Gold Mining and Proto–Urbanization:

Recent Evidence from Ghana[*]

Marcel Fafchamps

Stanford University, BREAD and NBER

Michael Koelle

University of Oxford

Forhad Shilpi

The World Bank

April 2016

Abstract

Central place theory predicts that agglomeration can arise from external shocks. We investigate whether gold mining is a catalyst for early stages of urbanization in rural Ghana. We call this phenomenon proto-urbanization. Using cross-sectional data, we find that locations with gold mines exhibit most of the tell-tale signs of proto-urbanization. They have higher population densities, and they are also sites where more sophisticated forms of economic activity agglomerate. These findings are consistent with agglomeration effects that induce non-farm activities to coalesce in a particular location. Over time, we find that an increase in gold production is associated with more specialization, but not with a clear sectoral transformation of employment. We also find that the changes arising from increasing gold production are not reversed when large gold mines shrink. Rather, they continue to become more consistent with processes of structural transformation.

1   Introduction

Much debate remains around the question of where town and cities locate. There is general agreement that the existence of agglomeration externalities is the reason why towns and cities exist. But the factors that predict the emergence of new towns and cities seldom predict their precise location, especially in the absence of strong geographical differentiation (e.g. mountains, rivers, sea). In such cases, some historical accidents are often presented as the reason why a particular town is where it is.

Many different types of historical accidents have been associated with the birth of a town – for example, the site of a Roman fort, the intersection of two ancient roads, or the existence of a shrine or sacred place. The discovery of mining deposits or other extractive resources is known to have led to the creation of new towns, notably in the American West. But these towns do not always survive when mines or resources are exhausted or lose their market value, and many become ghost towns. There are important exceptions, such as Johannesburg, the South African mega-city that grew out of gold mining. Other towns, such as Manaus, shrink once resource extraction declines (in the case of Manaus, natural rubber), but they do not disappear altogether.

A growing body of recent literature shows a substantial impact of investments in transport infrastructure on the concentration of economic activities over geographical space (Donaldson, 2014; Banerjee et al., 2012; Faber, 2014; Bird and Straub, 2014; Thompson, 2000; Michaels, 2008). There is however less empirical evidence on the role of primary production in the rise of cities. In the case of the USA, Glaeser et al. (2015) provide evidence on how proximity to historical mining deposits led cities to specialize in activities with significant economies of scale (e.g. larger firm sizes) but at the cost of fewer start-ups. Jedwab (2014) documents how, in Ghana and Côte d'Ivoire, cocoa production spawned many small towns and how these towns survived the gradual westward shift of cocoa plantations that took place over the last century. Using data from a large mine in Peru, Aragon and Rud (2013) find evidence of a positive effect of miners' demand for local inputs on incomes in non-mining sectors of surrounding areas. In Ghana, however, the same authors (Aragon and Rud, 2015) find that mining reduces agricultural productivity by almost 40% as a result of pollution. Using household data, the authors find that

mining activity is associated with an increase in poverty, child malnutrition and respiratory diseases. Focusing on female employment near mines in Sub-Saharan Africa, Kotsadamen and Tolonen (2016) find that women move from agriculture into the service sector when a new mine opens, but this reallocation does not survive mine closure. Using a similar dataset but at a global level in 44 developing countries, von der Goltz and Barnwal (2014) find that mining communities enjoy a substantial medium term rise in asset wealth but incur negative health consequences due to exposure to various pollutants.

In this paper, we use the opening and closure of formal gold mines and the rapid changes in mining output as a quasi-natural experiment to examine what mining does to the organization of economic activity across space and over time. We ignore informal gold mines that are responsible for much of the pollution discussed in the literature. Gold has long been mined in Ghana, so much so that the colonial name for the country was the Gold Coast. Gold exports dropped significantly after independence as a result of an over-valued exchange rate, under-investment in primary production, and stagnating gold prices after the end of the Bretton Woods agreements. But in the last two decades, gold mining has experienced a revival in Ghana. The sector has been liberalized and in the 2000's the gold price rose to a new all-time high. As a result, new entry took place and, in the recent past, several new formal mines have opened. Over the same period other mines have closed down when deposits became depleted. Ghana is a good country to study urbanization, given the rapid growth of both the population and the economy over the last two decades.

Using recent highly disaggregated location-specific data from a combination of sources, we investigate whether variation in gold mining is associated with tell-tale signs of what we dub 'proto-urbanization'. In astrophysics, a proto-star is a star in the making, a cloud of gas that is progressively aggregating in a single location through a process called accretion. Once the accumulated gas reaches a critical mass, it ignites and a star is born. Not all proto-stars become stars, however: sometimes they do not reach critical mass; other times a competing celestial body absorbs all or part of their matter. Being a manifestation of agglomeration forces, proto-urbanization can be thought of in the same way: for a location to become a town, it must attract people and economic activity and, like a proto-star, some clues of proto-urbanization must be observable even before it becomes a recognizable town. Moreover, like a proto-star, observing

early signs of urban formation in a particular place does not guarantee that a self-sustaining town will eventually arise: the location may be lose its appeal, or its people and industries may leave for a competing town. The purpose of this paper is to look for evidence that gold mining is associated with symptoms of agglomeration, that is, of the accumulation of non-agricultural economic mass near the mine, while locations further away lose their non-agricultural activities.

Using cross-sectional data from 2010, we find that, among locations classified as rural in 2000, those locations near gold mines exhibit most of the tell-tale signs of proto-urbanization. Employment shares in the service and trade sectors are higher, while they are smaller in agriculture. Population density and the density of non-agricultural jobs are higher too. Jobs (outside of mining) are more likely to be in the wage sector, and in formal enterprises or the public sector. A larger proportion of the working age population train as apprentices, search for work, or specialize in home production. The only important exception is that locations with gold mines do not have larger shares of manufacturing employment. We also find that most of these patterns reverse at distances of 20-30 km from gold mines. These cross-sectional findings are consistent with agglomeration effects that induce non-farm activities to coalesce in a particular location.

Over time, we find that an increase in gold production is associated with increasing population density in areas close to mines. Tasks with higher specialization (in wage employment, apprenticeships, unemployment or homemaking) and in more sophisticated forms of organization (in the formal private and public sectors) agglomerate close to mines, and move away from the surrounding hinterland. However, expanding mines crowd out manufacturing and trade and have ambiguous effects on other industrial sectors, and no effect on agricultural employment. We also find that the changes arising from increasing gold production are not reversed when large gold mines shrink. Population keeps growing where mines shrink. In addition, and in contrast to mine expansions, the total number of jobs and the manufacturing employment share respond positively to a mine contraction. Population inflows extend to the hinterland, but services and formal jobs coalesce at the site of the shrinking mine. In the hinterland, areas 20-30 km from contracting mines become more agricultural, and their workforce becomes less specialized.

To see if mining leads to increased income, we examine the pattern of income in relation to mining locations using two potential measures of income: the intensity of lights at night, and

the proportion of households with electricity. We find that the proportion of households using electricity for lighting increases after an expansion of mining output and even more so after the contraction of mining output. These results support our finding regarding employment patterns, namely that mines help to agglomerate non-farm activities in their immediate vicinity and that this tendency to agglomerate survives and strengthens when mine contracts. In contrast, we find that nightlight mimics the expansion and contraction of mines, suggesting that nightlight may be picking up mining intensity itself.

The rest of the paper is organized as follows. In Section 2 we briefly describe the conceptual framework. The gold mining sector in Ghana and various data sources used in the paper are presented in Section 3, together with descriptive statistics of the variables of interest. Our empirical strategy is outlined in Section 4. Estimation results are presented in Section 5.

2   Conceptual framework

The over-arching concept behind our analysis is central place theory: urban centers arise because certain economic activities benefit from agglomeration externality. The sectoral mix of cities indirectly reveals, for a given point in space and time, where these externalities are strongest. For instance, in the 19th and early 20th century, most of the manufacturing activities were concentrated in cities. But since the 1950's, manufacturing has been moving out of urban centers to make room for finance and other services (e.g. Desmet and Fafchamps, 2007). Agriculture, in contrast, requires a lot of land and thus tends to be displaced by urbanization – even though peri-urban agriculture often is the most productive because of its emphasis on high value crops and dairy production (Jacobs, 1969). The urban and peri-urban share of agricultural employment, however, is low because of the importance of non-farm production (e.g. Fafchamps and Shilpi, 2003). Mining, by definition, follows mineral deposits and mineral deposits typically occur for reasons unrelated with the urban potential of a particular location. Whether or not mineral deposits are exploited does, however, depend on accessibility and profitability. Formal mines create a significant amount of industrial employment, typically mostly for well-educated and highly specialized workers that migrate to the mining areas. However, miners may demand local, non-tradable services and perhaps source some local production inputs.

For countries at low levels of economic development, agglomeration externalities arise in part because spatial concentration of demand makes job specialization possible (Fafchamps and Shilpi, 2005; Fafchamps, 2012). Many goods and services that, in rural areas, are provided outside of the market realm fall within the market domain in cities (Fafchamps, 2011). As a result, many individuals are able to have a non-farm job as their primary occupation, a fact that should be reflected in occupational data. Because specialization allows the emergence of new and better goods and services, towns end up serving their surrounding rural hinterland with non-agricultural products and services – including trade and the transformation of agricultural products.

Specialization, market provision, and concentration of demand also mean that firms can increase in size to capture economies of scale. It follows that the share of people working for a wage (or running a firm with employees) is larger in towns and cities than in the surrounding countryside, where most people are self-employed or family workers, typically on a farm. Based on this, we expect urban centers to be characterized by a larger share of formal employment, a larger proportion of employees or employers, and a lower proportion of self-employed and family workers. There should also be more public sector employment because schools, health facilities, and government bureaucracies are often located in small urban centers from which they serve the surrounding countryside.

Since towns and cities have more people in specialized employment (either for a wage or on own account), we also expect them to have more people looking for employment, that is, a higher proportion of unemployed. In contrast, in a country like Ghana, people living on a farm are often underemployed, at least seasonally, but they rarely report themselves as formally unemployed. For similar reasons, women in rural Ghana often list farming as their primary occupation, while married women in urban areas are more likely to describe themselves as homemaker.

We expect the number and size of cities to grow with economic development. The intuition is that higher levels of income shift demand away from basic necessities such as food and shelter, towards more specialized goods and services best provided in an urban environment. In Africa, urbanization seems to also have been fueled by population growth, with much urbanization occurring even in periods with little economic growth. Based on this, we expect new towns to arise in Africa, particularly in countries with rapid economic growth.

Viewed in a dynamic context, central place theory predicts that towns should arise at more or less regular intervals that follow a honeycomb shape, and that they should structure into a hierarchy of small and large towns (e.g. Christaller, 1933; Isard, 1956).[1] According to this theory, towns surrounded by a rural hinterland arise as a part of a self-emerging relational structure. The position of each town relative to other matters, but not their absolute location. It follows that, on a relatively undifferentiated terrain, the precise location of newly emerging towns is not known beforehand, as there are typically several equally suitable potential sites. Given this, the location of new towns can be influenced by historical accident – such as a temporary gold mine. This observation forms the basis of our investigation. Specifically, we investigate whether gold mining affects the placement of new embryonic towns, a process that we dub 'proto-urbanization'.

Finding that gold mining leads to the emergence of urban-like activity does not, by itself, guarantees that the embryonic town will survive once the gold deposits are exhausted. There are indeed many examples of towns that were abandoned in this manner – so-called ghost towns. The critical test is whether the embryonic town survives – and thrives – after gold mining stops. Only then can we conclude that gold mining can be a catalyst for the formation of a self-sustaining agglomeration.

3   Empirical Framework

Our objective is to examine how employment and income patterns vary with proximity to mining activity in formal gold mines. Since we are interested in early signs of agglomeration, already established urban centers are omitted from the analysis. We focus on localities that were classified as rural at the beginning of our study period.

Let y_{it} denote an outcome of interest measured at the level of location i at time t. For example, y_{it} can be the proportion of male adults that are employed in agriculture in location i. Let the set $\{m_{jt}\}$ represent the production of all known gold mines in Ghana. Each variable m_{jt} is defined as the average amount of gold produced by mine j in a relevant time interval up

---

[1] The hierarchical structure of towns and cities has been used to account for the fact that the distribution of town sizes follows an exponential distribution – often called Zipf Law (e.g. Henderson, 1985). The evidence on towns following a honeycomb shape is less clear, but towns are typically not located right next to each other.

to year t. In our analysis, this time interval is typically 10 years, given that most of our outcome variables y_{it} are only available in 2000 and 2010.

We are interested in estimating $E[y_{it}|\{m_{jt}\}]$, that is, in estimating how the average value of y_{it} systematically varies with proximity to gold production.[2] To this effect, we wish to estimate a model of the form:

$$y_{it} = F(\{m_{jt}\}) + \text{ controls} + u_{it} \tag{1}$$

where function $F(.)$ is our object of interest. Controls are discussed later.

We think of the flexible function $F(.)$ as aggregating the predictive effect of gold mines located at various distances from i. To capture this idea in a compact way, we follow Fafchamps and Shilpi (2005) and let:

$$F(\{m_{jt}\}) = \int_0^\infty \gamma(\delta) \, m_t(\delta) \, d\delta \tag{2}$$

where $\gamma(\delta)$ is a parameter that varies with distance δ and $m_t(\delta)$ is the production in the time interval up to t of all mines j that are located at distance δ from location i. In practice, we discretize $\gamma(\delta)$ into a fixed number of distance intervals or 'bins' indexed by k. To be more specific, $k = 1$ corresponds to $0 \leq \delta \leq 10$ km, $k = 2$ corresponds to $10 < \delta \leq 20$ km, etc. Since we do not expect employment patterns to be affected by gold mines more than 100 km away, we assume that beyond 100 km, $F(.)$ is zero. With this parametric assumption, function $F(.)$ can be rewritten as:

$$F(\{m_{jt}\}) = \sum_{k=1}^{K=10} \gamma_k \, m_{ikt} \tag{3}$$

where m_{ikt} is the sum of m_{jt} located in distance interval k from location i. So, for instance, if mine $j = 3$ is located 5 km from i, mines $j = 7$ and $j = 9$ are located 12 km from i, and all other mines are more than 100 km from i, we have:

$$F(\{m_{jt}\}) = \gamma_1 \, m_{3t} + \gamma_2(m_{7t} + m_{9t})$$

We estimate three specifications of model (1). In the first specification, we regress our

---

[2]Alternatively, one may view this expression as a description of how y_{it} varies with proximity to a formal gold mine, whereby each mine is weighted by its average level of production in a time interval leading up to time t.

outcome of interest y_{it} on the function $F(\{m_{jt}\})$. In addition, we control for a vector of time-invariant geographical variables x_i and region dummies ν_r.[3]

$$y_{it} = F(\{m_{jt}\}) + x_i'\beta + \nu_r + u_{it} \tag{4}$$

In the second specification, we additionally control for proximity to towns and cities. The estimated model then is:

$$y_{it} = F(\{m_{jt}\}) + G(\{c_{n,t_0}\}) + x_i'\beta + \nu_r + u_{it} \tag{5}$$

where function $G(.)$ controls for how y_{it} varies with proximity to various urban centers of more than 10,000 inhabitants indexed by n. The size of each city is proxied by its population. Since our independent variable is gold output in the period leading up to t, city population at t might partially be an outcome of gold mining nearby. We therefore use past population c_{n,t_0} as proxy for city size. Conditioning on urban proximity identifies the pattern of correlation between outcome y_{it} and proximity to gold mines, net of any correlation between proximity to mines and proximity to urban centers. Function $G(.)$ is constructed in a manner similar to $F(.)$, i.e., we write:

$$G(\{c_{n,t_0}\}) = \sum_{k=1}^{K} \lambda_k \, c_{ik,t_0} \tag{6}$$

where c_{ik,t_0} is the sum of city population c_{n,t_0} located in distance interval k from location i.

Model (5) includes region fixed effects but not fixed effects at the level of the location i itself. This may affect inference if the placement of a gold mine is correlated with location specific invariant characteristics that also affect urbanization. To allow for this possibility, we

---

[3]The region is the highest level of sub-national administrative units in Ghana. Districts are the lower-level administrative units. However, due to the geographical clustering of mines in neighbourhoods of a relatively small number of districts (see maps in Figure 1), districts might also pick up part of the effect of mining otherwise captured in γ_k, because locations close to mines fall within only a few districts. For example, imagine a district that spans across locations situated 10-30 km from a mine. The dummy coefficient for this district will then pick up the average effect of mining at distances 10-30 km, while γ_2 and γ_3 will pick up deviations from this average for locations at distances 10-20 km and 20-30km, respectively. We therefore choose to control for the higher level of administration, region.

also estimate an expanded dynamic version of regression model (5):

$$\begin{aligned}
\Delta y_{it} &= \Delta F(\{m_{jt}\}) + \widetilde{G}(\{c_{j,t_0}\}) + x_i'\widetilde{\beta} + \widetilde{\nu}_r + \Delta u_{it} \\
&= \sum_{k=1}^{K} \gamma_k \, \Delta m_{ikt} + \sum_{k=1}^{K} \widetilde{\lambda}_k \, c_{ik,t_0} + x_i'\widetilde{\beta} + \widetilde{\nu}_r + \Delta u_{it}
\end{aligned} \tag{7}$$

In model (7) identification of the γ_k's relies solely on variation between 2000 and 2010. First differencing model (5) in principle removes time-invariant regressors c_{ik,t_0}, x_i and ν_r. We nonetheless include similar terms in (7) to allow for differential time trends depending on geographical characteristics, access to infrastructure, and proximity to cities.

The dynamic model becomes most useful when we use it to investigate whether, over the study period, gold mines served as catalyst for urban formation. If gold mining serves as catalyst, we would expect the urbanization effect of gold mines *not* to disappear once gold mines close or shrink. Urbanization forces may even grow stronger once the pollution from gold mining (e.g. mercury, tailings) is eliminated: once a mine leaves, the fledging town may expand further. The alternative is that, once the mine disappears, the fledging town shrinks or disappears.

To investigate this possibility, we estimate a model of the form:

$$\Delta y_{it} = \sum_{k=1}^{K} \gamma_k^+ \, \Delta^+ m_{ikt} + \sum_{k=1}^{K} \gamma_k^- \, \Delta^- m_{ikt} + \sum_{k=1}^{N} \widetilde{\lambda}_k \, c_{ik,t_0} + x_i'\widetilde{\beta} + \widetilde{\nu}_r + \Delta u_{it} \tag{8}$$

where

$$\begin{aligned}
\Delta^+ m_{ikt} &\equiv \max(0, \Delta m_{ikt}) \geq 0 \\
\Delta^- m_{ikt} &\equiv \min(0, \Delta m_{ikt}) \leq 0
\end{aligned}$$

Put plainly, we have split Δm_{ikt} into positive and negative changes $\Delta^+ m_{ikt}$ and $\Delta^- m_{ikt}$. If the creation or growth a formal gold mine in location i triggers proto-urbanization proxied by an increase in y_{it}, we expect $\gamma_k^+ > 0$. If proto-urbanization is reversed once a formal gold mine closes, we expect $\gamma_k^- > 0$ as well: remember that $\Delta^- m_{ikt} < 0$ in places where a mine closes or shrinks. If we find instead that $\gamma_k^- \leq 0$ this suggests that proto-urbanization does not stop once the mine shrinks or disappears, and it even increases if $\gamma_k^- < 0$.

3.1 Identification

Our identifying assumption is that the location and production volume of a gold mine are exogenous to the emergence of new urban centers. Our identification is aided by the relatively short time frame of our study (2000-2010) which limits exposure to other confounding changes. Another advantage of our empirical analysis is that our data are available at a high level of spatial disaggregation.[4] The validity of our identifying assumption however relies on the fact that new gold mines locate where gold is and are not attracted by other factors that have their own direct effect on urbanization.

Gold deposits in Ghana lie underneath chains of gentle hills in southwestern Ghana. Gold mines are foreign-owned, and gold output is exported and sold in the world market. Entry into gold mining takes place when gold prices are high, and costs are lower than revenues. Costs are partly determined by macroeconomic conditions, such as the exchange rate or an attractive business climate. Locally, a favourable geological profile is the most important determinant of production costs. The location of gold mines is determined by where gold deposits are, and whether exploitation of these ore deposits makes economic sense. The geological structure of the sub-soil – i.e. the kind of rock, or how deep underneath the ground and how densely packed the ore deposits are – is the main determinant of cost differences across locations. Similarly, the richness of deposits and the structure of sub-soil geology also determine marginal costs and therefore the intensity of production. Exit occurs when deposits are depleted, or the gold price falls and exploitation is no longer economically viable.

Our identifying assumption is violated if there are factors that determine both exploitation costs, and economic activity other than gold mining. Without proper control for these omitted factors, regression analysis may detect a spurious correlation between mining and outcomes of interest. The accessibility of the mine site is determined by the ruggedness of the terrain, and by the remoteness from existing transport infrastructure. Another factor is access to electricity, a key input into mining. Proximity to urban centers may also affect costs. On the one hand, urban amenities might make it easier to attract a qualified labor force. On the other hand, gold

---

[4]Two important confounding factors at higher level of aggregation (district or province) and longer time horizon are that mining may help shape local political economic factors (e.g. district is a local political entity) and affect historical institutional legacy. By focusing on shorter period and on finer locations, and by using fixed effects estimation, we avoid both of these confounding effects.

mining is a major source of pollution (Aragon and Rud, 2015) and competes for space with other urban uses. We expect the latter effect to dominate, and we do not expect new mines to locate close to already established towns and cities.

