THE JOURNAL OF TRANSPORT AND LAND USE http://jtlu.org
VOL. 6 NO. 1 [2013] pp. 73–88 http://dx.doi.org/10.5198/jtlu.v1.271

An Agent-Based Model of Origin Destination Estimation (ABODE)

Nebiyou Tilahun
University of Illinois at Chicago [a]

David Levinson
University of Minnesota [b]

Abstract: This paper introduces ABODE, an agent-based model for Origin-Destination (OD) demand estimation, that can serve as a work trip distribution model. The model takes residential locations of workers and the locations of employers as exogenous and deals specifically with the interactions between firms and workers in creating a job-worker match and the commute outcomes. It is meant to illustrate that by explicitly modeling the search and hiring process, origins and destinations (ODs) can be linked at a disaggregate level that is reasonably true to the actual process. The model is tested on a toy-city as well as using data from the Twin Cities area. The toy-city model illustrates that the model predicts reasonable commute outcomes, with agents selecting the closest work place when wage and skill differentiation is absent in the labor market. The introduction of wage dispersion and skill differentiation increases the average home to work distances considerably. Using data from Twin Cities area of Minneapolis-St. Paul, we also show that the model captures aggregate commute outcomes well. Overall, the results suggest that the behavior rules as implemented lead to reasonable patterns. Future improvements and directions are also discussed.

1 Introduction

The relationship between home and work has been an important research area for economists, urban geographers, sociologists and others. Traditionally, the approach taken by transportation professionals to match home and work locations has been to use trip distribution models. These models are part of the widely used four-step transportation planning process comprising steps of trip generation, trip distribution, mode choice and route assignment. Trip distribution models such as the gravity model, which use aggregate zonal variables to match home and work, are still widely used by many planning organizations. Together with the trip generation step, these models predict zonal interchange of trips at an aggregate level. While this framework is very useful, it overlooks much of what happens as the connection between people's home and work are established.

As output needs of transportation models have evolved, changes in the traditional modeling framework have become necessary. Today's planning questions involve a range of issues from air quality to congestion mitigation, evaluation of different demand management strategies, and providing a testing environment for different policy prescriptions. The response has been the development of different activity-based models that are grounded in travel behavior. These models continue to try to bring some realism in the choice of destinations, trip chaining behavior, scheduling within the constraints placed by households, other actors, and space and time. This change has also meant a shift in the unit of analysis from traffic zones to households and individuals, making it necessary that home and work locations also be known at a more disaggregate level.

In this paper we propose and test an agent-based model of worker and job matching. The connection between home and work is the outcome of interactions between employers and job seekers that have different goals. The employer is often motivated by increased productivity through the addition of new staff to perform particular duties (Holzer 1987), while the employee has aims of increased income and other long-term goals that have to do with their future ambitions. Each searches for their best fitting counterpart and hope the match they find takes them forward in the fulfillment of their respective goals. In addition to geographic locations of jobs and workers, this matching process is structured by how employers advertise and recruit, how workers search and weigh alternatives, and the limited amount of information that is available to each searcher about opportunities.

The model we propose takes residential locations of workers and the locations of employers as exogenous and deals specifically with the interactions between firms and workers in creating a job-worker match, and the commute outcomes. It is meant to illustrate that by explicitly modeling the search process and the interactions between firms and individuals, origins and destinations (ODs) can be linked at a disaggregate level that is reasonably true to the actual process. In doing so,

[a] ntilahun@uic.edu
[b] dlevinson@umn.edu

Copyright 2013 Nebiyou Tilahun, David Levinson.
Licensed under the Creative Commons Attribution – NonCommercial License 3.0.

the model seeks to develop home-work connections that can be used as part of either the traditional four-step model, replacing home based work trip distribution models, or as an input into activity based models, that generally take these longer term decisions as external inputs.

In the next section, we will discuss some of the background in matching residences to employment. Section 3 discusses the agent based model. That is followed by a test of the model using a hypothetical toy city in section 4 and a test of the model's behavior using different inputs in section 5. In Section 6, the model is applied to a subset of individual and job data extracted from the 2000 Travel Behavior Inventory (TBI) data for the Twin Cities area and the results are discussed. Section 7 provides a summary.

2 Background

The relationship between home and work has been an important research area for economists, urban geographers, sociologists and others who have each brought their field's experience to the understanding of how workers and jobs are matched.

Labor economist have used search theory (Stigler 1961, 1962) to explain the behavior of searchers in the labor market. The approach conceives of searchers that are faced with randomly arriving offers, which they can choose to accept or reject. Optimal stopping and searching decisions are often investigated and depend on what the searcher knows about wage distributions, offer arrival rates, and the marginal costs and expected gains of waiting for the next offer. Works by Rouwendal (1998, 1999), van Ommeren (1998); van Ommeren and Rietveld (2002); van Ommeren *et al.* (1999); van Ommeren and van der Straaten (2008) have used a search based approach to study the relationship between jobs, residence and commute where individuals maximize lifetime utility based on wages, place utility and commuting costs.

Urban economists on the other hand have proposed models that take work place locations as exogenous, and theorized about the choice of residential locations. Models of this type include those of Alonso (1964), Muth (1969) and Mills (1972). In the mono-centric city model, residential locations are assumed to be selected such that higher travel costs are compensated by cheaper land prices. Later works, for example by Hamilton (1982) and White (1988), have shown actual commutes to be higher than predicted by these models, though to different degrees.

van Ommeren *et al.* (1999) also show that models that assume transportation costs are fully compensated for by wages overestimate the wages required to compensate workers with longer commutes because they ignore residential relocation decisions. Similarly, they show that residential (place) utility need not fully compensate workers commute costs because of the possibility of future job changes.

While the possibility of future moves can account for some of the observed longer commutes, other possible reasons are also cited in the literature including within household effects of multi-worker households, increasing importance of non-work trips and other housing and neighborhood attributes (Giuliano and Small 1993). In addition, information imperfection in the job and residential search process, costs of relocation, as well as the very low rate at which employment offers reach job searchers have also been cited as possible explanations (van Ommeren and van der Straaten 2008).

There are several key elements in the matching process that are instrumental in increasing the chances for some to get employed while reducing the chances for others to get access to the same job even when they share similar skill sets. Let us consider a relatively simple system where employers advertise positions, searchers apply to these positions, employers then make an offer to the best candidate, and the employee decides to accept or reject the offer. For a match to be successful, the job seeker must have access to information that the position is open to begin with. Second the searcher must have skills that reasonably match the position. The searcher must also be able to meet the screening criteria of the employer. And finally for those that receive offers, the offer has to be better than what the searcher sees as their best alternative. Each of these steps creates both barriers and opportunities to different types of people from getting any given job that is available.

The path the employer follows in recruiting can effectively block out a segment of the population or create systematic bias as to who receives that information. For example, recruitment procedures that use employee referrals would focus on people who are in someway connected to existing employees (Holzer 1987); advertising in a subset of local labor markets can exclude others that may have similar skills but reside outside of the focus area (Neckerman and Kirschenman 1991). Once applications from job seekers are received, firm screening procedures also impact who gets particular opportunities. In addition to skills, education and experience, employers can use other ways to make judgements about the quality of the applicant's match. A referral is one way this is achieved, and other methods can include the use of demographic proxies to get at the same information. The latter is the focus of many studies that look at discriminatory behavior of employers in regards to a variety of factors ranging from race and gender to weight and smoking.

