City Location Choice and Household Productivity

Tan Sein Jone

1 Introduction

The location choice of households across cities is a fundamental question in urban economics. The choice of location is dependent on a variety of factors such as amenities, housing costs, and job opportunities. Quantitative spacial models such as the one developed by Redding (2016) have been able to explain the distibution of economic activity acorss cities by intergrating an amenity shifter into their model. We believe however that the distribution of economic activity across cities is also dependent on the distribution of job types across cities. With cities specializing in specific occupations attracting houeholds that are most productive in those occupations. This means that only cities that are most simillar in their technologies, relfected in their employment structure, will be able to attract hosueholds of specific types. In order to break the independence of irrelevant alternatives (IIA) assumption, we look to Lind and Ramondo (2023) who have developed an EK model with multiple technologies to produce correlated productivity draws across countries.

1.1 Literature Review

Eaton and Kortum (2022) present a quantitative trade model that incorporates Ricardian motives for trade by treating productivity as random draws acorss goods, countries and other observable units. This model however relies on indepndence asusmptions that don't capture the relavent substiution patterns for goods acorss countries. Specifically, the dynamic where countries with the most simillar technologies trade most with each other.

Lind and Ramondo (2023) overcome this limitation by developing an EK model with multiple technologies that produce correlated productivity draws across countries. This is achieved using a corss nested constant-elasticity-of-substitution (CES) structure for productivity, with latent nests having the interpretation of technologies. Sectors in countries can then share technologies, which allows for the correlation of productivity draws across countries.

2 Model of Production

$$Y_{ck}(\nu) = Z_{ck}(\nu) \tag{1}$$

Consider a closed economy consisting of N cities and a continnium of household types $\nu \in [0, 1]$. Each city c employs hosueholds of type ν in occupation k to produce output $Y_{ck}(\nu)$. We assume no trade costs between cities, and that the price of the good produced by occupation k is freely traded and priced under perfect competition. The price index of city c is Φ_c. The wage of a worker in sector k in city c is $w_{ck}(\nu) = \Phi_c Z_{ck}(\nu)$. As in EK, productivity is a random variable drawn from a max stable multivariate Fréchet distribution and is dependent on both the city and occupation. Output by a specifc occupation is assumed to be produced by a random sector.

3 Max Stable Multivariate Fréchet Distribution

We assume that the joint distribution of productivity across cities is given by

$$P[Z_{1k}(\nu) < z, \ldots, Z_{Nk}(\nu) < z] = exp[-\sum_{c}^{N}(T_{ck}H_k(\nu)z^{-\theta})^{\frac{1}{1-\rho_k}}]^{1-\rho_k} \tag{2}$$

Where Tck is the scale parameter for city c and occupation k, this represents a city's absolute advantage for occupation k. $H_k(\nu)$ is the occupation specific productivity shifter for household of type ν in occupation k, the heterogeneity of this shifter relfects the difference in human capital across different hosuehold types. $\theta > 0$ is the shape parameter, characaterizing the tail behavior of the distribution. ρ_k is the occupation specific correlation parameter, which dictates the extent to which productivity draws are correlated across cities for occupation k.

A household's schedule of productivities is characterized by a vector of draws from different Fréchet distributions for each city occupation pair. The realized productivity a hosuehold of type ν has in city c is the occupation that maximizes the producticity that particular household has in the city. This is given by the following:

$$Z_c(\nu) = \max_k \{Z_{ck}(\nu)\} \tag{3}$$

Unlike sequential games where households might pick a city before picking an occupation, this schedule of productivity already determines the ideal occupation for a household in a city. The joint probability of all productivities being less than some value z for all cities is then given by the following:

$$P[Z_1(\nu) < z, \ldots, Z_N(\nu) < z] = exp\{-\sum_k [\sum_c^N (T_{ck} H_k(\nu) z^{-\theta})^{\frac{1}{1-\rho_k}}]^{1-\rho_k}\} \tag{4}$$

