factors and can be readily accommodated.

\(^{39}\) Suppose the functional form of \(\xi_v(n_v(a'))\) is not linear in \(\lambda_v\), but rather given by \(\xi_v(n_v(a')) = f(\lambda_v, n_v(a')) \cdot n_v(a')\). Then equation 28 does not necessarily hold because \(f(\lambda_v, n_v(a'))\) is not constant over time. The \(n_v(a')\) surely changes at least once because the double swap is formed over a policy sequence that builds twice in a single village, such that \(n_v(a')\) is smaller in the first swap than it is in the second swap.

\(^{40}\) The strategy involves higher-order swaps. For example, consider actual policy \(\hat{a} = (1, \{2, 3\}, 1, \ldots)\). If \(\xi_v\) is concave in \(n_v\), then the double swap based on \((1,2,1)\) will be such that \(\lambda_{w_2(a)} > \lambda_{w_2(b)}\). But the same will be true for the double swap based on \((1,3,1)\), such that the these terms can be canceled out.
Like matched difference-in-differences, this relies on the similarity of unobservables within a matched pair. Here, the elimination of village unobservables $\lambda_v$ and time unobservables $\psi_t$ allows unbiased estimation of preference parameters $\tau$ given the assumption that village-time unobservables $\epsilon_{vt}$ are mean zero. That is, just as with parallel trends in difference-in-differences, the hope is that removing the village and time unobservables is sufficient for eliminating the selection bias. A violation of these parallel trends would occur if, for example, clinic placement decisions are made jointly with school placement decisions. In this case, the building of a clinic (which yields a swap inequality) would be associated with a non-zero value of $\epsilon_{vt}$, thereby violating the assumption that $E[\epsilon_{vt}] = 0$. The PODES dataset includes data on schools and other public goods, allowing me to address this concern by checking empirically for these joint placements and controlling for them where necessary.

Furthermore, as in difference-in-differences, this approach requires panel variation in the factors of interest. Cost factors in particular scale linearly in the number of facilities in a given village, and thus cost factors that vary only at the village level will be differenced out alongside the $\lambda_v$ terms.

I therefore focus on land costs and population density, which vary over time.

5.5 Prediction error and village-time shocks

For each identified double swap, I form a sample analogue to revealed-preference inequality 27 using predicted values for patient welfare and votes and observed data for costs.

\[
\frac{1}{R_a} \sum_{t=s_1(a)}^{s_2(a)-1} \beta^{t-1} \left( \Delta \tilde{W} VC(\hat{a}^t, a^t; \tau, \tilde{\omega}) \right) + \frac{1}{R_b} \sum_{t=s_1(b)}^{s_2(b)-1} \beta^{t-1} \left( \Delta \tilde{W} VC(\hat{a}^t, b^t; \tau, \tilde{\omega}) \right) \geq 0 ,
\]

(29)

where patient welfare, votes, and costs are

\[
\Delta \tilde{W} VC(\hat{a}^t, a^t; \tau, \tilde{\omega}) \equiv \Delta \hat{W}(\hat{a}^t, a^t; \tau, \tilde{\omega}) + \Delta \hat{V}(\hat{a}^t, a^t; \tau, \tilde{\omega}) - \Delta C(\hat{a}^t, a^t; \tau).
\]

The sample inequality fails to match the theory in two ways. First, sample inequality 29 may have error in its predicted values of patient welfare and votes as described by equations 14. That is, I obtain $\Delta \hat{W}$ and $\Delta \hat{V}$, not $\Delta W$ and $\Delta V$. Second, the sample inequality does not contain village-time shocks $\epsilon_{vt}$ as these terms are unobserved.

An aggregation step addresses these issues. Substituting prediction-error equations 14 into sample inequality 29, the inequality becomes

\[
\frac{1}{R_a} \sum_{t=s_1(a)}^{s_2(a)-1} \beta^{t-1} \left( \Delta W VC(\hat{a}^t, a^t; \tau, \omega^0) - \zeta_{at} \right) + \frac{1}{R_b} \sum_{t=s_1(b)}^{s_2(b)-1} \beta^{t-1} \left( \Delta W VC(\hat{a}^t, b^t; \tau, \omega^0) - \zeta_{bt} \right) \geq 0 ,
\]

