Online Appendix

"Good mine, bad mine: Natural resource heterogeneity and Dutch disease in Indonesia"

Paul Pelzl and Steven Poelhekke

Table of Contents

OA1 Additional summary statistics and results

OA1.1 Resources summary statistics by province (Table OA1)

Table OA1: Additional summary statistics: Natural resource endowments and mining techniques by province

Province	Districts	Mining Districts	Oil & Gas Districts	Mineral Ore Resources / Area, 1990	Mining Techniques	Minerals	Oil & Gas Production / Area, 1990
Bali	8	0	0	0			0
Bengkulu	4	1	0	0.41	UG,OP	Gold, Silver	0
Central Java	35	2	0	0.64	OP	Iron Ore, Manganese	0
Central Kalimantan	6	3	0	0.91	OP,Placer	Aluminum, Gold, Silver, Zirconium	0
Central Sulawesi	4	1	0	0.61	OP	Copper	0
Dista Aceh	10	1	2	27.37	OP	Copper	23.28
DI Yogyakarta	5	1	0	227.95	OP,Placer	Iron Ore	0
DKI Jakarta	5	0	1	0			197.12
East Java	37	1	3	0.10	OP	Iron Ore	0.40
East Kalimantan	6	3	4	152.31	UG,OP	Coal, Gold, Silver	335.86
East Nusa Tenggara	12	0	0	0			0
Irian Jaya (Papua)	9	2	1	41.85	UG,OP	Copper, Gold, Nickel, Silver	0.21
Jambi	6	0	3	0			1.58
Lampung	4	1	1	0.09	UG,OP	Gold	26.08
Maluku	4	2	1	8.21	OP	Copper, Gold, Nickel, Silver	0.06
North Sulawesi	6	3	0	43.86	OP	Copper, Gold	0
North Sumatra	17	0	2	0			0.44
Riau	7	2	5	5.04	OP	Coal, Tin	10.40
South Kalimantan	10	3	1	591.34	OP,Placer	Coal, Diamonds	0.23
South Sulawesi	23	1	0	8.19	OP	Nickel, Cobalt	0
South Sumatra	9	3	5	286.68	UG,OP	Coal, Gold, Silver, Tin	127.73
Southeast Sulawesi	4	1	0	2.00	OP	Nickel	0
West Java	24	3	5	1.21	UG,OP	Gold, Manganese, Silver	245.13
West Kalimantan	7	3	0	5.59	UG,OP	Aluminum, Uranium	0
West Nusa Tenggara	6	1	0	205.50	OP	Gold, Copper	0
West Sumatra	13	1	3	6.61	UG,OP	Coal	20.17
TOTAL	281	39	37				

Note: *Mining Districts* are those with positive 1990 mineral resources; *Oil & Gas Districts* are those with positive oil and/or gas production around the year 1990. *Mineral Ore Resources / Area, 1990* indicates mineral ore resources per province in thousand tons per square mile in 1990. *Mining Techniques* indicates the extraction technique(s) that were applied on mineral resources under development as of 1990 or techniques planned to be applied on resources that had been discovered but were not yet being developed in 1990. *UG* stands for underground, *OP* for open-pit and *Placer* for Placer mining. *Minerals* are those that were extracted, or planned to be extracted from discovered deposits, as of 1990. *Oil & Gas Production / Area, 1990* indicates the production of oil & gas per square mile in terms of barrels of oil equivalent, around the year 1990. The indicated number of mining districts is 39 instead of 40 since we are forced to treat Bangka and Belitung as one district; see Section OA2.1.

OA1.2 World-price elasticity of mining output and mining wages in boom times (Table OA2)

Table OA2 analyzes how the Indonesian mining sector responds to a change in global mineral prices. In Panel A we study the mine-level elasticity of ore production to a change in the price of the produced mineral(s), using annual production figures over 1990-2009 for a sub-sample of Indonesian mines for which data are available in the *RMD* database.[38] We regress the log of one plus ore production in megatons on the log price of the produced mineral(s), control for mine and year fixed effects and cluster standard errors at the district level.[39] The results show that mines significantly increase ore production by 33% when the price of the produced mineral(s) rises by 100 log points. Columns 2-4 show that this positive effect holds across different extraction techniques that imply different labor intensities and for both privately- and state-owned mines.

In Panel B we use annual data on wages paid by the Indonesian mining sector from the labor force survey SAKERNAS (see Section OA2.5 for details) to show that mining wages respond to mining booms. We regress the change in the log of the typical monthly wage received by a mining worker on the mining boom measure used in our main analysis, as well as our labor-intensity interactions. We control for year fixed effects, include district dummies to capture linear district-level trends, and cluster standard errors at the district level. The sample period is 2007-2015 since data are representative at the district level from 2007 onwards, and we have data until 2015. Column 2 shows that labor-intensive mines (which use underground techniques) significantly raise wages during a mining boom. In contrast, the baseline of open-pit mines, which are rather capital intensive in production, do not raise wages to increase production. Column 3 shows that the wage effect is strongest for districts in which all mines exclusively apply labor-intensive (underground) methods. Given that for the average pure labor-intensive mining district *Mineral Resources 1990* $= 0.018$, the reported marginal effect implies that mines in such districts raise the wage by $0.018 \times 12.711 \approx 23\%$ as the price of local minerals rises by 100 log points.

[38] For a given mine, we only consider years that lie within the opening year and closing year (including those years) as reported in our data. We disregard mines that do not produce at all within this period.

[39] Table OA9 reports descriptive statistics. We apply a within-transformation rather than take first differences to avoid losing observations, given that the panel is unbalanced. We normalize prices by the 1990-price before taking logs. For mines that produce multiple minerals we use a weighted price, using a mineral's share in total resources as weight; see Section OA2.1 for details.