We address issues of potential omitted variable and reverse causality bias in several ways. In the analysis based on cross-section data, we control for time-invariant measures of geography, transport infrastructure, and proximity to the electricity network. Although we expect mines to locate away from cities rather than close to them, we also control for distance to pre-existing cities as a robustness check. In addition, we include regional dummies to control for any regional variation that could arises due to differences in infrastructure, economic structures, sub-national government, or historical legacy.

In our panel analysis, we estimate a model in first differences. This removes the level effect of all time-invariant variables at the EA level, including all persistent determinants of both mines and urban agglomeration. Still, EAs closer to roads or electrical power, or EAs in a flat as opposed to hilly area could follow a different economic trend than more remote areas, and this could influence urbanization. At the same time, these factors could also determine the placement of new mines or the evolution of output of existing mines. We therefore include controls for baseline geography and infrastructure in equations (7) and (8). These controls now pick up differences in trends, as opposed to the level differences they controlled for in the cross-section analysis. As a robustness check, we also control for the possibility that trends vary with proximity to cities.

4 Data

To test our theoretical predictions, we select a country with rapid population and economic growth – Ghana – and a time-period during which international gold prices increased and much entry took place in gold mining – i.e. the period between 2000 and 2010. We start by briefly describing the gold mining sector in Ghana. We then detail the data sources that we use to study proto-urbanization.

4.1 Gold Mining in Ghana

During colonial times, Ghana was known as the Gold Coast. Gold mining dates back a long time in the pre-colonial era. The Ashanti kingdom, ruled from its historic capital Kumasi, controlled mining activities in many parts of the kingdom. At the end of the 19th century, French and British entrepreneurs obtained concessions and introduced modern, imported machinery to the industry (Taylor, 2006). Mining was largely nationalized after independence in 1957 until liberalization in the early 1990s. In the last two decades, the Ghanaian government has gradually reduced its stake in existing mining companies, and has awarded concessions to new prospective entrants, mostly foreign companies.

Gold mining is heavily concentrated in the southwest of the country, mostly between the cities of Sekondi-Takoradi on the coast, and Kumasi in the interior of the country. The maps in Figure 1 show the spatial distribution of mines and cities in 2000 and 2010, respectively. The vast majority of open pit and underground mines lie along the Ashanti Gold Belt, a geological formation that stretches from Sekondi-Takoradi on the coast in northeastern direction up to the region east of Kumasi (Carranza et al., 2009). Other mines are located in the Sefwi belt, parallel to the Ashanti belt (Pigois et al., 2003). Alluvial mining takes places along the rivers that run through this mineral-rich area: the Pra river and its tributary, and the Birim river.

We obtain mine-level production data from annual country reports published by the United States Geological Survey (USGS) for the years 1991-2003, and from corporate annual reports for the years 2004-2010. We cross-validate USGS and company data for the years preceding 2004. We exclude gold that is mined by artisans and traded by the Precious Minerals Marketing Corporation, because it cannot be traced geographically. Our measure of mining activity covers all formal enterprises, which are medium and large-scale mines. We code mining location using a variety of sources, including annual reports, other corporate publications, technical industry reports, and high-resolution satellite imagery.

Some 18 formal gold mines operated in Ghana at some point in time between 1991 and 2010. Of these, 16 were operating in the decade preceding the year 2000; 12 of them were still operating in the year 2000. In the 2000s, there were 12 operating mines, 8 of which were still active in the year 2010. Table 1 lists the average yearly production volume by decade for each

mine. In the 1990s, the mine at Obuasi produced on average as much gold as all the other mines together. The production volume of Obuasi subsequently halved, and overall gold production decreased until the opening of new mines. In recent years, a high proportion of production has come from a small number of mines, each of which is fairly large. The largest is the Tarkwa mine, with about a quarter of overall production in 2010.

Table 1 shows that there is substantial variation in mine production across sites. Some alluvial placer operations in the Eastern region are quite small, producing on average 26 kg of gold per year. Mid-size mines have an average yearly production of about 2,000-5,000 kg. The largest mines in our study produce between 8,000 and 20,000 kg of gold per year.

There is also considerable variation in gold production across years, shown in Figure 2. Production was generally rising throughout the 1990s. There was a particular acceleration in the four years preceding the 2000 Census. Aggregate production slowly declined afterwards, but this declining trend turned around in 2005 with the opening of the Chirano mine. The opening of the Ahafo mine in 2008, which immediately became the second largest mine in that year, led to another increase of aggregate production.

Gold mining provides for a very small share of employment in Ghana. The sector as a whole – including formal and informal, large and small mines– employed only 0.72 percent of the country's labor force in 2000, and 0.93 percent in 2010. But in the immediate vicinity of mines, gold mining accounts for a substantial share of local employment. For example, 15 percent of all employed males who live within 10 km of a large-scale mine worked in gold mining in 2000. This figure rises to almost 20 percent in 2010.

In Figure 1, we present a map of Southern Ghana that shows the location of mines and cities in 2000 and 2010. It also maps areas that are within 100 km of any mine. These are the areas that we focus on in this paper. The two maps show that urbanization in a large proportion of the country could potentially be affected by proximity to gold mining: the shaded area includes all the regions West of the mines up to Ivorian border and it stretches to the North past Kumasi. In 2000, it spread past Accra and Lake Volta to the East due to the presence of two small alluvial mines there. These mines had been discontinued by the year 2010.

14

4.2 Data sources

In addition to the mining data mentioned in the previous section, we make use of two data sources from which we construct proxies for proto-urbanization: the 2000 and 2010 population censuses; and night-time illumination data from satellite imagery. We discuss them in turn.

The census data have been made available to us at the enumeration area (EA) level. There were 26,708 EAs in the 2000 census and 37,625 EAs in the 2010 census. In addition, we have a complete map of the EAs from the 2000 census. We restrict our analysis to EAs that were rural in 2000, according to the classification of the Ghana Statistical Service.[5] With this restriction, we exclude from our analysis all areas that were already urbanized by 2000, and only focus on those EAs that can possibly be at the extensive margins of urbanization between 2000 and 2010. We extract information about the GPS coordinates of the center of each of these EAs. In total, we obtain 11,244 such coordinates. For the enumeration of the 2010 census, some of the EAs were split up, but none were merged. This allows us to link them to our map. We are able to link 10,708 of these EAs across the two censuses.[6]

Within each EA we have data on 10% of all households. In addition to gender, marital status, and relationship to the household head, the two censuses contain, for each adult, a broad characterization of their sector of activity (e.g. mining, agriculture, trade), labour market participation (e.g. employed, unemployed, homemaker), form of employment, if any (e.g. self-employed, wage worker, apprentice), and type of employer (e.g. private formal, private informal, public). In addition to population density, information on occupation and sector of activity is the focus of our empirical analysis. We also have data on some measures of housing and living standards, including the main source of electricity used for lighting.

The night-time illumination data come from satellite imagery. They were obtained from the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS). The dataset covers most of the earth at a 30 arc-second grid (approximately 1 square kilometer). The luminosity measure for each cell, or pixel, is a digital number in the range 0 to 63 that increases with luminosity. There are 280,000 pixels for Ghana. Recent studies have made use

---

[5]The GSS classification of urban EAs includes all EAs in localities – i.e., distinct population clusters that have a name or locally recognised status – with 5,000 or more inhabitants.

[6]Section A of the online appendix provides some more details about how the geo-referenced Census dataset was constructed. The appendix also describes the data sources for additional variables that we use as controls in our regressions.

of these data as a high-resolution, local measure of economic development (Bleakley and Lin, 2012; Henderson et al., 2012; Michalopoulus and Papaioannou, 2013). We map the nightlight data into our EAs, and are able to construct a panel dataset for 10,283 rural EAs over 19 years (1992-2010). For most years, we have observations from 2 different satellite readings. To reduce measurement error and noise, we use the average of the two as nightlight measure for that year.

4.3   Descriptive statistics

According to the census data, the population of Ghana was 18.85 million in 2000 and 24.66 million in 2010, corresponding to an annual rate of population growth of 2.7 percent – close to the African average. Urbanization is similarly rapid: in 2000, about 44 percent of the population lived in urban areas; in 2010 this figure stood at 51 percent. These two features make Ghana an ideal candidate to study town formation.

We restrict our attention to the EAs that were classified as rural in 2000. In these areas, the population increased from 10.47 million to 12.85 million, which is a lower rate of population growth than the national average. Of the 10,708 EAs that were rural in 2000 and that we could link over time, 257 had become urban localities by 2010, with a combined population of 1.10 million.

The main subject of our analysis is occupational data, summarized in Table 2. We focus on the working-age population, which we define as all adults aged 20 to 60 years. These made up 38.5 percent of the total population in 2000, and 39.9 percent in 2010.

In line with the rapid urbanization of the country, we observe a rapid transition of the economy out of agriculture in those initially rural areas: in 2000, 75 percent of total employment was in agriculture but by 2010, the employment share of agriculture had fallen to 62 percent. The difference was absorbed in the tertiary sector, with trade (from 6.3 to 10.5 percent) and services (from 10.7 to 16.8 percent) experiencing increasing employment shares. Manufacturing increased modestly from 6.8 to 8.5 percent. Unemployment halved from 6.5 to 3.1 percent. As documented in Fafchamps and Shilpi (2003, 2005), a higher share of employment in services is consistent with a more urbanized economy with higher levels of occupational specialization. Employment in manufacturing increased substantially less compared to services or trade activities, which is consistent with a pattern of urbanization without industrialization as documented in Gollin et

al. (2013).

The bottom part of Table 2 documents the types of occupation that Ghanaians hold and the types of firms and organizations by which they are employed. The largest occupational group are self-employed workers without employees, who accounted for 76 percent of employment in 2000 and 68 percent in 2010. Around 3 percent of the study population are employers, that is, they are self-employed with employees. Wage employment increased between 2000 and 2010 from 12 to 14 percent. Family workers made up 9 percent of the workforce in 2000 and 15 percent in 2010. The rest are apprentices and other kinds of workers. The large majority of employment is in informal private businesses, and the proportion rose over time – from 90.8% in 2000 to 92.3% in 2010. Over the same period, employment in formal private sector jobs fell slightly from 4.7 to 3 percent. Employment in the public sector accounts for a stable 4.3 percent of total employment.

The distribution of the nightlight measure over several time periods is documented in Table A1 in the online appendix. This measure is zero when a pixel is unlit at night and positive otherwise. Over our study period, about 93 percent of the country is unlit. This is in line with what is documented for poor countries with a medium population density (Henderson et al., 2012). The percentage unlit decreased from 97 percent in 1992 to 90 percent in 2010. Over time, we observe the entire nightlight distribution shifting upwards. This is in line with the general pattern of economic and urban growth in Ghana over this time period.[7]

5 Empirical analysis

What are the tell-tale signs of proto-urbanization that we can identify in our data? By definition, urban activity focuses on non-primary production, that is, on production other than agriculture or mining. If gold mines serve as catalyst for urbanization, we would observe that gold mines are associated with a rise in the share of non-primary employment in total employment. Urbanization is also associated with agglomeration effects that foster economies of scale in production. Given this, we expect to observe more employment in the (non-primary) wage

---

[7]The raw digital numbers are not strictly comparable over time, due to differences in calibration across satellites, and across years within a given satellite. We therefore include year dummies in our regressions to correct for this measurement error.

sector, either as worker or as employer.

For similar reasons, we expect to observe more people employed in the formal private sector. Wage employment in the public sector may similarly rise if government administration and services locate near gold mines, as this would also serve as catalyst for urbanization. Since apprentices are only observed in non-primary sectors, a rise in the share of apprentices would signal that a larger proportion of workers wish to learn skills that enable them to obtain employment in the non-primary sector.

Agglomeration leads to the concentration of jobs in a given location. This in turn makes it more appealing for people to move to an agglomeration in search of work. For this reason, we expect to find more self-declared 'unemployed' individuals near gold mines. If these individuals are looking for work primarily outside mining, we would also regard this as consistent with proto-urbanization. Another pattern we expect to find is the rise in the proportion of people (mostly women) who list themselves as homemakers. If they were located in a rural village where agriculture dominates activity, these homemakers would be spending part of their time in the fields, and would typically be counted as employed in agriculture (Fafchamps and Shilpi, 2005).

Agglomeration also manifests itself in an influx of population to the newly urbanizing area. For this reason, we would expect to find a higher population density close to gold mines. On the other hand, gold mining concessions compete for space with urbanizing areas, which limits the potential for population growth immediately next to mines. Recent literature has argued that urbanization in Africa, unlike in other parts of the world, might not be driven by industrialization, and that population agglomerates in 'consumption cities' (Jedwab, 2014; Gollin et al., 2016). This means that we should not necessarily expect to see higher shares of employment in manufacturing. Furthermore, the absolute level of employment (normalised by area) might not necessarily increase along with the population density. Finally, if gold mines trigger economies of agglomeration and higher returns from specialization, we also expect incomes to be higher – a feature that would manifest itself as an increase in a measure of income, which we proxy with nightlight and electricity usage by households.

5.1 Cross-Sectional Evidence

We start with the analysis of the census data. We report in Figures 3, 4, and 5 the regression estimates of γ_k from model (4) estimated using the 2010 data. The three figures correspond to different sets of dependent variables from the census data. Coefficient estimates γ_k are reported graphically, together with their 95% confidence interval based on standard errors adjusted for arbitrary spatial autocorrelation (Conley, 1999).[8] We run the analysis at the level of enumeration areas (EA). In all specifications, we control for location-specific, time-invariant characteristics, denoted x_i in equation (4), that might co-determine gold exploitation costs and economic activity. These controls are: average slope and elevation of the EA; distance of the EA to the nearest primary road and the nearest secondary road; and distance to the nearest power line of the national electricity grid. Given that mining is clustered within the south-western part of the country, we also include region dummies to account for systematic differences between locations in mining regions, and locations in regions further away from mines.

As explained earlier, we focus our attention on distances up to 100 km from a gold mine. Since distance is divided into intervals of 10 km each (see equation 3), we report ten γ_k coefficients for each regression. Coefficients represent the marginal effect on local employment shares (in percentage points) of a one ton increase in average annual gold production in 2001-2010 at a certain distance interval, relative to localities farther away than 100 km. When outcomes are densities instead of shares, each coefficient represents the marginal effect on population and employment density (inhabitants or jobs per square kilometer) of a one ton increase in average annual gold production in 2001-2010 at the distance interval k.

By construction, the number of locations that have positive m_{ikt} increases with distance from the mine. About 60 percent of all EA in 2010 are further than 100 km away from any active gold mine; these locations constitute the omitted category in our regressions. Table A2 in appendix provides an overview of the distribution of our regressor of interest m_{ikt}, tons of gold produced at distance interval j from EA i in the 10 years leading up to t.[9] Similarly, we express city population c_{ik,t_0} in thousands of inhabitants. Our definition of city includes all localities that are classified as urban by the Ghana Statistical Service, and have at least 10,000

---

[8]We implement the standard error correction with the x_ols program available from Conley's web site. We use a cutoff of 2 decimal degrees for the distance threshold after which spatial cross-correlation is restricted to zero.

[9]In 2000 about 45 percent of all locations are further away than 100 km from any active mine.

inhabitants.[10]

In Figure 3 we show the proportion of adults employed in different sectors of activity. We observe that the presence of a formal gold mine in the immediate vicinity (0-10 km) of the EA is associated with a significantly smaller share of agricultural employment and a higher share of employment in all other sectors except for the manufacturing sector. This includes of course gold mining. Localities more than 70 km away from a gold mine are not different from localities in the control group. This suggests that truncating $F(.)$ above 100 km is sufficiently conservative. We also note that individuals living around 20-40 km away from a gold mine seem to be less likely to be employed in a non-primary sector such as services and trade than individuals living farther away from a mine. If confirmed by subsequent analysis, this would indicate that these occupations 'agglomerate' in the immediate vicinity of the gold mine – a further indication of proto-urbanization.

In the top row of Figure 4 we report γ_k coefficient estimates for employment in the wage sector, employment in the non-wage sector, and the proportion of apprentices. Employment in the mining sector is omitted from the calculation.[11] Employment in the wage sector combines salaried workers and self-employed individuals with employees. Employment in the non-wage sector combines the self-employed and family workers. Apprentices constitute a much smaller category but are of interest in their own right because they signal a desire to acquire skills outside the primary sector. We find that areas in the immediate vicinity (within 0-10 km) of a gold mine have more employment in the wage sector, more apprentices, and less employment in the non-wage sector. EAs at intermediate distances (20-40 km) show the opposite effects. They have lower shares of wage and apprentice employment, and a higher share of non-wage employment. This too is consistent with proto-urbanization.

The middle row of Figure 4 splits employment into informal, private formal, and public formal.[12] Consistent with the results at the top of the figure, we find that immediate proximity to a gold mine is associated with significantly more formal and less informal employment in the

---

[10]We tried different population cutoffs, and they yield the same results. As we explain in the data appendix, our map contains all urban boundaries; that is, boundaries of contiguous groups of EAs that are together classified as an urban locality.

[11]More precisely, the employment share of the wage sector is the proportion of people employed in the wage sector among all those employed excluding those employed in the mining sector (gold or other).

[12]The public formal category includes workers in the parastatal sector.

private sector. At 30-40 km distance from mines, this pattern reverses: there is significantly more informal and less formal employment. Public sector employment is slightly higher in the immediate vicinity of mines (0-10 km), and slightly lower at 20-30 km distance, suggesting that government services agglomerate at mine locations. The bottom of Figure 4 shows results for the unemployment rate and the proportion of homemakers in the adult population. We see that they are both higher in the immediate vicinity of mines, and lower at 10-20 km from mines. As indicated earlier, this too can be interpreted as indicative of proto-urbanization.

Figure 5 shows results for population density, and for a measure of overall employment normalised by area ('Employment density'), and separately for employment in the agricultural and the non-agricultural sectors. As before, direct employment in mining is omitted from employment figures, so the agricultural sector is equal to the primary sector here. Areas 0-10 km from a mine have higher population density, and higher employment. The latter is entirely due to higher non-agricultural employment. The evidence for areas farther away from mines is less pronounced. At 20-30 km from mines we find a slightly higher concentration of population and agricultural employment, and at 10-20 and 30-40km distance from mine we find slightly less non-agricultural employment, but no significant differences in overall employment.

5.2 Robustness Checks for Cross-Sectional Regressions

To verify the robustness of our findings, we start by estimating Figures 3, 4, and 5 with the 2000 data. Results, shown in the online appendix, are similar but less pronounced. Next we control for proximity to urban centers. A look at the maps in Figure 1 indeed suggests that distance to mines may be correlated with distance to large cities. There are two major mining areas: the large mine near Obuasi, accounting for more than 50 percent of all mining output in the 1990s; and the complex of mines near Bogoso-Prestea-Tarkwa. There is a town close to each of them. Right next to the Obuasi mine is the town of Obuasi proper, Ghana's eighth largest city, and Obuasi itself is located about 50-60km from Kumasi, Ghana's second largest city. There is no large town in the immediate vicinity of the Bogoso-Prestea-Tarkwa mining complex, but it is located 50-60km from Sekondi-Takoradi, Ghana's third largest city. Since areas where much of Ghanaian gold production originates happen to be located 50-60 km from a major urban center, failing to control for urban proximity may affect our estimates.

21

We reestimate Figures 3-5 using the regression model (5), which controls for distance to cities. Results are again similar to those in Figures 3-5. They are available in the online appendix. Next, we truncate the function $F(.)$ at 60 km from mines, instead of 100 km and report 5 coefficients for γ_k. The omitted category that serves as a control group in this exercise are locations at 50-60 km from mines; that is, we also drop all locations further than 60 km from any mine. This does not affect the main results, all of which were only significantly different from zero for locations closer than 50 km to mines. In another robustness check, we redefine m_{ikt} to be the average mining output in the 5 years leading up to the 2010 Census, that is, from 2006 to 2010; instead of 10 years, from 2001 to 2010. This does not substantially affect our results. Finally, we redefine our function $F(.)$ to capture only the extensive margin. In particular, we redefine m_{ikt} as a dummy equal to one if there is at least one gold mine at a distance interval k from i that has been active in the decade leading up to year t. In this regression, the coefficients γ_k express the effect of having a gold mine at different distances from an EA versus not having a gold mine. Results are available in the online appendix. Again, the pattern that emerges is qualitatively similar to our reported main results, but some effects at the mine location itself are less pronounced.

5.3 Panel Data Evidence

Next we turn to a panel analysis of the census data. We first estimate the first-differenced model (7). We control for differential trends by elevation, slope, nearest distance to primary and to secondary roads, and nearest distance to the electricity grid. We document the results in Figures A2-A4 in the online appendix. They are surprising: many estimates of γ_k are not significant. For example, for specialization in Figure A3 and population and employment densities in Figure A4, all point estimates at 0-30 km are not significantly different from zero, except those for unemployment and homemaking. When coefficients are significant – such as the ones for employment shares across sector of activity in Figure A2 – they tend to have a sign opposite to that documented in Figures 3-5.