(30)

with disturbances

\[
\zeta_{at} \equiv \sum_{v \in V} \left( \epsilon^W_{vt}(\hat{a}^t) - \epsilon^W_{vt}(a^t) \right) + \tau \sum_{v \in V} \left( \epsilon^V_{vt}(\hat{a}^t) - \epsilon^V_{vt}(a^t) \right),
\]

\[
\zeta_{bt} \equiv \sum_{v \in V} \left( \epsilon^W_{vt}(\hat{a}^t) - \epsilon^W_{vt}(b^t) \right) + \tau \sum_{v \in V} \left( \epsilon^V_{vt}(\hat{a}^t) - \epsilon^V_{vt}(b^t) \right).
\]

\[41\]

The welfare terms survives the differencing both because welfare in village $v$ does not scale linearly in the number of facilities in $v$ and because it depends as well on the number of facilities in villages beyond $v$. 

22
which are mean-zero given assumptions 15, 42 These disturbances therefore drop out in aggregating over a group of double swaps, as long as the groupings are not selected on either $\Delta \hat{W}$ or $\Delta \hat{V}$, which are themselves functions of prediction errors $\hat{e}_{vit}$ and $\hat{e}_{vt}$. Similarly, the village-time shocks $\varepsilon_{vt}$ in revealed-preference inequality 27 aggregate to zero given assumption 13. Thus, the aggregated version of the sample inequality matches the aggregated version of the revealed-preference inequality delivered by the model.

5.6 Estimating $\tau^V$, $\tau^L$, and $\tau^P$

The estimation procedure identifies valid swaps, forms the corresponding aggregate sample inequalities, and chooses parameters that minimize violations of these inequalities. First, I identify $K$ valid double swaps and form an inequality value for each. I constrain double swaps to be either within the pre-decentralization data or within the post-decentralization in order to allow for the possibility that $\lambda_{V}^{pre} \neq \lambda_{V}^{post}$. 43 For a given double-swap pair $(a_k, b_k) \in D$, I define inequality value $I_k(\tau)$ as the left-hand side of inequality 29, which I rewrite as

$$I_k(\tau) = D \left( \Delta \text{surplus}(a_k^t, b_k^t) \right) + V^D \left( \Delta \text{votes}(a_k^t, b_k^t) \right) - \tau^L \left( \Delta \text{landcost}(a_k^t, b_k^t) \right) - \tau^P \left( \Delta \text{popden}(a_k^t, b_k^t) \right),$$

(31)

where

$$D \left( \Delta f(a_k^t, b_k^t) \right) = \frac{1}{R_{a_k}} \sum_{t=s_1(a_k)}^{s_2(a_k)-1} \beta^{t-1} \left( \Delta f(\hat{a}_k^t, a_k^t) \right) + \frac{1}{R_{b_k}} \sum_{t=s_1(b_k)}^{s_2(b_k)-1} \beta^{t-1} \left( \Delta f(\hat{a}_k^t, b_k^t) \right).$$

Besides government preference parameters $\tau$, this expression contains only predicted and observed values. The computationally intensive part of the estimation procedure is in calculating these inequality values for all $K$ double swaps. But the inequality value of each double swap is entirely independent of the other double swaps, so the process is readily parallelizable.

Second, I form aggregated moments by averaging these values in groups $G_{\text{land}+}^{\text{pre}}, G_{\text{land}+}^{\text{post}}, G_{\text{land}+}^{\text{pre}}, G_{\text{pop}+}^{\text{pre}}, G_{\text{pop}+}^{\text{pre}}, G_{\text{pop}+}^{\text{post}},$ and $G_{\text{pop}+}^{\text{post}}$.

$$M_g(\tau) = \sum_{k=1}^{K} \left( 1(k \in G_g) \cdot I_k(\tau) \right) / \sum_{k=1}^{K} \left( 1(k \in G_g) \right)$$

(32)

In order to identify upper and lower bounds for each parameter of interest, I group double-swap values based on whether their land cost term, $D(\Delta \text{landcost}(a_k^t, b_k^t))$, increases or decreases, and whether their population density term, $D(\Delta \text{popden}(a_k^t, b_k^t))$, increases or decreases. 44 I do not select on the patient surplus or votes terms because doing so involves selection on prediction errors.