Table OA2: Response of mining sector to changes in mineral prices

Panel A

Dependent variable →	ln(Ore Production + 1)			
Unit of observation →	Mine-year			
	(1)	(2)	(3)	(4)
ln(Price of produced mineral)	0.335***	0.337***	0.346***	0.292***
	(0.093)	(0.093)	(0.099)	(0.096)
ln(Price of produced mineral) × Labor-intensive mining		-0.190		
		(0.168)		
ln(Price of produced mineral) × Pure labor-intensive mining			0.024	
			(0.135)	
ln(Price of produced mineral) × Mixed labor-intensive mining			-0.300	
			(0.266)	
ln(Price of produced mineral) × State-owned mine				0.063
				(0.108)
Year FE	Yes	Yes	Yes	Yes
Mine FE	Yes	Yes	Yes	Yes
Observations	533	533	533	533
# Districts	29	29	29	29
# Mines	73	73	73	73

Panel B

Dependent variable →	Δln(Mining Wage)		
Unit of observation →	District-year		
	(1)	(2)	(3)
Mineral Resources 1990 × Δ ln(Mineral Price)	0.003	-0.019	-0.019
	(0.019)	(0.015)	(0.015)
Mineral Resources 1990 × Δ ln(Mineral Price) × Labor-intensive Mining		0.066***	
		(0.013)	
Mineral Resources 1990 × Δ ln(Mineral Price) × Pure labor-intensive Mining			12.730***
			(3.388)
Mineral Resources 1990 × Δ ln(Mineral Price) × Mixed labor-intensive Mining			0.066***
			(0.013)
Year FE	Yes	Yes	Yes
Linear district trends	Yes	Yes	Yes
Observations	713	713	713
# Districts	119	119	119
Marginal effect of labor-intensive mining boom		0.046***	
		(0.008)	
Marginal effect of pure labor-intensive mining boom			12.711***
			(3.390)
Marginal effect of mixed labor-intensive mining boom			0.047***
			(0.008)

Note: In Panel A we regress the log of one plus mine-level annual mineral ore production in megatons on the log normalized price (P_{1990}=100) of the produced mineral(s), between 1990 and 2009. Production data are from the *RMD* database. In Panel B the dependent variable is the change in the log of annual district-specific wages paid by the mining sector, from the labor force survey SAKERNAS, observed between 2007 and 2015. *Pure labor-intensive mining* is a dummy that equals one if resources are mined with underground techniques only. *Mixed labor-intensive mining* equals one if both underground and open-pit mining are applied. *State-owned mine* is a dummy that equals one for mines in which ownership of the state of Indonesia exceeds 50%. The difference-in-difference specification in Panel B absorbs district fixed effects. Standard errors in parentheses are clustered at the district level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

OA1.3 Labor intensity of resource extraction techniques (Tables OA3 and OA4)

Evidence using labor force survey data

We use data from Indonesia's labor force survey (SAKERNAS) and data on district population to further test the relative labor intensity of different mining techniques, and show that underground mining is more labor intensive than other types of mining, not only globally but also in Indonesia.

We use SAKERNAS data to infer the number of mining workers in each district-year between 2007 and 2015.[40] We pool the annual data, use district-year-specific log mining employment as dependent variable and regress it on mining technique dummies. In order to compare districts with similarly-sized mines, we also control for the district's total 1990 mineral resources, scaled by its average across districts with positive mineral endowment (but not by district size). We further include year fixed effects and cluster standard errors at the district level. The results in column 1 of Table OA3 suggest that the number of mining workers in underground mining districts is 114% larger than in non-underground mining districts with the same mining intensity. Column 2 shows that this result is driven by the districts in which only underground mining is used in all deposits, rather than the districts in which both underground and open-pit mining occur. In column 3 we replace year fixed effects by province-year fixed effects to account for differential regional wages and other labor-market characteristics. The coefficient on underground and open-pit mining is now statistically significant (and remains positive), but the ranking in terms of labor intensity is preserved.

In column 4 we test our hypothesis that oil & gas extraction is less labor intensive than mining. The dependent variable is the log sum of the number of mining *and oil & gas* workers in a given district. We pool the annual data between 2007 and 2015, include the district's total 1990 mineral resources, and additionally include its oil & gas production around 1990. Both variables are scaled by their respective average across districts with positive realizations (but not by district size). The results suggest that a district with two times the average 1990 mineral resources employs 39% more mining and oil & gas workers than a district with average 1990 mineral resources. In contrast, a district with two times the average 1990 oil & gas production employs only 7% more mining and oil & gas workers. This smaller coefficient cannot be explained

[40] The computation also involves population data from IPUMS International (see further below). See Section OA2.5 for details. For very few district-years, SAKERNAS does not report data.

by a difference in the overall relevance of mining compared to oil & gas extraction: an inspection of Indonesia's national accounts reveals that the average mining district only contributed 5% more to total GDP than the average oil & gas district over 2007-2014.[41] This corroborates that oil & gas extraction is less labor intensive than mining.

Table OA3: Indonesian district-level evidence on the labor intensity of extraction methods

Dependent variable →	ln(# District Mining Workers)			ln(# District Mining and Oil & Gas Workers)
	(1)	(2)	(3)	(4)
Underground Mining	1.143**			
	(0.473)			
100% Underground Mining		2.344***	1.953***	
		(0.283)	(0.278)	
Underground & Open-Pit Mining		0.242	1.212*	
		(0.548)	(0.640)	
Total Mineral Resources 1990	0.288**	0.381***	0.169	0.387***
	(0.133)	(0.132)	(0.121)	(0.086)
Total Oil&Gas Production ∼1990				0.072***
				(0.018)
Year FE	Yes	Yes	No	Yes
Province-Year FE	No	No	Yes	No
Observations	1,207	1,207	1,196	1,484
# Districts	247	247	243	262
p-value F-test $\beta_{(100\%\ UG\ Mining)} = \beta_{(UG\&OP\ Mining)}$		0.001	0.290	
p-value F-test $\beta_{(MinRes90)} = \beta_{(Oil\&Gas\ Prod.\ \sim 90)}$				0.002

Note: Data for the dependent variables come from Indonesia's labor force survey data (SAKERNAS) combined with population data, over 2007-2015. The unit of observation is a district-year. In columns 1-3 the dependent variable is the log of an approximation of the number of mining workers; in column 4 it is the log of an approximation of the number of mining and oil & gas workers. *Underground Mining* is a dummy equal to one if at least one of the 1990 deposits in the district is operated or planned to be operated by underground mining. *100% Underground Mining* is a dummy equal to one if all 1990 deposits are operated or planned to be operated by underground mining only. *Underground & Open-Pit Mining* is a dummy that equals one if both underground and open-pit mining are applied or planned to be applied to extract the district's 1990 mineral resources. *Total Mineral Resources 1990* equals mineral ore resources as of 1990 scaled by its mean across all districts with positive resources (but not by the district's surface area). *Total Oil&Gas Production ∼1990* equals the production of barrels of oil equivalent around the year 1990, scaled by its mean across producing districts (but not by the district's surface area). Standard errors in parentheses are clustered at the district level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

[41] Source: Indonesia's national statistical agency *Badan Pusat Statistik* (BPS).

Evidence using population data

An analysis of district-specific (working-age) population data over time offers another opportunity to test whether underground mining is more labor intensive than open-pit mining. The underlying idea is that if indeed underground mining is more labor intensive, we would also expect a stronger labor force response to a booming mining sector that employs more labor, relative to other mining districts. Analyzing population data also sheds light on the overall degree of labor mobility in Indonesia following local mining booms. If labor mobility is high, then there is less scope for upward wage pressure during a boom.

