Natural Resources and Local Communities: Evidence from a Peruvian Gold Mine
On-line appendix

Fernando M. Aragón Juan Pablo Rud

May 2012

A Data and main variables

Household data The empirical analysis uses data from repeated cross sections of the Peruvian Living Standards Survey (ENAHO), an annual household survey collected by the National Statistics Office. The survey consists of a stratified household sample representative at the regional level. The regions are defined for statistical purposes and consider both environmental conditions (coast, highlands and forest) and geographical location (north, center and south).[1] We focus on the North Highlands statistical region, the area where the mine is located and restrict attention only to households with an employed head.[2] Figure 1 shows the area of study and highlights in grey the districts in the survey's universe. Districts are the smallest political jurisdictions, usually composed by a main town and a surrounding rural area. The data set covers 10 years, between 1997 and 2006, and includes in total more than 7700 households located in 101 districts.[3]

The main purpose of the survey is to measure poverty and living standards. The survey contains detailed information on income, expenditure,

[1] In general, these statistical regions are larger than departments and do not necessarily share the same boundaries.

[2] This filter reduces the sample by just 46 observations and does not affect the results.

[3] It represents an average of 770 observations per year.

socio-demographics (such as gender, age, educational attainment of individuals), composition of the household, housing characteristics like access to public utilities and construction materials, and self-reported incidence of health problems and exposure to crime. The data set also has extensive information on prices and agricultural activity at household level.

To quantify exposure to the mine's center of activities, Cajamarca city, we construct a measure of the distance from the household's location to the city. This measure varies at district level. In particular, we measure distance as the length of the shortest route between the main town of the district and Cajamarca city using the existing road network.[4] We perform the calculation using the ArcGIS software and maps produced by the Ministry of Transport of Peru. The road map corresponds to the network available in 2001 and includes only tracks usable by motorized vehicles. The measure of distance ranges from 0 to 400 km, with an average value of 100 km. As we will discuss below, we use this threshold to define districts close and far from the city. Figure 1 shows the districts with households included in the survey sample, and highlights in dark grey the districts within 100 km to the city. Note that the sample includes districts in the vicinity of other cities, such as Chachapoyas.

Table 1 shows some summary statistics of the main variables from the household survey. We estimate the means and standard errors using sample weights and clustering by primary sampling unit to account for the sampling design.

Constructing a measure of real income Our main outcome variable is the real income per capita. To construct this variable we divide the net nominal income per capita by a local consumer price index. As a price index we use the value of the poverty line, as calculated by the National Statistics Agency (INEI).

To construct this index, the INEI sums the value of food and non-food minimum consumption baskets (INEI, 2010 p. 13-18). The food consumption basket reflects food required to meet minimum nutritional requirements. The

[4]The results are robust to alternative measures of distance.