42 I write the prediction errors as a function of the policies to make clear that, for example, $\hat{e}_{vit}^W(\hat{a}^t)$ and $\hat{e}_{vt}^W(a^t)$ do not necessarily cancel. By equations 14, the full set of arguments gives $\hat{e}_{vt}^W(\hat{a}^t, X_{vt}; \omega)$ and $\hat{e}_{vt}^W(a^t, X_{vt}; \omega)$.

43 If $\lambda_{V}^{pre} \neq \lambda_{V}^{post}$, then forming double swaps across the decentralization reform in 2001 will not result in the $\lambda_{V}$ terms canceling out.

44 Grouping inequalities with different types of identifying variation may eliminate it. For example, combining an inequality with an alternative that increases hospital distance and one with an alternative that decreases it will eliminate the identifying variation in hospital distance.
In revealed-preference inequality 27, the village-time shocks ε_{vt} and ε_{vt} will average out because I do not select on these values (in fact I cannot – they are unobserved) in the aggregation step.

Revealed preference implies that π(\tilde{a}) ≥ π(a) but is silent on the magnitude of the difference. Therefore, the criterion function penalizes violations of the revealed-preference inequalities, but it does not reward the compliance. The one-sided nature of this function is what necessitates averaging out mean-zero errors in the inequalities before forming the moments (otherwise each individual inequality could be taken as a single moment, and the method of moments would do the averaging).

To solve as an LP, I reformulate the problem as max_{\tau} \left\{ \frac{1}{G} \sum_{g=1}^{G} Z_g(\tau) \right\} s.t. Z_g(\tau) \leq M_g(\tau), Z_g(\tau) \leq 0 \forall g.

An alternative approach is to estimate village preferences \lambda_v alongside cost parameters \tau, in which case I can avoid the double-swap approach to differencing out the \lambda_v terms. This approach is simpler, but it comes with a cost. There is much less variation to identify \lambda_v, which differs from district to district, than there is to identify \tau, which is assumed to be common across districts. Obtaining unbiased estimates of \tau depends on accounting precisely for \lambda_v, either by differencing it out or by estimating it, such that noisy estimates of \lambda_v may result in biased estimates of \tau. By differencing out the \lambda_v terms, I obtain more reliable estimates of \tau, which in turn help in estimating \lambda_v.

5.7 Estimating \{\lambda_v^{\text{post}}\}

I return to the single-swap revealed-preference inequality given by inequality 25 to estimate the remaining \{\lambda_v^{\text{post}}\} terms. To do so, I use the single swaps identified in the post-decentralization period. I do not need to estimate the \{\lambda_v^{\text{pre}}\} terms because the counterfactuals of interest concern only placement in the post-decentralization period. Furthermore, I estimate \lambda_v^{\text{post}} only for villages that receive some facility by 2014. I cannot do more, as without any facilities there are no possible swaps with which to generate moment inequalities. In computing counterfactuals, I sidestep this issue by omitting these villages from the choice set of the government.

An alternative approach is to estimate village preferences \lambda_v alongside cost parameters \tau, in which case I can avoid the double-swap approach to differencing out the \lambda_v terms. This approach is simpler, but it comes with a cost. There is much less variation to identify \lambda_v, which differs from district to district, than there is to identify \tau, which is assumed to be common across districts. Obtaining unbiased estimates of \tau depends on accounting precisely for \lambda_v, either by differencing it out or by estimating it, such that noisy estimates of \lambda_v may result in biased estimates of \tau. By differencing out the \lambda_v terms, I obtain more reliable estimates of \tau, which in turn help in estimating \lambda_v.

45 In revealed-preference inequality 27, the village-time shocks ε_{vt} will average out because I do not select on these values (in fact I cannot – they are unobserved) in the aggregation step.

46 Revealed preference implies that π(\tilde{a}) ≥ π(a) but is silent on the magnitude of the difference. Therefore, the criterion function penalizes violations of the revealed-preference inequalities, but it does not reward the compliance. The one-sided nature of this function is what necessitates averaging out mean-zero errors in the inequalities before forming the moments (otherwise each individual inequality could be taken as a single moment, and the method of moments would do the averaging).

47 To solve as an LP, I reformulate the problem as max_{\tau} \left\{ \frac{1}{G} \sum_{g=1}^{G} Z_g(\tau) \right\} s.t. Z_g(\tau) \leq M_g(\tau), Z_g(\tau) \leq 0 \forall g.