District-level population over time is available from IPUMS International.[42] This source provides the micro-data from the Indonesian population census of 1990, 2000 and 2010, as well as the 1995 and 2005 SUPAS intercensal population surveys, all collected by the *BPS*.[43] We use these data to compute total population and working-age population (age 15-65) for the 1990-districts. Since population data are only collected every five years we study the change in log population during four periods: 1990-1995, 1995-2000, 2000-2005, and 2005-2010. In columns 1-2 of Table OA4 we focus on total population while in column 3 we look at working-age population, which is unaffected by changes in fertility and less affected by changes in mortality. We regress the dependent variables on the mining boom measure and the labor-intensity interactions, with the difference that the change in mineral prices is computed as simple average across the five annual price changes. We control for year fixed effects, initial (working-age) population, differential trends across districts with varying mining intensity and districts with varying oil & gas intensity, and (in columns 2-3) differential trends across districts with varying labor intensity in the mining sector. Standard errors are clustered at the district level.

The results suggest that while mining booms spur immigration overall, labor mobility during mining booms clearly depends on local extraction methods. When mining is more capital

42 See: `https://international.ipums.org/international/`

43 While annual population data would be preferred and is also reported by the World Bank's *Indonesia Database for Policy and Economic Research* (INDO DAPOER), these data appear unreliable since they are derived using predicted trends in fertility, mortality and migration between provinces and are not corrected ex-post using census or intercensal data. The IPUMS data misses population figures for Aceh in 2005 since no intercensal population survey was held in this province due to the Indian Ocean tsunami in 2004. For 1995, data are missing for 12 provinces: South Kalimantan (includes 3 districts with positive 1990 mineral resources), West Kalimantan (3), East Kalimantan (3), Central Kalimantan (3), South Sulawesi (1), Central Sulawesi (2), Southeast Sulawesi (1), North Sulawesi (3), Irian Jaya (now Papua) (2), and Maluku (2). For 1990, population data are missing for one district.

intensive, booms do not affect population, and nor do oil & gas booms. When mining is more labor intensive, population significantly increases in boom times, although the magnitude of the estimates suggests that labor mobility across districts as a response to mining booms is not large. Specifically, the marginal effect at the bottom of column 2 indicates that if the price of local minerals rises by 100 log points in each year over a period of five years, then district population increases by 6%, in the district with average mineral resources and where underground mining occurs. Column 3 shows that working-age population rises by 4.8% during such a sustained labor-intensive mining boom. The results confirm that underground mining is more labor intensive than other mining techniques.

Table OA4: Mining booms and population growth

Dependent variable →	$\Delta_5$ln(Population)		$\Delta_5$ln(Working-age Population)
	(1)	(2)	(3)
Mineral Resources 1990 × mean[Δln(Mineral Price)]	0.040**	-0.006	-0.006
	(0.020)	(0.033)	(0.030)
Mineral Resources 1990 × mean[Δln(Mineral Price)] × L-I Mining		0.066*	0.054*
		(0.033)	(0.030)
Oil&Gas Production ∼1990 × mean[Δln(Oil Price)]	-0.027	-0.029	-0.030
	(0.038)	(0.038)	(0.042)
ln(Population 1990)	-0.026***	-0.027***	
	(0.009)	(0.009)	
ln(Working-age Population 1990)			-0.026***
			(0.009)
Observations	941	941	941
# Districts	280	280	280
Marginal effect of labor-intensive mining boom		0.060***	0.048***
		(0.017)	(0.017)

Note: Data for the dependent variables come from IPUMS International. *L-I Mining* equals *Labor-intensive mining*. *mean[Δln(Mineral Price)]* equals the simple average across the current and past four annual log price shocks. The unit of observation is a district-period, and the dependent variable is the change in log population across the periods 1990-1995, 1995-2000, 2000-2005 and 2005-2010. In columns 1-2 we analyze changes in total population, while in column 3 we focus on changes in working-age population. All specifications contain dummies for the years 2000, 2005 and 2010, and the difference-in-difference specification absorbs district fixed effects. We also include the linear trend controls *Mineral Resources 1990*, *Oil&Gas Production ∼1990* and (in columns 2-3) *Labor-intensive Mining*$_k$ and [*Mineral Resources 1990* × *Labor-intensive Mining*$_k$], but do not show their coefficients. Standard errors in parentheses are clustered at the district level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

OA1.4 Labor-intensive mining: pure underground vs. mixed mining (Table OA5)

Sections 3.2 and OA1.3 show that underground mining is more labor intensive than other methods, and that mining is most labor intensive when only underground methods are applied. In Table OA5 we repeat Table 1 using the same sample of all plants and replace the labor intensity interaction with separate interactions for pure and mixed labor-intensive mining methods. The purpose is to gauge whether this finer distinction also translates into different effects on manufacturing-plant outcomes. The first row still captures the effect of capital-intensive mining booms and is thus identical to the even-numbered columns in the first row in Table 1. We find indeed that all interaction coefficients are much larger in absolute magnitude for pure labor-intensive mining. While these coefficients seem very large, they represent the effects for an endowment of *Mineral Resources 1990* = 1, while the average pure labor-intensive mining district has *Mineral Resources 1990* = 0.018. The key take-away is that if pure underground mines were as big as the average mine, then the labor market effects would be much larger.

Table OA5: Pure labor-intensive mining versus mixed labor-intensive mining

Dependent variable →	Δln # Employees	Δln Wage Bill / #Employees	Δln Unit Price	Δln Revenue
	(1)	(2)	(3)	(4)
MinRes 1990 × Δln(Mineral Price)	0.026***	-0.003	0.032	0.062*
	(0.005)	(0.021)	(0.044)	(0.036)
MinRes 1990 × Δln(Mineral Price) × Pure L-I mining	-2.079***	1.361**	15.246***	0.985
	(0.185)	(0.536)	(2.494)	(0.716)
MinRes 1990 × Δln(Mineral Price) × Mixed L-I mining	-0.037***	0.136***	0.196***	0.158***
	(0.006)	(0.021)	(0.044)	(0.036)
Observations	261,020	260,803	148,691	261,017
Marginal effect of pure labor-intensive mining boom	-2.053***	1.357**	15.278***	1.046
	(0.185)	(0.536)	(2.494)	
Marginal effect of mixed labor-intensive mining boom	-0.011***	0.133***	0.228***	0.220***
	(0.002)	(0.004)	(0.009)	(0.005)

Note: *MinRes 1990* equals *Mineral Resources 1990. L-I mining* equals *labor-intensive mining.* All regressions control for plant fixed effects, four-digit industry-times-year fixed effects, and district-specific linear trends. *Pure labor-intensive mining* equals one for districts where only underground methods are used or planned in all deposits in 1990. *Mixed labor-intensive mining* equals one for districts where both underground and open-pit methods are used or planned in 1990. The sample contains all privately-owned manufacturing plants with 20 or more employees, over 1990-2009. The oil & gas boom measure is included but not shown. Standard errors in parentheses are clustered at the district level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

