

# Herd Behavior in Occupational Choice\*

André Drost<sup>†</sup>

August 26, 2002

## Abstract

In this paper we examine why many professional labor markets are disturbed by cycles in the supply of new workers. We present a model where cycles in labor supply are a consequence of herd behavior in occupational choice. We also present evidence from nearly 150 West German labor markets which supports the herd behavior model of labor supply cycles.

JEL Code: I21, J24, J44

Keywords: analysis of education, occupational choice, professional labor markets

\*I am grateful to Thomas Weise from the Federal Statistical Office Germany for kindly providing the data analyzed in the empirical part of this paper.

<sup>†</sup>Address: Department of Economics, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany. E-mail: drost@wiso.uni-koeln.de.

# 1 Introduction

In many professional labor markets, as for example that for engineers, the supply of new workers follows a cyclical time path: periods when the supply of labor is high in these markets alternate with periods when it is low (Freeman and Leonard 1978, Drost 2002). Cycles in labor supply are problematic because wages seldom are perfectly flexible; therefore labor supply cycles tend to translate into a cyclical sequence of surplus phases when many workers are unemployed and shortage phases when many jobs are vacant. Moreover, when this type of sequence occurs in a market that is essential for research and development, the repeated shortage of R&D workers leads to a slowdown in economic growth: the annual growth rate of the United States, for instance, is reduced by 0.2 to 0.6 percentage points in any year in which 500 thousand R&D jobs cannot be filled with qualified professionals (Romer 2000). Motivated by these problems, we will present a theoretical model in this paper that can explain labor supply cycles. We will also present empirical evidence which supports the model.

Our model differs from previous ones in the way it determines the supply of new workers available in a given professional labor market. In the model developed by Zarkin (1982, 1983, 1985) the supply of new workers depends on the wages workers expect to earn in this market, where wage expectations are assumed to be rational. Since there is evidence that even personnel executives in large firms tend to badly forecast future salaries (Leonard 1982), the assumption of rational wage expectations is quite doubtful, though. In the model developed by Freeman (1971, 1972, 1975a, 1975b, 1976a, 1976b) the supply of new workers depends on wage expectations as well, but this time expectations are assumed to be myopic and not rational. Even the assumption of myopic wage expectations could be too ambitious, however, because it seems that only a minority of students is able to accurately rank different occupations by starting salary (Betts 1996). In our model we therefore take the position that workers have only vague wage expectations, if any, and this lack of information induces them just to enter the labor markets they expect to be entered by many other workers (according to the maxim that it cannot be wrong what many others do). The supply of new workers thus is determined by some sort of herd behavior in our model, a type of behavior which is not unusual in low-information environments.

Before we go into the details, let us first give a short overview of our theoretical model and the empirical test on it. The model is based on two assumptions, the first of which is that the number of new workers entering a professional labor market in some year  $t$  is increasing with the number of new workers expected to enter this market in that year (assumption of herd behavior in occupational choice). The second assumption of the model is that the expectation about the number of new workers entering a market in  $t$  is derived by extrapolation from the number of new workers entering the market in  $t - 1$  and  $t - 2$  (assumption of extrapolative expectations about occupational choice). As these two assumptions

imply, the number of new workers entering a market in  $t$  depends on the number of new workers entering the market in  $t - 1$  and the number of new workers entering the market in  $t - 2$ . The dynamics of the supply of new workers in a professional labor market thus is governed by a difference equation of order two, and as we will show in this paper, the time path solving this difference equation is determined by complex roots so that the path is characterized by cyclical ups and downs.

The empirical test of our model is based on three special properties of the difference equation: the coefficient of the lag-one term is larger than 0 and smaller than 2, the coefficient of the lag-two term is larger than  $-1$  and smaller than 0, and the relation between the lag-one and the lag-two coefficient is equal to  $-2$ . To find out whether these properties can be validated empirically, we analyze time series data on the supply of new workers in nearly 150 West German professional labor markets. For each of these markets, we estimate an autoregressive process of order two, so obtaining estimates for the lag-one and the lag-two coefficient as well as for the relation between them. According to these estimates, there is indeed an overwhelming number of markets with cycles in the supply of new workers where the lag-one coefficient is larger than 0 and smaller than 2, the lag-two coefficient is larger than  $-1$  and smaller than 0, and the relation between the two coefficients is in the neighborhood of  $-2$ . Some central predictions of the herd behavior model thus turn out to be correct, suggesting that we are on the right track when using the model to explain the cyclical behavior of worker supply observed in many professional labor markets. — In what follows, we will first present the theoretical model (section 2), then we will present the empirical evidence (section 3).