The joint probability above is a max-stable multivariate Fréchet distribution with a cross-nested CES correlation function. This distribution is a generalization of the Fréchet distribution to the multivariate case, and is used to model the joint distribution of extreme values. This is simillar to the GEV distribution, but with the added feature of a correlation function that allows for the dependence of extreme values across occupations. A houehold of type ν has realised productivity that is hence characterised by the maxmimum productivity draw across all cities scaled by the inverse of that city's price index Φ_c and is gievn by the following:

$$Z(\nu) = \max_{c=1,\ldots,N} \left\{ \frac{Z_c(\nu)}{\Phi_c} \right\} \tag{5}$$

3.1 Correlation Function

$$P[Z_{1k} < z, \ldots, Z_{Nk} < z] = exp[-\sum_c^N (T_{ck} z^{-\theta})^{\frac{1}{1-\rho_k}}]^{1-\rho_k} \tag{6}$$

$$P[Z_1 < z, \ldots, Z_N < z] = exp\{-\sum_k [\sum_c^N (T_{ck} z^{-\theta})^{\frac{1}{1-\rho_k}}]^{1-\rho_k}\} \tag{7}$$

When we intergrate over all houehold types for a particular occupation shifter, we get some average productivity shifter for that occupation. $\int_0^1 H_k(\nu)d\nu = 1$. With the case of a cross-nested CES function, we can approximate a correlation funciton. Assuming that productivity is distributed max-stable multivariate Fréchet, with scale parameter T_{ck}, shape parameter θ, and correlation parameter ρ_k, we can assume the following correlation function:

$$G(Z_1^{-\theta}, \ldots, Z_N^{-\theta}) = \sum_k [\sum_c^N (T_{ck} Z_c^{-\theta})^{\frac{1}{1-\rho_k}}]^{1-\rho_k} \tag{8}$$

$$G_c(Z_1^{-\theta}, \ldots, Z_N^{-\theta}) = \frac{\partial G(Z_1^{-\theta}, \ldots, Z_N^{-\theta})}{\partial Z_c^{-\theta}} \tag{9}$$

$$\gamma = \Gamma(\frac{\theta - 1}{\theta}) \tag{10}$$

Where G_c is the derivative of the correlation function with respect to city c. This correlation function is a generalization of the CES correlation function to the multivariate case, and is used to model the dependence of extreme values across cities. The Γ function which takes in the ratio of $\theta-1$ to θ gives us the substitutability between cities. Productivity is related to γ by the following funciton:

$$Z_c^{-\theta} = (\gamma \Phi_c)^{-\theta} T_c \tag{11}$$

Where T_c is the city's absolute advantage that applies to all occupations and is given by $T_c = \sum_k T_{ck} H_k$.

3.2 Choice Shares

Under perfect competition, wages are equal to marginal revenue product and is a direct function of productivity. As established in equation (5), the productivity for a worker ν in city c is given by the maximum productivity draw across all occupations. All we have to do is intergrate over all houshold types to get the productivity index for city c:

$$Z_c = \max_k \left\{ \int_0^1 Z_{ck}(\nu) d\nu \right\} \tag{12}$$

If productivity is distributed max-stable multivariate Fréchet and a continiously differentiable correlation function, then city c's choice shares for a given occupaiton k is given by the following:

$$\pi_{ck} = \frac{Z_c^{-\theta} G_c^k(Z_1^{-\theta}, \ldots, Z_N^{-\theta})}{G^k(Z_1^{-\theta}, \ldots, Z_N^{-\theta})} \tag{13}$$

Where $G^k(Z_1^{-\theta}, \ldots, Z_N^{-\theta}) = [\sum_c^N (T_{ck} Z_c^{-\theta})^{\frac{1}{1-\rho_k}}]^{1-\rho_k}$

The share of location choices for households has the same form as choice probabilities in GEV discrete choice models, with $Z_c^{-\theta}$ replacing choice specific utility. As in EK, the share of location choices of city c equals the probability that a worker is most productive in city c. Finally, the location choice share in each city is determined by the ratio of the expected productivity in that city to the expected productivity in all cities.

$$\pi_{ck} = \frac{(T_{ck} Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}{\sum_c^N (T_{ck} Z_c^{-\theta})^{\frac{1}{1-\rho_k}}} \tag{14}$$