OA1.5 Mining booms and local expenditure and infrastructure (Table OA6)

In Table OA6 we study the impact of a mining boom on local public expenditure and infrastructure. In Panel A we use district-level expenditure data provided by Indonesia's Ministry of Finance. Expenditure is indicated in current million Rupiahs and reported in two main categories: routine expenditure and development expenditure. We focus on the latter as well as its three most important sub-categories from the perspective of (more-traded) manufacturing plants, namely expenditure on: "Industry Sectors"; "Trade, Regional Business Development, Regional Finance and Cooperatives"; and the "Transportation Sector". We regress the change in the log of these variables on our standard mining boom measure, control for year fixed effects and differential trends across districts with varying mining intensity, and cluster standard errors at the district level. The sample period is 2000-2004 since before 2000 the reporting period was April 1 – March 31 and after 2004 a new reporting scheme was used.[44] We restrict the sample to districts that did not split over 1990-2004 such that we can meaningfully use the variable *Mineral Resources 1990*. The results show that an increase in the price of local minerals by 100 log points in the mining district with average endowments leads to an increase in overall development expenditure by 11%, expenditure on industry sectors by 27% and expenses in the category trade, regional business development, regional finance and cooperatives by 37%.[45] These developments clearly benefit manufacturing plants, and in particular more-traded goods producers.

In Panel B we use a data set of district-level scores on the availability and quality of local infrastructure taken directly from the Indonesian Regional Autonomy Watchdog *KPPOD*. A given score partly represents the results of a survey of the local business community and partly concrete and measurable indicators; see `https://www.kppod.org/` for details. We focus on the panel dimension of the data, which is available for 2002-2004. We restrict the sample to districts that did not split over 1990-2004 such that we can meaningfully use the variable *Mineral Resources 1990*. In column 1 we study the total state of infrastructure, which captures both availability and quality, while in columns 2-3 we separate the two. We regress the change in the log of these

[44] The reporting period over 2000-2004 was January 1 – December 31. The use of the new reporting scheme became mandatory as of 2006, but districts could volunteer to use it before 2006. As a consequence, the number of districts with available data decreases over time from 2000-2005, and in fact equals zero for 2005.

[45] Since for some district-years expenditure on one or more of the sub-categories is zero, the number of observations differs across columns. The results are highly robust to accounting for zero expenditure by taking the log of one plus expenditure rather than the log of expenditure when computing the dependent variable.

variables on our mining boom measure, control for year fixed effects and differential trends across districts with varying mining intensity, and cluster standard errors at the district level. Column 1 shows that a rise in the price of local minerals by 100 log points in the mining district with average endowments leads to an improvement of local infrastructure by 5.5%. Columns 2 and 3 show that this effect is driven by an improvement in infrastructure *quality*, which is intuitive given the annual horizon of our analysis. In column 4 we look at a more specific variable which is particularly important for more-traded goods producers given the geography of Indonesia, namely the quality of the local seaport, and again find a positive effect of local mining booms.

Table OA6: Mining booms and local expenditure and infrastructure

Panel A

Dependent variable →	Δln Total Development Expenditure	Δln Dev-Ex Industry Sectors	Δln Dev-Ex TradeBusFin Sector	Δln Dev-Ex Transport Sector
	(1)	(2)	(3)	(4)
Mineral Resources 1990 × Δln(Mineral Price)	0.108***	0.274**	0.374***	0.108
	(0.037)	(0.106)	(0.111)	(0.164)
Year FE	Yes	Yes	Yes	Yes
Observations	412	363	404	411
# Districts	182	169	181	182

Panel B

Dependent variable →	Δln Infra-str.	Δln Infra: Avail.	Δln Infra: Qual.	Δln Infra:Q. Seaport
	(1)	(2)	(3)	(4)
Mineral Resources 1990 × Δln(Mineral Price)	0.055***	0.009	0.103***	0.142*
	(0.017)	(0.018)	(0.017)	(0.085)
Year FE	Yes	Yes	Yes	Yes
Observations	190	190	190	190
# Districts	117	117	117	117

Note: In Panel A we use district-level expenditure data provided by Indonesia's Ministry of Finance. The sample period is 2000-2004. *TradeBusFin* stands for Trade, Regional Business Development, Regional Finance and Cooperatives. In Panel B we use district-level data on local infrastructure from the Regional Autonomy Watchdog *KPPOD*. The sample period is 2002-2004. We always include *Mineral Resources 1990* separately to capture differential linear trends across districts with varying mining intensity, but do not show the coefficient. The difference-in-difference specifications absorb district fixed effects. Standard errors in parentheses are clustered at the district level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

OA1.6 Alternative definition of more- vs. less-traded goods producers (Table OA7)

Table OA7: Alternative definition of more- versus less-traded goods producers

Panel A

Sample →	More-Traded Goods Producers (AK Cutoff)			
Dependent variable →	Δln # Employees	Δln Wage Bill / #Employees	Δln Unit Price	Δln Revenue
	(1)	(2)	(3)	(4)
Mineral Resources 1990 × Δln(Mineral Price)	0.023***	0.020	0.034	0.085
	(0.008)	(0.044)	(0.065)	(0.054)
Mineral Resources 1990 × Δln(Mineral Price) × L-I Mining	-0.044***	-0.044	-0.001	-0.089
	(0.009)	(0.046)	(0.068)	(0.055)
Oil&Gas Production ∼1990 × Δln(Oil Price)	0.001	0.001	-0.027**	0.007
	(0.002)	(0.003)	(0.013)	(0.006)
Observations	192,765	192,597	107,764	192,762
# Plants	31,527	31,514	22,225	31,527
# Districts	265	265	260	265
Marginal effect of labor-intensive mining boom	-0.021***	-0.024	0.033	-0.004
	(0.005)			(0.013)

Panel B

Sample →	Less-Traded Goods Producers (AK Cutoff)			
Dependent variable →	Δln # Employees	Δln Wage Bill / #Employees	Δln Unit Price	Δln Revenue
	(1)	(2)	(3)	(4)
Mineral Resources 1990 × Δln(Mineral Price)	0.036***	-0.070	0.033	-0.010
	(0.007)	(0.050)	(0.034)	(0.026)
Mineral Resources 1990 × Δln(Mineral Price) × L-I Mining	-0.037***	0.272***	0.279***	0.333***
	(0.007)	(0.050)	(0.036)	(0.027)
Oil&Gas Production ∼1990 × Δln(Oil Price)	0.001	-0.009***	0.035*	0.014
	(0.001)	(0.003)	(0.020)	(0.013)
Observations	68,223	68,175	38,838	68,223
# Plants	12,050	12,048	8,535	12,050
# Districts	267	267	260	267
Marginal effect of labor-intensive mining boom	-0.000	0.202***	0.312***	0.323***
	(0.002)	(0.006)	(0.015)	(0.008)

Note: In this table we re-define more- versus less-traded goods producers and thereby perform a robustness check on Table 2. We define more-traded goods producers as plants that export in at least one year over 1990-2009 and/or are plants that belong to a four-digit sector with a distance elasticity below 0.8. This cutoff is used by Allcott and Keniston (2018) and corresponds to an average shipment distance of around 500 miles. Standard errors in parentheses are clustered at the district level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

OA1.7 Accounting for outliers (Table OA8)

In this table we perform a robustness check on Table 1 by winsorizing the dependent variables at the 1% level to account for possible outliers. All results remain robust.