48 As in a fixed effects regression, selection bias is addressed by estimating the endogenous object. If the fixed effect is incorrectly estimated, then the endogeneity concern remains.
Table 5: Usage and votes by facility distance and congestion

<table>
<thead>
<tr>
<th></th>
<th>Usage</th>
<th></th>
<th>Votes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Standard Error</td>
<td>Estimate</td>
<td>Standard Error</td>
</tr>
<tr>
<td>Distance, public hospital</td>
<td>-2.295***</td>
<td>(0.0561)</td>
<td>-1.037***</td>
<td>(0.339)</td>
</tr>
<tr>
<td>Distance, private hospital</td>
<td>-1.127***</td>
<td>(0.0323)</td>
<td>1.093***</td>
<td>(0.233)</td>
</tr>
<tr>
<td>Distance, clinic</td>
<td>-2.456***</td>
<td>(0.233)</td>
<td>-4.223***</td>
<td>(1.214)</td>
</tr>
<tr>
<td>Distance, subclinic</td>
<td>-2.404***</td>
<td>(0.442)</td>
<td>-5.635***</td>
<td>(1.941)</td>
</tr>
<tr>
<td>Congestion, public hospital</td>
<td>-0.0247***</td>
<td>(0.00298)</td>
<td>-0.0395**</td>
<td>(0.0157)</td>
</tr>
<tr>
<td>Congestion, private hospital</td>
<td>-0.0172***</td>
<td>(0.00127)</td>
<td>-0.129***</td>
<td>(0.00597)</td>
</tr>
<tr>
<td>Congestion, clinic</td>
<td>-0.381***</td>
<td>(0.0389)</td>
<td>-0.125</td>
<td>(0.109)</td>
</tr>
<tr>
<td>Congestion, subclinic</td>
<td>-0.589***</td>
<td>(0.0246)</td>
<td>0.405***</td>
<td>(0.0924)</td>
</tr>
<tr>
<td>Price</td>
<td>-0.537***</td>
<td>(0.152)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For usage, each column is a single conditional multinomial logit regression with village and facility type-year fixed effects. The unit of observation is a village-year-facility type, where the set of facility types represents a village’s choice set in a given year. The outcome is usage by facility type, as recorded in the SUSENAS data. For votes, each column is a single conditional binary logit regression with village and year fixed effects. The unit of observation is a village-year, and the outcome is whether the village voted for the ruling party in the 1999 and 2004 legislative elections. The ruling parties were Golkar before the 1999 elections and PDI-P before the 2004 elections. Distance is to the closest facility of each type and is measured in units of 100 km. Congestion of the closest facility is the number of people for whom this facility is the closest of its type. This variable is measured in units of 100,000 people. Price is measured in units of $100 (in year 2000 USD). Additional controls include population and ruralness. Standard errors are clustered by village. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

6 Estimates

In this section, I present the results of the demand and moment-inequalities estimation procedures. For the latter, I also present the number of single and double swaps that I identify among the observed facility placement choices. I close by discussing the goodness of fit of the model.

6.1 Demand parameters \( \omega \)

Table 5 presents the results of specifications 19 and 10. Facility distance and congestion largely correspond with lower usage and fewer votes. For votes, there are two exceptions. First, distance to private hospitals has the opposite effect, perhaps because the arrival of private hospitals signals the government’s inability to provide public services. Second, subclinic congestion also takes a positive sign, but the magnitude of the congestion effect for subclinics is small compared to the distance effect. Villages that receive a nearby subclinic experience decreases in both distance and
Table 6: Number of single and double swaps for estimation

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Single swaps</th>
<th>Double swaps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PODES</td>
<td>Hospitals</td>
<td>Clinics</td>
</tr>
<tr>
<td>Centralized</td>
<td>1990, 93, 96, 2000</td>
<td>1,089</td>
<td>39,522</td>
</tr>
<tr>
<td>Decentralized</td>
<td>2003, 06, 08, 11, 14</td>
<td>5,013</td>
<td>45,267</td>
</tr>
</tbody>
</table>

Swaps are all within districts. After decentralization, district governments choose placements considering only in-district villages $V_n$. Before decentralization, the central government is the sole decision-maker and considers the full set of villages $V_{all}$.

Congestion, such that the net effect on votes is positive.\(^{49}\)

Table A2, appended, shows relatively little heterogeneity by population density. For usage, urban areas are more elastic in terms of distance and less elastic in terms of congestion, although only the congestion differences are statistically significant. For votes, the results are directionally similar, but the differences are not statistically significant. Thus, in the next stage of analysis I focus on the non-interacted demand estimates of table 5.