On the worker's side, what method of search one would use can lead to missed opportunities. In addition, their judgement as to whether the opportunities they know about are worth pursuing given their skill set and goals will impact the offers they eventually receive. Most search models reasonably assume searchers do not have complete information about all available jobs. While cost, as well as human capacity are part of the reasons for these, in some cases information barriers can be erected due to strategies adopted by firms. When considering commute outcomes, it becomes important to know of these systematic reasons especially when geographic targeting alters the possible application and offer rates that residents in some areas receive. Neckerman and Kirschenman (1991), for example, find that employment of inner-city blacks is hampered by recruitment practices which, when targeting neighborhoods, avoid inner cities. The economic benefits of hiring through referrals for firms is illustrated by Fernandez *et al.* (2000). Such a strategy would significantly lower the chances of individuals not connected to current employees to learn about opportunities. Other studies have also shown that finding a job is very much tied to whom one knows (Granovetter 1974). Finally, those who receive offers have to weigh among them by looking at wages, future prospects and other alternatives.

Given that both employer policies on advertising, screening, hiring, as well as worker strategies in searching are instrumental in forming the job-worker match, the model proposed below looks at both sides using an agent-based framework. The structure that is created by the interaction of searchers and firms suggests that even if minimizing commute time were a goal for searchers, it would have to be done within the context of what opportunities are known by the searcher. In brief, the model operates by having employers solicit applications to fill their open positions and unemployed workers (or others looking for new positions) searching and competing for employment. Having received interested applicants, firms undertake a screening step to filter out those that do not meet their hiring criteria, and select the best matching candidate from the remaining applicant pool. Workers, both employed and unemployed, search for opportunities, submit applications to those that they consider worth pursuing, and accept or reject offers based on what alternatives they have and their current state. Both workers and jobs are assumed to have skills and skill requirements respectively that are important in matching one another. Other social factors, such as the presence of contacts for example, are also assumed to have a role in promoting certain job-worker matches to occur. The model also provides a flexible platform that can be expanded to include different decision frameworks and policy variables. Subsequent sections will discuss the model and its performance using different data.

3 The Model

The model proposed here matches origins and destinations using employment search at the individual level. The outcomes depend on skills of the searcher, compensation, commute preferences, the locations of employment opportunities, and the willingness of firms to employ the searcher. The geographic plain on which the modeling is undertaken contains both employment locations which may be flexibly arranged into one or multiple employment zones or be randomly distributed, as well as residential locations which can also be distributed as desired. Firm and housing locations are assumed to be exogenous in the model. Workers will search for employment from fixed home locations.

The model contains both active agents which interact with one another through out the simulation and inactive agents which are mainly used to mark location and to house employment opportunities.The inactive agents in this model are job centers, where firms are located, and the firms where employment positions are housed. Job centers and firms are present to give structure to the location of employment opportunities. Job centers (which may be one or many) house firms, and firms house employment opportunities. The presence of job centers is optional. When job centers are not present firms can be distributed through out the modeled area randomly. All employment opportunities are housed within a firm to which they are randomly assigned. The active agents in this model are the workers and employment positions which interact with one another in determining job opportunities and pay scales, and negotiate agreeable arrangements for employment. Each of these agents are discussed below.

3.1 Job Centers

The purpose of job centers is to house firms. These are established as optional fields where a mono-centric, poly-centric, or a city with distributed employment can be modeled in the the home-job matching process. The location of the job centers can be at any location on the plain that is being modeled, though when mono-centric models are considered the location has been fixed at the center of the geographic area.

3.2 Firms

Firms house employment locations. When job centers are present, firms can locate in only one of the job centers. Assignment of firms to job centers is done randomly at the start of the simulation. In the current model, once a firm chooses a location, it does not relocate. Employment locations are also assigned to the firm randomly. Once employment positions are assigned to them, firms know how big they are what types of positions they have. Though the number of employees at a particular firm may change, the number of positions that are available at each of the firms does not change throughout the simulation.

3.3 Employment Positions and Workers

Employment positions are housed in Firms. Each employment position has characteristics that it requires fulfilled by potential employees (or a minimum skill set that is needed to be fulfilled). The skill set required by any position (J_q) is assigned as a randomly generated integer ranging from one to five. Each of these is assumed to be increasing in specialization and commands an average pay that is higher than the preceding level. Each position is assumed to have an amount that it is willing to pay an employee. At the start of the simulation, the pay that positions are willing to offer is assigned to the jobs by pulling from a uniform distribution whose mean is a function of the position's skill level as shown in equation 1 and whose range is \$10,000. Alternatively, wage dispersion can be set to 0, leaving the wage to be only a function of the desired skill.

$$W_o = J_q * 10000 + W_{disp} * Q_p \qquad \left\{ Q_p \sim Unif(0, 10000) \right. \tag{1}$$

where

J_q: the job-class for the position.

W_o: is the amount that the position offers to pay prospective employees.

Q_p: is a random draw from a uniform distribution with a range of \$10,000.

W_{disp}: indicates whether there is wage dispersion at a given skill level

At any given time, a positions can be open or taken (closed). When a position is open, it automatically advertises itself, and job seekers who encounter it can apply to occupy the position. When a position is already occupied by a worker, it is not searchable and does not take any applications. Employment positions know how well applicants as well as the person occupying them matches the requirements of the job. Each employment position acts as would a human resources department in real life, by accepting and screening applications as well as making offers, and negotiating a salary with qualified applicants. When they have difficulty attracting talent, positions increase their offer pay at each iteration.

Workers start out randomly assigned to residential locations from which they search for jobs. Workers residences are assumed to be stationary. Each worker is randomly assigned a skill class (S_c) that corresponds to the job-classes (J_q) for the employment positions. At the start of the simulation all workers are seeking employment. They search for open positions that fit their skills and put in applications reporting their qualifications. Each worker is also assumed to have a minimum wage (W_m) that they would want to accept any job offer. This is set at $W_m = S_c * 5000 + W_{disp} * Q_a$, where $Q_a \sim Unif(0, 5000)$. At the beginning of the simulation, each worker also has an expected wage (W_e^u) which is set 10 percent higher than W_m. Once the searcher is employed, their expected wage (W_e^e) will be set greater then or equal to their earnings at the time of search.

Workers are assumed to have limited information on available positions that match their skills. To find information, workers have to start searching for opportunities with some intensity I. Different workers can have different search intensities that describes how many applications they put in at any given time slice. A worker only receives offers from those positions to which it has applied.

Though skill matching is an important part of the model, workers are allowed to apply to positions for which they are slightly under or over qualified. They however are assumed to avoid applying to untenable positions by comparing their skill class (S_c) with the job class (J_q) of the employment position. Some searchers can use a contact to gain access to employment. A proportion of these contacts are assumed to be influential and can leverage their position to increase the match between the applicant and the open position even though the match of skills to criteria may not be perfect (or even when other better matches might be available).

The model allows for individuals to receive any number of offers at a given time given they have applied to the position, and they are selected as the best applicant for that position. When several job offers are made to the respondent within a given iteration, the model assumes they arrive such that they can be compared against one another simultaneously. Once an offer is made to a worker, searchers choose which offer is the best and decide to accept or reject the offer by comparing its offer wage and transportation costs to their current situation. The selection process may be specified so that a deterministic decision framework is adopted where the highest offer is chosen, or a probabilistic decision is made based on travel time and salary considerations within the Expected Utility framework. Decisions are also assumed to be made only on the basis of offers and current wages or reservation wages. Workers do not know what the likelihood of offers in the next time slice will be. Offers that improve the net present value of their net income (wages minus commuting costs discounted over expected tenure) are always accepted. Further all workers' residential locations are assumed to be fixed.

