A note on waiting

Gautier, Pieter A.

1997

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Gautier, P. A. (1997). *A note on waiting*. (Research Memorandum; No. 1997-37). Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam.

Faculteit der Economische Wetenschappen en Econometrie

Serie research memoranda

A note on waiting

Pieter A. Cautier

Research Memorandum 1997-37

August 1997

/a/l/e/r/t/

applied
labour
economics
research
team

vrije Universiteit amsterdam

A note on waiting

Pieter A. Gautier

Tinbergen Institute, Free University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

*When students apply to universities or unemployed workers search in the labor market, an increase of time input of one person (resulting in more applications) will increase the waiting time **of** the others. In this paper it will be show that in a decentralized market there will be excessive congestion Moreover, congestion increases with both the mean and the variance **of** the net benefits associated with waiting across the population.*

I. Introduction

Since the seminal work of Becker (1965), the economic importance of the allocation of time is widely accepted. Becker showed that consumers can be viewed as producers of commodities by combining inputs of goods and time according to the standard theory of the firm, and that in many occasions time and goods are substitutes. For major sport events or concerts for example, one can choose to spend a night at the ticket office or pay much more at the black market. For some composite goods, the time input increases the value of the product. The longer one is willing to spend in the waiting room, the more medical care one will get and the more application letters one writes, the larger the probability to get a job offer will be.

In many of those activities there are important externalities when an increase in the time input of one agent increases the waiting time of other agents. If for example filling a vacancy requires a positive amount of screening time, then any individual job searcher who decides to join one or more job queues, will increase his individual hiring probability but will also increase the average waiting time of the other job searchers. The probability to get a subsidy or rent controlled housing is also increasing in (time consuming) search effort. At the aggregate level however, high search intensity can lead to congestion. In other words, every time an agent joins a queue, he does not internalize the waiting costs and the "discouragement" effect he places on other agents.' Still, Barro and Romer (1987) showed that queues do not necessary reflect inefficiencies. In many cases the price per effective unit of service adjusts automatically if the explicit price is held fixed. One gets for example faster and better service in empty planes or restaurants than in full ones.

This paper gives some examples of cases in which there will be excessive waiting. We will see that congestion increases with both the mean and the variance of net rewards of waiting across the population. In the next section a small model of waiting will be presented in which we allow individuals to join more than one queue. Section III considers the case in which individuals differ in opportunity costs and rewards of waiting and section IV presents a labor market example.

II. The model

Consider an action A which can be completed by a limited amount of agents per time interval. Let the rate at which agents can complete A be exponentially distributed with parameter μ. We can think of A as e.g. applying to a university, buying a ticket, making a public phone call, or screening a job applicant. We will assume that the expected rewards of completing A are equal to y. In a subsidized goods model, y is equal to the price minus the money value of the good, while in a job search model y is the probability of a job offer times the expected wealth change of a (new) job. In any case, the larger the value of y, the larger the expected benefits of waiting are.* We will furthermore assume that the arrival rate of agents who want to undertake action A follows a Poisson process with parameter α. In many cases agents can apply for more than one action per time interval (e.g. a firm can apply for different subsidies, and a worker can apply for more than one job). Let ξ_i be the number of actions which agent i can undertake at a

[1] The author of this article appologizes for increasing the average waiting time of other submitters.

[2] Akerlof et al. show for example that unemployed workers have incentives to wait longer with accepting a job offers in recessions because wages are expected to be higher in booms.

cost $c\xi_i^{\beta}$. If in a certain time interval, the arrival of applications $(\xi\alpha)$ exceeds the number of actions which can be handled in a certain time interval (μ), queues of agents will be formed. With this information we are already able to derive the steady state queue line distribution and the equilibrium amount of waiting agents.

A. Steady state properties

Let P_n be the probability to find n agents in the system. Then in the steady state, the probability of being in state n and moving towards another state, within a very small time interval h, has to be equal to the probability of being in state n-l or state n-2 and moving to state n. Hence.

$$P_n \alpha h + P_n \mu h = P_{n-1} \alpha h + P_{n+1} \mu h \qquad (1)$$

For every possible state, a balance equation like (1) can be written. As $\sum P_n = 1$, there is a unique solution for the ergodic steady state queue line distribution which can be found by repeated substitution.[3]:

$$P_n = \left(\frac{\alpha}{\mu}\right)^n \left(1 - \frac{\alpha}{\mu}\right) \qquad (2)$$

If we allow ξ to be greater than 1, we can determine the steady state queue length in persons as follows.

$$L = \frac{1}{\xi}\sum_{n=11}^{\infty} n\left(\frac{\xi\alpha}{\mu}\right)^n\left(1 - \frac{\xi\alpha}{\mu}\right) = \frac{\alpha}{\mu - \xi\alpha} \qquad (3)$$

Note that we divided the amount of units in the system by the average value of ξ to get the queue length in persons.[4]. The total expected duration to complete A (D) is given

[3] Details can be found in any standard operations research text book, see e.g. Gross and Harris (1985).