## 2 The Model

In our model we examine how the supply of new workers in a professional labor market  $i \in \{1, 2, \dots, N\}$  evolves over time  $t \in \{1, 2, \dots\}$ , where any time period corresponds to one year. We denote the number of new workers who supply labor in market  $i$  at time  $t$  with  $n_{i,t}$  and assume, for simplicity, that  $n_{i,t}$  is a continuous variable. We think of this variable as being composed of a trend  $n_{i,t}^*$  and a residual  $n_{i,t}/n_{i,t}^*$ :  $n_{i,t} = n_{i,t}^* \cdot n_{i,t}/n_{i,t}^*$ . The trend component  $n_{i,t}^*$  is an exogenous variable and assumed to be a log-linear function of time. The residual component  $n_{i,t}/n_{i,t}^*$  is endogenous and interpreted as the de-trended number of new workers entering market  $i$  at time  $t$ . The de-trended number of new workers is the variable we concentrate on in the rest of this paper. We want to explain why this number is cyclical over time and develop an explanation which is based on two assumptions.

The first assumption is that there is herd behavior in occupational choice. Because new workers have no or only very poor information about the pecuniary and nonpecuniary differences between occupations, they consider it best simply

to imitate the occupational decisions of other new workers. Of course, when they make their choice of career, new workers do not know how the other ones will decide, and so all they can do is imitating the expected behavior of the others. The de-trended number of new workers entering market  $i$  at time  $t$  thus is an increasing function of the corresponding expected number  $(n_{i,t}/n_{i,t}^*)^e$ . Formally, the positive relation between these two numbers can be expressed as

$$\frac{n_{i,t}}{n_{i,t}^*} = h_i \left( \left( \frac{n_{i,t}}{n_{i,t}^*} \right)^e \right), \quad (1)$$

where  $h_i$  is a function of  $(n_{i,t}/n_{i,t}^*)^e$  on  $n_{i,t}/n_{i,t}^*$  with  $h_i(1) = 1$  and  $0 < h_i' < 1$ . By imposing the condition  $h_i(1) = 1$ , we express that the de-trended number of new workers is one whenever corresponding expected number is; this condition is necessary to ensure that  $n_{i,t}/n_{i,t}^*$  has a stationary state (see below). By requiring  $h_i' > 0$ , we indicate that the de-trended number of new workers is increasing with the expected number; this condition serves to introduce the herding effect into the model. By setting  $h_i' < 1$ , finally, we impose an upper bound to the herding effect; this upper bound is necessary to ensure that the stationary state of  $n_{i,t}/n_{i,t}^*$  is stable (see below). Note that there is an index  $i$  being associated with the function  $h_i$ ; this index indicates that the relation between the actual and the expected de-trended number of new workers need not be the same across professional labor markets — and it will not be the same, as the evidence presented in section 3 will show.

The second assumption our model is based on is that expectations about the de-trended number of new workers are derived by extrapolation from former de-trended numbers. If new workers observe that the de-trended number of new workers entering market  $i$  was rising in the past, then they will expect that the rise goes on; if, however, they observe that it was falling, then they will expect the fall to be continued. Expectations about the de-trended number of new workers are therefore derived according to the equation

$$\left( \frac{n_{i,t}}{n_{i,t}^*} \right)^e = \frac{n_{i,t-1}}{n_{i,t-1}^*} \cdot \frac{n_{i,t-1}}{n_{i,t-1}^*} \div \frac{n_{i,t-2}}{n_{i,t-2}^*}, \quad (2)$$