When searching, those that are already employed adjust their asking pay so that it is higher than their current salary. Those that are unemployed will lower their asking wage until it reaches their reservation wage for each iteration that they remain unemployed. To stay competitive employment positions also offer annual increases for their employees. In part these raises ensure that employers retain employees for longer durations than they would otherwise. The raise amount is randomly generated from a uniform distribution and implies a variability in the wages offered for similar positions. Researchers have empirically shown that similar workers receive markedly different wages for similar types of jobs (e.g., Krueger and Summers (1988); Murphy and Topel (1987)) whose existence has been theorized to arise from different reasons including employer wage policies, as well as unmeasured worker abilities (Christensen *et al.* 2005).

3.4 Job Search and Matching

The job search and matching process consists of a series of steps where worker agents seek employment, weigh offers, and decide on positions, while employers advertise, evaluate applicants and make hiring decisions. The overall flow of the decision process and interaction between these agents is shown in Figure 1. At the start of the simulation all workers do not have jobs and all positions are open. Workers start out deciding with what intensity I they will search for a job. The search intensity determines the number of applications a worker submits at a given iteration of the simulation. A global variable sets the minimum and maximum search efforts, and each agent selects their intensity within these limits. Since there is some cost to the application process, we assume that workers apply to positions that are reasonable fits for their skills. Once a searcher has selected the intensity I with which to search, applications are submitted to I open positions that satisfy equation 2. In cases where I exceeds the number of positions available that meet the application criteria, the worker applies to all available qualifying positions.

$$\left| J_q - S_c \right| \leq tol \tag{2}$$

where:

J_q: Job class for a position being considered

S_c: Skill class for the searching individual

tol: A global tolerance level among searchers for under or over-qualification

When the tolerance level in equation 2 is set to 0, workers only apply to jobs that are a perfect match to their skills. Greater numbers imply over or under-qualified applicants can also apply and compete for a position. There are at least two factors that can make slightly under-qualified applicants appealing to the employer. First is that their asking salary may be lower as compared to those with more skills. Second, job searchers may use contacts to find employment, a proportion of whom may be able to influence the outcome of the hiring process. Two global variables in the model P_c and P_{cl} control the proportion of people who find jobs through contacts and the proportion of those whose contacts can leverage their relationship with the employer to assist in better matching respectively. Currently whether a contact is used and whether the contact is influential is set randomly according to probabilities equal to P_c and P_{cl} at each iteration, and no information as to the identity of the contact is generated.

3.5 Evaluating Applicants and Making Offers

Open positions which have received applications evaluate each applicant using two steps. In the first step employers remove unqualified applicants from the pool and retain only those applicants that satisfy equation 3. Positions then randomly sample one of the qualified applicants and extend an offer.

$$S_c + C_l \geq J_q \tag{3}$$

where

S_c: Skill class for the applicant j

C_l: Whether applicant j used an influential contact (1 if true, 0 otherwise)

J_q: Job class for the position being considered

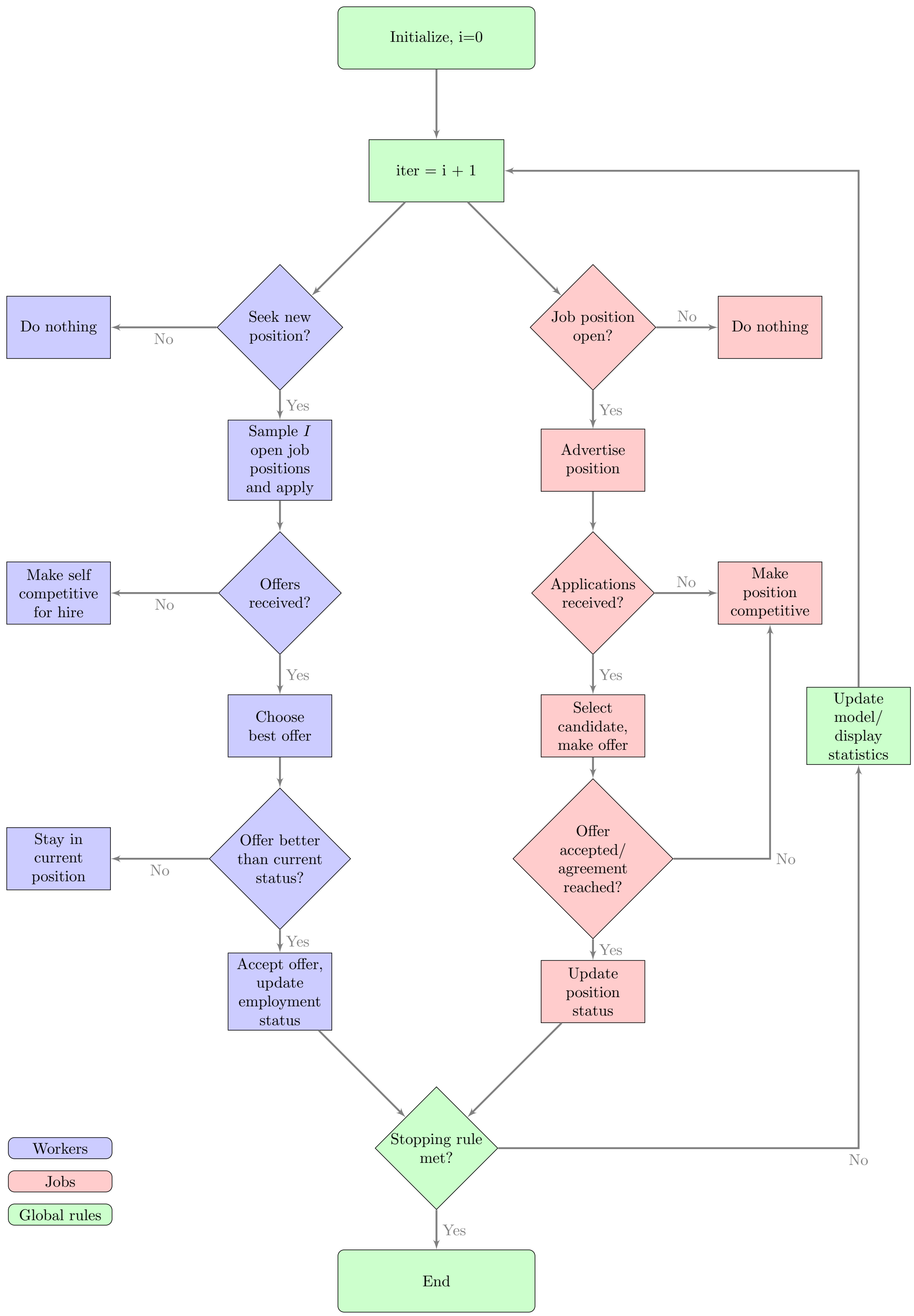


Figure 1: Flowchart of the model. The sequence of decisions for each agent is shown. Interactions between workers and jobs (e.g., putting in applications, receiving offers) are not linked by arrows for clarity.