6.2 Government preference parameters $\tau$

Table 6 shows how many valid swaps can be identified based on the actual policy $a$ observed in the data. I form double swaps separately on the pre- and post-decentralization data.\(^{50}\) Swaps are constrained to be within districts both before and after decentralization.\(^{51}\) Where redistricting occurs, I omit swaps that span redistricted borders.\(^{52}\) Double-swap sequences require multiple construction in a single village, so there fewer double-swap sequences before decentralization given fewer waves of data and less construction. I am working on getting new hospitals data that should

\(^{49}\) Congestion is defined as the number of people in a facility’s catchment region, which covers the villages for which that facility is the closest of its type. A village can experience a decrease in congestion without any change in distance if a new facility is constructed far from the village but within the vicinity the village facility’s catchment region. Distance is unchanged because the original facility is still closest to the village, but congestion decreases because other villages that once visited the original facility now go instead to the new facility. The estimates somewhat counterintuitively predict that votes will decrease in this case. The reason may be that these catchment regions are small for subclinics, so a decrease in subclinic congestion without a corresponding decrease in subclinic distance means that a relatively nearby village received a new subclinic. In this case, the village gets passed over in receiving public resources, and the proximity of the new, non-local subclinic may be salient enough to overwhelm the direct benefit of lower congestion.

\(^{50}\) Adding restrictions on village-specific unobservables $\lambda$ can increase the number of valid swaps. For example, if $\lambda$ is assumed to be constant within a subdistrict, then double-swaps across subdistricts and single-swaps within subdistricts are each valid sources of revealed-preference inequalities.

\(^{51}\) Swaps compare the actual policy $a$ to some alternative policy $a$ within the decision-maker’s choice set. If alternative $a$ is not within the decision-maker’s consideration set, then the resulting revealed-preference inequality will not necessarily hold. After decentralization, district governments choose facility placement within their districts. Before decentralization, in principle the central government chooses over the full set of villages. The full set of swaps is therefore very large. In practice, however, population-based rules govern the allocation of facilities to districts, so placement choices may still be constrained to be within districts. As such, I consider only the subset of swaps that occur within districts.

\(^{52}\) In matching swap pairs to form double swaps, I ignore redistricting. This choice amounts to assuming that $\lambda^{\text{pre}}$ and $\lambda^{\text{post}}$ vary only by village and do not change when a village experiences a change in district government. Allowing for this variation would require restricting double swaps not to span a district split, further reducing the number of valid double swaps. Note that I do not ignore redistricting in identifying the initial set of single swaps.
Table 7: Government preference parameter estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Votes ($\tau^V$)</td>
<td>17.55</td>
</tr>
<tr>
<td>Land costs ($\tau^L$)</td>
<td>3.17</td>
</tr>
<tr>
<td>Population density costs ($\tau^P$)</td>
<td>2.01</td>
</tr>
<tr>
<td>“R-squared”</td>
<td>0.62</td>
</tr>
</tbody>
</table>

increase the number of pre-decentralization swaps.

Table 7, shows estimates of government preference parameters $\tau$, which I generate using the demand estimates from table 5. For now, I present these estimates without standard errors. Estimation yields point estimates, which occur when not all inequalities can be simultaneously satisfied. The government values one additional vote as much as $17 of patient welfare. This effect is large in magnitude: among single swaps in Jakarta, the average absolute difference between placements is 777 for the welfare term and 46 for the votes term. If votes are valued at $17 of welfare, then the government places roughly equal weight on patient welfare and votes ($777 \approx 46 \times 17$). Furthermore, the government values an additional dollar of spending on land costs at $3 of patient welfare. That is, the government requires a welfare benefit of $3 in order to justify an additional dollar of government spending. Finally, figure A3, appended, maps the estimated village preferences $\lambda_v$ for Jakarta in the post-decentralization period. The majority of villages have a large, negative values of $\lambda_v$, which reflects that they do not receive many facilities despite potentially large welfare and vote benefits.