Now for a city's aggreagte choice shares. We make the simplifying asusmption that $\rho_k = \rho$ for all occupaitons k. This allows us to begin with the following expressions:

$$G(Z_1^{-\theta}, \ldots, Z_N^{-\theta}) = \sum_k [\sum_c^N (T_{ck} Z_c^{-\theta})^{\frac{1}{1-\rho}}]^{1-\rho} \tag{15}$$

$$G_c(Z_1^{-\theta}, \ldots, Z_N^{-\theta}) = \frac{\partial G(Z_1^{-\theta}, \ldots, Z_N^{-\theta})}{\partial Z_c^{-\theta}} \tag{16}$$

$$\pi_c = \frac{Z_c^{-\theta} G_c(Z_1^{-\theta}, \ldots, Z_N^{-\theta})}{G(Z_1^{-\theta}, \ldots, Z_N^{-\theta})} \tag{17}$$

Evaluating expression (17), we get the following:

$$\pi_c = (Z_c^{-\theta})^{\frac{1}{1-\rho}} \left[\frac{\sum_k T_{ck}^{\frac{1}{1-\rho}} \lambda_k^{-\rho}}{\sum_k \lambda_k^{1-\rho}} \right] \tag{18}$$

3.3 Substitutability

In the case of a cross nested CES function, we can approximate the substitutability between cities within a given occupation. We can do this by taking the log derivative of the choice share of city c with respect to the city's absolute advantage T_{ck}. Solving for the derivative of Equation (14) with respect to $ln(T_{ck})$, we get the following:

$$\frac{\partial ln\pi_{ck}}{\partial lnT_{c'k}} = -\frac{1}{1-\rho_k}\frac{1}{T_{c'k}}\pi_{c'k} \tag{19}$$

We can see that for any given occupation, the substitutability between cities is determined by the correlation parameter ρ_k, the city's absolute advantage $T_{c'k}$, and the choice share of that city $\pi_{c'k}$. As occupations are more correlated, cities become more substitutable. As any given city's share increases, the substitutability increases. This illustrates a gravity effect where cities that are more productive in a given occupation and hence have a larger share of houeholds are more likely to further attract more households for that occupation.

When we evaluate the own price "own price" elasticity of city c's choice share with respect to the city's absolute advantage T_{ck}, we make the simplifying assumption that each city is small and therefore $\frac{\partial \lambda_k}{\partial T_{ck}} = 0$. This gives us the following expression:

$$\frac{\partial ln\pi_{ck}}{\partial lnT_{ck}} \approx \left(\frac{1}{1-\rho}\right) \left[\frac{\partial lnZ_c^{-\theta}}{\partial lnT_{ck}} + \frac{T_{ck}^{\frac{1}{1-\rho}} \lambda_k^{-\rho}}{\sum_k T_{ck}^{\frac{1}{1-\rho}} \lambda_k^{-\rho}} \right] \tag{20}$$

Simmilarly, we can get the cross price elasticity of city c's choice share with respect to that city's absolute advantage $T_{c'k}$:

$$\frac{\partial ln\pi_c}{\partial lnT_{c'k}} \approx -\pi_{c'k}\left[1+\left(\frac{\rho}{1-\rho}T_{ck}^{\frac{1}{1-\rho}}\right)\right] \tag{21}$$

4 Simulations

Assume now that we have two occupations trades and services. We will also assume three cities, Detriot, Chicago and New York with Detriot and Chicago specializing in trades and New York specializing in services. We will also assume that the correlation parameter ρ_k is 0.7 for trades and 0.5 for services and the shape parameter θ is 2. This will give us the folliwing scale parameters for cities and hosuehold types:

	Detriot	**Chicago**	**New York**
Trades	7	8	3
Services	2	1	8

Table 1: City Scale Parameters T_{ck}

These shifters will determine the Frechet distributions from which city and occupation specific productivities are drawn for each household type. For context, this is the pdf for each of the Frechet distributions:

$$P[Z_{ck} < z] = exp[-(T_{ck}z^{-\theta})^{\frac{1}{1-\rho_k}}] \tag{22}$$

For each household, the draws will produce a $K \times N$ matrix which reflects that hosuehold's schedule of productivities across all cities and occupations. The realized productivity for each household is then the maximum productivity draw across all cities and occupations. Location choice is hence the column which contains that realised productivity.