3.6 Weighing Offers and Searcher Decisions

Workers that the employer selects as best meeting their criteria receive offers of employment. A searcher may receive more than one offer depending on where they applied and whom they competed against. The decision on the part of the searcher to accept any given offer proceeds in two steps. First, the applicant selects which offer they regard as the best among the competing offers. Second this best offer is weighed against the current state that the searcher is in. A decision is made whether to accept the offer or remain in the current position by comparing the net present value of their income stream less transportation costs over their expected tenure. The selection of the best offer among competing offers is done by weighing the travel distance and the offer wage in an expected utility framework. The probability that a person considers offer k the best among the competing offers made to them is calculated according to equation 4.

$$p_{ij}^{b} = \frac{e^{\beta_{1i} * T_{dj} + \beta_2 * W_{oj}}}{\sum_{k=1}^{n_{offers}} e^{\beta_{1i} * T_{dk} + \beta_2 * W_{ok}}} \tag{4}$$

where

p_{ij}^{b}: The probability that worker i selects offer j as the best among competing offers

T_{dj}: The daily travel cost (round trip travel time to the offered job) of offer j (in hours)

W_{oj}: The offer pay that position j is willing to pay (\$/day)

β_{1i}: The coefficient for travel cost for person i.

β_2: The coefficient for wages (set at 1).

The parameters for the utility function are selected so that individual level differences are allowed in the travel cost variable. The marginal rate of substitution between travel cost and wages is selected so that it represents the agent's expected wage when searching. With β_2 fixed at 1 for all individuals, this would mean that -β_{1i} = W_{ei} for person i. The selection of β_2 to be equal to 1 is arbitrary. Any other selection would not alter the order of preferences for any given set of alternatives though the probabilities may change.

Once a searcher determines the best offer, this offer is then compared to their current position using the net present value of staying where the agent is currently and the new offer over the expected tenure of the new position. The net present value calculation takes into consideration income and transportation costs only, and at a yearly discount rate γ, it is calculated as shown in equation 5.

$$NPV_{t_e} = \sum_{t=0}^{t_e} \frac{W_k - VOT * T_{dk} * N_w}{(1+\gamma)^t} \tag{5}$$

where

NPV_{kt_e}: the net present value as a result of being at job k for an expected tenure of t_e

W_k: annual wage at job k

VOT_k: value of commute time for person j at job k (assumed W_k)

T_{dk}: The daily travel cost (round trip travel time to the offered job) of offer j (in hours)

N_w: Number of working days in a year

γ: the annual discount rate

t_e: the expected tenure at the new position

This value is calculated both for the current job that the individual holds and the new best offer that the individual receives. If the NPV of the best offer is higher than that for the current job, then the worker chooses to accept the new position; otherwise, they will stay and continue searching. Searching and job relocation costs are assumed negligible in the NPV calculation. Workers will accept the offered job as is if it is above their expected pay. When the offered salary is less than what is expected by the searcher, a negotiated salary is assumed that is randomly determined to be between the offered salary and the expected salary. This new salary serves as the basis of future search when the worker decides to search again.

3.7 Changing Employment

Each year after a worker has found employment, they reevaluate whether or not they should search for an alternative position. While several reasons may motivate searching for another position while already employed (on-the-job search), here we assume the probability to search for another job is a function of tenure. Farber (1999) makes two observations that are important here about the U.S. labor market. First is that long-term employment relationships are common, and that beyond a certain point, workers will choose to stay with their current employer; and second, that most new jobs end early. The overall effect is that with increasing tenure, the probability of job change declines (Farber 1999). Here, the probability for on-the-job searching is modeled so that it is decreasing with tenure. Each year the probability that a person would want to change employment is given by equation 9. If a search at any given time was not successful, the workers will have to decide whether or not to search in the next cycle.

$$p_{rel} = 1 - \frac{e^{(t_r - \bar{t_r})}}{1 + e^{(t_r - \bar{t_r})}} \tag{6}$$

where

p_{rel}: is the probability of wanting to change jobs

t_r: tenure at current employment

$\bar{t_r}$: the tenure for the population beyond which the probability to start searching for another job declines below 0.5 (a global variable in the model).

3.8 Competition

In trying to match one another, both employers and workers can make themselves attractive to each another. Each year that a position remains open, it can increase its offer price to attract applicants. Similarly workers that are unemployed can make themselves attractive by reducing their asking price each iteration they remain unemployed until they reach their minimum pay. Those that are currently employed and searching can increase their expected wage by adding a few percentages to their existing salary. Similarly employers can also increase the wages paid to their current employees each year to counteract any incentives that rising offers at open positions may have. These adjustments are assumed to be random draws from a distribution whose maximum is specified at the beginning of the simulation. The magnitude of these changes are assumed to decrease as the length of search or tenure increases. In reality, employed individuals retire and limits exist for how much a firm is willing to pay for a particular position. By implementing these caps, a position at any given skill level increases the wage offered only to a point when it remains unfilled. Expected wages also decline to a minimum beyond which the person is unwilling to work. Alternatively, by setting these percentages at zero, the search can be conditioned on fixed wages. These cases are compared in the following section using a simplified urban landscape.

$$W_{o,t} = W_{o,t-1} * (1 + \gamma_o^{f(u_o)}) \tag{7}$$

$$W_{e,t}^u = max(W_{a,t-1}^u * (1 - \gamma_c^{f(u_s)}), W_{min}) \tag{8}$$

$$W_{e,t}^e = W_{t-1} * (1 + \gamma_r^{f(t_r)}) \tag{9}$$

where:

$W_{o,t}$: Wage offer at time t

$W_{e,t}^u$: The expected wage at time t, when the person is unemployed

$W_{e,t}^e$: The expected wage at time t, when the person is employed

W_t: The wage received at time t by the searcher

γ_o: The rise in offer wages by employers when a position remains open $\gamma_o \sim U(0, \gamma_{o,max})$

γ_c:The decline in expected wages when unemployed, $\gamma_c \sim U(0, \gamma_{c,max})$

γ_r: The rise in wages each year a person remains employed, $\gamma_r \sim U(0, \gamma_{r,max})$

u_o: the number of iterations the offering positions has stayed open

u_s: the number of iterations the unemployed worker has been searching

t_r: tenure at the current position

$f(.)$: function, here equal to $f(x) = \frac{x}{10}$.

3.9 Stopping rules

Currently the model stops when employment conditions have not changed for any agent for 50 iterations (the equivalent of 4.2 years in the tenure calculations) or if the total number of iterations has reached 500. Under the simplest conditions, the model stops well ahead of the 500 count.

4 Testing with simplified models

The model as described was implemented using NetLogo (Wilensky 1999). A screen shot is given in figure 3. To see if the model gives reasonable outcomes, tests are performed using simple urban structures where workers are randomly generated and employment is concentrated at a finite set of locations. The simplest scenario is the mono-centric urban model where all employment is at one location. However, since neither workers' residences or workplaces are allowed to change, this model would have everyone working at the core and travel times would be determined by where housing is located relative to the center. Hence we start with a slightly more complicated landscape where there are four employment zones, each located at the center of a quadrant that divides the modeled area into four zones.

To simplify the model, the total number of employment opportunities available at each employment zone are made equal. A total of 520 workers are also randomly generated and distributed through out the plane, but with each quadrant having equal number of residents (see figures 2a to 2d). Several versions of this layout are tested while changing different variables in the model but without altering the landscape (the location of employment and workers' residences). In the simplest cases tested, it is also assumed that there is no skill differentiation, and no wage dispersion at the given skill level. Each of these are relaxed and the commute outcomes of the model are compared. In addition, different combinations of minimum and maximum search effort, and the possibility of contacts enabling matching are also modeled.