The explanation for these puzzling results comes to light when we estimate model (8) where we interact changes in gold output with their sign. Model (7) is only appropriate if the effects of an expansion of gold production by a certain amount are reversed when production decreases

by the same amount. Model (8) relaxes this restriction and allows for asymmetric responses to expansions and contractions. Estimates for γ_k^+ are shown in Figures 6, 7 and 8 while estimates for γ_k^- are shown in Figures 9, 10 and 11. The γ_k^+ reported in Figures 6-8 represent the changes in economic activity associated with an expansion in formal gold mining, compared to locations with no change in gold production at distance k.

In Figure 7, we obtain basically the same picture for an expansion of mining output as we did in Figure 4 for the cross-section. All of the coefficients at 0-10 km from mines have the same sign as before and are all significant. Coefficients at distances farther away from mines are significant for most outcome variables. Consistent with Figure 5, in Figure 8 we find that population density at the mine location increases together with gold production, and that non-agricultural employment further away from mines decreases. We do not find significant effects of changes in mine production on agricultural employment.

The picture that emerges from changes in employment composition by industrial sector associated with an expansion of a mine, in Figure 6, is mixed. Employment shares in gold mining and in services in the immediate vicinity of a mine do not significantly change when a mine expands. At 10-20 km of a mine, gold mining and service sector employment increase significantly, and agricultural employment decreases. For gold mining, employment effects extend up to a distance of 70 km to a mine. We can only speculate that these contradictory effects at 0-10 versus 10-20 km of a mine are related to the fact that mining areas tend to be very large, and that increased gold output is associated with a territorial expansion of a mine. The coefficient estimates of manufacturing and trade activities, at 0-20 and 0-40 km respectively, are negative, suggesting a crowding out of some tradable sectors.

Estimates of the γ_k^- coefficients add a completely new dimension to the above findings. If urbanization is reversed once the gold mine disappears, we should find that γ_k^- coefficients have the same sign as γ_k^+ and as those found in Figures 3 to 5. Instead we find that, when the γ_k^- coefficient estimates are significant, they have the opposite sign to those found earlier. This is particularly striking in Figure 10, where all graphs of γ_k^- coefficients have the opposite shape to those in Figures 4 and 7. Most of the coefficients in the immediate vicinity of mines are not individually statistically significant, but the coefficients at 10-40 km distance are. This is also true for population density (in the upper left corner of Figure 11), and for the density (lower left

corner of Figure 11) and the share (upper left corner of Figure 9) of agricultural employment. This means that locations that experienced shrinking gold production between 2000 and 2010 have a higher *increase* in population density, and a higher *decrease* in the share and density of agricultural employment between 2000 and 2010, compared to locations that did not. We further find that services concentrate at the site of a contracting mine, and that manufacturing employment growth is stronger at 0-40 km from a contracting mine than elsewhere. If anything, this finding suggests that, on average, proto-urbanization continues after the gold mine shrinks or closes.

Perhaps the most striking finding is that employment in gold mining increases once a formal mine shrinks or closes. To investigate this issue further, we split gold mining employment into formal and informal, wage and non-wage employment. We report results in Figures 12 and 13. Figure 13 indicates that a reduction in mining output is associated with an increase in informal wage and non-wage (self-) employment in gold mining close to the shrinking mine. This is suggestive of a substitution by informal and small-scale mining activities once a large formal mine closes – which is likely to be a temporary phenomenon, driven by attempts to extract small amounts of gold from tailings and closed mines. This phenomenon does not, by itself, constitute evidence of a durable urbanization legacy.

To summarize our results so far: we find that gold mining is associated with certain tell-tale signs of urbanization in the immediate vicinity of mines. This is the case for: population density; higher employment in the wage, apprentice, private formal enterprise, and public sectors; lower employment in the non-wage and informal sectors; and higher rates of unemployed and homemakers. The effects for employment composition (but not for population density) are reversed at distances further away from mines (at 20-40 km distance). All of these findings from the cross-section also hold true for an expansion of mines, and interestingly also for a reduction of mine activity. Results for industrial composition of employment are mixed. We find a lower share of agricultural and a higher share of manufacturing, services, and trade employment in the cross-section and when mines contract. On the other hand, when mines expand we observe a crowding out of employment in trade and manufacturing in the immediate vicinity of mines.

5.4 Robustness Checks for Panel Regressions

We perform several robustness checks, similar to the ones we carried out in the cross-section regressions. First, we control for distance to pre-existing cities using function $G(.)$ that takes the distance to population in cities over 10,000 inhabitants in the year 2000 as its argument. Our results are robust to this control. Second, we truncate the function $F(.)$ at 60 km from mines. This does not affect results. Third, we let function $F(.)$ express changes in mining at the extensive margin. In the first-differenced model, the coefficients γ_k^+ now describe the effect of a mine opening at distance interval k. Similarly, the coefficients γ_k^- express the effect of a mine closure. This is a low-powered test because we observe only six mine closures and two mine openings between 2000 and 2010. Of the six mine closures, five correspond to the smallest mines in our sample and only one affects a larger mine, Teberebie. The mine openings took place in 2005 and 2008, only a few years before the year 2010. Nevertheless, we observe effects of a greenfield mine opening that are mostly consistent with what we observe for a mine expansion, but are less pronounced. For closures of small mines, we observe an increase in population density, informal mining activities, and a decrease in agricultural employment, but no spillover effects to other sectors.

5.5 Nightlight and household electricity use

To see if gold mining is associated with higher incomes, we rely on two proxies for income: nightlight; and the proportion of households with electricity. Several studies have used nightlight as a proxy for income or GDP ((Bleakley and Lin, 2012; Henderson et al., 2012; Michalopoulus and Papaioannou, 2013). In our setting, however, nightlight that is directly emitted from mines constitutes an important confounding factor that makes it difficult to interpret brighter nightlight as higher income. We thus look at a more direct measure of household electricity consumption: the proportion of households who use electricity as their main source of lighting. This variable indicates more directly the extent to which households – as opposed to street lighting and firms, including mines – use electric power, and it can be taken as a proxy for households' increasing wealth.

As before, we report results from three regression specifications: (1) a cross-section spec-

ification; (2) a first-difference specification, which corresponds to model (7); and (3) a first difference specification (8) in which we allow γ_k to be different for contracting and expanding gold production. All regressions include year dummies that control for annual differences in sensor sensitivity. In addition to the same geography and infrastructure controls as before (including distance to the electricity grid), we include distance to the nearest power plant as a control for power supply. This additional control is important in view of the fact that mining activities may lead to an increase in electricity supply and thus improve household's access to electricity. Controlling for the supply of electricity ensures that the proportion of households with electricity picks up the demand for electricity arising from higher income. The mining data, as before, come from formal gold mines. Spatially clustered standard errors are used throughout following the method proposed by Conley (1999).

As before, coefficient estimates γ_k are reported graphically, together with their 95% confidence intervals. We limit our analysis to the years 2000 and 2010, and use the average nightlight for each EA polygon. This is done so as to obtain results as comparable as possible to those reported so far. As before, mining variables represent average mining output in the preceding decade.

From the upper-left corner graph of Figure 14, we see that locations in the immediate proximity (0-10 km) of a mine have substantially more nightlight. Locations at 30-40 km from mines have marginally less nightlight. The first-difference regressions indicate that an increase in local mining activity is also associated with an increase in nightlight, but only at the mine site itself. We do not find a difference between expansion and contraction of a mine. Nightlight increases when a mine expands, and decreases by the same amount when it shrinks.

Are these results due to increased economic activity and prosperity, or are they entirely due to light emissions from mines? In Figure 15 we report results for electricity use by households as the outcome variable. In the cross section, we find that households in locations up to 30 km from mines are more likely to use electricity as their main source of lighting. The coefficients at 0-10 km of an expansion and a contraction have opposite signs: more households use electricity when a mine expands, but also when a mine contracts. Interestingly, the coefficient on contraction is twice as large in absolute magnitude than the coefficient on expansion. Households at locations at 30-40 km from an expanding mine use less electricity than they would otherwise.

The results for nightlight and the proportion of household using electricity partially contradict each other. Electricity use, a proxy for household material wellbeing, increases with a mine expansion, but even more so with a mine contraction. Effects on electricity use can also be felt further away from mines. These results are consistent with the pattern found for employment analysis. Nightlight, on the other hand, changes monotonically with changes in mine output. This could be due to a decrease in general economic activity that uses electricity intensively, but it could also be entirely due to the direct effect of lights from mines. However our finding that manufacturing – which is expected to use electricity more intensively – increases when mines contract does suggest that the nightlight results are perhaps driven by the intensity of mining itself.

6   Conclusion

Central place theory predicts that, when conditions are suitable for the creation of new urban centers, agglomeration effects driven by external shocks can determine where new urban centers locate. We are interested in testing the idea that a permanent urban settlement can arise as a result of a shock that temporarily generates a concentration of population in an arbitrary location. The external shock we examine here is gold mining, which tends to locate where minerals are found and is largely unaffected by other geographical features that make a site suitable for urban location. This is particularly true in southwestern Ghana where the geography is fairly undifferentiated – i.e. no mountains, no navigable rivers.

We investigate whether gold mines in Ghana are associated with tell-tale signs of proto-urbanization. Although gold has been extracted from Ghanaian soil for centuries, the last two decades have seen a large increase in production after the sector was liberalized and foreign investors entered gold mining. The Ghanaian economy has been growing steadily, and the population is expanding and urbanizing rapidly. These are ideal conditions to study town formation.

Using spatially disaggregated cross-sectional data, we find that EAs with gold mines have a larger proportion of people employed in services and trade, and a larger proportion of people employed in the wage, formal and public sectors, as apprentices, looking for work or being

homemakers. Population density and the density of non-farm jobs are higher close to mines. The presence of a gold mine is also associated with more nightlight and a higher proportion of households that light their homes with electricity, which suggests higher incomes and a more urban setting with electrification. We also find evidence suggesting that the agglomeration of non-farm employment (and several other tell-tale signs of urbanization) around gold mines is accompanied by a decrease of non-farm employment at distances of 20-30 km from gold mines. These findings are consistent with agglomeration effects that induce non-farm activities to coalesce in a particular location.

Slightly different results are found when we seek to identify the effect of gold mining over time. We find that an increase in the output of formal gold mines is associated with a higher proportion of the workforce in wage employment and apprenticeship, a higher proportion in the formal and public sectors, and a lower share in the non-wage and informal sectors. We also find a higher population density when a mine expands, but no difference in the number of jobs. These findings are broadly consistent with the cross-sectional analysis. However, we also find that the share of employment in manufacturing and trade decreases when a mine expands. Results for the share of agriculture, services, and even gold mining itself are mixed.

We also find that the changes associated with an expansion in formal gold production are not reversed when a formal gold mine shrinks or closes. To the contrary, they seem to be even stronger. The share of agricultural employment reduces after formal mines decrease gold production, and the share of trade and manufacturing employment increase, as well as the share of formal sector jobs. The same is true for population density, and for the number of jobs per square kilometer. Employment in informal gold mining increases too. The latter effect is likely to be a temporary phenomenon associated with the extraction of gold from tailings and closed mines using less technology-intensive methods. This phenomenon does not, by itself, constitute evidence of a durable urbanization legacy. Although the older and larger mines in our data are associated with large urban settlements, more time is needed to ascertain whether gold mining in Ghana can, like a grain of sand in a pearl, trigger the formation of self-sustaining agglomerations.

Bibliography

Fernando M. Aragon and Juan Pablo Rud (2013). "Natural Resources and Local Communities: Evidence from a Peruvian Gold Mine", *American Economic Journal: Economic Policy*, 5(2): 1-25.

Fernando M. Aragon & Juan Pablo Rud (2015). "Polluting Industries and Agricultural Productivity: Evidence from Mining in Ghana", *Economic Journal*, forthcoming.

Abhijit Banerjee, Esther Duflo and Nancy Qian (2012). "On the Road: Access to Transportation Infrastructure and Economic Growth in China", NBER working paper No. 17897.

Julia Bird and Stéphane Straub (2014). "The Brasília Experiment: Road Access and the Spatial Pattern of Long-term Local Development in Brazil", TSE Working Papers 14-495, Toulouse School of Economics.

Nathaniel Baum-Snow, Loren Brandt, J. Vernon Henderson, Matthew A. Turner and Qinghua Zhang (2014). "Roads, Railroads and Decentralization of Chinese Cities", mimeo.

Hoyt Bleakley and Jeffrey Lin (2012). "Portage and Path Dependence", *The Quarterly Journal of Economics*, 127(2):587-644.

Emmanuel J.M Carranza, Emmanuel A. Owusu and Martin Hale (2009). "Mapping of prospectivity and estimation of number of undiscovered prospects for lode Gold, southwestern Ashanti Belt, Ghana". *Miner Deposita*, 44: 915-938.

Amitabh Chandra and Eric Thompson (2000). "Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. ", *Regional Science and Urban Economics*, 30(4): 457-490.

Walter Christaller (1933). *Die zentralen Orte in Süddeutschland* (The Central Places in Southern Germany), Gustav Fischer, Jena.

Timothy Conley (1999). "GMM Estimation with cross-sectional dependence", *Journal of Econometrics*, 92(1): 1-45.

David Donaldson (2014). "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure", *American Economic Review*, forthcoming.

Klaus Desmet and Marcel Fafchamps (2006). "Employment Concentration across US Counties", *Regional Science and Urban Economics*, 36(4): 482-509.

Benjamin Faber (2014). "Trade Integration, Market Size, and Industrialization: Evidence

from China's National Trunk Highway System", *Review of Economic Studies*, 81(3):1046-1070.

Marcel Fafchamps and Forhad Shilpi (2003). "The Spatial Division of Labor in Nepal", *Journal of Development Studies*, 39(6): 23-66.

Marcel Fafchamps and Forhad Shilpi (2005). "Cities and Specialization: Evidence from South Asia", *Economic Journal*, 115(503): 477-504.

Marcel Fafchamps (2011). "Development, Social Norms, and Assignment to Task", *Proceedings of the National Academy of Sciences (PNAS)*, 108(Supplement 4): 21308-15.

Marcel Fafchamps (2012). "Reprint of Development, Agglomeration, and the Organization of Work", *Regional Science and Urban Economics*, 42(5): 765-78.

Edward L. Glaeser, Sari Pekkala Kerr, and William R. Kerr (2015). "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines", *Review of Economics and Statistics*, 97(2): 498-520.

Douglas Gollin, Rémi Jedwab, and Dietz Vollrath (2016). "Urbanization with and without Industrialization", *Journal of Economic Growth*, 21(1):35-70

Jan von der Goltz and Prabhat Barnwal (2014). "Mines: The Local Welfare Effects of Mineral Mining in Developing Countries", Discussion Paper Number 1314-19, Department of Economics, Columbia University.

J. Vernon Henderson (1985). *Economic Theory and the Cities*, 2nd edition, Academic Press.

J. Vernon Henderson, Adam Storeygard, and David Weil (2012). "Measuring Economic Growth from Outer Space", *American Economic Review*, 102(2): 994-1028.

Walter Isard (1956). *Location and Space-Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, and Urban Structure*, John Wiley and Sons, New York.

Jane Jacobs (1969). *The Economy of Cities*, Random House, New York.

Rémi Jedwab (2014). "Urbanization without Structural Transformation: Evidence from Consumption Cities in Africa", George Washington University, mimeograph.

Andreas Kotsadam and Anja Tolonen (2016). "African Mining, Gender, and Local Employment", *World Development*, forthcoming.

Guy Michaels (2008). "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System", *The Review of Economics and Statistics*, 90(4): 683-701

Stelios Michalopoulus and Elias Papaioannou (2013). "Pre-Colonial Ethnic Institutions and

Contemporary African Development", *Econometrica*, 81(1): 113-152.

Jon-Philippe Pigois, David I. Groves, Ian R. Fletcher, Neal J. McNaughton and Lawrence W. Snee (2003). "Age constraints on Tarkwaian paleoplacer and lode-gold formation in the Tarwa-Damang district, SW Ghana", *Mineralium Deposita*, 38: 695-714.

Ayowa Afrifa Taylor (2006). *An Economic History of the Ashanti Goldfields Corporation, 1895-2004: Land, Labour, Capital, and Enterprise.* PhD Thesis, London School of Economics.

United States Geological Survey, (1994-2010). "The Mineral Industry of Ghana" in: *Minerals Year Book*, ed. United States Geological Survey, various years.

Figures and Tables

Figure 1: Gold Mines and Urban Centers in Southern Ghana

(a) In 2000
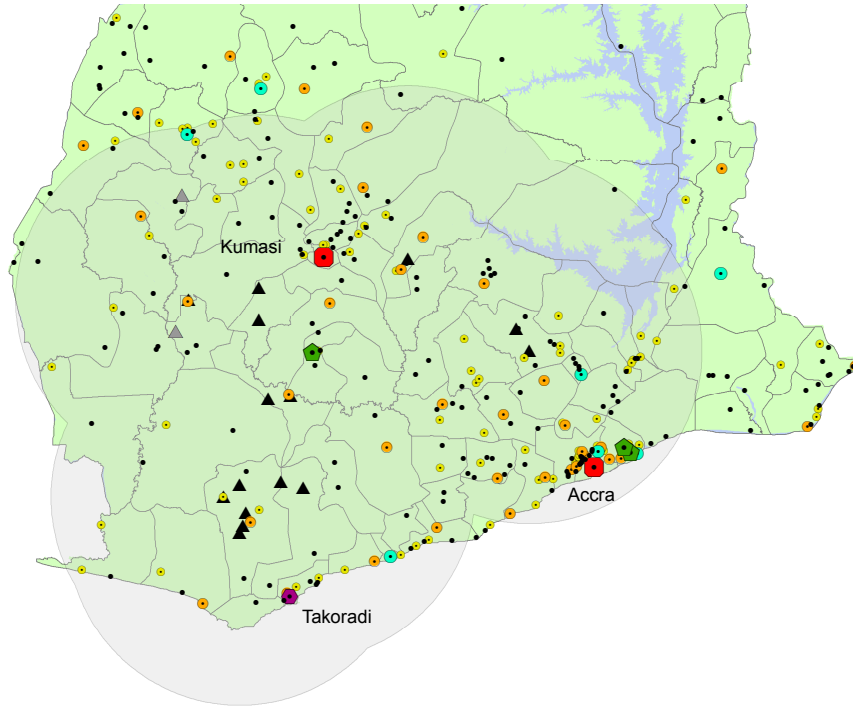
(b) In 2010
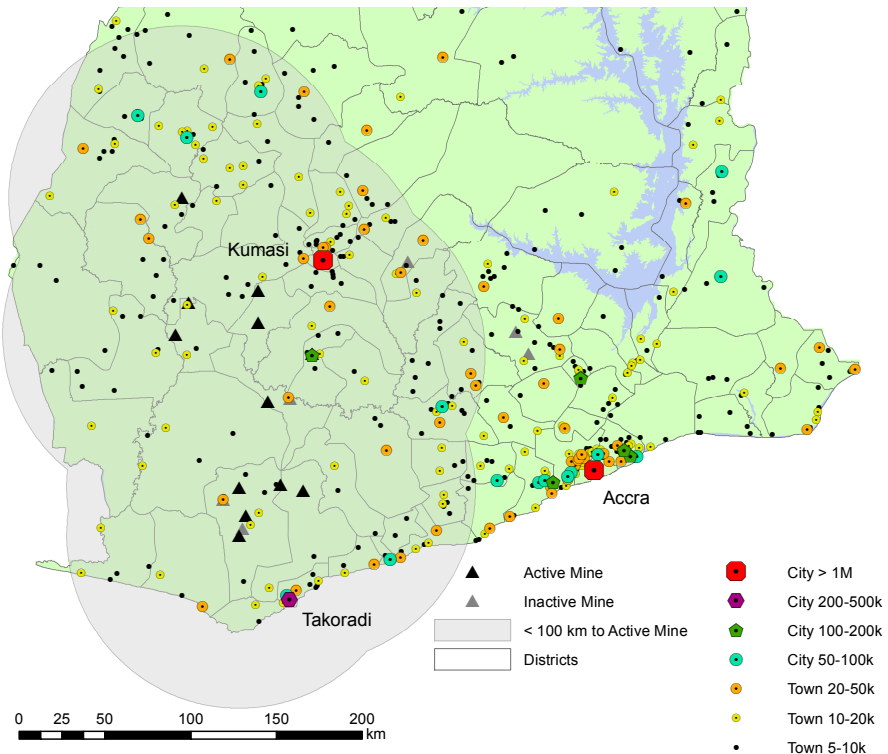
Figure 2: Gold production of 10 largest mines in Ghana, 1991-2010
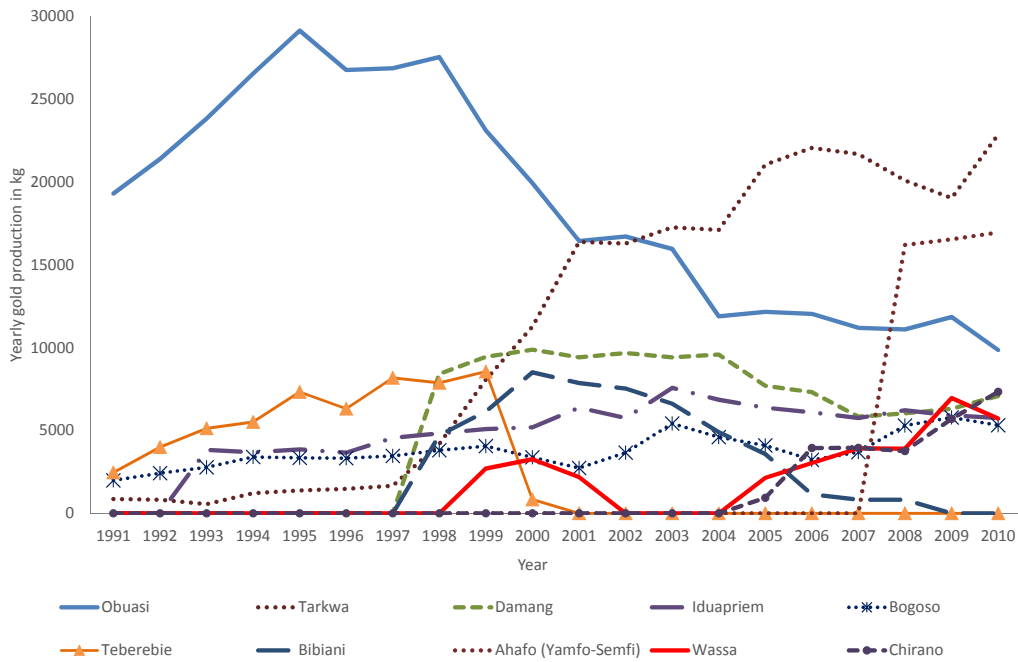
Notes: This figure shows yearly variation in production volumes of the ten largest (by average production volume) gold mines in Ghana from 1991 to 2010. Data sources are United States Geological Survey (USGS) annual reports on the Mineral Industry in Ghana, and individual mining company annual reports.