[4] The necessary condition for an equilibrium to exist is that the average screening rate exceeds the average arrival rate of applications ($\mu > \xi\alpha$). In what follows, we will assume that this condition will be met. If this condition is not met, waiting time will go to infinity.

by L/α or[5]:

$$D = \frac{1}{(\mu - \xi\alpha)} \tag{4}$$

When agents are "served" according to a FIFO (first in first out) principle, there are no direct incentives for ξ to exceed one[6]. But under a SIRO system (service in random order) there are potential benefits for agents to choose larger ξ's , An example of a SIRO system is search in the labor market.

The probability to finish A for an individual agent (p_i) can be written as $(\alpha\xi_i/\xi^c L)$, where ξ^c is the average value of ξ chosen by the other agents. We will assume that the action of agent *i* has a neglectable influence on the aggregate waiting rate and that agent *i* takes ξ^c as given, hence:

$$p_i = \frac{\xi_i}{\xi^c}(\mu - \alpha\xi^c) \tag{5}$$

So we see that the individual hiring rate is increasing in ξ_i, but that the aggregate hiring rate is decreasing in ξ because higher ξ's only lead to more congestion. The value of "waiting" in a SIRO system can be expressed as an asset:

$$rB_{Si} = p_i y - c\xi_i^{\beta} \tag{6}$$

Where rB_{Si} is the expected income stream of joining the queue and p_i is given by (4). Note that we expressed the costs of waiting as a flow variable but of course any flow variable can be transformed into a stock variable. The agent's problem is now to choose a value for ξ_i which maximizes rB_{Si}. To simplify things we will set β equal to 2. ξ_i^* can now be obtained by maximizing (6) with respect to ξ_i. This yields:

[5] See e.g. Little (1961).

[6] It is however easy to extend the model to allow for multiple service points. In that case agents can choose to join more queues so that also under a FIFO system there are incentives to join more than one queue. For an example in the labor market see e.g. Gautier (1996).

$$\xi_i^* = MAX\left[\frac{(\mu-\alpha\xi)y}{2\xi c}\ ,\ 0\right] \tag{7}$$

The socially optimal value for ξ^* will differ because a social planner will internalize the waiting externalities. Hence, ξ^* will be equal to one if the benefits of joining the queue exceed the costs or in other words if y exceeds its reservation value y^r:

$$y \geq \frac{2\xi c}{\mu-\alpha\xi} = y^r \tag{8}$$

and zero otherwise.

In the decentralized economy, ξ_i will be larger than or equal to one when condition (8) holds. If y is for example equal to $4\xi c/(\mu\text{-}\alpha\xi)$, all agents will choose ξ_i to equal 2. The probability to be selected to complete A does not change but the expected waiting time will be longer than in the social optimum where ξ^* would be 1. We see thus that the larger the value of y is, the more excess congestion there will be in equilibrium.

III. Differences in the rewards of waiting across the population

In this section we will allow for heterogeneity in the sense that either the opportunity costs of waiting or the rewards of waiting differ over the population. Not surprisingly, we will see that agents with low opportunity costs of waiting will be overrepresented in the queues.7

In the labor market for example, different workers have different expected rewards and opportunity costs of joining a queue. We will therefore consider y to be a drawing from a distribution F(y) with pdf f(y). Since ξ_i depends on y, ξ_i will be left censored at y $=y^r$.[8] The conditional mean (ξ_w) of ξ_i is given by:

[7] A similar point was made by Becker (1965).

[8] We will abstract from integer issues regarding ξ,here.

$$\xi_w = E(\xi_i) = \frac{(\mu - \alpha\xi^c)\int y f(y|y \geq y^r) dy}{2\xi^c c} \qquad (9)$$

Where:

$$f(y|y \geq y^r) = \frac{f(y)}{\bar{F}(y^r)} I(y > y^r) \qquad (10)$$

I is an indicator function which has value one if y $\geq y^r$ and zero otherwise. We see that when the mean of y increases (while the other moments do not change), ξ will be larger and consequently there will be more congestion. A change in the (unconditional) variance of y has an ambiguous effect on the unconditional mean (calculated over the entire population) of ξ_i.

Now let $H(y)$ be a mean preserving spread of F(y), where H is more spread than F. Then if y is drawn from H instead of F, both the fraction of agents who benefit a lot from joining the queue and the fraction of agents who have better things to do than waiting, will increase. The first group will choose large ξ_i's but the last group will cause the mass at the lower support of $H(y)$ to be larger than the mass at the lower support of F(y). The average ξ_i, over the agents who have decided to wait (ξ_W), will however unambiguously increase. Consequently, ξ_w will be larger when y is drawn from H than when y is drawn from F.

