Assumptions :

- One household per house
- Preferences are identical across households
- Time periods are independent so developers have to choose whether to develop land independently in each period (probably accurate for shacks, but not for formal houses)

Households choose which house to live in with utility given by

$$U = \delta_l - \lambda(k) - \theta R_{j,k} + \epsilon_j$$

where $l$ indexes the location, $j$ indexes the plot, and $k$ indexes the number of houses on the plot. $\delta_l$ indicates the amenity value of living in location $l$ and can vary by area (project, spillover, and other) and time (pre post). $\lambda(k)$ captures the congestion disutility from having multiple houses on the same plot and is therefore increasing in the number of houses, $k$. $R_{j,k}$ is the price charged for each house. $\epsilon_j$ is the property specific preference shock that each property receives. Households can also choose to live outside the city and receive reservation utility, $\overline{U}$.

Developers observe household preferences and choose whether to build houses on each plot. Profit per house is given by

$$\pi_{j,k} = R_{j,k} - C_j$$

where the cost of constructing each house may depend on geographic features such as land slope, $C_j = c + \psi S_j$ where land slope is given by $S_j$.

Developers maximize profits by setting rents just high enough to ensure households are indifferent between living in or out of the city according to the following expression

$$R^*_{j,k} = \frac{\delta_l - \lambda(k) + \epsilon_j - \overline{U}}{\theta}$$

Developers will choose to build houses, $k$, on a plot while they are able to earn positive rents for each house, $R^*_{j,k} \geq C_j$. The optimal number of houses as a function of the plot-specific preference shock can be written as

$$k^* = \begin{cases} 0 & \text{if } \epsilon_j \leq \theta C_j - \delta_l + \lambda(1) + \bar{U} \\ 1 & \text{if } \theta C_j - \delta_l + \lambda(1) + \overline{U} < \epsilon_j \leq \theta C_j - \delta_l + \lambda(2) + \bar{U} \\ \dots & \dots \\ k & \text{if } \theta C_j - \delta_l + \lambda(k) + \overline{U} < \epsilon_j \leq \theta C_j - \delta_l + \lambda(k+1) + \bar{U} \end{cases}$$

Assuming that plot specific preferences $\epsilon_j$ are distributed normally, the probability of observing the number of houses on each plot of land is given by

$$
\begin{aligned}
Pr[k^* = 0] &= \Phi\left(\frac{\theta C_j - \delta_l + \lambda(1) + \overline{U}}{\sigma}\right) \\
Pr[k^* = 1] &= \Phi\left(\frac{\theta C_j - \delta_l + \lambda(2) + \overline{U}}{\sigma}\right) - \Phi\left(\frac{\theta C_j - \delta_l + \lambda(1) + \overline{U}}{\sigma}\right) \\
&\cdots \\
Pr[k^* = k] &= \Phi\left(\frac{\theta C_j - \delta_l + \lambda(k) + \overline{U}}{\sigma}\right) - \Phi\left(\frac{\theta C_j - \delta_l + \lambda(k+1) + \overline{U}}{\sigma}\right)
\end{aligned}
$$

where $\Phi()$ is the standard normal cumulative distribution function. This expression fits within an ordered probit framework and can be estimated with maximum likelihood.

First, we estimate the differential amenity values of being in project and spillover areas before and after scheduled construction for constructed and constructed projects. We expand the location preference term, $\delta_j$ to include terms for each of these locations in the following way

$$
\begin{aligned}
\delta_l = {} & \text{Proj}_l \left(\alpha_1 \,\text{Post}_t \times \text{Const}_l + \alpha_2 \,\text{Post}_t + \alpha_3 \,\text{Const}_l + \alpha_4\right) + \\
& \text{Spill}_l \left(\beta_1 \,\text{Post}_t \times \text{Const}_l + \beta_2 \,\text{Post}_t + \beta_3 \,\text{Const}_l + \beta_4\right) + \\
& \gamma_1 \,\text{Post}_t \times \text{Const}_l + \gamma_2 \,\text{Post}_t + \gamma_3 \,\text{Proj}_l
\end{aligned}
$$

where $\alpha_1$ and $\beta_1$ capture the triple-difference coefficients of interest capturing the effects of housing project construction on amenity values within and nearby project. $\lambda(k)$ is estimated in terms of a series of cut points in the distribution of preference shocks necessary to explain building density patterns. We also estimate a constant term which includes unobserved outside utility as well as average construction costs. Without variation in construction costs, this specification is unable to separately identify the marginal disutility of rent, $\theta$. The variance term, $\sigma$ captures plot-specific variation in preferences as well as possible variation in construction costs, which are not directly measured in this specification. We also allow $\sigma$ to vary according location, $l$, interacted with time, $t$.

- 231 average
- construction costs (38%): 231x0.38 = 88
- insfrastructure costs (23%): 231x0.21 = 53
- med slope : [infra.] 53x0.25 + [cc.] 88x0.05 = 18
- high slope : [infra.] 53x0.50 + [cc.] 88x0.15 = 40

Second, we estimate the marginal disutility of rent, $\theta$ using an instrumental variables strategy that generates variation in construction costs, $C_j$, which are passed on to households in the form of higher rents. We proxy for construction costs by observing that in a competitive market for land, the price difference between plots with and without houses should be equal to construction costs. We leverage the intuition that steeply sloping land is more costly for construction. We assume that land slope only affects demand for housing through its effect on construction costs. This assumption would be violated if households have direct preferences for living on sloped land. The Johannesburg metro area features a variety of ridges and plateaus creating substantial variation in land gradients.