6.3 Goodness of fit

I examine the goodness of fit of the model in two ways. First, I evaluate the set of single-swap inequalities at the estimated parameters, and I report the percentage that are positive in table 7. To do so, I compute

$$
\sum_{t=s_1(a)}^{s_2(a)-1} \beta^{t-1} \left( \Delta \hat{WVC}(\hat{a}^t, \hat{\omega}; \hat{\tau}, \hat{\omega}) + \hat{\lambda}_{w_2(a)} - \hat{\lambda}_{w_1(a)} \right) \geq 0
$$

for every identified single swap as the sample analogues to inequality 25. The percentage that are positive reflects the degree to which the model and the estimated parameters explain the observed placement, at least relative to its one-step deviations.53

Second, I compute the optimal placement predicted by the model and compare it to the observed placement. Given computational constraints, in identifying the optimal placement I restrict attention to placements that are resequencings of the observed placement. The set of single swaps is a subset of this space, as single swaps are resequencings that are at most one permutation away from the observed placement. To compare the predicted and observed placements, I compare the timing with which each village receives its facilities.

53 I do not estimate the $\varepsilon_{vt}$ parameters; the omission of this term means that, even at the true parameter values, not all single-swap inequalities will necessarily be positive.
7 Welfare exercises

The estimated model yields the social planner’s objective function. I solve for the allocation that this welfare-maximizing agent would have chosen (subject to the observed budget), and I compare this allocation to the actual allocation. By this metric, an “efficient” allocation is one that coincides with the social planner’s allocation. I make this comparison before and after the reform, and I find that the post-reform allocation is more efficient. I draw on the structure of the model to decompose mechanisms, and I find that electoral concerns drive the change.

7.1 Facility placements

Obtaining the counterfactual placements requires solving a high-dimensional combinatorial optimization problem. Note that estimation sidestepped this challenge by using moment inequalities, which evaluate only a subset of the choice space. To make progress here, I only consider placements that satisfy budget-spending assumption 18 – namely, if a district receives $N_f$ facilities in a given period, then all possible placement choices maintain $N_f$ facilities in that period – and this restriction does help to reduce the choice space. Even so, the large number of facilities placed in each period generates a choice space that is still of intractibly high dimension.

One way forward is to restrict attention to resequencings of the observed placement, as I did in evaluating the model’s goodness of fit. An expanded choice set might involve an arbitrary allocation of facilities across villages, but focus only on villages that eventually receive a facility as in Zheng (2016). A further step would further expand attention to villages close to those that eventually receive a facility, even if these villages never receive a facility themselves. The computation burden increases in each of these cases, so I approximate the dynamics in the objective function using forward induction with some pre-specified lookahead depth, and where necessary I further rely on stochastic heuristic such as simulated annealing.

7.2 Social welfare

In order to quantify welfare effects, I define social welfare function

$$\pi^S(a) = \sum_{t=1}^{\infty} \beta^{t-1} \left( W(a^t) - C(a^t) \right),$$

where the social planner does not consider votes or unobservables. Rather, for a given placement, the social planner only considers welfare gains net of costs. In making counterfactual predictions, I hold constant the facility budget and focus instead on the location of facilities. Conditioning on the budget in this way eliminates first-order cost differences, while the model further adjusts for cost differences associated with land prices and population density.

Using this metric, I evaluate the actual allocations observed under the pre- and post-reform periods. To do so, I calculate social welfare under the social planner’s allocation and compare it to the social welfare generated by the actual allocation. Table 8 reports the results. In the pre-reform period, the actual allocation generated $10 million in social welfare, while the social planner would have generated $13 million. The observed allocation therefore generated 77% of the efficiency frontier. In the post-reform period, the actual allocation generated $14 million of a potential $15 million, and this 16 percentage point increase over the pre-reform ratio suggests an improvement. Figure 3 visualizes the difference between the two allocations, which result in gains.
Table 8: Post-reform misallocation and channels (Jakarta)

<table>
<thead>
<tr>
<th></th>
<th>$a_{\text{actual}}$</th>
<th>$a_{\text{SP}}$</th>
<th>(1 − ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misallocation</td>
<td>$14m$</td>
<td>$15m$</td>
<td>0.07</td>
</tr>
<tr>
<td>Effects of:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elite capture (relative to pre-reform)</td>
<td>-2 p.p.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All channels</td>
<td>-10 p.p.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Measured relative to no new construction. Effect of elite capture relative to zero elite capture (as opposed to pre-reform levels) is 2 p.p.

for some regions and loses for others (although in sum the net gain is positive). Decomposing the channels in the post-reform period, I find that electoral concerns explain 12 percentage points of the 16-point change. The positive effect of electoral concerns is much larger than the negative effect of uninternalized spillovers, which is consistent with the reforms’ having a positive effect on net.

