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. δ_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 assumed to be increasing and convex in the number of houses, k. $R_{j,k}$ is the price charged for each house. ϵ_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 indexed by k, on plot j while they are able to earn positive rents for each house, $R_{j,k}^* \geq C_j$. Total profit per plot is equal to

$$\pi_j = k\left[R_{j,k} - C_j\right]$$

The optimal number of houses, k^* as a function of the plot-specific preference shock

can be written as

$$k^* = \begin{cases} 0 & \text{if } \epsilon_j \leq H + \lambda(1) \\ 1 & \text{if } H + \lambda(1) < \epsilon_j \leq H + 2\lambda(2) - \lambda(1) \\ \dots & \dots \\ k & \text{if } H + \Lambda(k) < \epsilon_j \leq H + \Lambda(k+1) \end{cases}$$

Where

$$H = \theta C_j - \delta_l + \bar{U}$$
$$\Lambda(k) = k\lambda(k) - (k-1)\lambda(k-1)$$
$$\lambda(0) = 0 \text{ and } \lambda(.) \text{ is increasing and convex}$$

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

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

$$Pr[k^* = 1] = \Phi\left(\frac{\theta C_j - \delta_l + \Lambda(2) + \overline{U}}{\sigma}\right) - \Phi\left(\frac{\theta C_j - \delta_l + \Lambda(1) + \overline{U}}{\sigma}\right)$$

$$\dots$$

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

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, δ_j to include terms for each of these locations in the following way

$$\delta_l = \text{PROJ}_l \left(\alpha_1 \text{ POST}_t \times \text{CONST}_l + \alpha_2 \text{ POST}_t + \alpha_3 \text{ CONST}_l + \alpha_4\right) +$$

$$\text{SPILL}_l \left(\beta_1 \text{ POST}_t \times \text{CONST}_l + \beta_2 \text{ POST}_t + \beta_3 \text{ CONST}_l + \beta_4\right) +$$

$$\gamma_1 \text{ POST}_t \times \text{CONST}_l + \gamma_2 \text{ POST}_t + \gamma_3 \text{ PROJ}_l$$

where α_1 and β_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, θ. The variance term, σ 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 σ to vary according location, l, interacted with time, t.

Second, we estimate the marginal disutility of rent, θ 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 land slope. First, recent research on housing market dynamics in Africa suggests that construction costs are comparable to property market values consistent with a competitive market in property development (CAHF). 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. CAHF also puts average construction cost for a fully serviced, 46 m2 house in Pretoria at 336,000, which is also consistent with our estimate given that we are looking at the bottom 20% of the housing market.

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 the Center for Affordable Housing Finance in Pretoria, Gauteng estimates that out of total construction costs, 62% is attributable to building costs while 12% is attribute to infrastructure costs. We create a construction cost proxy by combining average price data with data on local elevation.

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 multiple elevation-levels necessary to calculate slope. We 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 where we are able to calculate a change in slope (35% of squares). We exclude areas that overlap with housing project footprints for this analysis in order to focus on endogenous behavior in housing markets. We arrive at a final sample of 1,921 squares and around 3,700,000 observations composed of 25 m^2 grid cells pre and post con.

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

$$\dots$$

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

The coefficient on construction costs gives us ψ, 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 α_1 and β_1 by ψ. This exercise suggest that housing projects produce $\frac{0.4}{0.0000102} = $ R39, 216 in welfare gain within project areas, which is equal to around 17% of the average housing prices. Meanwhile, housing projects produce $\frac{0.2}{0.0000102} = $ R19, 608 in lost welfare within spillover areas (0-500m from project boundaries), which is equal to around 8.5% of average housing prices and is comparable to what we found in the housing price regressions.

Table 1. Ordered Probit Triple-Difference

	Amenity δ_l	Variance $\log(\sigma)$
inside $\times$ constr $\times$ post	0.40	-0.17
	(0.26)	(0.11)
0-500m away $\times$ constr $\times$ post	-0.20[c]	0.01
	(0.11)	(0.06)
inside $\times$ post	-0.07	0.22[b]
	(0.24)	(0.09)
0-500m away $\times$ post	0.28[a]	0.02
	(0.10)	(0.06)
constr $\times$ post	0.03	-0.03
	(0.05)	(0.03)
inside $\times$ constr	0.51[b]	-0.18[c]
	(0.24)	(0.11)
0-500m away $\times$ 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

First Stage: Construction Costs	
Land Gradient (%)	183463.79[a]
	(6295.03)
Estimate	224501.23[a]
	(510.12)
Reduced Form: Total Buildings	
Construction Costs	-0.0000102[a]
	(0.0000020)
Cut Point 1	
Estimate	-1.09[b]
	(0.49)
Cut Point 2	
Estimate	-0.70
	(0.49)
Cut Point 3	
Estimate	-0.39
	(0.49)
Cut Point 4	
Estimate	-0.15
	(0.49)
Cut Point 5	
Estimate	0.18
	(0.49)
Variance σ	
Estimate	8.76[a]
	(0.03)
N	3665766

Standard errors are clustered at the project by square level.