Figure 1: Districts in sample, by distance to Cajamarca city

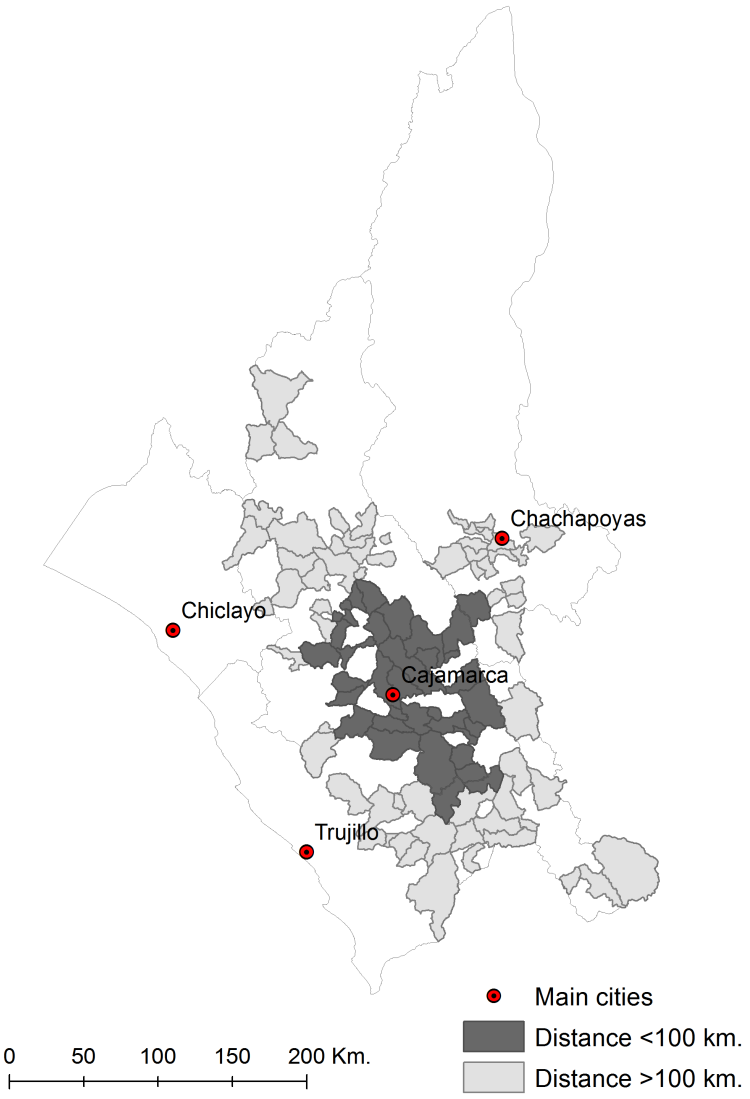


Table 1: Summary statistics of household data

Variables	Mean N=7,738	Standard error
Household head		
Years of education	5.4	0.1
Age	47.4	0.2
% female	15.7	0.4
Household		
Income per capita	212.0	3.4
Consumption per capita	190.2	2.2
Poverty line	173.3	0.3
% poor	63.5	0.5
% extreme poor	33.9	0.5
% urban	36.5	0.5
% access to electricity	38.1	0.6
% access to piped water	59.1	0.6
Nr. Household members	4.7	0.03
Nr. Income earners	2.0	0.01
Distance to Cajamarca city (km)	97.0	0.7

Note: The mean and its standard error are calculated using sample weights and clustering by primary sampling unit. Income, consumption and poverty line are measured in Nuevos Soles. In the period of analysis, the average exchange rate was 1 US dollar=3.2 Nuevos Soles.

composition of the food basket remains stable over the period of analysis, but the prices are updated every year. The prices are obtained from the household survey and are calculated as the average of each department's urban and rural area.[5] The resulting value of this minimum food basket corresponds to the extreme poverty line threshold.

The non-food consumption basket includes goods from major consumption groups such as clothing, transportation, health services, entertainment, and housing. The value of this consumption basket is calculated using prices collected in main cities (like Cajamarca, Trujillo, Chachapoyas and Chiclayo).[6] The rural prices are assumed to be the same as urban prices.

A main concern is that the poverty line may fail to capture the actual change in local cost of living. Nonetheless, there are two reasons that justify the use of the poverty line as a price deflator. First, the poverty headcount in the sample is 65 percent. This implies that the median household is poor and hence its consumption basket may not be too different from the one used to calculated the poverty line.

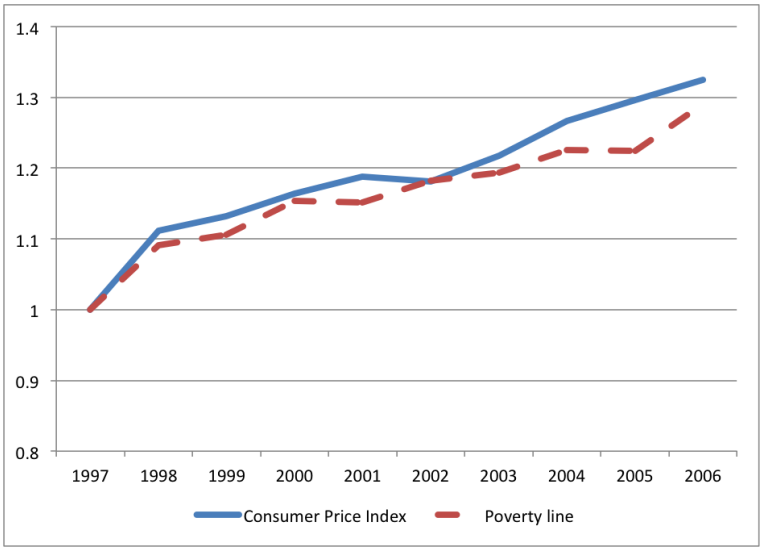
Second, we compare the evolution of the poverty line in Cajamarca city with the official consumer price index used by the National Statistics Office. This price is used to report city's inflation and it is only available for major cities (such as Cajamarca). Figure 2 depicts both variables, normalized to be equal to 100 in year 1997 for Cajamarca city. Note that the poverty line has a similar trend than the official consumer price index. This evidence suggests that the poverty line captures a relevant dimension of the local cost of living.