which states that the expected value of  $n_{i,t}/n_{i,t}^*$  is equal to the past value of  $n_{i,t}/n_{i,t}^*$  times the past growth factor of  $n_{i,t}/n_{i,t}^*$ . The reason why we model expectations this way is that other common ways to model expectations will not result in a cyclical pattern of  $n_{i,t}/n_{i,t}^*$ . Assume, for example, that expectations about the de-trended number of new workers are formed myopically so that  $(n_{i,t}/n_{i,t}^*)^e = n_{i,t-1}/n_{i,t-1}^*$ . Then combining this equation with equation (1) results in  $n_{i,t}/n_{i,t}^* = h_i(n_{i,t-1}/n_{i,t-1}^*)$  with  $h_i' > 0$ , implying that the time path of  $n_{i,t}/n_{i,t}^*$  is monotonic and not cyclical. Assume, as an alternative, that expectations about the de-trended number of new workers are formed rationally so that  $(n_{i,t}/n_{i,t}^*)^e = n_{i,t}/n_{i,t}^*$ . Then combining this equation with equation (1) yields

$n_{i,t}/n_{i,t}^* = h_i(n_{i,t}/n_{i,t}^*)$ , a static equation implying that the time path of  $n_{i,t}/n_{i,t}^*$  does not exhibit any dynamics at all.

The assumptions of herd behavior expressed in equation (1) and of extrapolative expectations expressed in equation (2) are the only assumptions our model is based on. By taking them together, we obtain the reduced form of the model

$$\frac{n_{i,t}}{n_{i,t}^*} = h_i \left( \left( \frac{n_{i,t-1}}{n_{i,t-1}^*} \right)^2 \div \frac{n_{i,t-2}}{n_{i,t-2}^*} \right), \quad (3)$$

a nonlinear difference equation of order two, which determines the dynamic behavior of  $n_{i,t}/n_{i,t}^*$ . Before we start to examine the dynamics, however, let us first analyze the stationary state of  $n_{i,t}/n_{i,t}^*$ . The stationary-state value of  $n_{i,t}/n_{i,t}^*$  follows from the fact that the number of new workers  $n_{i,t}$  must be equal to its trend component  $n_{i,t}^*$  in the stationary state; so the de-trended number of new workers must be equal to one in the stationary state, or  $n_{i,t}/n_{i,t}^* = 1$ . Note that the stationary state will exist if and only if  $h_i(1^2 \div 1) = 1$ ; this is the reason why we have imposed the condition  $h_i(1) = 1$  on the function of  $h_i$ .

We are now going to examine the dynamics of  $n_{i,t}/n_{i,t}^*$ , but because the difference equation in (3) is nonlinear, we restrict the analysis to the local dynamics in the neighborhood of the stationary state. To begin with, we log-linearize the difference equation in (3) around the stationary state, so obtaining

$$\log \left( \frac{n_{i,t}}{n_{i,t}^*} \right) = 2 \cdot h'_i(1) \cdot \log \left( \frac{n_{i,t-1}}{n_{i,t-1}^*} \right) - h'_i(1) \cdot \log \left( \frac{n_{i,t-2}}{n_{i,t-2}^*} \right), \quad (4)$$

where  $\log$  denotes the natural logarithm. This equation is a linear difference equation of order two which determines the local dynamics of  $\log(n_{i,t}/n_{i,t}^*)$ , and because it is linear, we can analyze it by applying the usual methods. Before we will do this, we want to draw the reader's attention to three special features of this equation, however. Note first that, because we have assumed  $h'_i$  to satisfy the inequality  $0 < h'_i < 1$  (otherwise there would be no herding or the model would be unstable, see below), the coefficient of the lag-one term of the equation,  $2 \cdot h'_i(1)$ , is larger than 0 and smaller than 2. Note second that, due to the same inequality, the coefficient of the lag-two term of the equation,  $-h'_i(1)$ , is larger than  $-1$  and smaller than 0. Note finally that the relation between the lag-one coefficient  $2 \cdot h'_i(1)$  and the lag-two coefficient  $-h'_i(1)$  is equal to  $-2$ . If the model presented here is true, these three features of the difference equation should be observed in the majority of professional labor markets that exhibit cycles in the supply of new workers. We can therefore use the difference equation to evaluate the model empirically, which is exactly what we will do in the following section.