Table OA8: Robustness check: Winsorizing the dependent variables

Sample →	All Plants							
Dependent variable →	Δln # Employees		Δln Wage Bill / #Employees		Δln Unit Price		Δln Revenue	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mineral Resources 1990 × Δ ln(Mineral Price)	0.014	0.024***	0.033	0.002	0.095*	0.036	0.095***	0.060*
	(0.009)	(0.003)	(0.027)	(0.022)	(0.054)	(0.040)	(0.033)	(0.033)
Mineral Resources 1990 × Δ ln(Mineral Price) × L-I Mining		-0.034***		0.104***		0.193***		0.115***
		(0.003)		(0.022)		(0.041)		(0.033)
Oil&Gas Production ∼1990 × Δ ln(Oil Price)	0.000	0.000	-0.002	-0.002	-0.008	-0.009	0.005	0.005
	(0.001)	(0.001)	(0.003)	(0.003)	(0.012)	(0.012)	(0.007)	(0.007)
Observations	261,020	261,020	260,803	260,803	148,691	148,691	261,017	261,017
# Plants	42,210	42,210	42,196	42,196	30,055	30,055	42,210	42,210
# Districts	274	274	274	274	272	272	274	274
Marginal effect of labor-intensive mining boom		-0.010***		0.106***		0.230***		0.176***
		(0.002)		(0.004)		(0.009)		(0.006)

Note: Dependent variables have been wisorized at the 1% level. Standard errors in parentheses are clustered at the district level. ***Significant at 1% level; **Significant at 5% level; *Significant at 10% level.

OA1.8 Summary statistics for variables only used in Online Appendix (Table OA9)

Table OA9: Summary statistics for variables only used in Online Appendix

Variable	Sample is districts with:	Mean	p(50)	s.d.	Min	Max	N (non-missing)
District data							
Total Mineral Resources 1990	MRes90>0	1	0.048	2.103	0.000	9.601	39
	MRes90>0, L-I Mining	2.308	0.017	3.632	0.001	9.601	9
Total Oil&Gas Production ∼1990	O&G Prod∼90 >0	1	0.013	4.204	0.000	25.717	37
District-year data							
ln(Mining Workers)	All	7.446	7.301	1.584	3.503	12.104	1,207
ln(Mining and Oil&Gas Workers)	All	7.512	7.432	1.549	3.553	12.104	1,484
$\Delta_5$ln(Population)	All	0.068	0.057	0.160	-2.224	1.088	941
	MRes90>0	0.105	0.097	0.116	-0.161	0.690	109
	MRes90>0, L-I Mining	0.115	0.092	0.156	-0.161	0.690	30
Δln(Mining Wage)	All	0.086	0.075	0.594	-2.646	2.944	713
Δln(Development Expenditure)	All	0.566	0.450	0.865	-3.128	7.762	412
Δln(Dev-Ex on Industry Sectors)	All	0.537	0.420	1.353	-4.535	8.550	363
Δln(Dev-Ex on TradeBusFin Sector)	All	0.745	0.607	1.228	-2.482	9.651	404
Δln(Dev-Ex on Transport Sector)	All	0.514	0.417	1.100	-4.080	8.464	411
Δln(Infrastructure)	All	0.068	0.065	0.340	-0.820	0.799	190
Δln(Infra: Availability)	All	0.057	0.067	0.369	-0.928	1.112	190
Δln(Infra: Quality)	All	0.092	0.063	0.412	-0.841	1.156	190
Δln(Infra: Quality of Seaport)	All	0.151	0	0.698	-1.642	2.335	190
Mine-year data							
ln(Ore Production + 1)	n/a	1.599	1.232	1.293	0	5.017	533

Note: This table provides summary statistics for variables that are only used in the Online Appendix. *Total Mineral Resources 1990* equals mineral ore resources in 1990 scaled by its mean across districts with positive resources (but not by the district's surface area). *Total Oil&Gas Production ∼1990* equals the production of barrels of oil equivalent around 1990 scaled by its mean across districts with positive production (but not by surface area). *MRes90>0* holds for districts with positive mineral resources in 1990. *L-I Mining* stands for *Labor-intensive Mining*. *O&G Prod∼90>0* holds for districts producing oil & gas around 1990. *TradeBusFin* stands for Trade, Regional Business Development, Regional Finance and Cooperatives.

OA2 Online Data Appendix

OA2.1 Mining

Combining RMD *and* MinEx *data*

The data sources we use to compute district-specific mineral resources as of 1990 are *Raw Materials Data* (RMD) and *MinEx Consulting* (MinEx). Both data sets claim full coverage, and the majority of deposits are indeed listed in both. We double-checked the reported deposits with public data from the *Mineral Resources Data System* (MRDS) of the *United States Geological Survey* (USGS), which however lists fewer deposits. To build a complete data set we match deposits across sources using a deposit's name. For each unmatched deposit, we use additional variables such as location and ore resources to verify if it corresponds to a deposit in the other data set. We identify 82 mineral deposits with positive mineral resources in 1990. 49 of these are listed in both sources, while the remaining 33 are only listed in one. These 33 deposits have statistically significantly lower 1990 mineral resources than the deposits listed in both data sets. 24 of the 33 deposits are unique to *MinEx* and nine are unique to *RMD*. For matched deposits we use the *MinEx* data because we are more confident about its accuracy, based on a test in Google Earth revealing that the *MinEx* location data are more precise.

Location of mineral deposits

Both *RMD* and *MinEx* report the location of a deposit in terms of latitude and longitude. For the set of deposits that are operated by a mine over our sample period and for which different latitude and longitude data are reported by *MinEx* and *RMD*, we entered the location data into Google Earth and regard the location displaying a mine as the correct one. For three deposits, our sources do not provide location data; we retrieved these via Internet search (sources are available on request). Using latitude and longitude, we identify the home district of the deposit as of 2016 using Google Maps. We then identify the corresponding 1990-district, using district proliferation tables provided by the *BPS* and information provided by Bazzi and Gudgeon (2020).