We further explore the robustness of the results to alternative price deflators in two ways. First, we compare Cajamarca city to other cities. In that case we use the official consumer price index as a price deflator (see Section B.3.3). Second, we combine information from the poverty lines with proxies of self-reported housing rents to construct price deflators at district level. The results are similar to the ones using the poverty line (see Table 5 in the Appendix A2).

[5] In our sample, it means there are 7 different values of the poverty line each year.

[6] The data used to calculate this index, however, is not available in the household survey.

Figure 2: Consumer price index and poverty line for Cajamarca city



Firm data To measure the expansion of the mine activities, we collect data from Yanacocha reports on total payment to workers, local purchases and total production (Minera Yanacocha, 2006). The frequency of this data is annual and covers the period 1993 to 2006. Local purchases include goods and services bought to local suppliers and contractors. This variable includes the wages of workers that work for mine contractors. The wage bill includes all work-related payments to Yanacocha's directly employed workers. This includes wages, bonuses and a share of the mine's profits.[7] We include this last item as part of the wage bill, since it is effectively part of the total remuneration of workers.[8]

We measure the mine's demand of local inputs as the sum of the wage bill and local purchases. Panel A in Table 2 presents summary statistics for the firm level data over the period 1997 to 2006. The value of wage bill and local purchases is measured in millions of US dollars while the quantity produced is measured in millions of ounces.

[7] 8 percent of the mine profits are distributed among mine workers. This benefit is defined by law and accrue only to workers directly employed by the mine, not to workers employed through contractors.

[8] The results are robust to the exclusion of the workers share of mine's profits of the measure of wage bill.

Municipal data We complement the household and firm data with data at the municipal level.[9] Municipalities are the lowest tier of autonomous local government with jurisdiction over districts. We obtain annual data about revenues and expenditures for each municipality in the North Highlands region and within 400 km from Cajamarca city.[10] This geographical scope corresponds to the distance range observed in the household data. The data set covers information on 102 municipalities over the period 1998 to 2006, and contains detailed information about the sources of revenue, including the amount of mining transfers (canon) received. This information provides a reliable measure of the magnitude of the revenue windfall experienced by each local government.

Panel B in Table 2 displays some summary statistics. The average municipality has an annual budget of 4 million Nuevos Soles (approximately US$ 1.25 million), but a slightly smaller expenditure. The difference is kept by the local government and rolled forward to subsequent periods.

[9]We use this data in Section V.A. to evaluate the role of the fiscal revenue windfall as an alternative explanation of the observed phenomena.

[10]This data comes from municipal budgetary reports collected by the Peruvian Ministry of Economy and Finance.

Table 2: Summary statistics of firm and municipal data

Variables	Mean	Standard deviation
A. Firm data N=10		
Wage bill	42.3	27.7
Local purchases	55.5	33.9
Gold production	2.2	0.8
% local purchases	0.12	0.05
B. Municipal data N=522		
Total revenue	4.0	9.0
Canon minero	0.9	3.7
Total expenditure	3.2	6.1
Capital expenditure	1.8	3.6

Note: The value of wage bill and local purchases is measured in million of US$ while the quantity produced is measured in millions of ounces. The municipal data is measured in millions of Nuevos Soles. In the period of analysis, the average exchange rate was 1 US dollar=3.2 Nuevos Soles.