For the time being we stick to the theoretical analysis, however, and determine the time path of  $\log(n_{i,t}/n_{i,t}^*)$  that solves the difference equation in (4). This time path is  $\log(n_{i,t}/n_{i,t}^*) = c_1 \cdot r_1^t + c_2 \cdot r_2^t$ , where  $c_{1,2}$  are constants which can

be specified by initial conditions and  $r_{1,2}$  are characteristic roots which can be derived from the characteristic equation  $r^2 - 2 \cdot h'_i(1) \cdot r + h'_i(1) = 0$ . As the solution of the characteristic equation,  $r_{1,2} = h'_i(1) \pm \sqrt{h'_i(1) \cdot (h'_i(1) - 1)}$ , shows, the time path of  $\log(n_{i,t}/n_{i,t}^*)$  will be unstable if  $h'_i(1) > 1$ , a case that cannot be observed in practise; to avoid this problem, we have chosen  $h'_i(1) < 1$  as one of the properties of  $h_i$ . With  $h'_i(1) < 1$ , the characteristic roots are complex conjugates, and this implies that we can apply De Moivre's theorem and some elementary trigonometry to the time path  $\log(n_{i,t}/n_{i,t}^*) = c_1 \cdot r_1^t + c_2 \cdot r_2^t$ , so being able to express it as

$$\log\left(\frac{n_{i,t}}{n_{i,t}^*}\right) = k_1 \cdot \left(\sqrt{h'_i(1)}\right)^t \cdot \cos\left(k_2 + \arccos\left(\sqrt{h'_i(1)}\right) \cdot t\right), \quad (5)$$

where  $k_{1,2}$  are functions of the constants  $c_{1,2}$ . As we can see by expressing the time path in this way, there are cyclical fluctuations in the motion of  $\log(n_{i,t}/n_{i,t}^*)$  and, thus, in the motion of  $n_{i,t}/n_{i,t}^*$ . The herd behavior model can therefore explain why the de-trended number of new workers is cyclical over time.

By having shown that our model is able to explain cycles in the supply of new workers, we have made our main point. One additional point is worth mentioning, however, and this point refers to the question how the intensity of herd behavior affects the cyclicity of labor supply. As equation (5) shows, the damping factor  $1/\sqrt{h'_i(1)}$  of labor supply cycles is decreasing and the period  $2 \cdot \pi / \arccos(\sqrt{h'_i(1)})$  of labor supply cycles is increasing with the intensity  $h'_i(1)$  of the herding effect. Labor markets where the herding effect is strong are therefore characterized by strong supply cycles — and, thus, by serious employment and growth problems —, while labor markets where the herding effect is weak do only exhibit weak supply cycles. Due to this linkage between the intensity of herd behavior on the one hand and the cyclicity of labor supply on the other hand, we should examine whether and, if yes, why herd behavior is more intense in some and less intense in other markets. This problem will be tackled in another paper, however; in this paper we will continue with an empirical evaluation of the model, instead.

### 3 The Evidence

As we have indicated above, we will evaluate the herd behavior model of labor supply cycles by empirically checking whether the lag-one coefficient, the lag-two coefficient, and the ratio of the two lag coefficients from the difference equation in (4) are indeed element of  $(0, 2)$ ,  $(-1, 0)$ , and  $\{-2\}$ . In order to carry out this check, we will run an OLS regression and estimate the autoregressive process

$$\log\left(\frac{n_{i,t}}{n_{i,t}^*}\right) = a_i \cdot \log\left(\frac{n_{i,t-1}}{n_{i,t-1}^*}\right) + b_i \cdot \log\left(\frac{n_{i,t-2}}{n_{i,t-2}^*}\right) + u_{i,t} \quad (6)$$

for a large number of professional labor markets  $i$ , so obtaining estimates for the lag-one coefficient  $a_i$ , the lag-two coefficient  $b_i$ , and their ratio  $a_i/b_i$ , with  $u_{i,t}$  being a stochastic error term satisfying the conditions of the classical linear regression model. Having obtained these estimates, we will examine in how many markets where the characteristic roots  $a_i/2 \pm \sqrt{a_i^2/4 + b_i}$  of the autoregressive process are complex so that the time path of  $\log(n_{i,t}/n_{i,t}^*)$  is cyclical the lag-one coefficient  $a_i$  is situated in the interval  $(0, 2)$ , the lag-two coefficient  $b_i$  is element of the interval  $(-1, 0)$ , and the coefficient ratio  $a_i/b_i$  is found in a small interval around the number  $-2$ , say  $(-3, -1)$ . If the number of markets characterized by these properties turns out to be large, we will take this result as evidence supporting the model presented in the last section.