We now draw 10000 households, simulating the dtstribution of location choices across cities. As shown in figure 1, the city with the highest population is New York, with the majority of service focused households choosing to locate in the city. Detriot and Chicago have a more even distribution of trades focused households.

We then shock Detriot's productivities, reducing it by 50% and observe the change in location choices. As

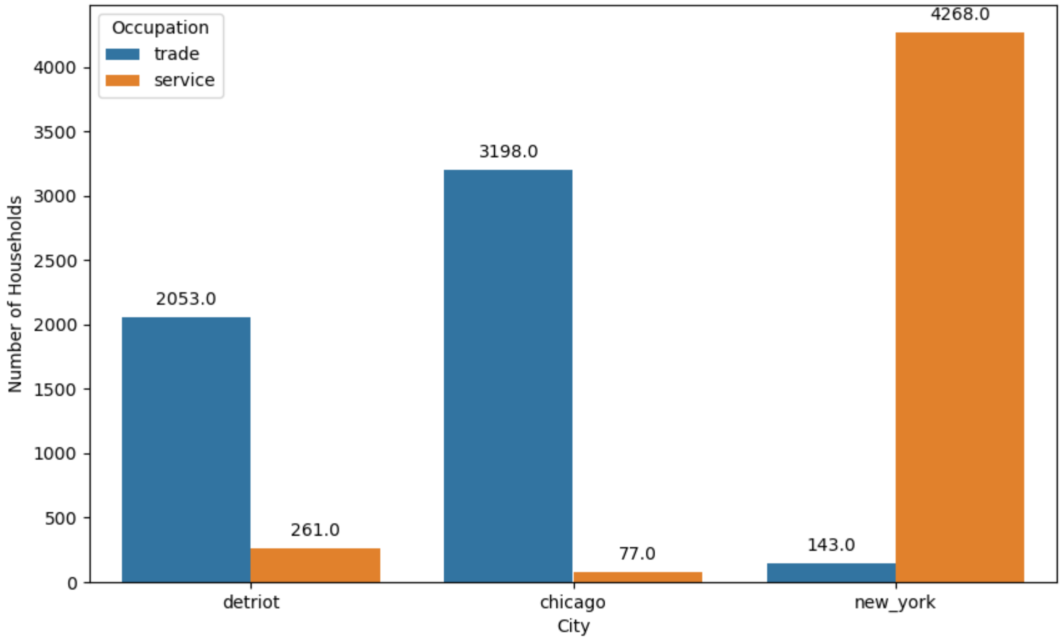

Figure 1: Initial Distribution of Populations

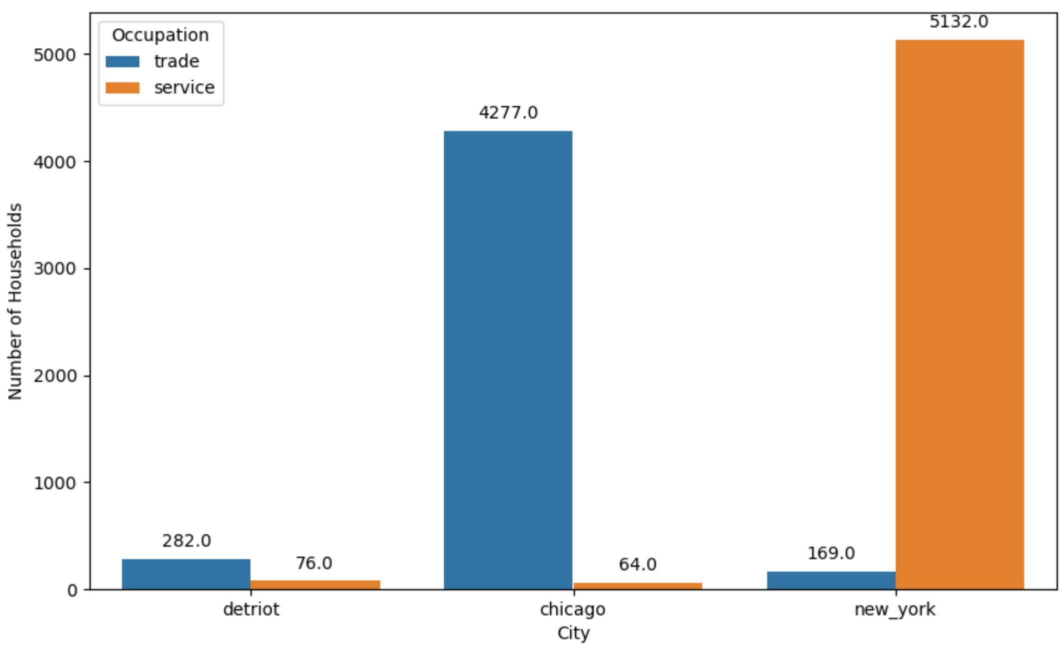

Figure 2: Post Detroit Shock

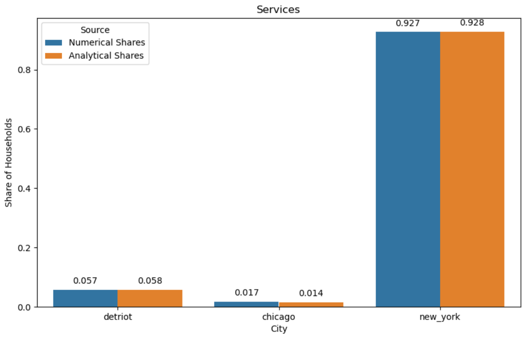

Figure 3: Service Choice Shares

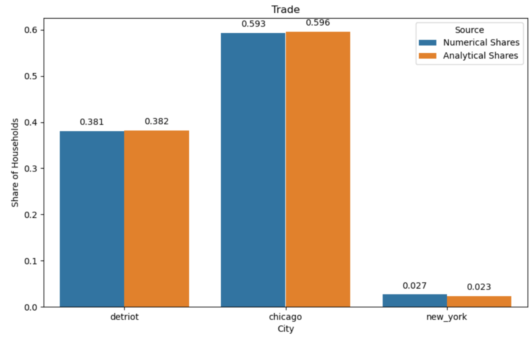

Figure 4: Trades Choice Shares

shown in figure 2, there is a redistibution in location choices with the majority of trades focused households now choosing to locate in Chicago. Some households have also chosen to locate in New York due to higher productivity draws in that city.

4.1 Numerical vs Analytical Results Comparison

Given the generation of synthetic reuslts, this gives us the opportunity to compare the numerical results with the analytical results. We can compare the choice shares of each city as given by the analytical model with the choice shares of each city as given by the numerical model. This will allow us to validate the analytical model and see if it is able to capture the distribution of location choices across cities.

As you can see in figures 3 and 4, the analytical model is able to capture the distribution of location choices across cities. The choice shares of each city as given by the analytical model are very simillar to the choice shares of each city as given by the numerical model. This validates the analytical model and shows that it is able to capture the distribution of location choices across cities.

The same excercise can be done for the choice shares for each city. We compare the choice shares for each city as given by the analytical model with the choice shares for each city as given by the numerical model in order to validate the analytical model.

As shown in figure 5, the analytical model is able to capture the distribution of location choices across cities. While the shares are not exactly the same, the analytical model is able to capture the general distribution of location choices across cities. This validates the analytical model and shows that it is able to capture the distribution of location choices across cities.