Figure 3: Gold Mining and Sector of Activity, 2010 Cross-Section
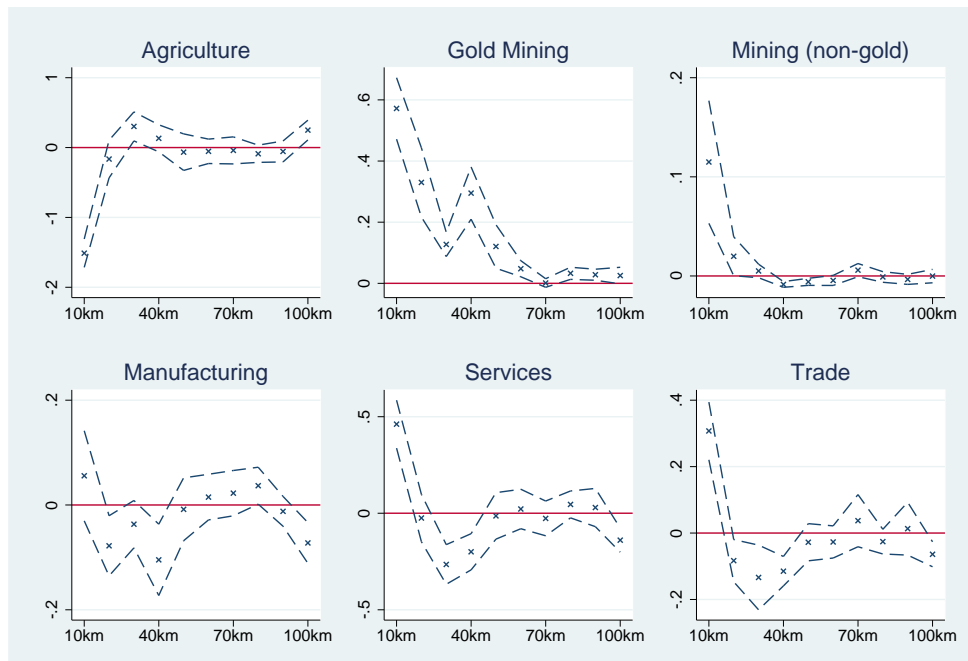
Notes: This figure shows the effect of gold mining on employment shares of different sectors at various distances from mines. Graphed are coefficients γ_k from regression of model (4), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of 1000kg gold production on employment shares (in range 0-100) at distance k. Omitted category are areas further than 100km from any mine. Regression controls for nearest distance to primary road, secondary road, and electricity grid; average slope and elevation of EA; region dummies; and a constant.
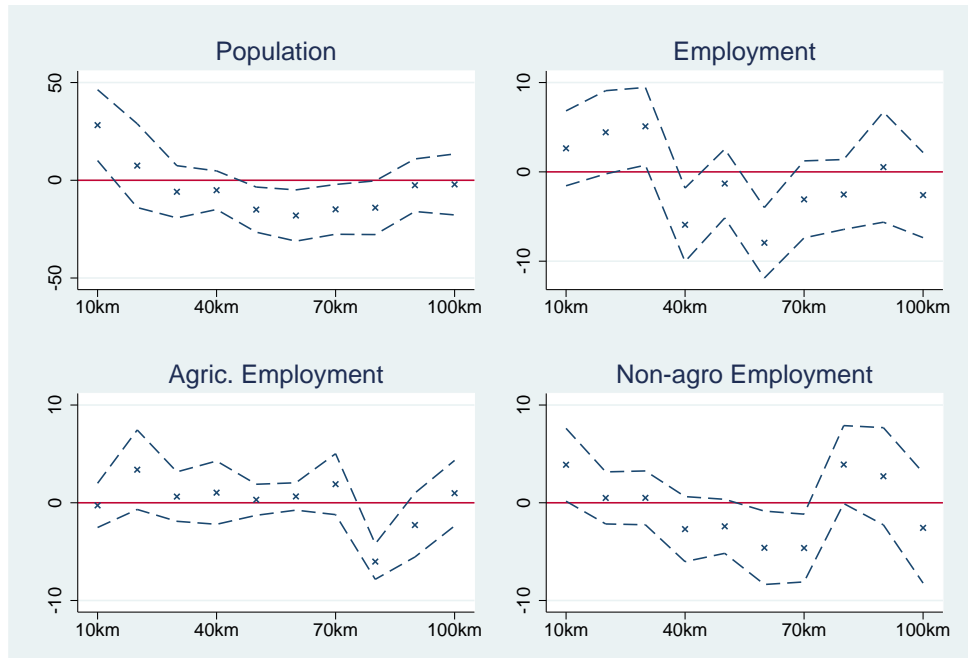
Figure 4: Gold Mining and Specialization, 2010 Cross-Section

Notes: This figure shows the effect of gold mining on employment shares of different forms of organization and specialization at various distances from mines. Graphed are coefficients γ_k from regression of model (4), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of 1000kg gold production on employment shares (in range 0-100) at distance k. Omitted category are areas further than 100km from any mine. Regression controls for nearest distance to primary road, secondary road, and electricity grid; average slope and elevation of EA; region dummies; and a constant.

Figure 5: Gold Mining and Population and Employment Density, 2010 Cross-Section
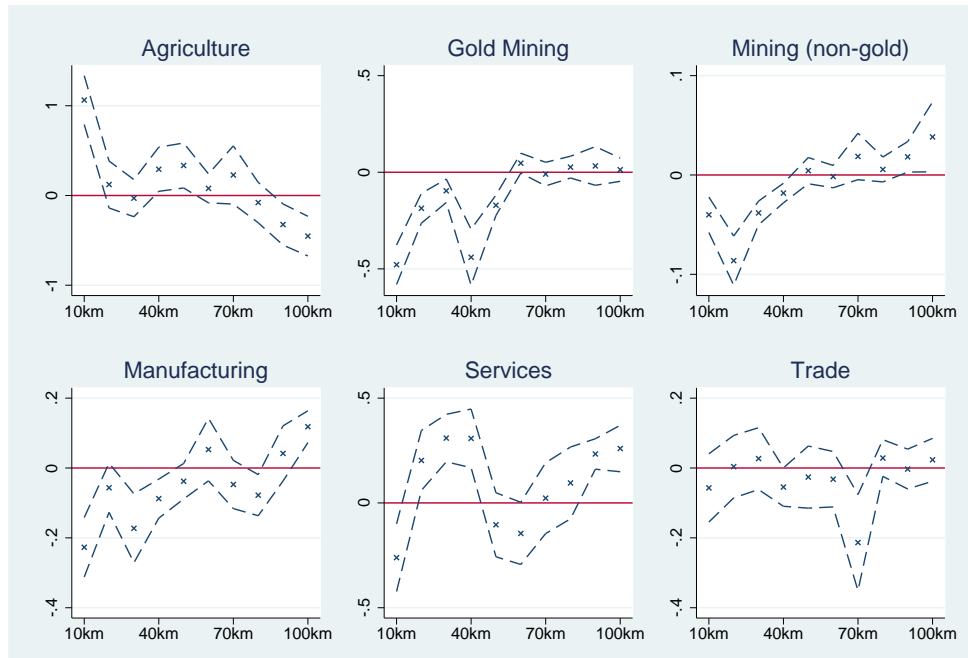

Notes: This figure shows the effect of gold mining on population and employment densities (number of people / area in km^2) at various distances from mines. Graphed are coefficients γ_k from regression of model (4), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of 1000kg gold production on densities at distance k. Omitted category are areas further than 100km from any mine. Regression controls for nearest distance to primary road, secondary road, and electricity grid; average slope and elevation of EA; region dummies; and a constant.

Figure 6: Gold Mining and Sector of Activity, Expanding Mine

Notes: This figure shows the effect of an *expansion* of gold mining on employment shares of different sectors at various distances from mines. Graphed are coefficients γ_k^+ from regression of model (8), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg increase in gold production on change in employment shares at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure 7: Gold Mining and Specialization, Expanding Mine

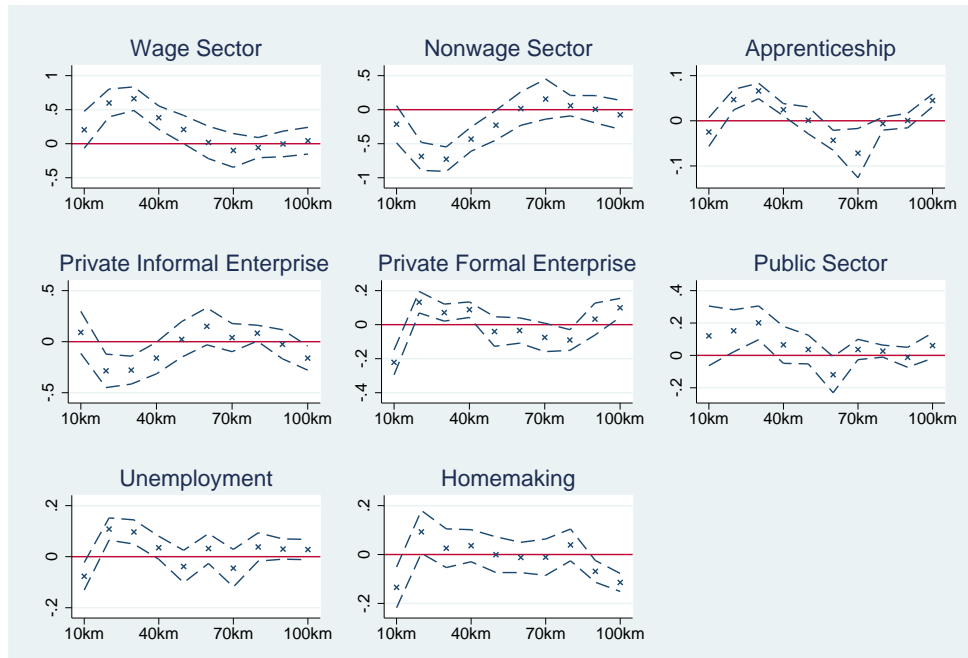*Notes:* This figure shows the effect of an *expansion* of gold mining on employment shares of different forms of organization and specialization at various distances from mines. Graphed are coefficients γ_k^+ from regression of model (8), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg increase in gold production on change in employment shares at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure 8: Gold Mining and Population and Employment Density, Expanding Mine

Notes: This figure shows the effect of an *expansion* of gold mining on population and employment densities (number of people / area in km^2) at various distances from mines. Graphed are coefficients γ_k^+ from regression of model (8), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg increase in gold production on change in densities at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure 9: Gold Mining and Sector of Activity, Contracting Mine

Notes: This figure shows the effect of a *contraction* of gold mining on employment shares of different sectors at various distances from mines. Graphed are coefficients γ_k^- from regression of model (8), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg decrease (a *negative* number) in gold production on change in employment shares at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure 10: Gold Mining and Specialization, Contracting Mine

Notes: This figure shows the effect of a *contraction* of gold mining on employment shares of different forms of organization and specialization at various distances from mines. Graphed are coefficients γ_k^- from regression of model (8), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg decrease (a *negative* number) in gold production on change in employment shares at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure 11: Gold Mining and Population and Employment Density, Contracting Mine

Notes: This figure shows the effect of a *contraction* of gold mining on population and employment densities (number of people / area in km^2) at various distances from mines. Graphed are coefficients γ_k^- from regression of model (8), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg decrease (a *negative* number) in gold production on change in densities at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure 12: Breakdown of Gold Mining Employment, Expanding Mine
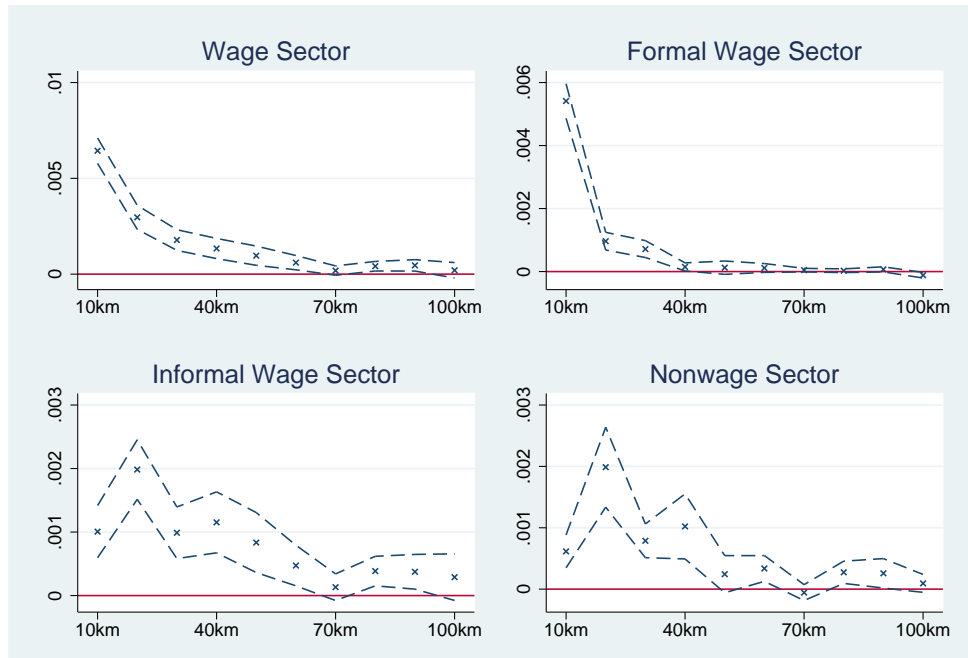
Notes: This figure breaks down the effect of an *expansion* of gold mining on the employment share in gold mining (from Figure 6) into different organizational and legal forms. Graphed are coefficients γ_k^+ from regression of model (8), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg increase in gold production on change in employment shares at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure 13: Breakdown of Gold Mining Employment, Contracting Mine

Notes: This figure breaks down the effect of a *contraction* of gold mining on the employment share in gold mining (from Figure 9) into different organizational and legal forms. Graphed are coefficients γ_k^- from regression of model (8), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg decrease (a *negative* number) in gold production on change in employment shares at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure 14: Gold Mining and Night Lights
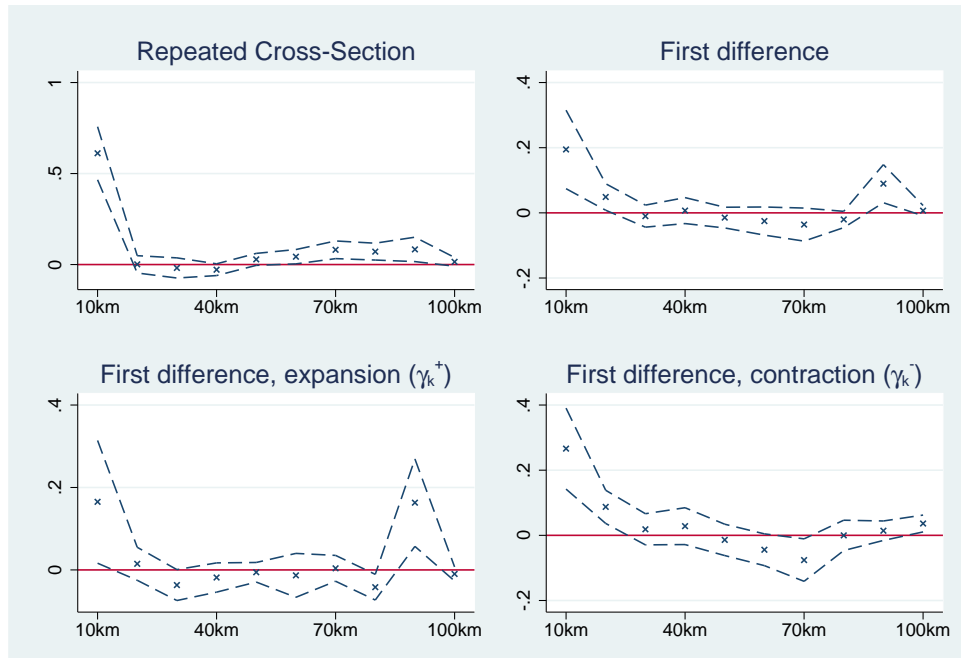
Notes: This figure shows the effect gold mining on night lights. Outcome variable is average luminosity (in digital numbers, range 0-63) of EA. Graphed are coefficients γ_k from regression of model (4) (top right), γ_k from model (7) (top left), and coefficients γ_k^+ (bottom left) and γ_k^- (bottom right) from model (8); labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Omitted category are areas further than 100km from any mine. All regressions control for (in case of first differences, differential trends by) nearest distance to primary road, secondary road, electricity grid, and electrical power plant; by average slope and elevation of EA; by region; and a constant.

Figure 15: Gold Mining and Household Use of Electricity

Repeated Cross-Section

First difference

First difference, expansion (γ_k^+)

First difference, contraction (γ_k^-)
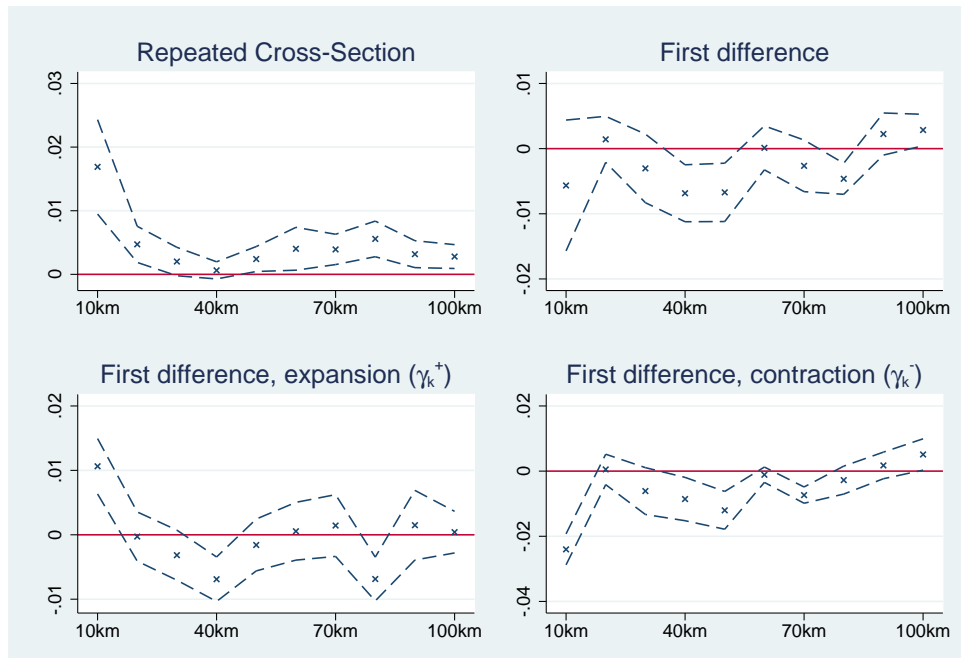
Notes: This figure shows the effect gold mining on household use of electricity. Outcome variable is percentage of households using mains electricity as primary source of lighting, obtained from census data. Graphed are coefficients γ_k from regression of model (4) (top right), γ_k from model (7) (top left), and coefficients γ_k^+ (bottom left) and γ_k^- (bottom right) from model (8); labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Omitted category are areas further than 100km from any mine. All regressions control for (in case of first differences, differential trends by) nearest distance to primary road, secondary road, electricity grid, and electrical power plant; by average slope and elevation of EA; by region; and a constant.