B Additional empirical results

B.1 Alternative Measures of Distance

In the baseline regressions, we use the shortest route by road and the average distance (i.e. 100 km) as the threshold. We check the robustness of our results to alternative measures of distance. In column 1 of Table 3 we show that results hold when the threshold is defined by the median, i.e. 92.6 km.

Additionally, we obtain two alternative measures of distance: a topographic measure and a straight line. The topographic measure is calculated using the ArcGIS package by minimizing the sum of the normalized values of altitude and gradient, regardless of the existence of a road. It can be interpreted as a proxy for where a road may be located or alternative transportation routes in the absence of roads. The straight line measure is calculated as the Euclidean distance between the district capital town and the city of Cajamarca. In order to distinguish district closer and farther from the city, we use as a threshold the median value of the measure of distance. Columns 2 and 3 show that the effects are similar, irrespective of the measure of distance used to tell apart districts that are far and close to Cajamarca.

Finally, we explore in more detail the monotonic decline of the effect by distance, which is a crucial feature of our identification strategy. To do that, we estimate the baseline regression including the interaction between the mine's demand of local input and different functions of distance. Column 1 in Table 4 displays the results with the linear measure of distance as a benchmark. Columns 2 and 3 allow for non-linearities by including the logarithm and inverse of distance. In all cases, the results support the claim that the effect of the mine expansion on real income declines with distance to the city.

Table 3: Alternative measures of distance

	Ln(real income)		
	(1)	(2)	(3)
Demand for local inputs	0.135*	0.146**	0.174**
× distance < median	(0.074)	(0.070)	(0.075)
Measure of distance	Shortest path by road	Topographic	Straight line
Median distance (km)	92.6	75.8	65.9
Observations	7,738	7,738	7,738
R-squared	0.524	0.524	0.524

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include year and district fixed effects and the same control variables of the baseline regression. Topographic distance is calculated as the length of the shortest path that minimizes the normalized sum of altitude and gradient. Straight line distance is the Euclidean distance between two points.

Table 4: Exploring the decrease in the effect by distance

	Real income		
	(1)	(2)	(3)
Demand for local inputs	-0.115**		
× distance	(0.053)		
Demand for local inputs		-0.026**	
× Ln(distance)		(0.011)	
Demand for local inputs			0.001**
× distance^{-1}			(0.001)
Observations	7,738	7,738	7,738
R-squared	0.524	0.523	0.523

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include year and district fixed effects and the same control variables of the baseline regression. *Distance* is equal to the length of the shortest route by road from the main town of the district where the household lives to Cajamarca city, expressed in hundreds of kilometers.

B.2 Alternative price deflators

The main results use the value of the poverty line as a price deflator to construct a measure of real income. As previously mentioned, a main concern is that the poverty line may fail to capture relevant changes in local prices such as house rents.

To evaluate the robustness of the main results, we replicate them using alternative price deflators. In particular, we construct a local consumer price index (*local CPI*) combining information of the poverty lines (total and extreme) and actual house rental prices.

The local CPI is defined as $\sum s_k p_k$ with $k \in \{\text{food, housing, other goods}\}$. s_k is the consumption share of good k and p_k is a measure of the local price of the good. In order to sum these three prices, we normalize then to be equal to 100 in the base year of 1997.

As a proxy of local food prices, we use the value of the extreme poverty line. Recall, that this threshold reflects the value of buying a minimum food basket. As a proxy of other non-food non-housing goods, we use the difference between the total and extreme poverty line. This difference corresponds to the value of a non-food consumption basket that include, among other expenditures, housing. The value of this non-food consumption basket is calculated annually using prices from major cities (like Cajamarca). Rural prices are assumed to be equal to urban prices.

As a proxy of housing prices, we use actual house rental prices from the household survey. This information is only collected for tenants and hence the sample size reduces considerably. We aggregate this measure of housing prices by taking the weighted average at district level. Finally, we calculate consumption shares s_k using survey information in 1997 and aggregating household shares at district level. The share of food is obtained directly from information on household expenditure, while the share of housing expenditure is calculated as the value of house rental relative to the household total expenditure. On average, the share of food expenditures is 66.1 percent, while the share of housing expenditure is 8.8 percent.