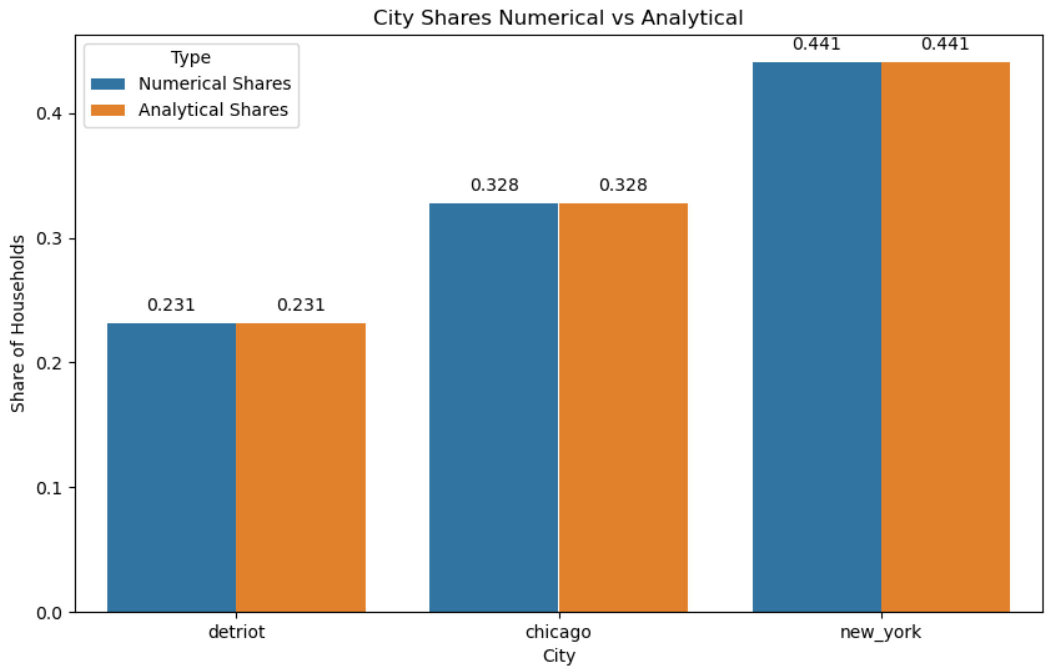

Figure 5: City Choice Shares

5 Proofs

5.1 Proof of Equation (14)

$$
\begin{aligned}
G^k(Z_1^{-\theta},\ldots,Z_N^{-\theta}) &= [\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}]^{1-\rho_k} \\
G_c^k(Z_1^{-\theta},\ldots,Z_N^{-\theta}) &= \frac{\partial G(Z_1^{-\theta},\ldots,Z_N^{-\theta})}{\partial Z_c^{-\theta}} \\
&= (1-\rho_k)[\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}]^{-\rho_k}(\frac{1}{1-\rho_k}(T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}-1}T_{ck}) \\
&= (1-\rho_k)\frac{1}{1-\rho_k}\frac{T_{ck}}{T_{ck}Z_c^{-\theta}}(T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}[\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}]^{-\rho_k} \\
Z_c^{-\theta}G_c^k(Z_1^{-\theta},\ldots,Z_N^{-\theta}) &= \frac{(T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}{\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}[\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}]^{1-\rho_k} \\
\frac{Z_c^{-\theta}G_c^k(Z_1^{-\theta},\ldots,Z_N^{-\theta})}{G^k(Z_1^{-\theta},\ldots,Z_N^{-\theta})} &= \frac{(T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}{\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}} \\
\pi_{ck} &= \frac{(T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}{\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}
\end{aligned}
$$

5.2 Substitutability Proof

$$
\begin{aligned}
\frac{\partial ln\pi_{ck}}{\partial lnT_{ck}} &= \frac{1}{1-\rho_k} - \frac{\frac{1}{1-\rho_k}(T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}-1}}{\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}} \\
&= \frac{1}{1-\rho_k}\left[1 - \frac{1}{T_{ck}}\frac{(T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}{\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}\right] > 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial ln\pi_{ck}}{\partial lnT_{c'k}} &= -\frac{1}{1-\rho_k}\frac{1}{T_{c'k}}\frac{(T_{c'k}Z_{c'}^{-\theta})^{\frac{1}{1-\rho_k}}}{\sum_c^N (T_{ck}Z_c^{-\theta})^{\frac{1}{1-\rho_k}}} \\
&= -\frac{1}{1-\rho_k}\frac{1}{T_{c'k}}\pi_{c'k} < 0
\end{aligned}
$$

5.3 Estimation Equation

$$\pi_{ck} = \frac{(T_{ck} Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}{\sum_c^N (T_{ck} Z_c^{-\theta})^{\frac{1}{1-\rho_k}}}$$

$$ln\pi_{ck} = \frac{1}{1-\rho_k}\left[ln(T_{ck}) + ln\left\{(\gamma\Phi_c)^{-\theta}\sum_k T_{ck}\right\}\right] - ln\left[\Lambda_k\right]$$

Where $\Lambda_k = \sum_c^N (T_{ck} Z_c^{-\theta})^{\frac{1}{1-\rho_k}}$. And $Z_{ck}^{-\theta} = (\gamma\Phi_c)^{-\theta} T_c$.