Table 1 - Gold Mines in Ghana, 1991-2010

Mine Name	District (in 2000 System)	Region	Average yearly gold production in kg		Opening	Closure
			1991-2000	2001-2010		
Obuasi	Adansi West	Ashanti	24,442	12,921		
Obotan	Amansie West	Ashanti	1,599	657	1997	2003
Bonte (alluvial)	Amasie West	Ashanti	759	482		
Obenemasi / Konongo	Asante Akim North	Ashanti	188	0	1998	
Ahafo	Asutifi	Brong Ahafo	0	4,970	2008	
Ayanfuri	Upper Denkyira	Central	936	36		2002
Dunkwa	Upper Denkyira	Central	97	0		1999
Asikam (alluvial)	East Akim	Eastern	27	0		2000
Goldenrae (alluvial)	East Akim	Eastern	26	0	1993	
Bibiani	Bibiani Anwhiaso-Bekwai	Western	1,938	3,330		
Chirano	Bibiani Anwhiaso-Bekwai	Western	0	2,564	2005	
Wassa	Mpohor Wassa	Western	598	2,787	1999*)	
Damang	Wassa West	Western	2,775	7,841		
Iduapriem	Wassa West	Western	3,473	6,268		
Prestea	Wassa West	Western	945	0		
Bogoso	Wassa West	Western	3,200	4,386		
Tarkwa	Wassa West	Western	3,154	19,383		
Teberebie	Wassa West	Western	5,621	0		2001

Notes: Production volume, openings and closures of formal gold mines in Ghana, 1991-2010. Data sources: USGS Minerals Industry in Ghana, annual reports 1992-2010; company annual reports. *) Wassa was not operating from 2002 to 2004.

Table 2: Labor Market in rural Ghana in 2000 and 2010

	2000		2010	
	Number	**%**	**Number**	**%**
A. Sectoral Composition				
Agriculture	257,921	75.28	283,386	62.82
Gold Mining	1,813	0.53	4,934	1.09
Other Mining	1,609	0.47	1,005	0.22
Manufacturing	23,264	6.79	38,434	8.52
Services	36,560	10.67	75,741	16.79
Trade	21,429	6.25	47,603	10.55
Total	342,596	100.00	451,103	100.00
B. Employment Sector				
Wage sector				
Wage-employed	30,611	8.94	51,008	11.46
Employers	12,891	3.76	14,409	3.24
Non-wage sector				
Self-employed	262,026	76.48	304,347	68.37
Family workers	29,943	8.74	65,761	14.77
Apprentice	4,265	1.26	6,200	1.39
Other	2,661	0.78	3,439	0.77
Total	339,174	100.00	445,164	100.00
C. Employment by type of firm / organization				
Public (including parastatal)	14,605	4.31	19,262	4.33
Private formal	15,873	4.68	13,540	3.04
Private informal	308,076	90.83	410,664	92.25
Other	620	0.18	1,698	0.38
Total	339,174	100.00	445,164	100.00
D. Labor Supply				
Working	342,596	81.93	451,103	83.28
Unemployed	27,269	6.52	16,738	3.09
Individuals who are homemakers	23,506	5.62	29,879	5.52
Other (student, pensioner, etc.)	24,801	5.93	43,981	8.11
Total	418,172	100.00	541,701	100.00

Notes: Data sources are Ghana Population and Housing Census 2000 and 2010, 10% sample stratified at the level of enumeration area (EA). This table shows employment for all individuals of working age (20-60 years old) living in rural areas as defined in the Census.

Appendix for online publication

Gold Mining and Proto-Urbanization: Recent Evidence from Ghana

Marcel Fafchamps, Michael Koelle, and Forhad Shilpi

A   Data Appendix

Geo-referencing of Census data

We obtained a map of the enumeration areas (EA) of the 2000 census from the Ghana Statistical Service (GSS). This map shows the boundaries for all individual rural EA, and the boundaries for all urban localities, defined as distinct population clusters with more than 5000 inhabitants that have a name or locally recognised status. Our map contains the boundaries of 12,472 such areas. We further know which EA from the Census belongs to each urban locality.

Furthermore, we have a list that produces a mapping from the 2010 to 2000 EA. That is, we know for each EA from the 2010 Census the corresponding EA from the 2000 census that covers the geographical area of the 2010 EA. This allows us to merge the 2010 EAs onto our map. We are able to link 11,698 of areas (rural EA and urban localities) across the two Census years, corresponding to 96.17% of all households in 2000 and 98.51% of all households in 2010. We subsequently restrict our geo-referenced dataset to EA that were rural in 2000. By 2010, these areas were either rural, or belonged to newly formed urban localities, in 2010.

We were able to link 10,708 of 11,244 of these initially rural EAs across Census years. Missing geo-referenced observations are due to any of the following three reasons. Firstly, there are some polygons on the map for which the link to a Census EA is missing. Secondly, there are some 2000 Census EAs for which we do not know the corresponding polygon. Thirdly, there are 2010 Census EAs for which we do not know the corresponding 2000 Census EA.

Mapping of nightlights into EA

We obtain annual nighlights data from the Defense Meteorological Satellite Program Operational Linescan System, using the Average Visible, Stable Lights, and Cloud Free Coverages series. We overlay the gridded night lights dataset with our map of enumeration areas. For each satellite-

year, we assign to each EA the average value of nightlight within the EA polygon boundaries. In years with two satellite observations we then take the average across satellites. The result we obtain is our EA-level nightlights measure.

Slope and elevation

We obtained elevation and slope data from the NASA shuttle radar topography mission that were constructed at a 3 arc-second (approximately 90m) grid. Our elevation and slop measures were constructed as EA averages.

Distance to nearest primary and secondary road

We calculate the distance (in km) from each EA centroid to the nearest primary and secondary road of the 2006 road network in Ghana. Distance to electricity grid and power plants. We obtained the location of the electricity transmission grid, and of electrical power plants, from the World Bank Africa Infrastructure Country Diagnostic (AICD), accessed and downloaded from http://www.infrastructureafrica.org/.

B   Additional tables and figures

Figure A1: Night Lights in Ghana, Mining Region
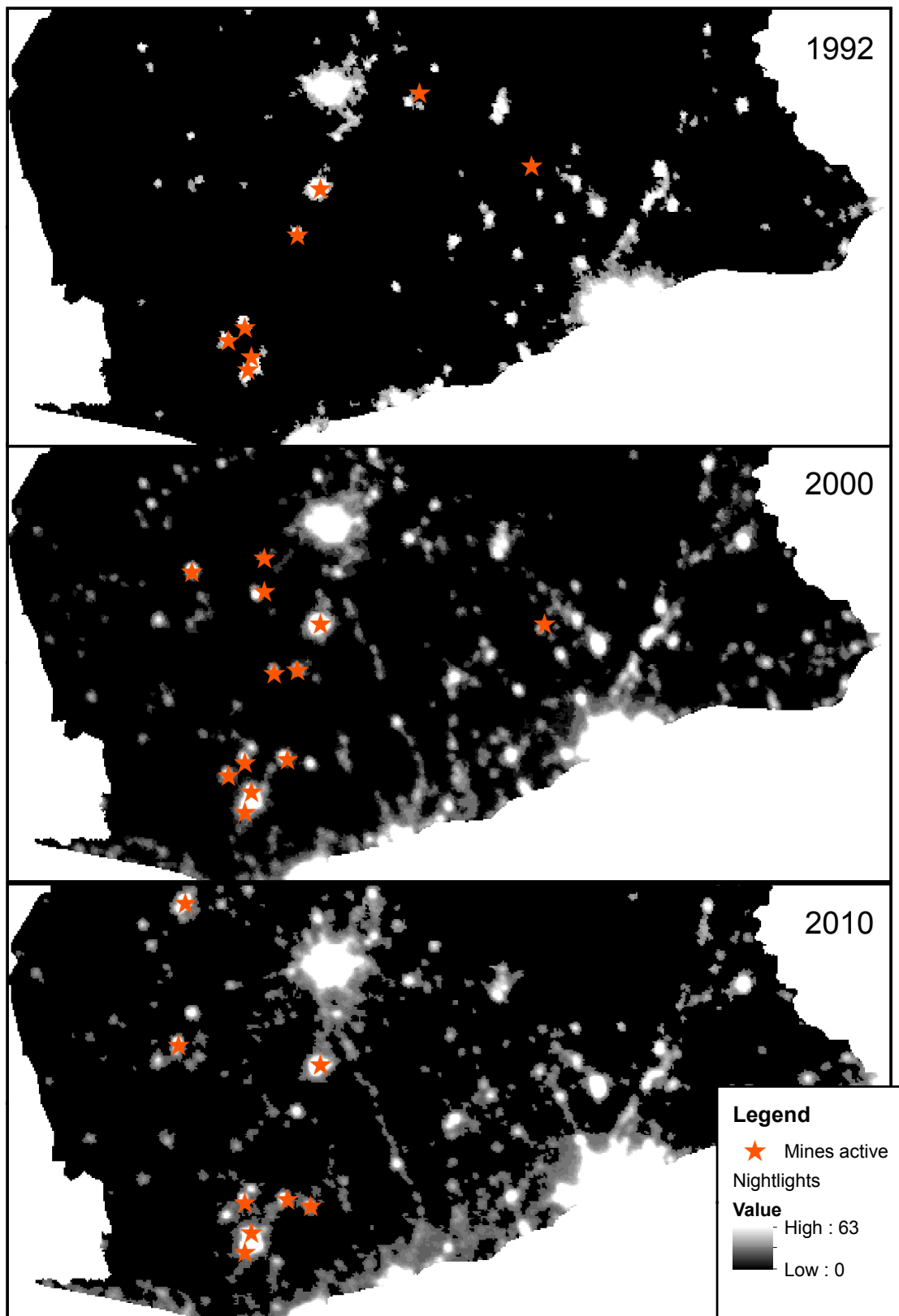
Figure A2: Gold Mining and Sector of Activity, First Differences

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.

Notes: This figure shows the effect of gold mining on employment shares of different sectors at various distances from mines using a first-differenced model. Graphed are coefficients γ_k from regression of model (7), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg change in gold production on change in employment shares at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure A3: Gold Mining and Specialization, First Differences

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.

Notes: This figure shows the effect of gold mining on employment shares of different forms of organization and specialization at various distances from mines using a first-differenced model. Graphed are coefficients γ_k from regression of model (7), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg change in gold production on change in employment shares at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Figure A4: Gold Mining and Population and Employment Density, First Differences

Notes: This figure shows the effect of gold mining on population and employment densities (number of people / area in km^2) at various distances from mines using a first-differenced model. Graphed are coefficients γ_k from regression of model (7), labeled with the upper end of the distance bin (e.g. $k = 1$ collecting gold production at 0-10km from EA is labeled '10km'). Dashed lines are 95% confidence intervals, with spatial standard errors using the method of Conley (1999). Each coefficient represents the marginal effect of a 1000kg change in gold production on change in densities at distance k. Omitted category are areas further than 100km from any mine. First-differenced regression controls for differential trends by nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant.

Table A1: Night Lights Data for Several Periods

	1992-2010	1990s	2000s	1992	2000	2010
DN 0 (Unlit)	92.90%	94.07%	91.82%	96.88%	91.34%	89.62%
DN 1-5	5.19%	4.22%	6.07%	1.91%	6.61%	5.79%
DN 6-10	1.04%	0.92%	1.15%	0.65%	1.14%	2.88%
DN 11-20	0.40%	0.36%	0.44%	0.26%	0.42%	0.87%
DN 20-62	0.46%	0.41%	0.50%	0.30%	0.48%	0.76%
DN 63	0.01%	0.01%	0.01%	0.01%	0.01%	0.07%

Notes: This table tabulates the distribution of night lights, measured in Digital Numbers (DN) that ranges from 0 to 63, across EA in Ghana for different years and time periods. Data were obtained from DMSP-OLS. When multiple satellite-year observations were available, the average across satellites was taken as the value for a given year.

Table A2: Distribution of Gold Production by Distance Bin

	0-10km	10-20km	20-30km	30-40km	40-50km	50-60km	60-70km	70-80km	80-90km	90-100km	>100km
Average yearly gold production 1991-2000											
Mean	0.12	0.29	0.45	0.69	0.95	0.94	1.15	1.20	1.36	1.47	41.14
SD	1.43	2.29	2.83	3.47	4.15	4.00	4.46	4.41	4.76	4.97	14.43
Max	24.44	26.04	27.07	27.22	32.02	34.48	36.42	34.89	37.35	40.83	49.78
% Zero	97.3%	93.6%	89.6%	84.0%	80.6%	76.8%	75.5%	74.9%	75.1%	73.0%	0.0%
Average yearly gold production 2001-2010											
Mean	0.14	0.32	0.48	0.71	1.04	1.29	1.34	1.51	1.65	1.79	55.36
SD	1.40	2.00	2.40	2.80	3.78	4.41	4.11	4.28	4.47	4.84	16.66
Max	25.65	34.39	36.28	30.04	36.28	38.31	43.48	40.63	40.14	42.99	65.62
% Zero	97.7%	94.7%	92.6%	88.8%	85.8%	82.8%	80.8%	78.4%	77.3%	76.1%	0.0%

Notes: All statistics in this table refer to a variable at the EA level measuring production in metric tons at different distances from the EA. This variable is denoted m_{ikt} in the paper. Distribution taken across all rural EA for the average production of the periods 1991-2000 and 2001-2010, respectively. Rows labeled '% Zero' refer to percentage of EA for which there is no mine producing gold at the respective distance interval.

C   Robustness checks

We perform a number of robustness checks that are described in sections 5.2 and 5.4 in the paper. Here we report the results in the same way we report the main results in the paper, by graphing coefficients γ_k, γ_k^+ and γ_k^- as appropriate. All regressions control for (in case of first differences, differential trends by) nearest distance to primary road, secondary road, and electricity grid; by average slope and elevation of EA; by region; and a constant. Any additional controls are specified in the figure headers.

Figure A5: Robustness - Gold Mining and Sector of Activity, 2000 Cross-Section

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
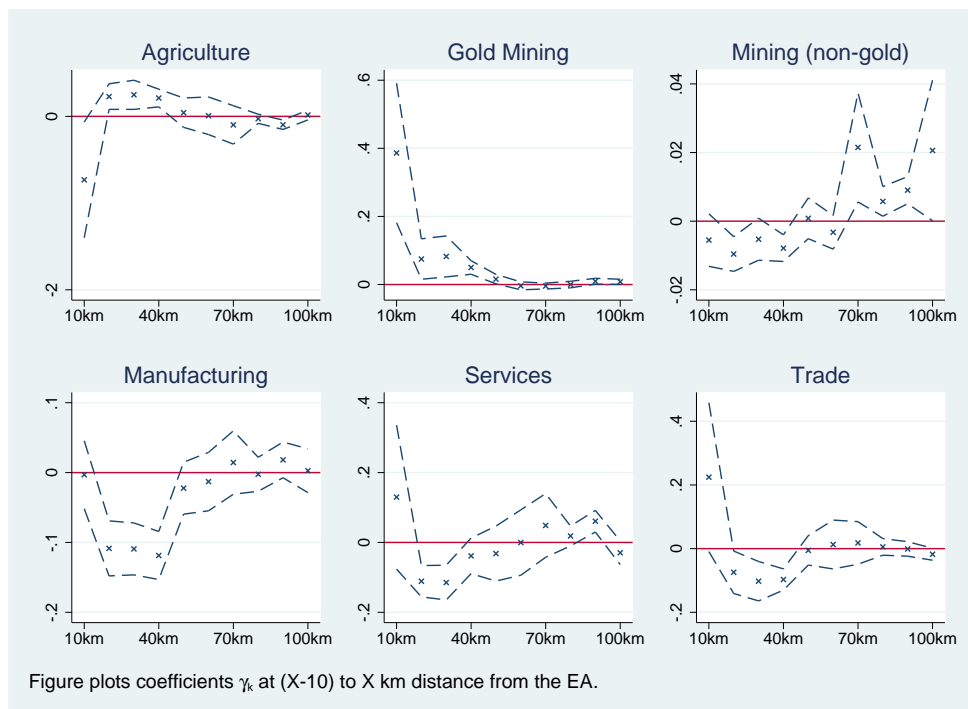
Figure A6: Robustness - Gold Mining and Specialization, 2000 Cross-Section

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
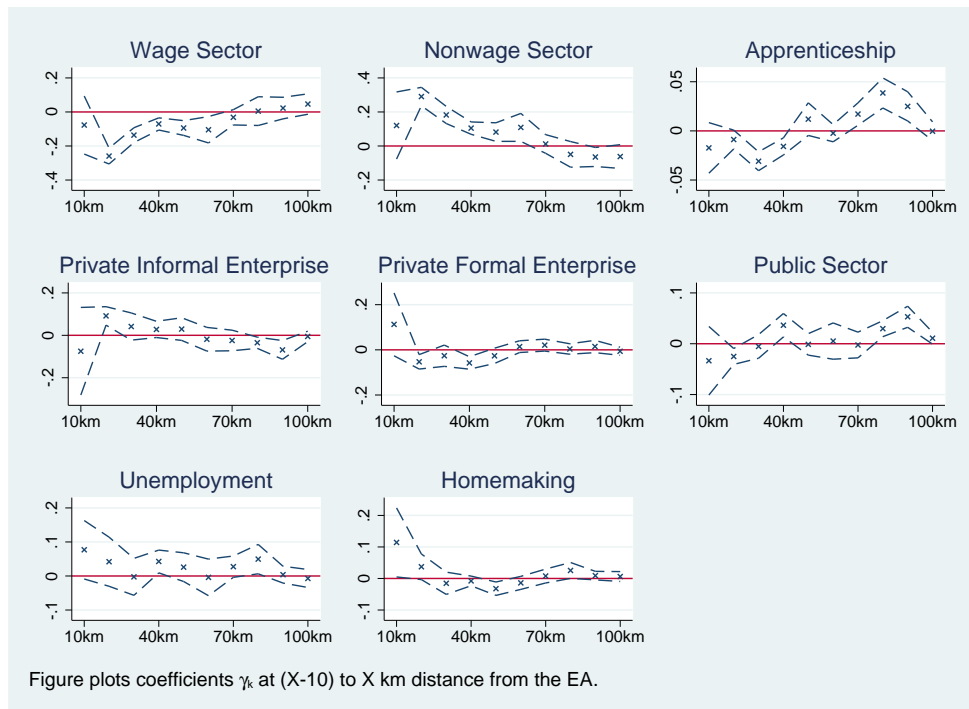
Figure A7: Robustness - Gold Mining and Population and Employment Density, 2000 Cross-Section

58
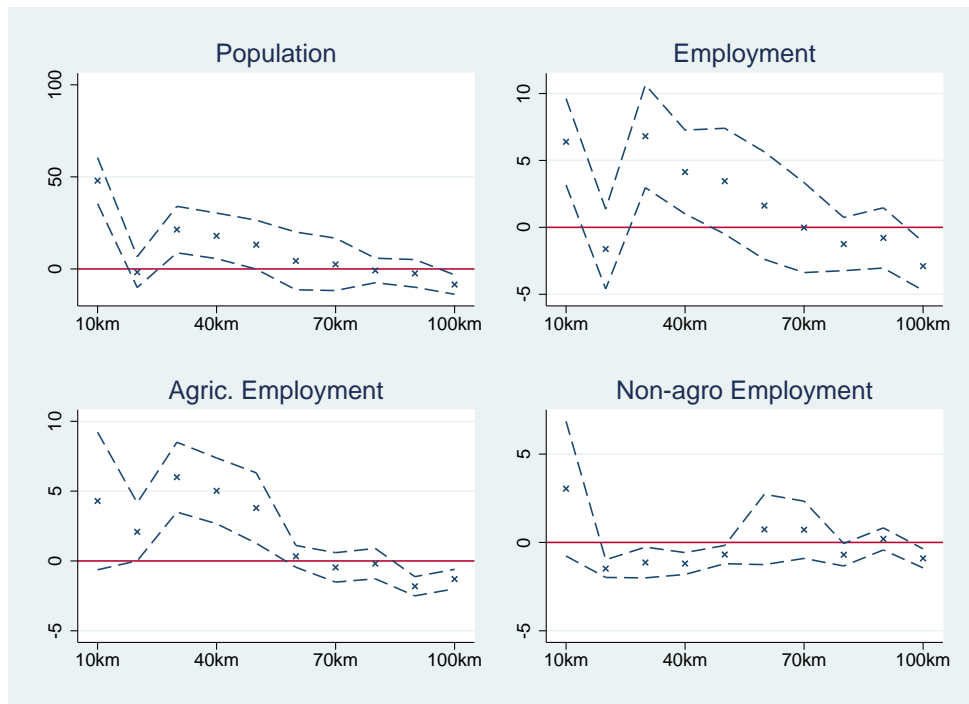
Figure A8: Robustness - Gold Mining and Sector of Activity, 2010 Cross-Section controlling for distance to cities