To estimate the autoregressive process specified in equation (6) for a large number of professional labor markets, we need time-series cross-section data on the logarithmic de-trended number of new workers  $\log(n_{i,t}/n_{i,t}^*)$  and, hence, on the number of new workers  $n_{i,t}$  the logarithmic de-trended number is derived from. Before we can examine where to obtain these data, we have to answer the question which is the optimal way of measuring the number of new workers in the context of our model. Because the model focuses on occupational choice, which is more or less equivalent to educational choice in the case of professional occupations, it appears to be best to measure the number of workers newly entering a labor market by the number of freshmen newly enrolling in the field of study corresponding to the market. Time-series cross-section data on this number can be obtained from the Federal Statistical Office Germany: though this institution does not provide very detailed data on the enrollment of freshmen in its publications, it is ready to provide an exhaustive data set on freshmen enrollment on request. The data set that was provided to us consists of semiannual time series on the number of freshmen enrolling at German universities for 278 fields of study, with the longest series covering the period between summer term 1972 and winter term 2000/2001.

Due to various statistical problems, we cannot use the time series provided by the Federal Statistical Office directly in our empirical analysis. A first problem is that the time series refer to enrollment at West German universities until summer term 1992 and to enrollment at West as well as East German universities afterwards; as a consequence, we can only use the sub-series ending in summer term 1992 in the empirical analysis. A second problem is that winter enrollment is much larger than summer enrollment in most fields so that there are considerable seasonal fluctuations in the series provided by the Statistical Office; we therefore have to convert the sub-series ending in summer term 1992 from a semiannual to an annual frequency before they can be analyzed empirically. As a third problem, there are several among the converted sub-series that do not cover the whole period between 1972 and 1991; these series have to be omitted from the empirical analysis because they are too short for a reliable estimation of the lag-one coefficient  $a_i$  and the lag-two coefficient  $b_i$ . A fourth problem, finally, is