In the next section a labor market example will be given in which this ξ_W is the relevant variable.

IV. An example for the labor market

If we interpret y as the improvement in wealth associated with a job offer and μ as the screening rate, then the above described results imply that less agents will join job queues when y is drawn from $F(y)$ than when y is drawn from $H(y)$. Those agents who do decide to search will however search more intensive (choose higher ξ_i's) and will therefore cause more congestion on each other. The conditional mean of ξ_i will accordingly be larger under ***H.*** We will assume now that employers can choose how many vacancies to open, given that the expected rents of a vacancy (rW_V) are given by:[9]

$$rW_V = \frac{\alpha}{V}s - \gamma = 0 \qquad \textbf{(11)}$$

Where α/V is the probability that the vacancy will be filled, s is the employer's share of the match surplus and γ is the flow cost of keeping the vacancy open. Employers realize that only agents (and not applications) can fill up vacancies so the supply of vacancies will be independent of ξ_w.[10] Under free entry and when all profit opportunities are explored, rW_V will be zero in equilibrium. Hence $V=\alpha s/\gamma$. The easiest way to picture this job search process is by introducing a single labor office which screens all applications and offers jobs at a rate μV. Unemployment (U) can now be defined to be equal to the amount of waiting agents. Adjusting (2) yields:

$$U = \frac{\alpha}{\mu V - \xi_w \alpha} = \frac{1}{\mu s/\gamma - \xi_w} \qquad \textbf{(12)}$$

Since there are no contact delays in this model, increases in α will be absorbed by an increase in V and will not affect unemployment nor the hazard rate". An Increase in the application intensity of the workers will lead to higher unemployment. From (9) we see that an increase in the mean of y leads to a larger value of ξ_w and hence to more unemployment. Finally, when y is drawn from $H(y)$ instead of F(y), there will be: (1) Less unemployment inflow because the mass at the lower support will be larger than under $F(y)$. (2) Less unemployment outflow because V will fall (thus employment will also fall). (3) An increase in the unemployment stock because the average number of applications

---

[9] This is a common assumption in the search equilibrium literature, see e.g. Pissarides (1990).

[10] The larger contact probability, associated with a larger value of ξ_w will be exactly offset by a lower acceptance probability.

[11] Mortensen (1982) discusses externalities which arise when an action by one of the players not only decreases the probability to get a future reward for other players but also ends the game. In matching models for example, the match surplus is shared after a worker and vacancy have found each other. Afterwards the contact probabilities for other workers and vacancies change (see also Diamond 1982). The model in this paper is different because there are screening delays rather than contact delays, the supply of vacancies is always proportional to the amount of job searching agents and the game does not end when the action is completed.

per applicant, ξ, has increased and therefore there will be more congestion.

V. Final remarks

In this paper we showed that actions which can be completed by a limited amount of agents per unit of time can give rise to negative externalities. Those externalities will not only be larger when the mean of the outside opportunities increases but also when the variance increases. In this case the representative agent framework (which only deals with first moments) is not sufficient[12].

References

Akerlof G.A.,A.K.Rose, and J.L.Yellen (1990), Waiting for work, *NBER Working Paper No.3385.*

Barro R.J. and P.M. Romer (1987), Ski-Lift Pricing, with Applications to Labor and Other Markets, *American Economic Review,* vol. 77, no. 5.

Becker G.S. (1965), A theory of the allocation of time, *Economic Journal,* pp. 493-517.

Caballero R. J. (1992), A Fallacy of Composition, *American Economic review 82-5,* pp. 1279-92.

Diamond P.A.(1982) Wage determination and efficiency in search equilibrium, *Review of Economic Studies,* 49, pp. 217-227.

Gautier P.A.(1996), Search Externalities in a Queuing Model of the Labor Market, mimeo, Free University Amsterdam.

Gross D. and C.M. Harris (1985), *Fundamentals of Queuing Theory,* Wiley and sons, NY.

Little J.D.C.(1961), A Proof of the Queuing Formula $L=\lambda W$, *Operations Research 9 , 383-87.*

Mortensen, D.T. (1982), Property Rights and Efficiency in Mating, Racing, and related Games, *American Economic review,* vol. 72 no. 5 ,pp. 968-979.

Pissarides, C.A. *Equilibrium Unemployment Theory,* Basil Blackwell, London.

Acknowledgement

The author wishes to thank Pierre Koning for helpful comments

---

[12] Caballero R.J. (1992) also shows that in many important cases, statements which hold on the individual level vanish at the aggregate level.