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
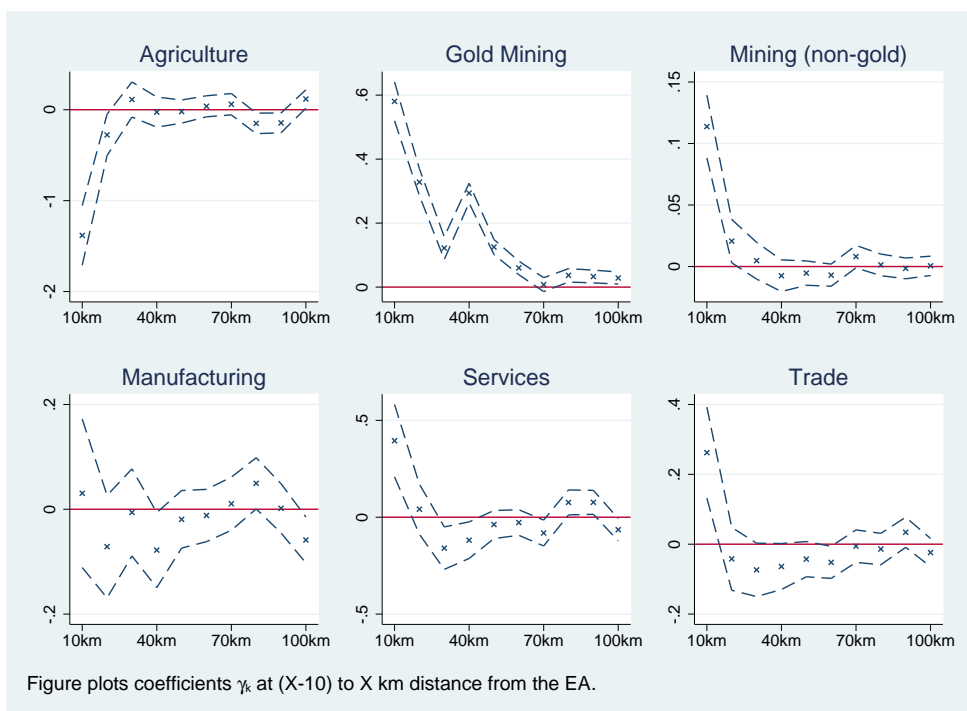
Figure A9: Robustness - Gold Mining and Specialization, 2010 Cross-Section controlling for distance to cities

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
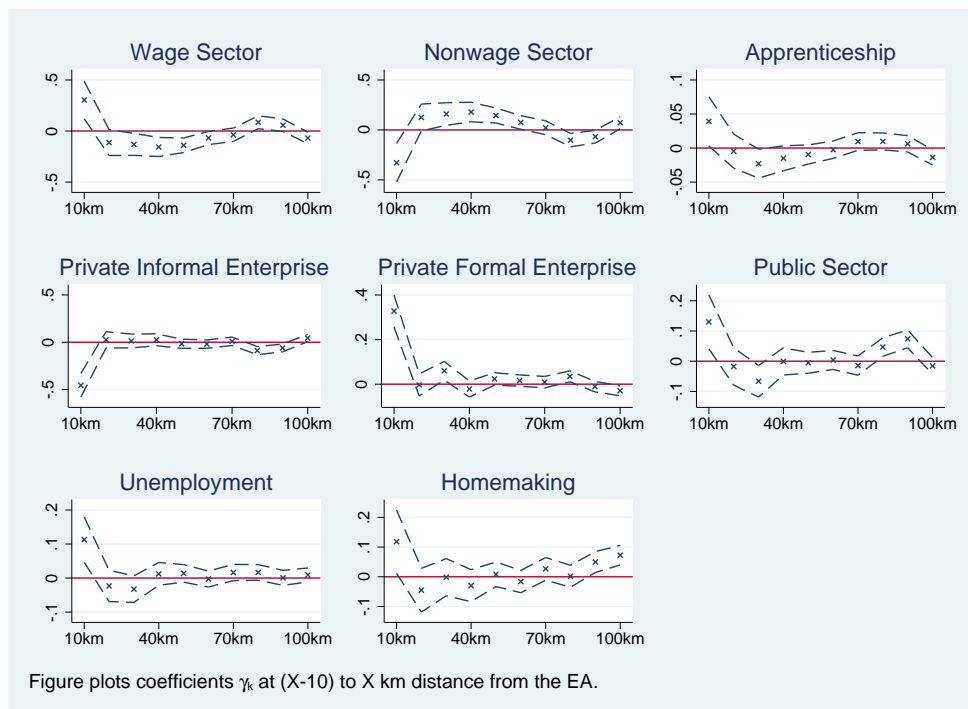
Figure A10: Robustness - Gold Mining and Population and Employment Density, 2010 Cross-Section controlling for distance to cities
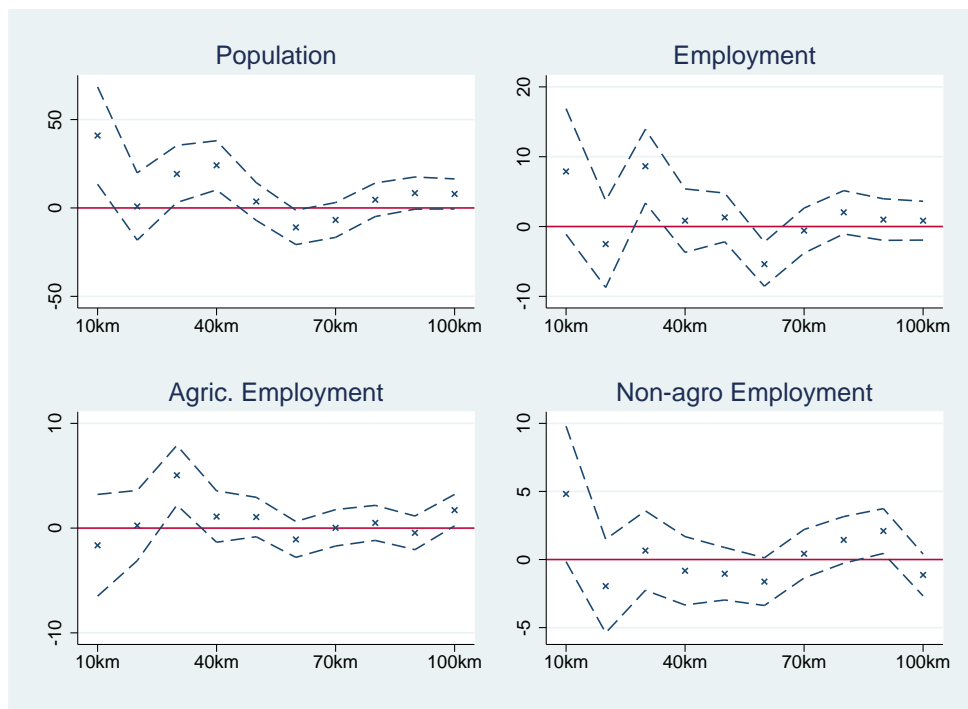
Figure A11: Robustness - Gold Mining and Sector of Activity, 2010 Cross-Section truncated at 50 km

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
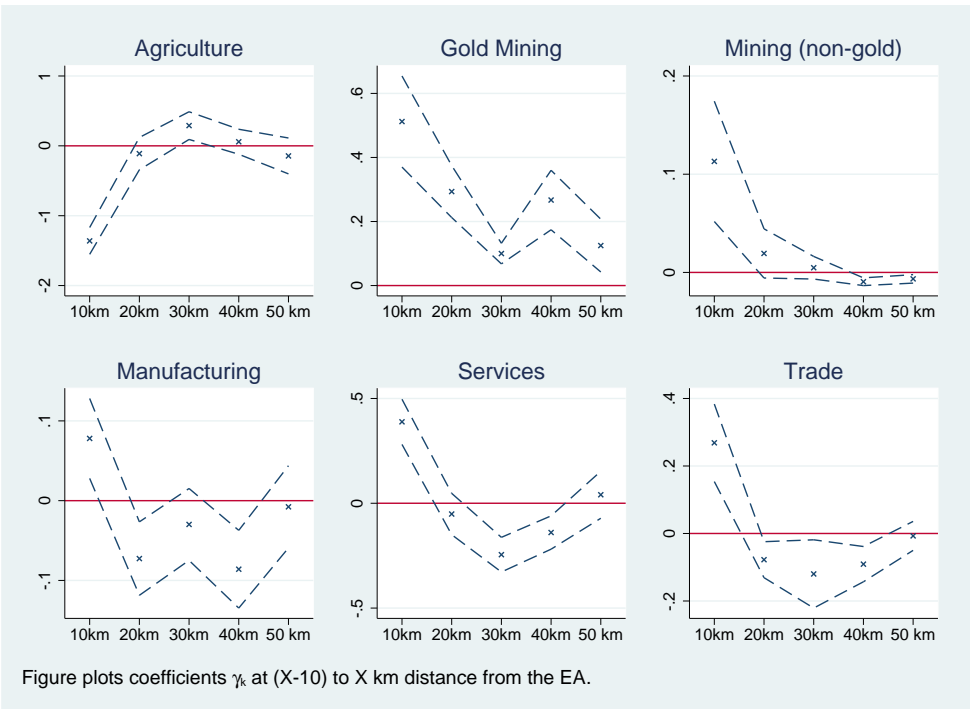
Figure A12: Robustness - Gold Mining and Specialization, 2010 Cross-Section truncated at 50 km

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
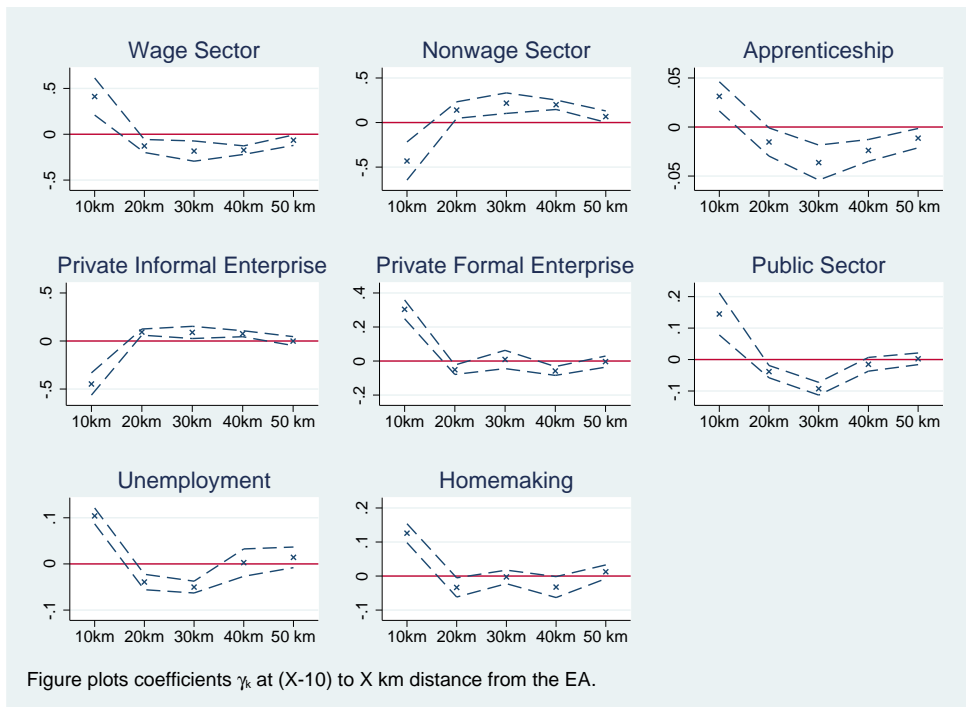
Figure A13: Robustness - Gold Mining and Population and Employment Density, 2010 Cross-Section truncated at 50 km
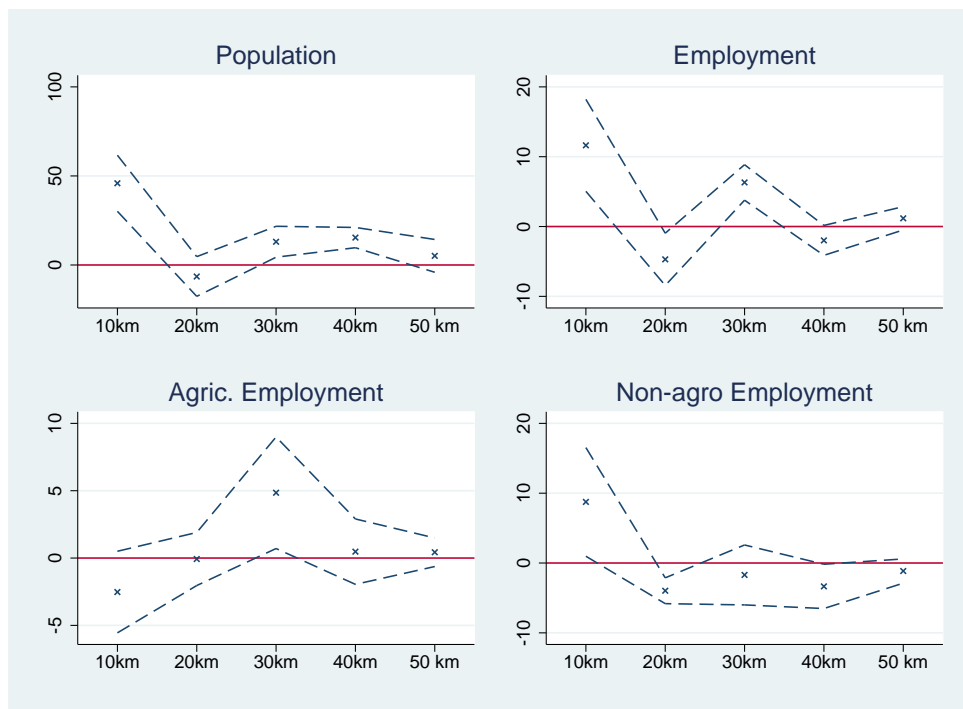
Figure A14: Robustness - Gold Mining and Sector of Activity, 2010 Cross-Section with average gold production 2006-2010

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
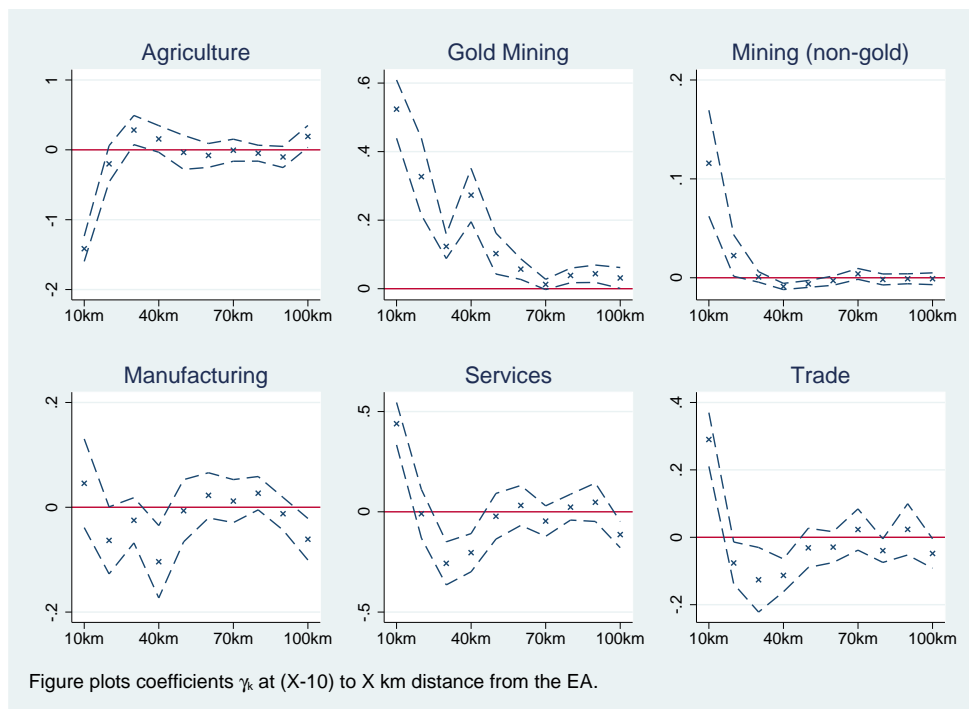
Figure A15: Robustness - Gold Mining and Specialization, 2010 Cross-Section with average gold production 2006-2010

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
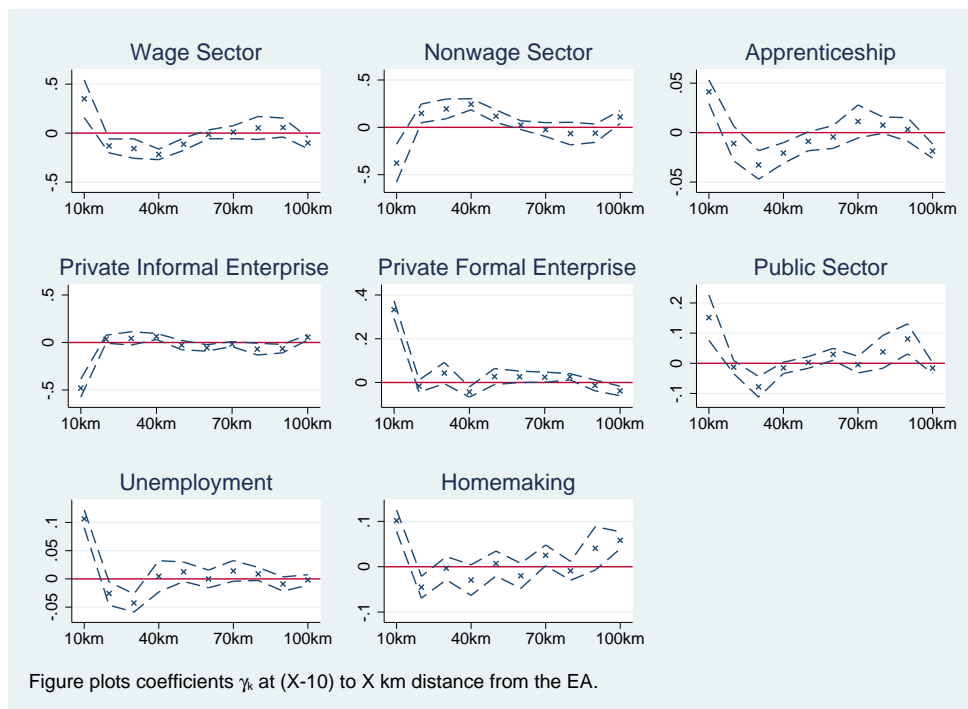
Figure A16: Robustness - Gold Mining and Population and Employment Density, 2010 Cross-Section with average gold production 2006-2010
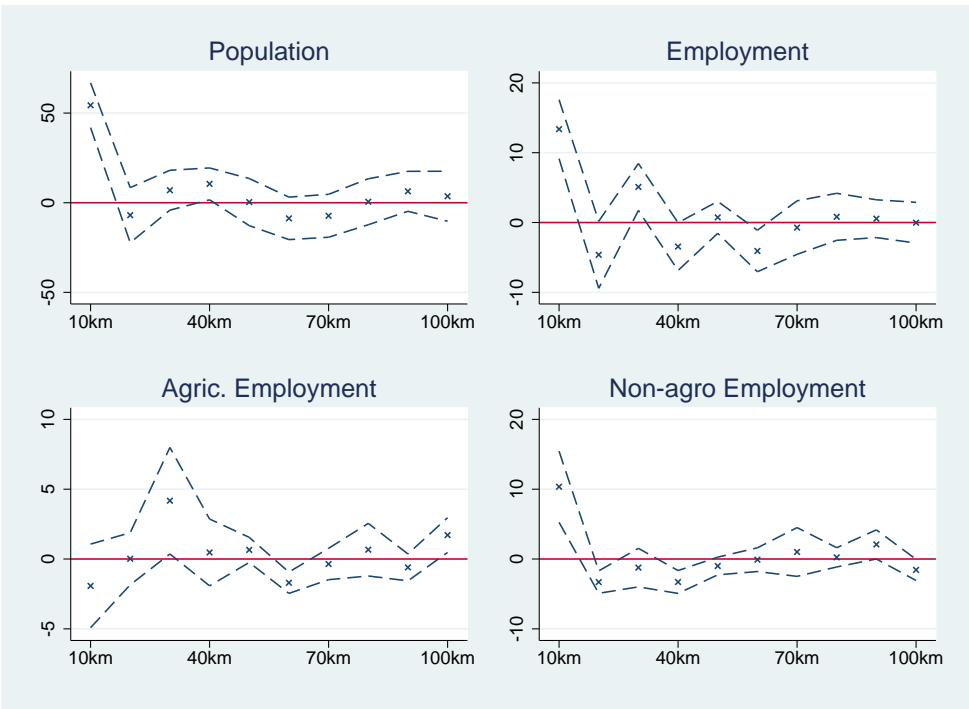
Figure A17: Robustness - Extensive Margin of Gold Mining and Sector of Activity, 2010 Cross-Section