We construct a measure of construction costs as a function of slope using . First, recent research on housing market dynamics in Africa suggests that construction costs are comparable to property market values. The average property value in our data is equal to R231,000. To benchmark this value, civil engineers estimate construction costs to be around R3,500 per square meter of floor space for the lower end of the housing market. Dividing construction costs by costs per square meter implies that the floor space of the average house in the low end of the housing market is around 66 m2. This size seems reasonable given that housing project houses are constructed have 40 m2 of floor space.

South African National construction guidelines indicate that constructing homes on slopes with 6% to 12% gradients increases infrastructure costs by 25% and building costs by 5% and constructing on slopes with greater than 12% gradients increases infrastructure costs by 50% and building costs by 15%. Recent research from South Africa suggests that out of total construction costs, 38% is attributable to building costs while 23% is attribute to infrastructure costs.

To measure these features, we construct 1.8km$^2$ squares that fall within 4 km of a housing project and are large enough to capture many housing transactions and elevation-levels.

Average property values our data .

Housing researchers in Africa demonstrate that construction costs are an important

by taking the difference between average deeds property prices for plots with houses and plots without houses in 1.8km$^2$ squares where houses are measured by the 2001 building census. We find mean construction costs equal to around R45,000. By contrast, recent research suggests that construction costs are equal to around 38% of housing prices in South Africa, which given an average housing price of R231,000 in our data, suggests average construction costs of R88,000. In the analysis, we include only squares where we are able to calculate construction costs (17% of squares).

We then measure average land gradient within a 1.8km$^2$ square by taking the difference between the highest and lowest recorded elevation and dividing by half of the width of each square. To leverage this variation, we include only squares with a positive change in slope (35% of squares). Our elevation measure is relatively sparse, which means that we are likely to underestimate the degree of slope change within an area. We exclude housing project footprints for this analysis because we do not observe housing prices for these areas. We arrive at a final sample of 258 squares and around 1,000,000 25 $m^2$ grid cells.

We estimate an instrumental variables specification according to the following

First Stage: Linear Regression

$$C_j = c + \phi S_j + \upsilon_j$$

Second Stage: Ordered Probit

$$Pr[k^* = 0] = \Phi\left(\frac{\theta C_j - \delta_l + \lambda(1) + \overline{U}}{\sigma}\right)$$

$$\ldots$$

$$Pr[k^* = k] = \Phi\left(\frac{\theta C_j - \delta_l + \lambda(k) + \overline{U}}{\sigma}\right) - \Phi\left(\frac{\theta C_j - \delta_l + \lambda(k+1) + \overline{U}}{\sigma}\right)$$

The coefficient on construction costs gives us $\psi$, which tells us household marginal disutility from paying rent. We can use this coefficient to translate the amenity effects in the first step terms of rands by dividing triple difference coefficients of interest $\alpha_1$ and $\beta_1$ by $\psi$. This exercise suggest that housing projects produce $\frac{0.4}{0.0000078} = R51,282$ in welfare gain within project areas, which is equal to around 22% of the average housing prices. Meanwhile, housing projects produce $\frac{0.2}{0.0000078} = R25,641$ in lost welfare within spillover areas (0-500m from project boundaries), which is equal to around 11% of average housing prices and is comparable to what we found in the housing price regressions.

**Table 1.** Ordered Probit Triple-Difference

| | Amenity $\delta_l$ | Variance $\log(\sigma)$ |
|---|---|---|
| inside × constr × post | 0.40 | -0.17 |
| | (0.26) | (0.11) |
| 0-500m away × constr × post | -0.20[c] | 0.01 |
| | (0.11) | (0.06) |
| inside × post | -0.07 | 0.22[b] |
| | (0.24) | (0.09) |
| 0-500m away × post | 0.28[a] | 0.02 |
| | (0.10) | (0.06) |
| constr × post | 0.03 | -0.03 |
| | (0.05) | (0.03) |
| inside × constr | 0.51[b] | -0.18[c] |
| | (0.24) | (0.11) |
| 0-500m away × constr | -0.08 | -0.04 |
| | (0.21) | (0.10) |
| post | -0.13[a] | 0.14[a] |
| | (0.04) | (0.03) |
| inside | 0.10 | -0.01 |
| | (0.17) | (0.08) |
| 0-500m away | 0.40[b] | -0.02 |
| | (0.17) | (0.08) |
| constr | 0.19 | 0.08 |
| | (0.14) | (0.07) |
| Cut Point 1 | | |
| Estimate | 1.35[a] | |
| | (0.06) | |
| Cut Point 2 | | |
| Estimate | 1.69[a] | |
| | (0.07) | |
| Cut Point 3 | | |
| Estimate | 2.08[a] | |
| | (0.07) | |
| Cut Point 4 | | |
| Estimate | 2.42[a] | |
| | (0.09) | |
| Cut Point 5 | | |
| Estimate | 2.87[a] | |
| | (0.10) | |
| N | 6718738 | |

Standard errors are clustered at the project by location level.

**Table 2.** Ordered Probit IV

| | |
|---|---|
| Total Buildings | |
| Construction Costs | -0.0000089[a] |
| | (0.0000018) |
| Construction Costs | |
| Land Gradient (%) | 207458.40[a] |
| | (7246.33) |
| Estimate | 224044.46[a] |
| | (592.20) |
| Cut Point 1 | |
| Estimate | -0.81[c] |
| | (0.43) |
| Cut Point 2 | |
| Estimate | -0.42 |
| | (0.43) |
| Cut Point 3 | |
| Estimate | -0.11 |
| | (0.43) |
| Cut Point 4 | |
| Estimate | 0.14 |
| | (0.43) |
| Cut Point 5 | |
| Estimate | 0.46 |
| | (0.43) |
| Variance $\sigma$ | |
| Estimate | 8.91[a] |
| | (0.03) |
| N | 3665766 |

Standard errors are clustered at the project by square level.