Time of discovery of deposits

Only *MinEx* reports the year of discovery, which refers to "when the deposit was recognized as having significant value". Data are missing for around one third of deposits. Since we are only interested in whether the discovery took place before 1990, for several of these deposits we use the fact that production started before 1990. For all remaining deposits we carried out an Internet search. We found the discovery year for 42 deposits, mostly through annual reports of the companies operating the deposits or via mining information websites such as *mining-atlas.com*.[46] For some remaining deposits, we infer that the discovery took place after 1990 if in 2016 (the vintage of the *MinEx* data) the deposit's status is either "Advanced Exploration", "Emerging Project" or different categories containing the term "Feasibility Study". For all deposits that are only listed in the *RMD* data, we also use the pre-1990 production start-up rule, Internet search (23 deposits) and the deposit's status to infer the discovery date, in this order. Regarding the deposit status, we infer that the discovery (if at all) took place *after* the most recent year for which the deposit's status is either "Project, no specification", "Conceptual", "Feasibility", "Prefeasibility", "Abandoned Project" or "Abandoned". For the remaining deposits from both data sets with missing discovery date, we infer the year of discovery as the year of production start-up minus the median difference between discovery year and production start-up year across all deposits for which we have information on both variables, which is eight years.[47]

Inferring missing ore resources data

Ore resources data are missing for some deposits. We infer ore resources as ore *reserves* times the mineral-specific average ratio of resources and reserves.[48] In case there is no other deposit of the same mineral with non-missing resources and reserves data, we infer resources as reserves times the average ratio of resources and reserves across all deposits and minerals. If both reserves and resources data are missing for a given deposit, we retrieve data using Internet search. There are no deposits that were discovered by 1990 for which we were unable to retrieve resources data.

Ore reserves and resources data are missing for all tin deposits in both *RMD* and *MinEx*. We

[46] For some deposits, we proxy discovery with the year of establishment of the company (or branch) which operated the deposit, if the name of the company or branch contains the name of the deposit. Since for all these deposits that year is after 1990, this turns out to be equivalent to dropping the deposits from our sample.

[47] We drop one single (small) deposit from our sample for which neither the discovery year nor the production start-up year is reported.

[48] Resources are "the concentration or occurrence of material of intrinsic economic interest in or on the Earth's crust in such form and quantity that there are reasonable prospects for eventual economic extraction" (Raw Materials Data Handbook, p.57). Reserves are defined as "the economically mineable part of a measured or indicated mineral resource" (p.58). The ratios of resources and reserves are obtained from *RMD*, since *MinEx* only reports ore resources.

retrieved the missing data via Internet search. Since we could not obtain resources data at the deposit level, we use resources data of state-owned operator *PT Timah*, which virtually has a monopoly on tin mining in Indonesia. Total tin resources of *PT Timah* equaled 1.06 megatons of tin in 2008, according to the 2009 PT Timah annual report. We were unable to retrieve ore resources data for an earlier year. In order to infer Indonesian 1990 tin resources, we add total tin production over 1990-2008 to the 2008 figure, using annual production data of all Indonesian tin mines from *Indonesia's Department of Mines and Energy*, which is made available by the *U.S. Bureau of Mines*. Since *RMD* and *MinEx* do not contain grade information for Indonesian tin deposits, we convert the resulting number to tons of *ore* rather than tons of tin using the average ratio provided by different sources. Specifically, according to *earthsci.org*, "Indonesia produces tin mainly from alluvial deposits" (`http://earthsci.org/mineral/mindep/tin/tin.html`), and the ratio of ore and tin from alluvial deposits ranges between 0.01 and 0.015 per cent across different sources; we thus infer a ratio of 0.0125 for our analysis.

Since *PT Timah* annual reports do not indicate the spatial distribution of tin resources across Indonesia, we infer the share of the different 1990-districts using annual production data from *Indonesia's Department of Mines and Energy*. While data on annual aggregate tin production in Indonesia are available from 1949-2008, data at the sub-national level are only available for the period 1978-1988 (Wu, 1989), thus we compute the production shares using the data from this period. Since with these data we cannot attribute tin deposits that are located in the districts Bangka and Belitung to either of the two districts, we treat these two 1990-districts as one district in our analysis. 91% of Indonesian tin production took place in deposits located in the Bangka-Belitung archipelago between 1978-1988. We thus infer the tin resources of Bangka-Belitung as this percentage times our measure of total tin resources as of 1990. The remaining 9% of tin production over 1978-1988 took place in deposits in the Riau archipelago; we thus inferred 1990 tin resources of the 1990-district Riau as 9% of total 1990 tin resources.

Computation of district-specific 1990 ore resources

With the exception of tin, we first compute mineral ore resources as of 1990 for each deposit. We then sum 1990 resources across all deposits in a district.

If a deposit was discovered before 1990 but did not start production before that year, the deposit's 1990 resources equal its initial resources. If a deposit was operated by a mine before

1990, we deduct the mine's pre-1990 ore production from the initial resources. For all deposits contained in *RMD*, this is done using annual production data whenever available. Since *MinEx* does not report production data, for all deposits unique to *MinEx* annual production data are unknown. For these deposits we infer total production before 1990 as *average* annual production times the number of production years before 1990, both of which can be inferred using *MinEx*.[49]

In the *RMD* data, for some deposits pre-1990 production is only reported in terms of metal rather than ore. In these cases we compute the average ratio of ore and metal production of the specific deposit and metal for each year in which both are available, and use this ratio to infer pre-1990 ore production. If ore production is not available for any year, we use the mine- and metal-specific *grade* to infer ore production from metal production. If the grade is not reported, we retrieve it via Internet search. For five deposits in which production started before 1990, pre-1990 production data are entirely unavailable. In these cases, we infer pre-1990 production as the average yearly (post-1990) production across years in which production data are reported, multiplied by the number of pre-1990 production years. In one case we do not have any information on production, and therefore infer 1990 ore resources as initial resources.

Multi-mineral deposits

RMD reports deposits' annual production figures per extracted mineral, with maximum coverage 1975-2011. 11 deposits in our final sample (thus with positive 1990 ore resources) that are listed in *RMD* produced more than one mineral at any point in time between 1975 and 2011. These 11 deposits are spread across 11 districts. Unfortunately, we do not know the share of each mineral in total ore resources for the 11 deposits. We thus infer the share of mineral m in total resources using the average ratio of ore production of mineral m over total ore production of the respective deposit, using all years in which the deposit is operating and production data are available. When production is only reported in terms of metal rather than ore output, we infer ore production using the average mine-specific ratio of metal to ore production across all years for which the ratio can be computed, and otherwise with the use of mine-metal-specific grade data. In the 11 districts that contain at least one multi-mineral deposit with positive resources in

[49] *MinEx* reports both "initial resources", the year of production commencement and "current resources". The year as of which current resources are reported varies by deposit. We compute annual average production as the difference between initial resources and current resources, divided by the number of years between production commencement and the year in which current resources are reported.