Table 5 replicates the main results using these alternative price deflators.

As a benchmark, column 1 presents the results using the poverty line as a price deflator while column 2 uses the local CPI. Note that the results using this alternative price deflator are even larger than the baseline results.

Column 3 calculates the CPI with a larger share of housing expenditure. In particular, it uses all house-related expenditures as reported in the survey expenditure data, which includes rent, energy consumption and home maintenance. The mean of this alternative share of housing expenses is 21.9 percent. Since house rents increase with the mine expansion, this approach overestimates the increase in local prices. Note that in this case, the estimate of the effect of the mine is smaller and noisier, but still positive and significant.

Column 4 uses self-reported house rents as a price deflator. In this case, the effect of the mine is insignificant and suggests that nominal income has increased in similar proportion than house rental prices. This result is consistent with the finding that both workers and home owners benefit from the mine expansion, and that the benefits of the demand shock from the mine are transmitted through general equilibrium effects.

We report this last result for completeness only. Note that it should not be interpreted as the effect of the mine on real income, since it would require assuming that housing corresponds to 100 percent of a household's consumption. Alternatively, we would need to assume that all local prices increase accordingly to the house rental price. This would not be the case, for example, in the presence of tradable goods.

B.3 Additional checks

B.3.1 Heterogeneous trends

In this section, we perform additional checks to evaluate the robustness of the main results. A first concern is that there are some systematic differences between areas close to and far from Cajamarca. Areas closer to the city have larger population, are more urbanized and dense, are located at higher

Table 5: Effect of Yanacocha's expansion on real income, using alternative price deflators

	Ln(nominal income/price deflator)			
	(1)	(2)	(3)	(4)
Demand for local inputs × distance < 100 km	0.169**	0.346***	0.307***	0.118
	(0.077)	(0.095)	(0.100)	(0.248)
Price deflator	Poverty line	Local CPI	Local CPI with larger share of house exp.	House rental price
Observations	7,738	3,919	3,919	3,919
R-squared	0.524	0.601	0.597	0.608

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include year and district fixed effects, and control variables as in the baseline regression.

altitude, and have a slightly better educated population.[11]

In the main specification, this is dealt with by using district fixed effects and controlling for household characteristics. These different initial conditions, however may also lead to different trends of income. This would violate our identification assumption and imply that the estimated effect on real income could not be attributed to the mine expansion.

To address this concern, we include a non-parametric trend interacted with dummies related to observable characteristics. In particular, we use an indicator of urbanization (a dummy equal to 1 if the household is located in an urban area), and dummies equal to 1 if the population size, population density, altitude or average schooling of the district are above the sample median. In addition, we add to this specification the measure of demand for local input interacted with distance to other cities (i.e. Chachapoyas, Chiclayo and Trujillo). This check explores whether the previous results are driven by proximity to Cajamarca city or could be attributed to proximity

[11]We perform a mean comparison to identify variables which are not balanced between both areas, see Table 9 in the Appendix A3.

to any city.

Columns 1 and 2 in Table 6 show the results of these robustness checks. In all cases, the results are similar to those found in the baseline regression.[12] Note that the interaction between the mine demand for local inputs and proximity to other cities is insignificant, while the interaction with proximity to Cajamarca city remains positive and significant. These findings reduce concerns that the baseline results are driven by differential trends between areas close and far from Cajamarca city.

A second concern is that the results are mechanically reflecting transfers from the mine to some groups instead of positive spillovers. We identify three groups that may benefited directly from the mine expansion: mine workers, public workers, and communities near Yanacocha's mine site.[13] Then, we exclude from the sample households with at least one mining or public worker[14] as well as households living in the districts targeted by Yanacocha's development projects. In our sample, these three groups of households represent around 16 percent of total observations. The results of this check are shown in column 3 in Table 6

B.3.2 Falsification test

The previous checks suggest that our results are not driven by different trends based on some observable characteristics. However, there might be unobservable shocks contemporaneous to the mine expansion that affect areas differently according to their proximity to any city.