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
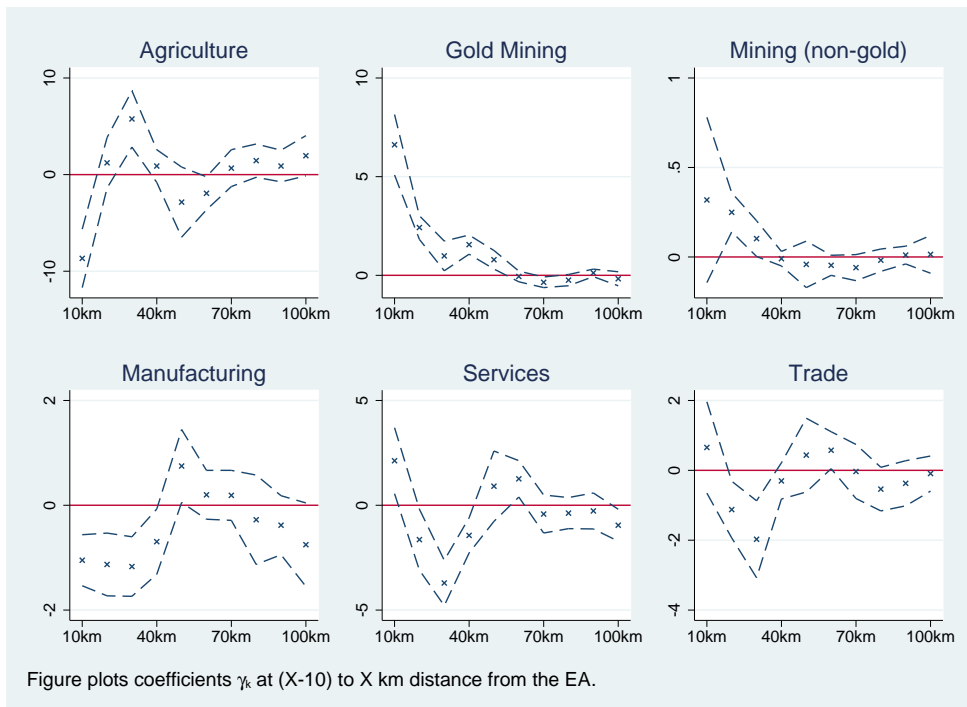
Figure A18: Robustness - Extensive Margin of Gold Mining and Occupational Choice, 2010 Cross-Section

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
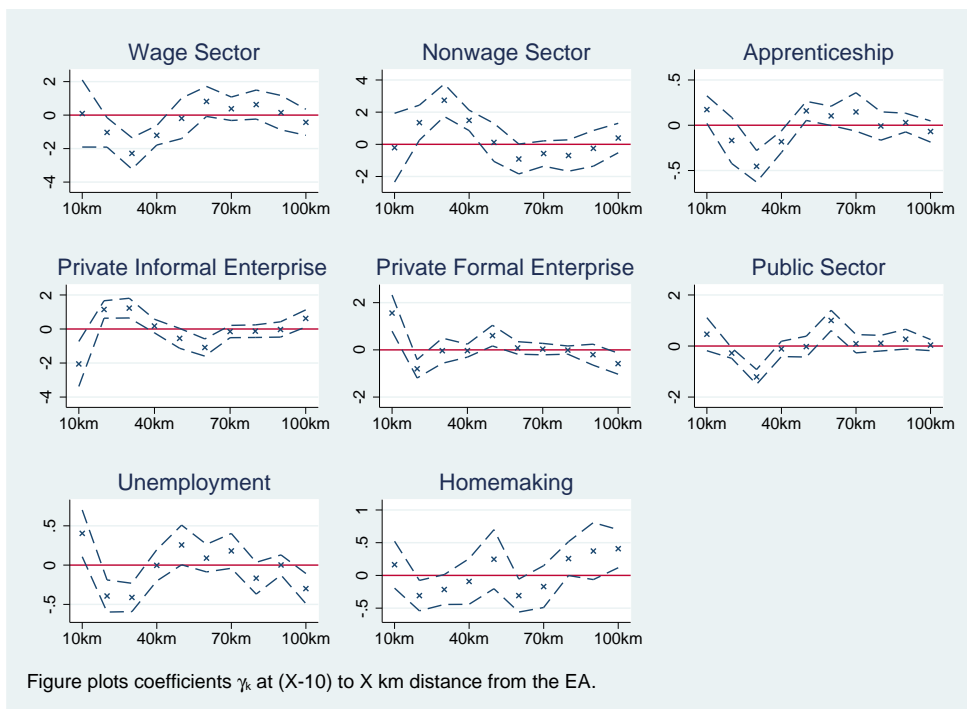
Figure A19: Robustness - Extensive Margin of Gold Mining and Population and Employment Density, 2010 Cross-Section

Figure A20: Robustness - Gold Mining and Sector of Activity, Mine expansion, controlling for distance to cities

Figure plots coefficients γ_k^+ at (X-10) to X km distance from the EA.
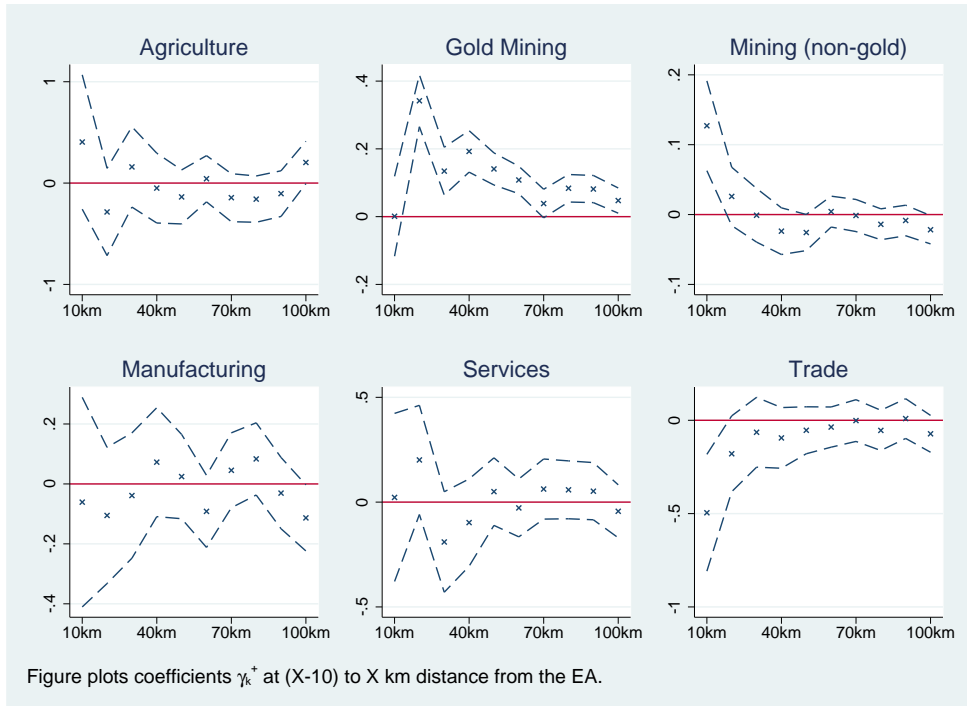
Figure A21: Robustness - Gold Mining and Specialization, Mine expansion, controlling for distance to cities

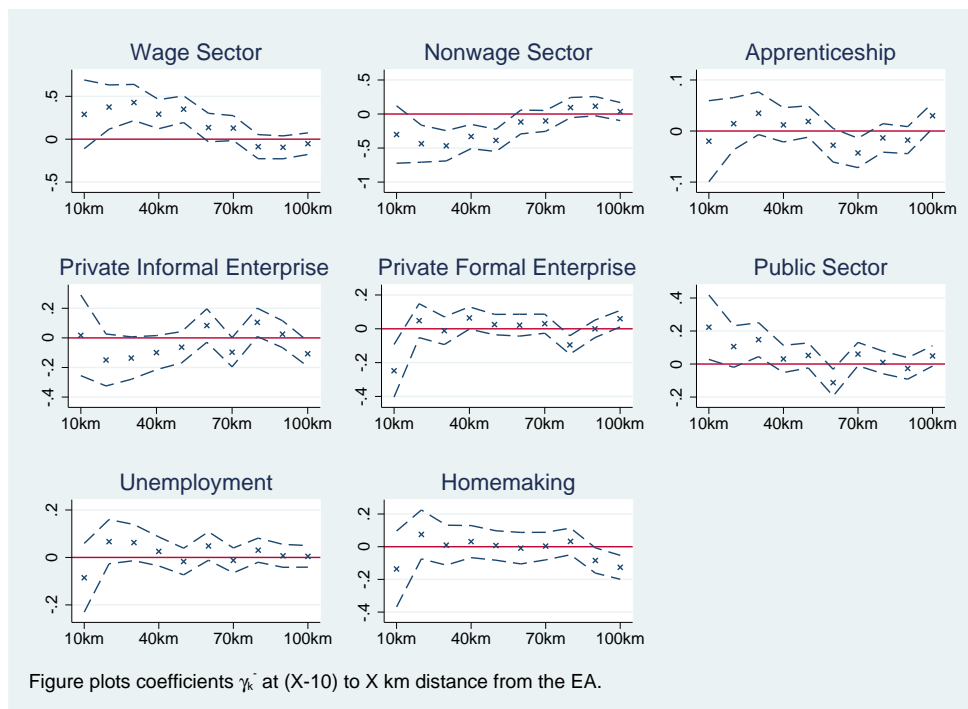Figure plots coefficients γ_k^+ at (X-10) to X km distance from the EA.
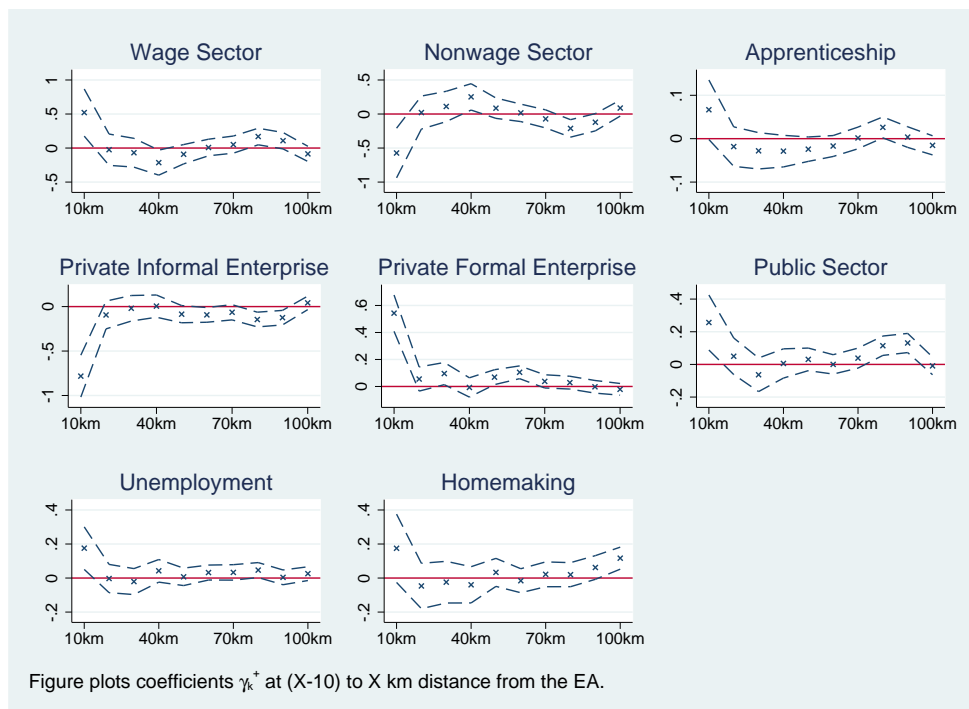
Figure A22: Robustness - Gold Mining and Population and Employment Density, Mine expansion, controlling for distance to cities


Figure plots coefficients γ_k^+ at (X-10) to X km distance from the EA.
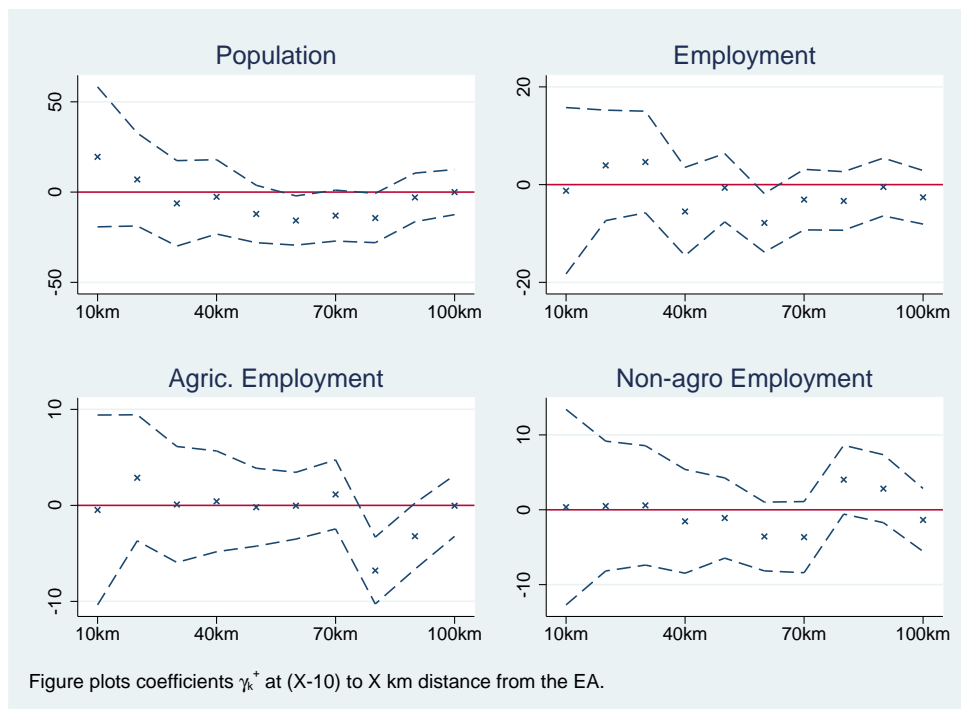
Figure A23: Robustness - Gold Mining and Sector of Activity, Mine contraction, controlling for distance to cities

Figure plots coefficients γ_k^- at (X-10) to X km distance from the EA.
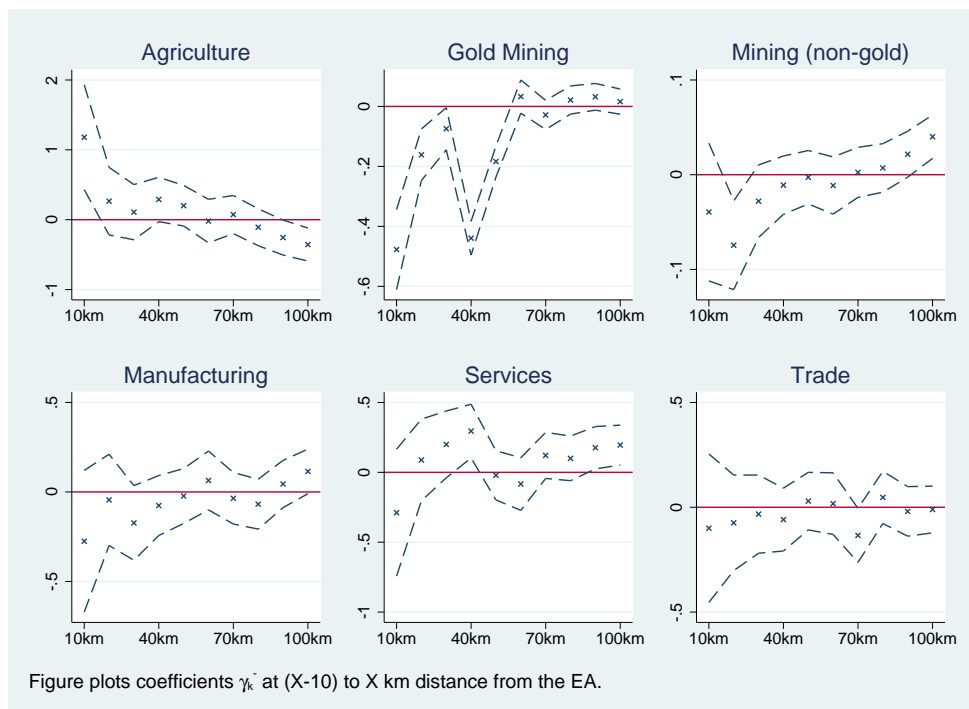
Figure A24: Robustness - Gold Mining and Specialization, Mine contraction, controlling for distance to cities

Figure plots coefficients γ_k^- at (X-10) to X km distance from the EA.

Figure A25: Robustness - Gold Mining and Population and Employment Density, Mine contraction, controlling for distance to cities

Figure plots coefficients γ_k^- at (X-10) to X km distance from the EA.
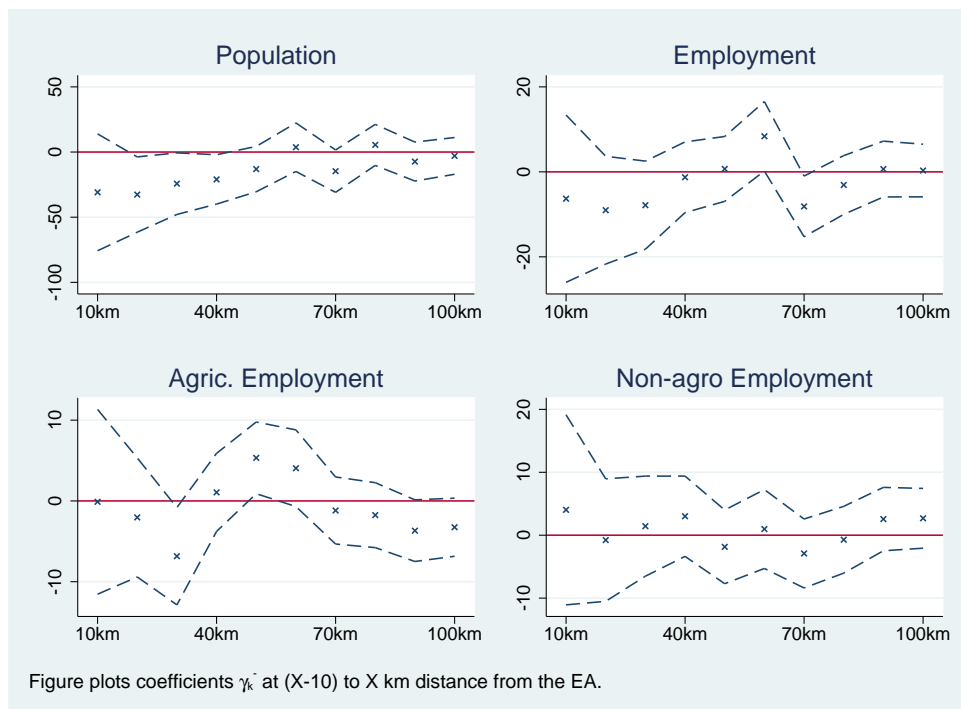
Figure A26: Robustness - Gold Mining and Sector of Activity, Mine expansion, truncated at 50 km

Figure plots coefficients γ_k^+ at (X-10) to X km distance from the EA.
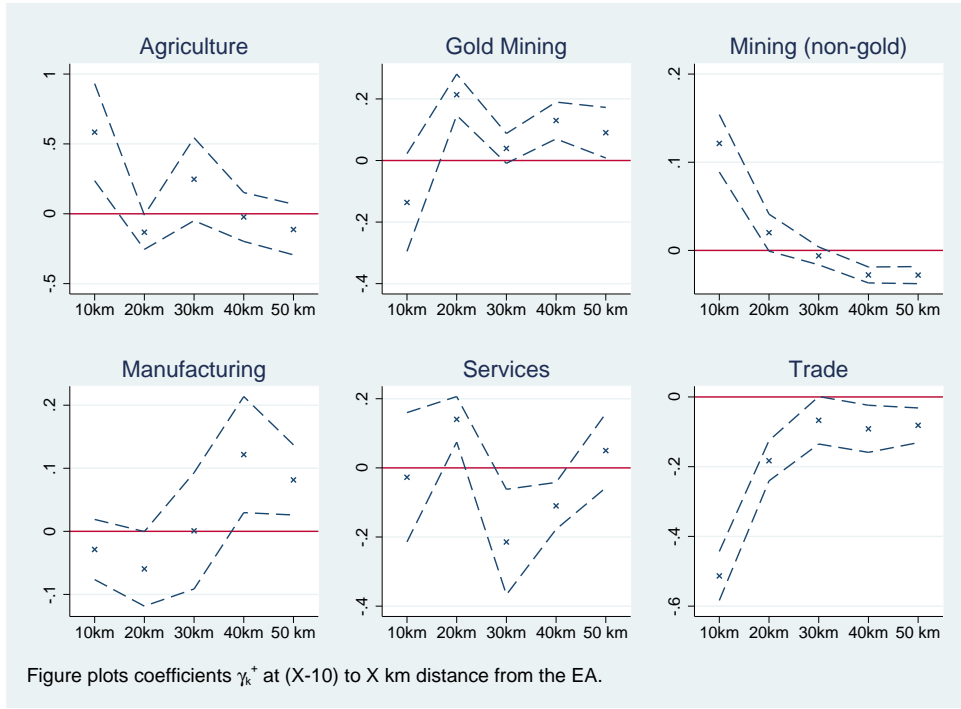
Figure A27: Robustness - Gold Mining and Specialization, Mine expansion, truncated at 50 km