Four test cases that illustrate how the model matches workers and jobs are discussed next. Under each of these cases one or two variables are changed to illustrate how the OD matching is affected. In case 1, all jobs and workers have no skill differentiation and all wages are the same. In case 2, wages are assumed to have a uniform distribution but no skill differentiation exists. In case 3, skill differentiation exists, wages are also an increasing function of skills, however there is no wage dispersion at any skill level. In case 4, skill differentiation exists and for 25 percent of workers searching for work, contacts can improve their chances of hire if they are under-qualified for a position. Three skill levels are assumed but there is no wage dispersion at at any skill level. In each of these cases the search intensity for each agent when searching is a random draw from a uniform distribution ranging from 0 to 5. Each model is run until no employment changes occur for any agent for 50 consecutive iterations. Table 1 summarizes the main attributes of these cases.

Figures 2a to 2d show the matched home and work outcomes from one typical run under the described case. For clarity only the home-work connections for the jobs located in one quadrant are shown. The large black dots are the employment centers, and the small red dots are the workers (located at their residences).

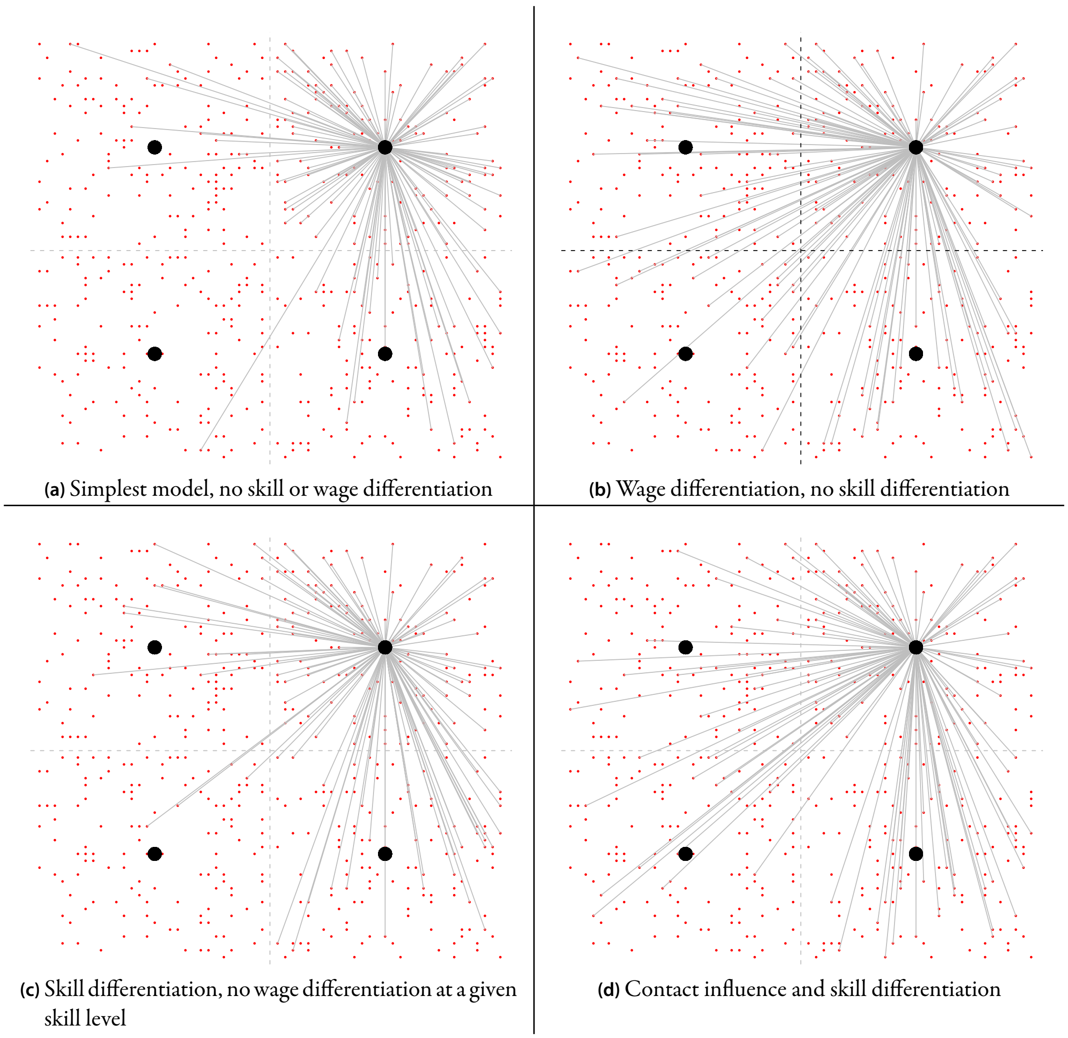
The first case, as expected, leads to the shortest home to work distances. Most of the workers in the quadrant also live in the same quadrant. The outcome is similar in each of the other quadrants. The average home to work distance in this case is 16.3 units. In case 2, all attributes of the model 1 are kept the same, with only wage differentiation allowed. In this case the wage offered by any one position is a random draw from a uniform distribution between \$10K-\$20K. In this case the average distance travelled rises significantly to 23.3 units. In case 3, each of the jobs as well as workers is randomly assigned one of three job classes. Here again, matching the differentiated skills leads to higher home to work distances. Case 4, also leads to higher distances as positions relax their requirements and allow under-skilled but socially connected individuals, but much of the rise is accounted for by skill differentiation. Indeed there is no a-priori reason for jobs found through contacts to lead to shorter commutes unless the process occurs through neighbors who themselves have short commutes. Based on 100 simulations each, in moving from the simplest model (case 1) to a model with wage dispersion (case 2), the sum of commute distances rises by about 76 percent as those with attractive wages become less likely to take on lower paying but closer jobs. This value reduces to 48 percent when both wage dispersion and skill differentiation are both present. The model with skill differentiation but no wage dispersion at a given skill level (case 3) leads to a rise of about 5.5 percent.

Table 1: Model parameters for illustrative cases

case	tol	I_{min}	I_{max}	Job classes	$P_c * P_{cl}$	W_{disp}
1	0	0	5	1	0	0
2	0	0	5	1	0	1
3	0	0	5	3	0	0
4	1	0	5	3	0.25	0

5 Model Stability

To test the model's sensitivity, the commute outcomes to different combinations of the input variables is tested. Two general cases are studied. In case 1, wages are taken as exogenous to the search process and adjustments do not occur as described in section 3.8. In this case γ_o, γ_c, and γ_r are all set to 0, which means only the expected wage when employed rises to the wage currently being paid. In the second case, each of the rise proportions (γ_o,γ_c,γ_r) are assumed to be random draws from a uniform distribution between 0 and 0.05. The input variables in each case are the existence of wage dispersion, skill differentiation, the use of influential contacts, and the skill



(a) Simplest model, no skill or wage differentiation

(b) Wage differentiation, no skill differentiation

(c) Skill differentiation, no wage differentiation at a given skill level

(d) Contact influence and skill differentiation

Figure 2: Toy model commute outputs under different assumptions

mismatch tolerance in considering applicants. In each of these tests the minimum and maximum search intensities are set to be random draws from a uniform distribution between 0 and 5. Each of these factor combinations are set to two levels for a 2^5 factorial design. The 32 different combinations are run 25 times each, for a total of 800 simulations over which the commute outcome was evaluated. In each of these runs, the employment locations are fixed. Residential locations on the other hand are freely chosen within each quadrant while each quadrant still has equal number of workers. Residents are allowed to locate freely since we are mainly concerned with the overall pattern of commute in these tests.