1990, we incorporate the inferred mineral shares in multi-mineral deposits into our computation of the mineral price index (MPI) of the district.

MinEx only lists the *main* mineral of a given deposit, thus for deposits unique to *MinEx* we have to assume that the main mineral is the only mineral. Given the low occurrence of multi-mineral deposits in *RMD* and the fact that deposits only listed in *MinEx* have small ore resources, we do not expect this to affect our results.

Deposit ownership

For each deposit in our sample, we determine whether the state of Indonesia's ownership share exceeds 50% and assign the value one to the dummy *stateowned* if this condition is met. For deposits that are listed in *RMD*, time-varying data on ownership is available, but in practice there is no time variation in *stateowned* for any deposit in our sample over 1990-2009. For all deposits that are only listed in *MinEx*, we use ownership status at the time of discovery, as no other data is available. For 12 deposits data is missing in our data sources; for these we fill *stateowned* using Internet search.

OA2.2 Oil & Gas

The *Indonesia Oil and Gas Atlas* is divided into six volumes, each of which covers a certain geographic area. Specifically, these are North Sumatra and Natuna (Volume 1, 1989), Central Sumatra (Volume 2, 1991), South Sumatra (Volume 3, 1990), Java (Volume 4, 1989), Kalimantan (Volume 5, 1991) and Eastern Indonesia (Volume 6, 1988). We assign a field producing oil and/or gas to its 1990-district using data on the field's latitude and longitude provided in the data source. If a field is located offshore, we assign it to the closest district in terms of geographic distance.

OA2.3 Prices

We use prices reported by Platts Metals Week and the USGS for: copper (U.S. producer cathode, 99.99-percent-pure copper), nickel (London Metal Exchange cash price for primary nickel of minimum 99.80% purity), tin (New York composite), aluminum (99.7-percent-pure aluminum ingot, U.S. market spot price) and cobalt (99.8-percent cobalt cathode, U.S. spot price).[50] For gold and silver, we use the prices determined on the London Bullion Market, which

[50] Source: USGS. `https://pubs.usgs.gov/sir/2012/5188/tables/`

is a wholesale over-the-counter market.[51] Due to availability and data quality, the prices we use for manganese, diamonds, chromium, zirconium and uranium are those paid domestically in the United States.[52] For iron ore and coal, it is somewhat harder to identify an observed price that comes close to a single world price. For iron ore, we use the price China pays per imported metric ton on average in a given year, since China is by far the world's largest importer of iron ore and geographically close to Indonesia.[53] For coal, we use the price of Australian coal instead of other coal types, due to data quality and given that price changes are likely most aligned with Indonesian coal, since China is a key importer of both Australian and Indonesian coal.[54] For crude oil, we use the spot price of West Texas Intermediate (WTI), which is a benchmark for the prices of other crude oil sorts.[55]

OA2.4 Manufacturing Census

Data cleaning

We drop plant-years in which production worker employment is larger than total employment, as well as plant-years in which the reported number of employees is below 20.[56] We drop six plants that have a district ID that does not correspond to any of the district IDs in the *BPS* list. Around 6% of plants are reported to operate in different (two or more) 1990-districts in different years. This could be due to changes in district borders that are not explained by district splits, by the plant actually moving to another district or, arguably most likely, by measurement error. The plant fixed effects that we control for only nest district-specific fixed effects if plants are always recorded as in the same 1990-district. We therefore keep the plant's district-years of the 1990-district that is reported for the longest consecutive period.

Defining more- versus less-traded goods producers

For each of the 473 six-digit industries of the 1997 *North American Industry Classification System* (NAICS 1997), Holmes and Stevens (2014) estimate a (constant) distance elasticity, which equals

[51] Source: London Bullion Market Authority (LBMA). `https://www.quandl.com/data/LBMA`

[52] Uranium prices are from the IMF (`http://www.imf.org/external/np/res/commod/index.aspx`), all other prices from the USGS.

[53] Source: IMF.

[54] Source: IMF.

[55] Source: Energy Information Administration (EIA). `https://www.eia.gov/dnav/pet/hist/rwtcA.htm`

[56] Consistent with the plant-size threshold of 20 employees, only for a few plants the data reports less than 20 employees, which we treat as typos.

the percentage reduction in trade volume as distance increases by one percent. For this purpose the authors use data from the 1997 *U.S. Commodity Flow Survey* (CFS), which documents the destination, product classification, weight and value of a broad sample of shipments. Holmes and Stevens (2014) estimate the distance elasticity via a standard log-log specification. The higher the trade costs of a specific industry, the shorter its optimal average shipment distance (equivalently, the higher its distance adjustment). Ready-Mix Concrete (4.2), Ice (3.0) and Asphalt (2.9) have the highest estimated distance elasticity. 29 industries have an estimated distance elasticity of zero, including Semiconductors, Analytical laboratory instruments and Aircraft, in which transportation costs are very low relative to product value.

We use the estimates of Holmes and Stevens to classify Indonesian manufacturing plants into more- versus less-traded goods producers, using the four-digit sector of each plant, as defined by the 2000 version of the *Klasifikasi Baku Lapangan Usaha* (KBLI 2000). This nearly corresponds to Revision 3.1. of the *International Standard Industry Classification* (ISIC Rev.3.1), however not one-to-one. Therefore, we first use KBLI 2000 and ISIC Rev.3.1 documentation files to assign to each KBLI 2000 industry code its corresponding ISIC Rev.3.1 code. Next, we walk from ISIC Rev.3.1 to NAICS 1997 using correspondence tables provided by the *United States Census Bureau*. Since our sample contains 123 four-digit (ISIC Rev.3.1) industries, in the great majority of cases, one four-digit ISIC Rev.3.1 industry code matches with more than one NAICS 1997 code. In all these cases, we compute the ISIC-realization of the distance elasticity as the average realization across all the NAICS industries matching with the particular ISIC industry.

Defining upstream plants

We use the 2007 U.S. Input-Output tables of the *Bureau of Economic Analysis* (BEA) to identify upstream plants. These tables distinguish between more sectors than any Indonesian Input-Output table does, which thus allows a finer evaluation of an industry's linkage to the mining sector. Because formal mining is done in a very standard way across the globe, we can confidently use Input-Output tables of another country for the mining sector.