Figure plots coefficients γ_k^+ at (X-10) to X km distance from the EA.
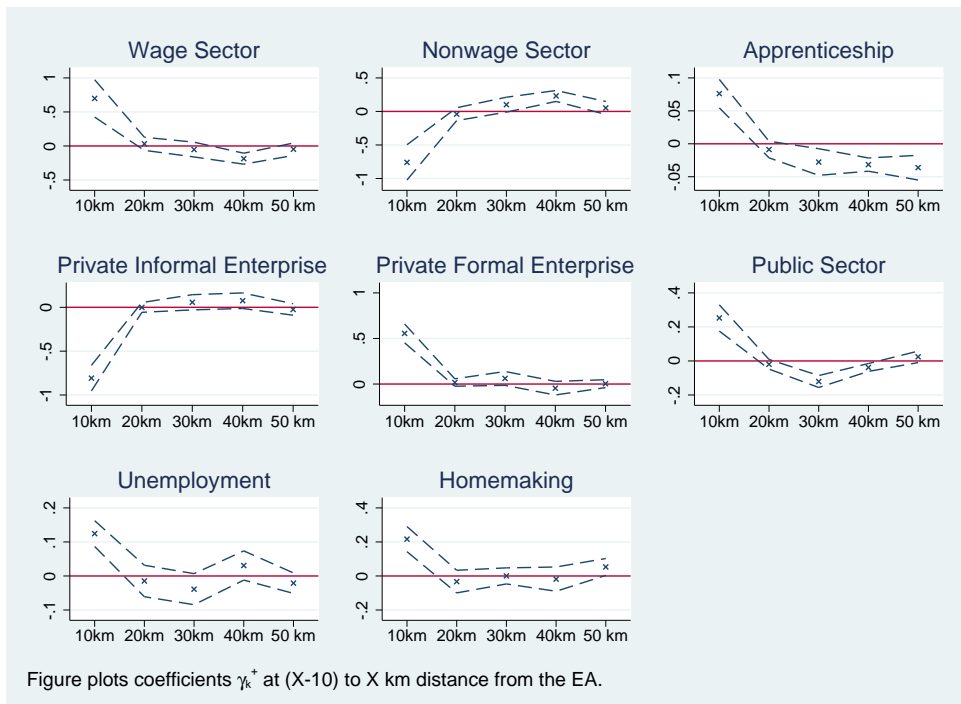
Figure A28: Robustness - Gold Mining and Population and Employment Density, Mine expansion, truncated at 50 km

Figure plots coefficients γ_k^+ at (X-10) to X km distance from the EA.
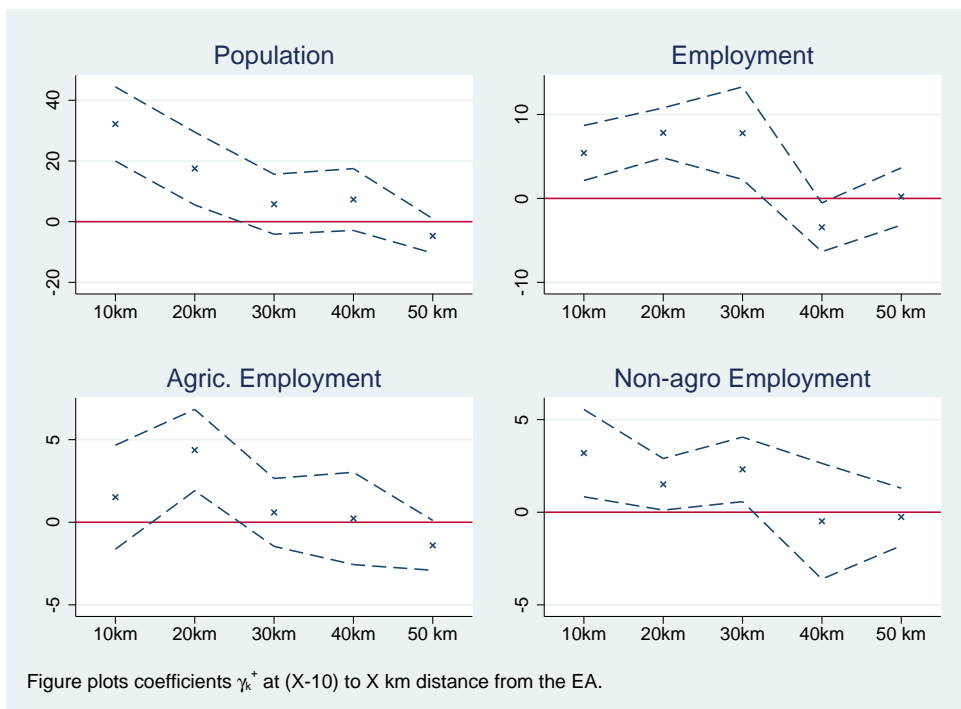
Figure A29: Robustness - Gold Mining and Sector of Activity, Mine contraction, truncated at 50 km

Figure plots coefficients γ_k^{-} at (X-10) to X km distance from the EA.
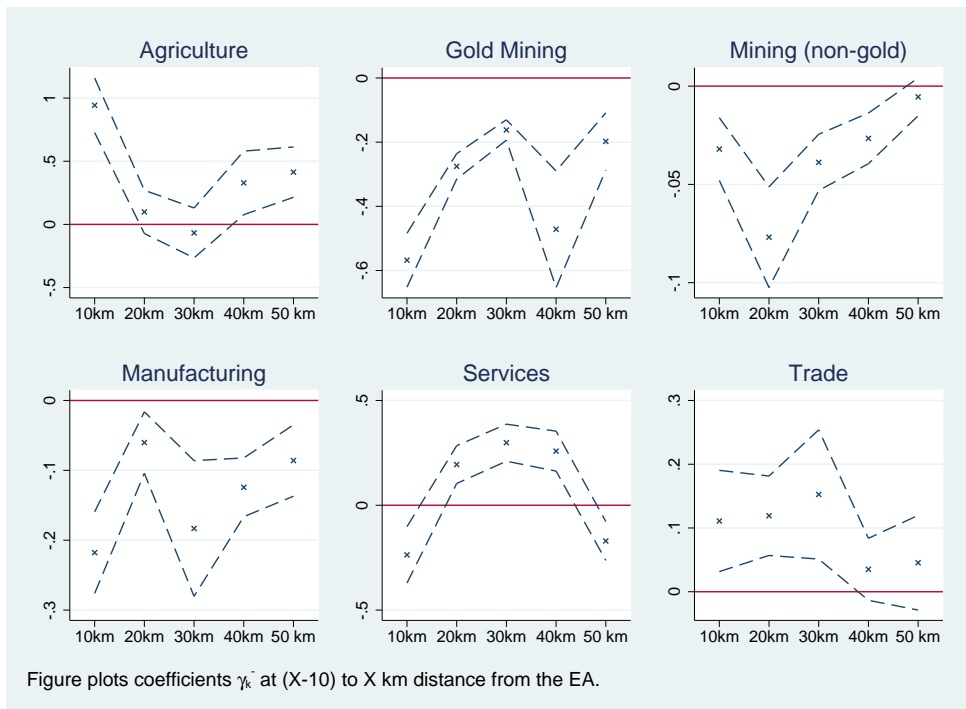
Figure A30: Robustness - Gold Mining and specialization, Mine contraction, truncated at 50 km

Figure plots coefficients γ_k at (X-10) to X km distance from the EA.
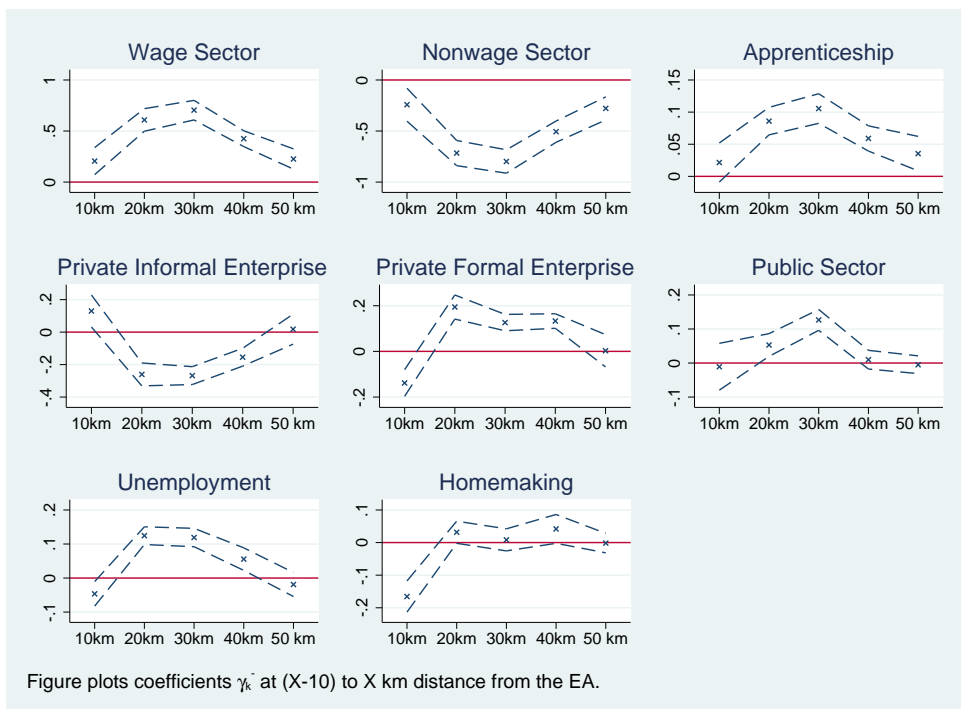
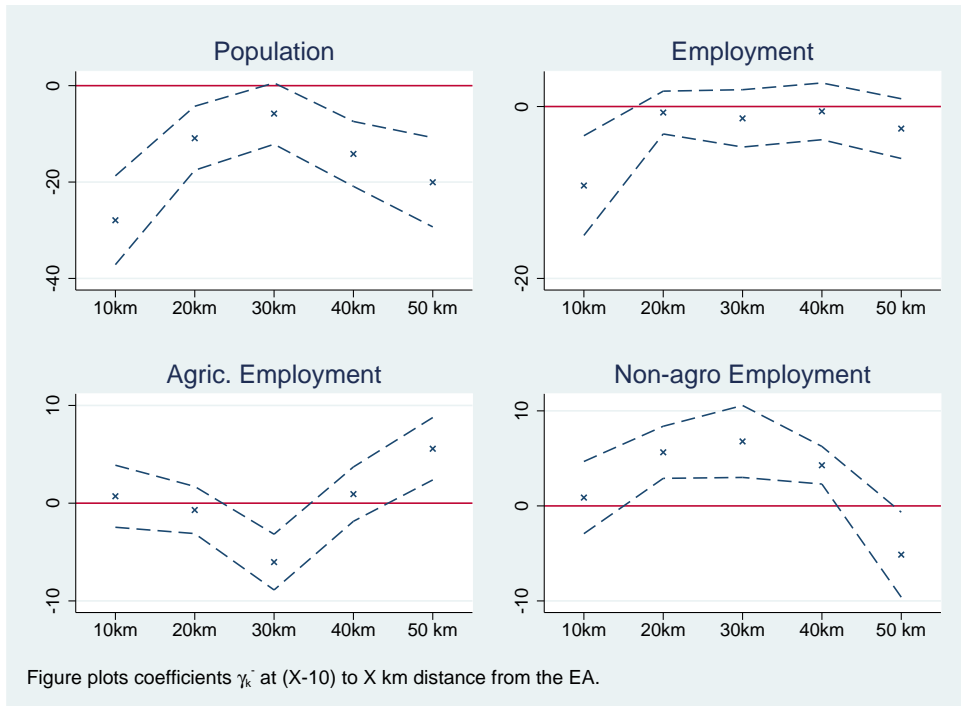Figure A31: Robustness - Gold Mining and Population and Employment Density, Mine contraction, truncated at 50 km

Figure plots coefficients γ_k^- at (X-10) to X km distance from the EA.

Figure A32: Robustness - Extensive Margin Gold Mining and Sector of Activity, Mine openings

Figure plots coefficients γ_k^+ at (X-10) to X km distance from the EA.
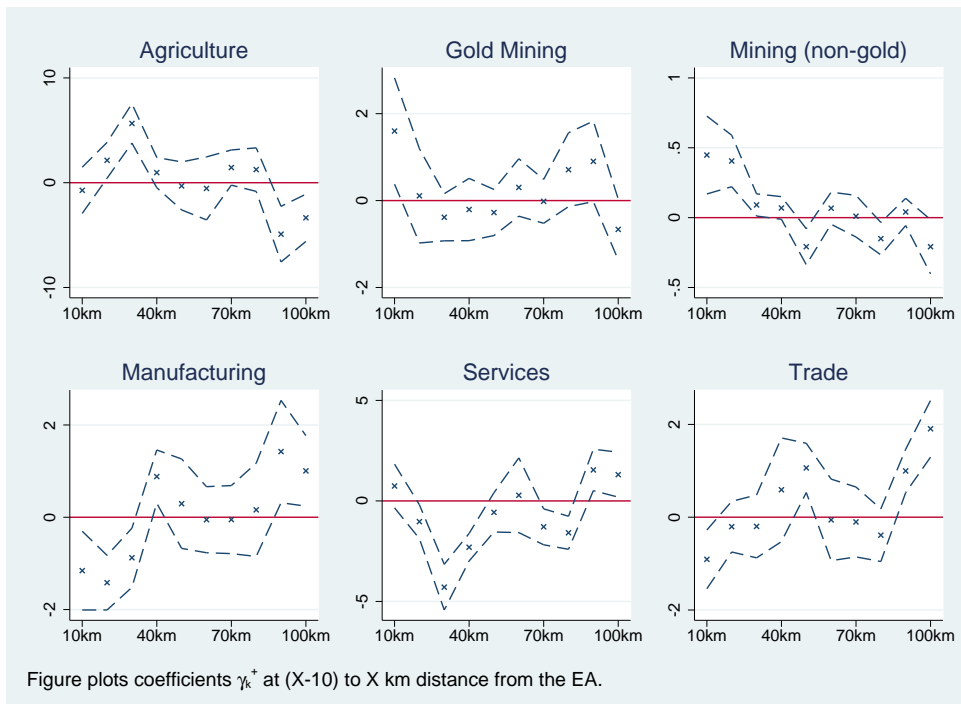
Figure A33: Robustness - Extensive Margin of Gold Mining and Occupational Choice, Mine openings

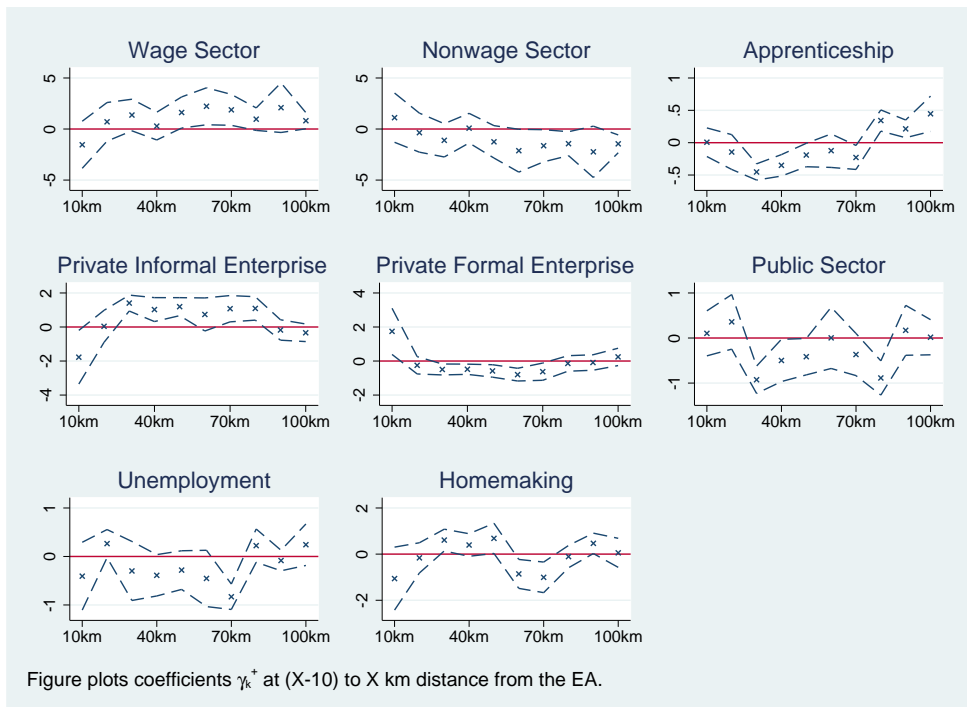Figure plots coefficients γ_k^+ at (X-10) to X km distance from the EA.

Figure A34: Robustness - Extensive Margin of Gold Mining and Population and Employment Density, Mine openings

Figure plots coefficients γ_k^+ at (X-10) to X km distance from the EA.

Figure A35: Robustness - Extensive Margin of Gold Mining and Sector of Activity, Mine closures

Figure plots coefficients γ_k^- at (X-10) to X km distance from the EA.
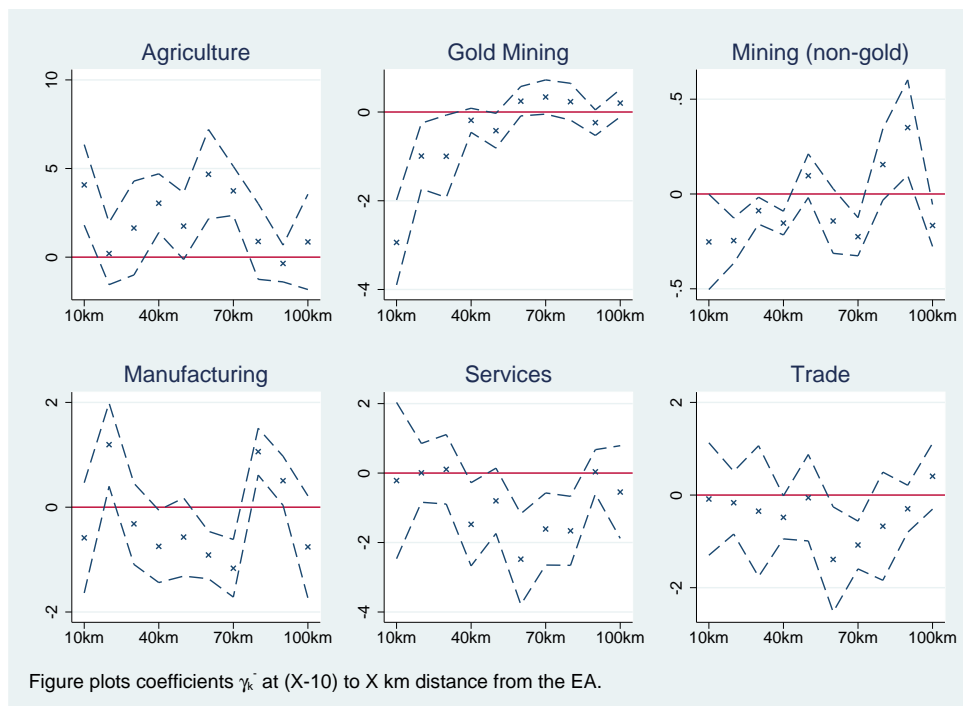
Figure A36: Robustness - Extensive Margin of Gold Mining and Occupational Choice, Mine closures

Figure plots coefficients γ_k^- at (X-10) to X km distance from the EA.
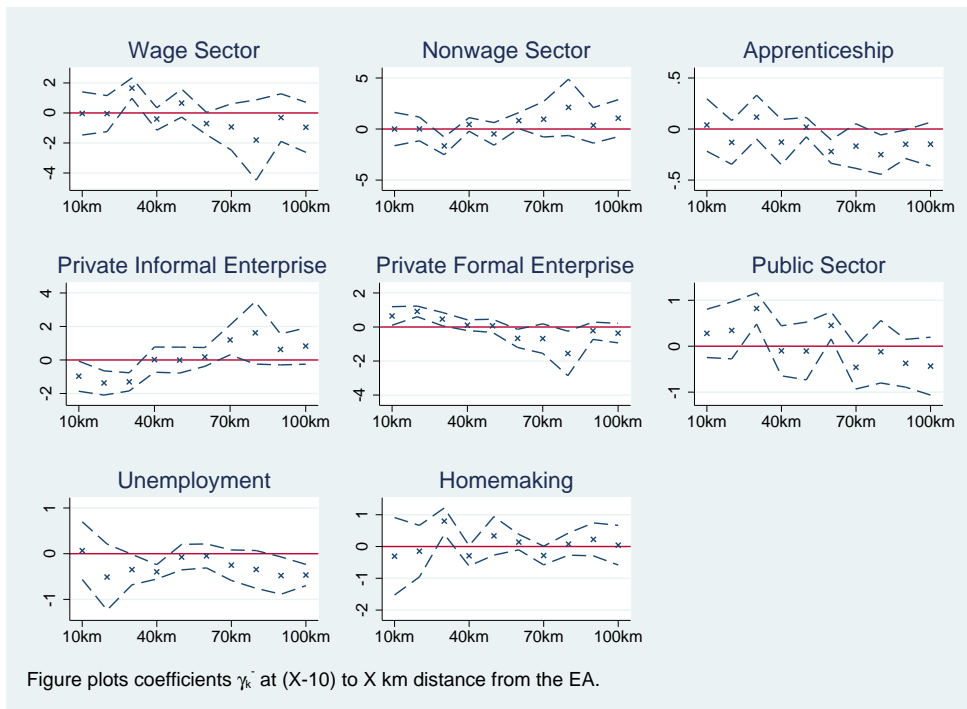
Figure A37: Robustness - Extensive Margin of Gold Mining and Population and Employment
Density, Mine closures

Figure plots coefficients γ_k^- at (X-10) to X km distance from the EA.