The results from these tests illustrate the model is stable within replications, and responds in a consistent manner to adjustments of the variables in the model. Table 2 presents an analysis of variance on the mean home to work distance under the tested combinations. Changes in the variables of the model account for the majority of the behavior that is observed in successive runs of the model. About 80.4 percent of the variance in the average commute outcome is accounted for by the levels of the factors. Wage dispersion accounts for a significant amount of the variance. In transitioning from the simplest model (Case 1) to one where there is wage dispersion, the average home to work distance rises by 5.5 units as searchers compete for the best fitting job available.

6 Testing the Model with Minnesota Data

The model was applied to a subset of the 2000 TBI (Travel Behavior Inventory) data for the Twin Cities region. This data set was chosen because it includes disaggregated individual data for workers and the location of their of employment. Consideration was also given to using the Longitudinal Employer-Household Dynamics (LEHD) data set. However, while the LEHD includes blockgroup level data for both income and sector of employment, there is no link between these two variables at both the residential and workplace ends. It is therefore not possible to know which sectors are paying what amount to workers in a blockgroup. The downside to using the TBI is that the data does not contain employment sector information or more details that could be used on the workplace. Nonetheless, the disaggregate detail allows the individual level identification of each worker and their demographic and education characterstics. In addition, the necessary education level and salary levels for the employment locations can also be approximated by the characteristics of the person currently occupying these positions.

Since income is reported at the household level in the TBI, for this test application we use only single worker households that reported one job and lived and worked in the metropolitan area. While it is possible that respondents reported income includes non-wage earnings, this approach at least avoids additional assumption about how income should be subdivided among workers in multi-worker households or between jobs when a second employment is reported. A total of 805 individuals were extracted from the data who, in addition to the above criteria, had also reported their education level. The education level was used to classify the skill level of the respondent as well as the job-class of the respondent's place of employment into one of five categories. Once these data are extracted from the TBI, the link between home and work are severed, and the challenge is to re-establish the link using the agent-based model.

Before running the simulation, certain modifications had to be made to the model to make it work with the TBI. First the definition of job classes as well as skill classes had to be adjusted. Since the data includes the education level of respondents and where they work, each job is marked as requiring the skill class of the person who currently occupies it. The workers skill class is also marked by their education level. The five categories used for skill classification correspond to "Below High school," "High school graduate," "2 years of college/Associate's Degree," "4 years of college/Bachelor's Degree" and "Post-graduate." The initial expected wage of each worker is set at 70 percent of their reported income. In addition, the minimum pay below which agents will not work is set to 70 percent of their initial expected wage. The offer pay at the start of the simulation for the positions is set at the 70 percent of the income that the respondent currently working there reported. Each position is allowed to offer up to 75 percent above what its occupant reports as income in making itself competitive. The function $f(.)$ in equation 9 is set to 1 so that wage change percentages are a random draw from the uniform distributions defined by $\gamma_{o,max}$, $\gamma_{c,max}$, and $\gamma_{r,max}$. In addition to these modifications, parameters of the simulation were set to reflect the limitations imposed by the data. The matching tolerance (tol, equation 2) as well as the proportion of contacts that are influential (P_{cl}) are both set to zero.

Other aspects of the basic model remain the same. Different combinations of I_{min}, I_{max}, $\gamma_{o,max}$, $\gamma_{c,max}$, and $\gamma_{r,max}$ were used to run the simulation to replicate the distribution of distances as observed in the TBI data. For the Twin Cities case, the model is run until all working individuals from the simulation have been at their last jobs at least 100 iterations

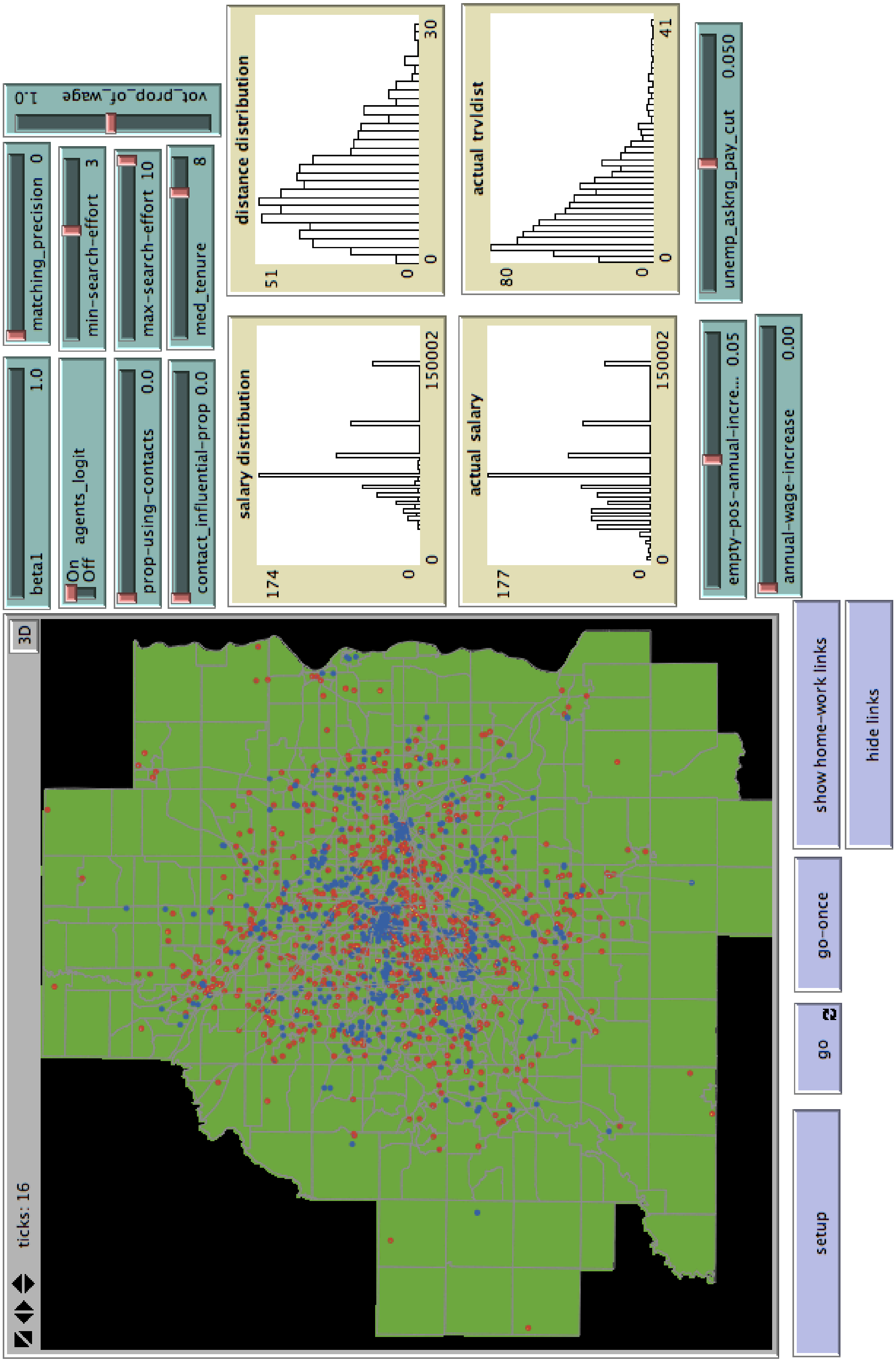


Figure 3: Screen shot of ABODE. Red dots are workers and blue dots are employment positions.