To explore this issue, we perform a falsification test replicating the estimates of the effect of the mine on real income but using as reference points

[12]The results are similar if we add the non-parametric trends one by one.

[13]Mine workers receive a wage premium. For example, in 1997, the average salary for a Yanacocha employee was almost three times the salary for a similar job in Cajamarca city (Pascó-Font et al., 2001, p.165). Similarly, public workers may have received part of the revenue windfall associated with mining, as documented by Caselli and Michaels (2009) in he Brazilian case. Finally, Yanacocha implements small-scale social development projects among communities in the vicinity of the mine site.

[14]The industry of occupation is based 2-digit International Standard Industry Code of the main activity of working individuals. We classify education workers as part of the public sector given that most teachers are in the government payroll.

Table 6: Effect of the mine on real income, controlling by heterogeneous trends and excluding potential direct beneficiaries

	Ln(real income)		
	(1)	(2)	(3)
Demand for local inputs × dist. to Cajamarca < 100 km	0.186***	0.205***	0.180*
	(0.068)	(0.065)	(0.092)
Demand for local inputs × dist. to Chachapoyas < 100 km		-0.034	
		(0.129)	
Demand for local inputs x dist. to Chiclayo < 100 km		-0.089	
		(0.152)	
Demand for local inputs × dist. to Trujillo < 100 km		0.105	
		(0.111)	
Non-parametric trends × district characteristics	Yes	Yes	No
Sample excludes potential direct beneficiaries	No	No	Yes
Observations	7,738	7,738	6,570
R-squared	0.543	0.543	0.482

Notes: Robust standard errors in parentheses. Standard errors are clustering at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include year and district fixed effects and the same controls as the baseline regression. The non-parametric trends × district characteristics are year fixed effects interacted with indicators whether the district's population, density, altitude or schooling are above the sample median, and with a dummy equal to 1 if household lives in urban area. Potential direct beneficiaries include: households with at least 1 mining or public workers, and households living in districts targeted by Yanacocha's social development programs.

other cities instead of Cajamarca. Finding a similar effect of the mine expansion on other cities would suggest that the observed effect on real income observed in areas closer to Cajamarca city is just reflecting a broader urban-rural phenomenon and would raise concerns about the validity of the identification assumption

We select the other main cities around the North Highlands region: Chachapoyas, Chiclayo and Trujillo (see Figure 1 for a localization map of the cities). All these cities are, like Cajamarca, a departmental capital, and have a similar governmental status. Chachapoyas is located in the highlands and have a similar size as Cajamarca. In contrast, Chiclayo and Trujillo are much larger cities located on the coast. For each city, we calculate proximity using the same algorithm as in the baseline results.

Table 7 displays the results of the falsification test using two alternative samples. Panel A uses the same sample as in the baseline results: households in the North Highland region. Panel B includes households within 200 km of the cities, regardless of the geographical region.[15] In all cases, the effect of mine wages and purchases becomes insignificant or even negative. Note that results in Panel A show that the main results presented in the baseline regression (table 2 in the paper) are only explained by proximity to Cajamarca (and not other cities). Results in Panel B, using households closer to each of the other cities, show that there are no contemporaneous positive effects in areas surrounding other cities. The lack of effect on this falsification exercise weakens the explanation that there is some confounding factor which affects large cities, which is driving the results.

B.3.3 Comparing Cajamarca to other cities

Our identification strategy compares the evolution of households living in different proximities to Cajamarca city. This strategy relies on the assumption that the impact of backward linkages from the mine diminishes with distance. A complementary approach is to compare the evolution of households in the proximity of Cajamarca city relative to households near other cities[16].

[15]The results are similar using larger areas of influence, e.g. 400 km

[16]We thank an anonymous referee for suggesting this approach.