The tables distinguish between three mining industries which we together refer to as the "the mining sector": *Coal mining*; *Iron, gold, silver and other metal ore mining*; and *Copper, nickel, lead and zinc mining*. Details on the concordance of the ISIC Rev.3.1 codes inferred from the manufacturing census and the BEA codes used in the Input-Output tables are described further below. For each of the 389 industries j that are distinguished in the 2007 Input-Output tables of the BEA, we compute its 'upstreamness' to the mining sector as the ratio of the (weighted) sum of its direct and indirect sales to the mining sector (as defined above) and its total sales:

$$Upstream_{jk} = \frac{\sum_m Sales_{j,m} \times (R_{km}/R_k)}{\sum_j Sales_j} + \sum_{-J} \left[\frac{Sales_{j,-j}}{\sum_j Sales_j} \times \frac{\sum_m Sales_{-j,m} \times (R_{km}/R_k)}{\sum_j Sales_{-j,j}} \right] \in [0,1]$$

where $-J$ denotes the set of all industries apart from j; k is the district identifier as usual; and m={*Coal mining; Iron, gold, silver and other metal ore mining; Copper, nickel, lead and zinc mining*}. R_{km} equals the total 1990 resources of the minerals contained in group m in district k and R_k equals the total 1990 mineral resources in district k. $Upstream_{jk}$ takes into account which minerals are found locally, which makes it industry- and district-specific rather than only industry-specific. For example, if industry j is only upstream to the coal mining sector and there are no coal deposits but only gold deposits in 1990 in district k, then we do not classify plants in industry j in district k as upstream ($Upstream_{jk} = 0$). The reasoning behind this choice is that in our empirical analysis, we try to test whether any effect of a local mining boom is driven by plants that are upstream to the *local* mining sector. Using our previous example, we do not expect plants that sell to the coal sector to benefit or suffer more from a gold boom in their home district than plants in the same district that do not sell to any of the three mining sectors, since neither group of plants sells to the sector *Iron, gold, silver and other metal ore mining.* On the other hand, if coal deposits were present in district k, then the plants selling to the coal sector might perform differently, and more so if the coal mining sector is in district k is more important.

We first walk from the BEA Input-Output table codes to the 2002 NAICS codes, and then match those with the ISIC Rev.3.1 codes, using correspondence tables provided by the *United States Census Bureau.* In the census data, 133 four-digit ISIC Rev.3.1 manufacturing industries are represented, while the BEA tables report 389 industries. As a consequence, in the great majority of cases, one four-digit ISIC industry code matches with more than one BEA code. In all these cases, we compute the realization of $Upstream_{jk}$ as the average realization across all the BEA industries matching with the particular ISIC code. We argue that the inferred value provides a reasonable approximation, since the realizations of $Upstream_{jk}$ are very similar across BEA codes that match with the same ISIC code.

Total Factor Productivity (TFP)

The calulation of TFP is based on the method by De Loecker and Warzynski (2012) and Ackerberg et al. (2006). First, a separate translog production function for each two-digit ISIC sector is estimated, relating the log value added to (the log of) capital, labor, and materials (including squared terms and all interactions) and year and four-digit-ISIC-industry fixed effects. Input coefficients are allowed to vary by exporter and foreign ownership status. Demand for materials proxies for unobservable productivity shocks. This yields expected industry-level output, which then results in plant-year level deviations from expected output. In the second step, these are regressed using GMM on its lag, capital and labor input where current labor is instrumented with lagged labor as suggested by Ackerberg et al. (2006). Finally, the innovations of this regression capture TFP. Value added equals output net of inputs of material and energy. Capital is proxied with fixed assets, labor with the number of employees. All variables are expressed in Indonesian rupiahs, deflated using five-digit industry producer price indices.

OA2.5 SAKERNAS Labor Force Survey

We use the August round of the survey for the years 2007-2015, because these are representative at the district level, unlike other rounds or the years 1976-2006. This sample of SAKERNAS covers all 1990-districts except in the years 2013 (five districts missing) and 2015 (one district missing), and includes data on between 490,468 (in 2014) and 953,172 (in 2010) individuals.[57] This implies a coverage of between 0.2 and 0.4%.

In Panel B of Table OA2 we use annual district-level mining wage data from SAKERNAS. The variable is computed as a weighted average of the typical monthly wage across the sectors *Coal Mining and Peat Excavation*; *Uranium and Thorium Mining*; and *Metal Mining* in a given 1990-district and year, using the sample weight assigned to an individual respondent in the data.

For our analysis in Table OA3 we approximate the number of workers employed in the mining sector and the number of workers employed in the combined mining and oil & gas sectors in a given 1990-district and year. To compute the latter variable, we first compute the weighted *share* of surveyed individuals who reported to work in the mining or oil & gas sector. The numerator of this share is the weighted number of respondents in the district-year who state that their main

[57] In a given district, certain census blocks are selected, in which 16 households are sampled (10 from 2011 onwards). All individuals sampled in a certain census block obtain the same weight, which depends on the relative importance of the census block in terms of overall district representation.

activity in the past week was working *and* who report to work in one of the following sectors: *Coal Mining and Peat Excavation*; *Uranium and Thorium Mining*; *Metal Mining*; *Oil & Gas*. The denominator is the weighted number of respondents in the district-year. We then multiply the ratio by the most recent available population figure from a given year's perspective.[58] To approximate the number of mining workers, we repeat the above exercise, but exclude oil & gas workers from the share's numerator.

For illustrative purposes (see Section 3.1), in Appendix-Table A1 we report descriptive statistics on the fraction of mining workers to total workers across district-years (based on districts with mineral resources as of 1990), over 2007-2015. For a given district and year, the computation of the numerators of these shares is done as described above, conceptually. The denominator of both shares is the weighted number of surveyed individuals who state that their main activity in the past week was working.

References

Ackerberg, D., K. Caves, and G. Frazer (2006). Structural identification of production functions. UCLA, mimeo.

Allcott, H. and D. Keniston (2018). Dutch disease or agglomeration? The local economic effects of natural resource booms in modern America. *The Review of Economic Studies 85*(2), 695–731.

Bazzi, S. and M. Gudgeon (2020). The political boundaries of ethnic divisions. *American Economic Journal: Applied Economics, forthcoming.*

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *The American Economic Review 102*(6), 2437–2471.

Holmes, T. J. and J. J. Stevens (2014). An alternative theory of the plant size distribution, with geography and intra- and international trade. *Journal of Political Economy 122*(2), 369–421.

Wu, J. C. (1982-1989). The mineral industry of Indonesia. In *Minerals yearbook area reports: International*. Washington: U.S. Bureau of Mines.

[58] See Section OA1.3 for details on population data. We multiply the share of mining workers in 2015 with the population data from 2010, since the results of the 2015 intercensal population survey have not been published by the MPC yet.