Table 2: Analysis of variance for test model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Wage rise	1	1228.92	1228.92	675.94	0.000
Wage rise:Match tolerance	2	391.12	195.56	107.56	0.000
Wage rise:Skill differentiation	2	39.55	19.78	10.88	0.000
Wage rise:Contact use	2	57.20	28.60	15.73	0.000
Wage rise:Wage dispersion	2	4181.93	2090.97	1150.09	0.000
Residuals	790	1436.29	1.82		

(which is to say no one has changed their job in the last 100 iterations) or unemployment levels are at 4 percent or less and the model has run for 1500 iteration. In general the first condition tended to dominate and achieved employment for 93-95 percent of the searching agents.

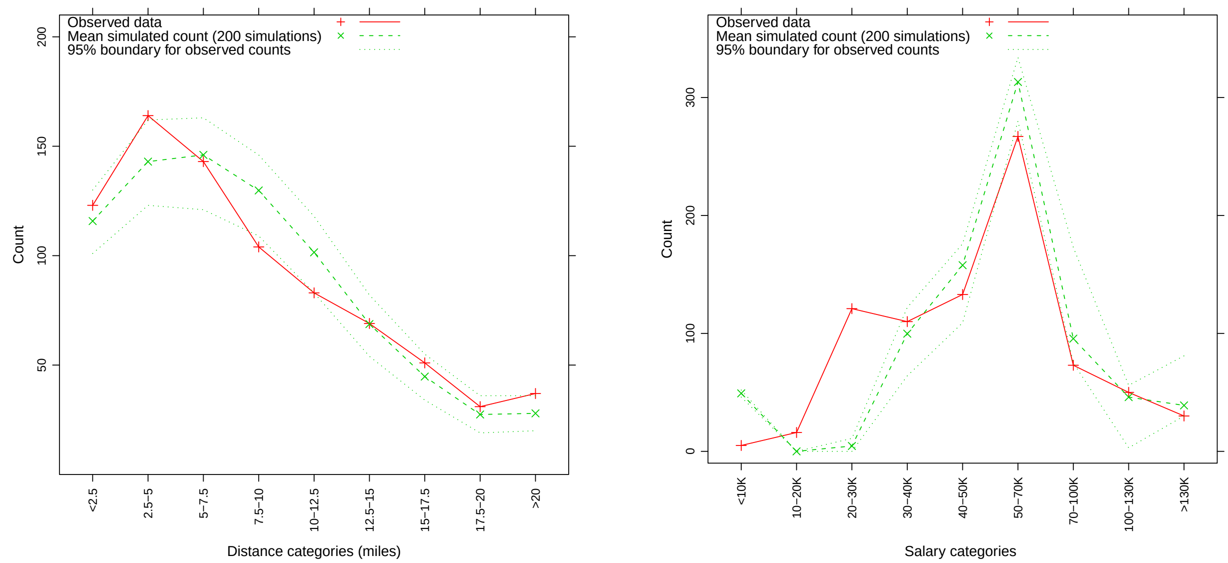
The tested combinations of the model parameters and the values that replicated the commute distribution for the workers adequately are shown in table 3. A combination of longer runtimes with moderate wage raises annually for employed persons ($\gamma_{r,max}$ = 0.02), latitude for aggressive recruitment by increasing wages for positions that are not taken ($\gamma_{o,max}$ = 0.1), and yearly declines in expected wages for the unemployed ($\gamma_{c,max}$ = 0.05), leads to on aggregate reasonable distance and wage distribution. Figures 4a and 4b show the counts of individuals in different commute and wage categories from 200 simulations using the best fit parameters. Boundaries around the mean simulated count for each distance and income category reflect the range of 95 percent of the counts from these simulations. Visual examination indicates that the overall distance distribution provides a good fit for the observed data. While the wage distribution doesn't replicate the observed wages as well as the distances distribution especially in the lower income ranges, the simulated results are more acceptable for income ranges of above $30,000.

The job matching mechanism in the model leads to aggregate distributions that are in tune with observed distributions. However, individual level matching is much less precise. These challenges can be seen in figures 5a and 5b. Each plot shows the distribution of home to work distances and salaries from the simulation for those individuals whose actual distance and salary falls in the categories on the horizontal axis. Though the increasing trend points in the right direction, both plots suggest that there is considerable variance in the predictions for those in each category. There are several potential reasons why these mismatches appear. Primarily this is because the mechanism for matching skills and wages currently uses only education levels. Because this leads to jobs being identified as requiring a 'high school diploma', 'a bachelor's degree' and so on, employment opportunities that are not similar in the type of skill they require are assumed to be equivalent in the model. Since the link between education and income is not very strong in the data, a person who has a high school diploma may be offered the job currently being done by someone who has a similar education background but whose skill sets are much broader and command higher wages. Improved matching would be possible if information on sector, job type, and experience were available. These would make it easier for the employer agents to better filter qualified applicants.

Overall, despite challenges at the individual level, the model replicates the aggregate distributions quite well. It should be noted that the model is using data that is not very well suited for its needs. The use of education as a proxy for skills particularly is too general and allows persons to consider and take positions for which they are not qualified. As currently implemented, the model also assumes that people's commute decisions are permanent. In reality people can accept longer commutes with the hope of reducing their commute through a residential move. For others a shorter commute may allow them to choose housing at farther distances. Future direction for addressing such challenges and identifying data needs are discussed in the next section.

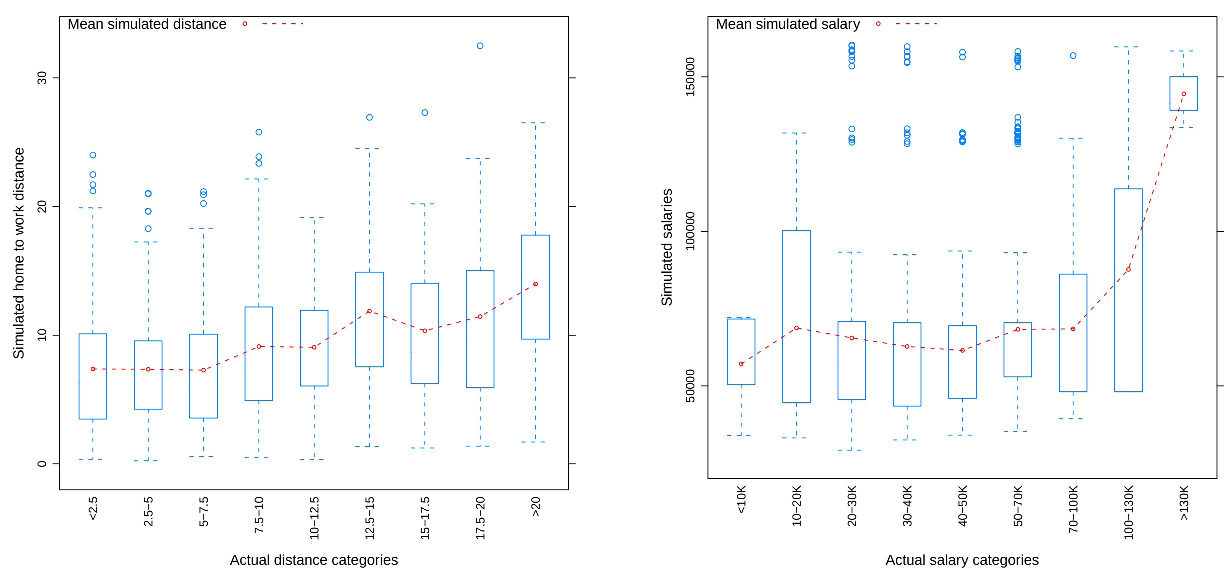
Table 3: Model parameters

Variable	Values	Best fit
I_{min}	0, 3, 5	5
I_{max}	5, 10, 15	10
$\gamma_{c,max}$	0.02, 0.05	0.05
$\gamma_{r,max}$	0.02, 0.05	0.02
$\gamma_{o,max}$	0.02, 0.05, 0.1	0.1
tol	0	0
$\bar{t}_r$	4	4
$P_{cl}*P_c$	0	0