8 Conclusion

This paper studies the welfare benefits and costs of decentralized planning in public service provision. To do so, I examine the expansion of healthcare infrastructure in Indonesia before and after the country’s big-bang decentralization reform in 1999. I find that decentralization, both administrative and political, has a significant effect on the placement of health facilities, and I use a structural model to quantify the welfare implications of these effects. I estimate the model applying moment-inequality techniques from the industrial organization literature. I show the post-reform allocations were better in terms of social welfare than the pre-reform allocations, and that this change is driven by the introduction of electoral concerns.
**Figure 3:** Consumer surplus gains under the social planner’s allocation (post-reform)
References


A Appendix

Table A1: Variation in other-district exposure by redistricting

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2003-2008</td>
<td>Exposure</td>
<td>0.415</td>
<td>0.415</td>
<td>0.417</td>
<td>0.446</td>
<td>0.471</td>
<td>0.534</td>
<td>0.596</td>
<td>0.641</td>
<td>0.644</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>17,948</td>
<td>17,948</td>
<td>17,948</td>
<td>17,948</td>
<td>17,948</td>
<td>17,948</td>
<td>17,948</td>
<td>17,948</td>
<td>17,948</td>
</tr>
<tr>
<td>Never</td>
<td>Exposure</td>
<td>0.812</td>
<td>0.812</td>
<td>0.812</td>
<td>0.812</td>
<td>0.812</td>
<td>0.812</td>
<td>0.812</td>
<td>0.812</td>
<td>0.812</td>
</tr>
</tbody>
</table>

The unit of observation is a village in a given year. Exposure to other districts is calculated as the proportion of people within 100 km of a village (downweighting linearly by distance) who reside in a district different from that of the village.

Figure A1: Facility construction by exposure to other districts

Each figure plots pre-trends for an event-study regression with village and year fixed effects. I restrict the sample to villages in districts that undergo redistricting in the post-decentralization period. The figure examines the impact of exposure to other districts on facility construction, where the change in exposure is generated by redistricting. Exposure to other districts is calculated as the proportion of people within 100 km of a village, downweighting linearly by distance, who reside in a district different from that of the village. The outcome is the number of facilities in a village, where the outcome facility is denoted by the column labels. Controls include include population, ruralness, and the number of facilities for non-outcome facility types, including private hospitals. The unit of observation is a village-year. Standard errors are clustered by village, and the standard error bars are at the 95% significance level.
Figure A2: Facility construction by exposure to other districts, pre-decentralization

Each figure plots pre-trends for an event-study regression with village and year fixed effects. I restrict the sample to villages in districts that undergo redistricting in the pre-decentralization period. The figure examines the impact of exposure to other districts on facility construction, where the change in exposure is generated by redistricting. Exposure to other districts is calculated as the proportion of people within 100 km of a village, downweighting linearly by distance, who reside in a district different from that of the village. The outcome is the number of facilities in a village, where the outcome facility is denoted by the column labels. Controls include population, ruralness, and the number of facilities for non-outcome facility types, including private hospitals. The unit of observation is a village-year. Standard errors are clustered by village, and the standard error bars are at the 95% significance level.

Figure A3: Estimated village preferences $\lambda_v$, Jakarta post-decentralization
A.1 Specifying treatment

As Donaldson and Hornbeck (2016) discuss, properly specifying our treatment variables is key. In their context, using a binary treatment variable implies a comparison between counties that receive railroads to those that do not. They avoid this formulation given the presence of treatment spillovers. When “untreated” countries also benefit from railroad construction in neighboring counties, estimates from the direct comparison will be biased downward. Instead, they construct a market access measure that captures both direct and indirect effects of railroad construction, allowing for a more precise measure of relative treatment intensity.

In a similar spirit, I take distance\(_{v_{ft}}\) and congestion\(_{f_{t}}\) as the treatment variables. New facilities would change each of these variables, giving some predicted effect on usage. The following example illustrates why both are necessary for capturing treatment intensity.