Table 7: Falsification test using distance to other cities

	Ln(real income)		
	(1)	(2)	(3)
Panel A: North Highlands sample			
Demand for local inputs	-0.191	-0.060	-0.136
× distance to city < 100 km	(0.148)	(0.165)	(0.132)
Observations	7,738	7,738	7,738
R-squared	0.523	0.523	0.523
Panel B: Households within 200 km			
Demand for local inputs	-0.036	0.003	-0.080*
× distance to city < 100 km	(0.087)	(0.041)	(0.043)
Observations	6,978	10,801	12,073
R-squared	0.554	0.536	0.606
City	Chachapoyas	Chiclayo	Trujillo

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include year and district fixed effects and the same control variables of the baseline regression. Each column measures distance to a different city. The reference point is the city indicated in the last row.

We focus on Cajamarca city and the three major cities around the North Highlands region: Chachapoyas, Chiclayo and Trujillo. We restrict the sample to households living in the vicinity of each city and estimate the following regression:

$$y_{hdt} = \alpha_d + \eta_t + \beta(\ln M_t \times cajamarca\_vicinity_d) + \mathbf{X}_{hdt}\gamma + \varepsilon_{hdt}, \quad (1)$$

where $cajamarca\_vicinity$ is a dummy equal to 1 if household lives in the vicinity of Cajamarca city and 0 if household lives in the vicinity of other cities such as Chachapoyas, Chiclayo or Trujillo. The rest of the specification is similar to the baseline regression.

We define a city's vicinity in several ways. First, we define it as the area within 5 km of the city center.This is the narrowest definition and corresponds to Cajamarca's metropolitan area. The second definition takes a city's vicinity as the province where the city is located. Provinces are jurisdictions intermediate between districts and departments. Finally, we use a broader definition and define a city's vicinity as the area within 100 km of the city, in line with the previous findings that the effect of the mine is circumscribed to this distance.

Note that this identification strategy is effectively a difference in difference. The treatment is the expansion of the mine's demand for local inputs, while the treated and control groups are the households living in the vicinity of Cajamarca city and other cities, respectively.[17] An advantage of this specification is that, by focusing on cities, we can use the official consumer price indices to construct the measure of real income, at least with a narrow definition of a city's vicinity.[18] Thus, this strategy also allows us to evaluate the robustness of the results to an alternative price deflator.

Table 8 estimates regression (1) using these alternative definitions of a city's vicinity. Columns 1 and 2 use the narrowest definition: households

[17]In contrast, the baseline specification compares households living within 100 km of Cajamarca to households living further away. While the treated group in both identification strategies is similar, the control groups are different.

[18]Recall that these price indices are calculated by the National Statistics Agency and are only available for major cities.

living within 5 km of a city. Column 1 uses the same measure of real income as in the baseline results (i.e. using poverty line as a price deflator). In contrast, column 2 uses the official consumer price index of each city as a price deflator. Columns 3 and 4 expand the definition of city's vicinity to the province and the area within 100 km of the city, respectively.

Note that in all cases, the estimates of β are positive and significant, suggesting that the expansion of the mine is associated with the increase of real income in areas close to Cajamarca city relative to other surrounding cities. Note that the point estimates have values similar to those obtained in the baseline results.

Table 8: Comparing Cajamarca city to other cities

	Ln(real income per capita)			
	(1)	(2)	(3)	(4)
Demand for local inputs	0.177***	0.175**	0.101**	0.136***
× Cajamarca vicinity	(0.054)	(0.054)	(0.047)	(0.049)
Definition of city's vicinity	≤ 5 km of city	≤ 5 km of city	City's province	≤ 100 km of city
Price deflator	Poverty line	Consumer Price Index	Poverty line	Poverty line
Observations	4,439	4,439	6,767	14,170
R-squared	0.475	0.487	0.504	0.555

Notes: Robust standard errors in parentheses. Standard errors are clustered at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include year and district fixed effects and the same control variables of the baseline regression. *Cajamarca vicinity* is a dummy equal to 1 if household lives in Cajamarca's vicinity and 0 if in the vicinity of other cities such as Chachapoyas, Chiclayo or Trujillo. Column 2 uses the official consumer price index of each city to calculate real income. Columns 1, 3 and 4 use the poverty line as price deflator, as in the baseline specification.