(a) Distance comparison **(b)** Wage comparison

Figure 4: Mean simulated aggregate distance and wage distributions as compared to actual data



(a) Distance comparison **(b)** Wage comparison

Figure 5: Box plots of simulated distance and wage distributions for persons whose actual distance and incomes are shown on the horizontal axes. Plot is based on one random run of the model

7 Summary

This paper develops an agent-based job matching model. The model explicitly considers the job-search process and simulates a commute distance and wage distribution for workers. Using a toy urban area we show that the model leads to reasonable outcomes, with agents selecting the closest work place when wage and skill differentiation is absent. Relaxing these assumptions increases the observed commute. The introduction of wage dispersion at given skill levels in the model specially increases the average home to work distance significantly. The model is also tested using data from the Minneapolis - St. Paul Metropolitan area. The commute results from this simulation on aggregate capture the trends in the observed data and illustrate that the behavior rules as implemented lead to reasonable patterns. The wage distributions of the simulation also provided a reasonable match though to a lesser extent. But weaknesses were present in replicating individual level distances and wages. In part these weaknesses are the result using data collected to be used in a traditional gravity model that did not include (nor require) details on the job seeker and employment opportunities that are key to the current model.

As the demand for more disaggregate models increases, we believe the agent based approach holds several benefits that make it attractive. Primary among these is that it is easily extendable. One could easily move from modeling relatively long-term decisions at the individual or household level to using them as constraints to other short-term decisions. Individual-level and household-level tastes could be incorporated in a straightforward manner giving the agents qualities that are difficult to achieve in aggregate models. Once calibrated for jobs, the model could be used to model other shorter term location decisions such as where to shop or meet friends. The approach also provides an environment in which the environment and the agent are simultaneously affected by one another.

At least two lines of further work are essential. The first is to improve the matching by incorporating more detailed information on both searchers and jobs so that matching can further be refined. In the current data, for example, the connection between reported education levels and wages tended to be weak. Therefore, better ways of gauging the skill requirements of positions are needed. The data needed to more fully test the proposed model would need to include at a minimum data on skills and experience that go beyond education levels and ask specific streams of study or areas of expertise. In addition, wage data for different skills are also required. On the employer side, details on jobs would need to be more specific than the educational level of those currently occupying the position. The type of job, the industrial class of the firm, as well as the wages that are paid would help improve the match that is created.

A second line is to incorporate the effect of residential location (and relocation) in affecting what a person accepts as an acceptable commute. In addition, the tradeoffs multi-worker households make between each other's commute should be considered. The current model considers only individuals and assumes that they do not relocate to adjust their commute. However, some of the home-work commute observed in the test data is likely to have been measured after a residential relocation.

These extensions can strengthen the model and make it useful not only for OD matching, but also for testing different policy prescriptions that may impact both job search and commute outcomes. By working at the individual level, the approach also makes possible other extensions which look at the persons non-work activities while carrying over their personal attributes and constraints.

References

Alonso, W. 1964. *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge, Mass.: Harvard University Press.

Christensen, B., R. Lentz, and D. Mortensen. 2005. On-the-job search and the wage distribution. *Journal of Labor Economics*. URL http://www.journals.uchicago.edu/doi/abs/10.1086/425432.

Farber, H. S. 1999. Mobility and stability: The dynamics of job change in labor markets. In D. C. Orley Ashenfelter, ed., *Handbook of Labor Economics*, volume 3b of *Handbooks in Economics*. Elsevier Science.

Fernandez, R., E. Castilla, and P. Moore. 2000. Social capital at work: Networks and employment at a phone center. *American Journal of Sociology*. URL http://www.journals.uchicago.edu/doi/abs/10.1086/210432.

Giuliano, G. and K. Small. 1993. Is the journey to work explained by urban structure? *Urban Studies*. URL http://usj.sagepub.com/cgi/content/abstract/30/9/1485.

Granovetter, M. S. 1974. *Getting a job: a study of contacts and careers*. Cambridge, Mass.: Harvard University Press.

Hamilton, B. W. 1982. Wasteful commuting. *The Journal of Political Economy*, 90(5):1035–1053.

Holzer, H. J. 1987. Hiring procedures in the firm: Their economic determinants and outcomes. Working Paper 2185, National Bureau of Economic Research. URL http://www.nber.org/papers/w2185.

Krueger, A. and L. Summers. 1988. Efficiency wages and the inter-industry wage structure. *Econometrica*. URL http://www.jstor.org/stable/1911072.

Mills, E. S. 1972. *Studies in the Structure of the Urban Economy*. Johns Hopkins Press.

Murphy, K. M. and R. H. Topel. 1987. Unemployment, risk, and earnings: Testing for equalizing wage differences in the labor market. In J. S. L. Kevin Lang, ed., *Unemployment and the structure of the labor market*. New York: Blackwell.

Muth, R. F. 1969. *Cities and Housing; The Spatial Pattern of Urban Residential Land Use*. The University of Chicago Press.

Neckerman, K. and J. Kirschenman. 1991. Hiring strategies, racial bias, and inner-city workers. *Social Problems*. URL http://www.jstor.org/stable/800563.

Rouwendal, J. 1998. Search theory, spatial labor markets, and commuting. *Journal of Urban Economics*. URL http://linkinghub.elsevier.com/retrieve/pii/S0094119096920331.

Rouwendal, J. 1999. Spatial job search and commuting distances. *Regional Science and Urban Economics*. URL http://linkinghub.elsevier.com/retrieve/pii/S0166046299000022.

Stigler, G. 1961. The economics of information. *The Journal of Political Economy*. URL http://www.journals.uchicago.edu/doi/abs/10.1086/258464.

Stigler, G. 1962. Information in the labor market. *The Journal of Political Economy*. URL http://www.jstor.org/stable/1829106.

van Ommeren, J. 1998. On-the-job search behavior: The importance of commuting time. *Land Economics*, 74(4):526–540.

van Ommeren, J. and P. Rietveld. 2002. Commuting, spatial search and labour market bargaining.

van Ommeren, J., P. Rietveld, and P. Nijkamp. 1999. Job moving, residential moving, and commuting: A search perspective. *Journal of Urban Economics*, 46(2):230–253.

van Ommeren, J. and J. van der Straaten. 2008. The effect of search imperfections on commuting behaviour: Evidence from employed and self-employed workers. *Regional Science and Urban Economics*. URL http://linkinghub.elsevier.com/retrieve/pii/S0166046208000045.

White, M. 1988. Urban commuting journeys are not "wasteful". *The journal of political economy*. URL http://www.jstor.org/stable/1837250.

Wilensky, U. 1999. Netlogo. Center for connected learning and computer-based modeling, Northwestern University. Evanston, IL. URL http://ccl.northwestern.edu/netlogo/.