Recall that villages use their closest facility, and that facilities are homogeneous except in location. A binary treatment variable is clearly inaccurate, as all three villages have access to a facility despite not having received one within their village borders. Distance captures the first-order difference in treatment intensity: villages \(v^1\) and \(v^2\) are closer to their closest facilities than \(v^3\) is, and as such are more intensely treated. Congestion also matters, as facilities face binding capacity constraints. Villages \(v^1\) and \(v^2\) are equidistant to their closest facilities (facilities \(f^1\) and \(f^2\), respectively). But \(f^2\) serves two villages, while \(f^1\) only serves one. In this sense village \(v^2\) has better health access than \(v^1\) – they are not equally treated.
### Table A2: Usage and votes by facility distance and congestion, population-density interaction

<table>
<thead>
<tr>
<th></th>
<th>Usage Estimate (Standard Error)</th>
<th>Votes Estimate (Standard Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance, public hospital</td>
<td>-2.058*** (0.0568)</td>
<td>-0.890** (0.355)</td>
</tr>
<tr>
<td>Distance, private hospital</td>
<td>-0.950*** (0.0298)</td>
<td>0.842*** (0.241)</td>
</tr>
<tr>
<td>Distance, clinic</td>
<td>-2.817*** (0.241)</td>
<td>-4.140*** (1.222)</td>
</tr>
<tr>
<td>Distance, subclinic</td>
<td>-4.467*** (0.447)</td>
<td>-5.830*** (1.967)</td>
</tr>
<tr>
<td>Distance, public hospital × population density</td>
<td>-0.0424 (0.0353)</td>
<td>-4.590 (2.885)</td>
</tr>
<tr>
<td>Distance, private hospital × population density</td>
<td>-0.0303 (0.0210)</td>
<td>4.153*** (1.019)</td>
</tr>
<tr>
<td>Distance, clinic × population density</td>
<td>-0.155 (0.137)</td>
<td>-9.791 (15.79)</td>
</tr>
<tr>
<td>Distance, subclinic × population density</td>
<td>-0.803 (0.646)</td>
<td>9.863 (8.764)</td>
</tr>
<tr>
<td>Congestion, public hospital</td>
<td>-0.0206*** (0.00298)</td>
<td>-0.0411** (0.0166)</td>
</tr>
<tr>
<td>Congestion, private hospital</td>
<td>-0.0133*** (0.00123)</td>
<td>-0.129*** (0.00608)</td>
</tr>
<tr>
<td>Congestion, clinic</td>
<td>-0.281*** (0.0346)</td>
<td>-0.119 (0.114)</td>
</tr>
<tr>
<td>Congestion, subclinic</td>
<td>-0.468*** (0.0380)</td>
<td>0.374*** (0.121)</td>
</tr>
<tr>
<td>Congestion, public hospital × population density</td>
<td>0.00268 (0.00171)</td>
<td>0.0199 (0.0480)</td>
</tr>
<tr>
<td>Congestion, private hospital × population density</td>
<td>0.00196** (0.000900)</td>
<td>0.0117 (0.00867)</td>
</tr>
<tr>
<td>Congestion, clinic × population density</td>
<td>0.0172** (0.00738)</td>
<td>0.0619 (0.208)</td>
</tr>
<tr>
<td>Congestion, subclinic × population density</td>
<td>0.0426*** (0.0136)</td>
<td>0.0307 (0.0548)</td>
</tr>
<tr>
<td>Price</td>
<td>-0.764*** (0.152)</td>
<td></td>
</tr>
<tr>
<td>Village FE</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Facility type-year FE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year FE</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>202,668</td>
<td>118,337</td>
</tr>
</tbody>
</table>

For usage, each column is a single conditional multinomial logit regression with village and facility type-year fixed effects. The unit of observation is a village-year-facility type, where the set of facility types represents a village’s choice set in a given year. The outcome is usage by facility type, as recorded in the SUSENAS data. For votes, each column is a single conditional binary logit regression with village and year fixed effects. The unit of observation is a village-year, and the outcome is whether the village voted for the ruling party in the 1999 and 2004 legislative elections. The ruling parties were Golkar before the 1999 elections and PDI-P before the 2004 elections. Distance is to the closest facility of each type and is measured in units of 100 km. Congestion of the closest facility is the number of people for whom this facility is the closest of its type. This variable is measured in units of 100,000 people. Price is measured in units of $100 (in year 2000 USD). Population density is measured in units of 10,000 people per square kilometer. Additional controls include population and ruralness. Standard errors are clustered by village. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.