B.4 Mean comparison and main results collapsing data at district level

Table 9: Mean comparison between areas close and far from Cajamarca city

Variable	Mean		Mean
	Far	Close	difference
	(1)	(2)	(1)-(2)
Altitude (masl)	2379.2	2629.9	-250.8
	(88.4)	(70.0)	(127.8)*
Area (km2)	289.2	264.5	24.6
	(32.8)	(30.6)	(49.0)
Population	8332.6	15513.8	-7181.1
	(1095.4)	(3467.3)	(3000.6)***
Population density	38.5	63.0	-24.5
	(4.0)	(9.2)	(8.8)***
% urban	24.0	16.5	7.6
	(2.7)	(2.8)	(4.1)*
% farmers	79.4	79.8	-0.3
	(2.9)	(3.2)	(4.5)
% female HH heads	16.6	18.7	-2.1
	(1.3)	(1.7)	(2.1)
Age of HH head	48.6	50.1	-1.5
	(0.6)	(0.7)	(0.9)
Years of education of HH head	5.1	4.4	0.7
	(0.2)	(0.2)	(0.3)**

Notes: Standard errors in parenthesis. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. HH stands for household head. Data aggregated at district level, N=102.

Table 10: Effect of Yanacocha's expansion on real income, consumption and poverty - annual data collapsed at district level

	Ln(real income			Ln(household consumption)	Poor	Ln(real income)
	(1)	(2)	(3)	(4)	(5)	(6)
Mine activity × distance < 100km	0.114**	0.143***	0.201***	0.130**	-0.085***	
	(0.048)	(0.050)	(0.068)	(0.029)	(0.032)	
Mine activity × continuous distance						-0.089**
						(0.038)
Measure of mine activity	Demand for local inputs	Nr. total workers	Gold production	Demand for local inputs		
Observations	522	522	522	522	522	522
R-squared	0.840	0.841	0.842	0.801	0.734	0.839

Notes: Robust standard errors in parentheses. Standard errors are adjusted for clustering at district level. * denotes significant at 10%, ** significant at 5% and *** significant at 1%. All regressions include year and district fixed effects. The annual data is collapsed at district level by taking weighted averages. The weights are the sample inflation factors. The regression includes as controls the weighted average of the same control variables used in the baseline regression. Continuous outcome variables, such as log of income or consumption, are collapsed by taking the log of the district's weighted average.

References

Caselli, Francesco, and Guy Michaels. 2009. "Do Oil Windfalls Improve Living Standards? Evidence from Brazil." National Bureau of Economic Research Working Paper 15550.

Instituto Nacional de Estadística e Informática (INEI). 1993. "Censos Nacionales IX de Población y IV de Vivienda." http://iinei.inei.gob.pe/iinei/RedatamCpv1993.asp?ori=C (accessed February 2008).

Instituto Nacional de Estadística e Informática (INEI). 2010. "Perú: Perfil de la Pobreza por Departamentos 2005-2009." http://www.inei.gob.pe/biblioineipub/bancopub/Est/Lib0906/libro.pdf.

Instituto Nacional de Estadística e Informática (INEI). n.d.. "Sistema de Información Regional para la Toma de Decisiones- SIRTOD." http://webinei.inei.gob.pe:8080/SIRTOD/ (accessed January 15 2011).

Minera Yanacocha. 2006. "Social and Environmental Balance." http://www.yanacocha.com.pe/publicaciones/descar/pdf/balance_2006_e.pdf.

Ministerio de Economía y Finanzas (MEF). 1997-2006*a*. "Consulta de Transferencias a los Gobiernos Nacional, Locales y Regional." http://ofi.mef.gob.pe/transferencias/gl/default.aspx and http://ofi.mef.gob.pe/transferencias/gl/default_hist.aspx (accessed November 2008).

Ministerio de Economía y Finanzas (MEF). 1998-2006*b*. "Sistema de Integración Contable de la Nación - SICON."

Pascó-Font, Alberto, Alejandro Diez Hurtado, Gerardo Damonte, Ricardo Fort, and Guillermo Salas. 2001. "Chapter 4. Peru: Learning by Doing." In *Large Mines and the Community: Socioeconomic and Environmental Effects in Latin America, Canada and Spain.* , ed. Gary McMahon and Felix Remy. Washington D.C.: The World Bank.