Table 1: Data I

<i>i</i>	Field	<i>i</i>	Field
1	Shop/technology	75	Catholic theology and religious studies
2	Mining	76	Philosophy
3	Smelting and foundry engineering	77	Religion
4	Mine surveying	78	Archeology
5	Chemical engineering	79	History
6	Printing and reproduction technology	80	Library science
7	Power engineering (excluding electrical engineering)	81	Media studies and communication science
8	Light engineering	82	General literary studies
9	Industrial and production engineering	83	General linguistics and Indogermanic languages and literature
10	Medical technology	84	Byzantine studies
11	Glass and ceramics technology	85	Greek
12	Plastics technology	86	Latin
13	Mechanical engineering	87	German language and literature
14	Metallurgy	88	Dutch language and literature
15	Engineering physics	89	Nordic/Scandinavian languages and literatures (Nordic philology, individual languages n.o.i.b.n.)
16	Textiles and clothing technology	90	American studies/literature
17	Process engineering	91	English language and literature
18	Materials sciences	92	French language and literature
19	Electrical and electronic engineering	93	Italian language and literature
20	Vehicular engineering	94	Portuguese language and literature
21	Aerospace engineering	95	Romance languages and literatures (Romance philology, individual languages n.o.i.b.n.)
22	Navigation	96	Spanish language and literature
23	Naval architecture	97	Finno-Ugric languages and literatures
24	Architecture	98	Russian language and literature
25	Interior decorating	99	Slavic languages and literatures (Slavic philology)
26	Regional planning	100	Egyptology
27	Civil engineering	101	African languages and literatures
28	Surveying (geodesy)	102	Arabic language and literature
29	Mathematics	103	Non-European languages and cultures in Southeast Asia, Oceania, and the Americas
30	Computer science	104	Hebrew language and literature
31	Astronomy and astrophysics	105	Indology
32	Physics	106	Islamic studies
33	Biochemistry	107	Japanology
34	Chemistry	108	Oriental and Old Oriental language and literature
35	Food chemistry	109	Chinese and Korean languages and literatures
36	Pharmacy	110	Central Asian languages and cultures
37	Anthropology (human biology)	111	Ethnology
38	Biology	112	Folklore
39	Geology/paleontology	113	Psychology
40	Geophysics	114	Education
41	Meteorology	115	Education of the blind and visually handicapped
42	Mineralogy	116	Education of pupils with learning disorders
43	Oceanography	117	Education of the deaf and hard-of-hearing
44	Geography	118	Education of the mentally handicapped and slow learners
45	Economic and social geography	119	Education of the physically handicapped
46	Medicine (general medicine)	120	Education of pupils with learning disabilities
47	Dental medicine	121	Special education
48	Veterinary medicine	122	Speech pathology
49	Landscape architecture	123	Education of the emotionally disturbed
50	Agriculture	124	Sports and physical education
51	Brewing and beverage technology	125	Art education
52	Horticulture	126	Art and art history
53	Food technology	127	Fine and graphic arts
54	Dairy farming	128	Sculpture
55	Viticulture and vinting	129	Painting
56	Silviculture	130	Commercial art
57	Wood industry	131	Graphic design and communication design
58	Domestic science and nutrition	132	Textile design
59	Elementary school social studies education	133	Handicrafts
60	Political science	134	Performing and dramatic arts, directing
61	Social studies	135	Motion pictures and television
62	Social sciences	136	Dramaturgy
63	Sociology	137	Conducting
64	Social work	138	Instrumental music
65	Social-welfare education	139	Church music
66	Law	140	Composition
67	Public administration	141	Music education
68	Labor and business studies	142	Music studies and history
69	Business administration	143	Sound engineer
70	Economics		
71	Economics education		
72	Economic sciences		
73	Industrial engineering		
74	Protestant theology and religious studies		

n.o.i.b.n. = not otherwise indexed by name (in the original list received by the Federal Statistical Office Germany)







Table 3: Results I

$i$	$a_i$	$b_i$	$a_i/b_i$	$a_i^2/4 + b_i$	$i$	$a_i$	$b_i$	$a_i/b_i$	$a_i^2/4 + b_i$
1	0.9320	-0.1418	-6.5747	0.0754	73	1.1001	-0.4061	-2.7090	-0.1035
2	0.9141	-0.2412	-3.7903	-0.0323	74	1.1487	-0.4180	-2.7480	-0.0881
3	0.9050	-0.3825	-2.3661	-0.1777	75	1.0057	-0.3699	-2.7188	-0.1171
4	0.6683	-0.1471	-4.5449	-0.0354	76	0.6564	-0.4007	-1.6380	-0.2930
5	1.1075	-0.5338	-2.0748	-0.2271	77	0.3718	0.1515	2.4539	0.1860
6	0.2722	-0.4546	-0.5987	-0.4361	78	0.9617	-0.3488	-2.7576	-0.1175
7	0.5559	-0.0534	-10.4157	0.0239	79	0.3149	0.0832	3.7862	0.1080
8	0.9403	-0.6512	-1.4441	-0.4301	80	0.6198	-0.0529	-11.7222	0.0432
9	0.5357	0.1622	3.3020	0.2340	81	0.4571	-0.3098	-1.4754	-0.2576
10	1.0031	-0.4716	-2.1267	-0.2201	82	0.6923	-0.2907	-2.3816	-0.1709
11	0.8140	-0.2600	-3.1303	-0.0944	83	0.9259	-0.3857	-2.4007	-0.1714
12	0.9284	-0.3225	-2.8789	-0.1070	84	0.1471	0.1958	0.7511	0.2012
13	0.9300	-0.8178	-1.1371	-0.6016	85	0.7250	-0.1961	-3.6970	-0.0647
14	1.0657	-0.4464	-2.3872	-0.1625	86	0.9665	-0.3347	-2.8874	-0.1012
15	0.8168	-0.3000	-2.7227	-0.1332	87	1.0163	-0.3933	-2.5841	-0.1351
16	0.3995	0.0955	4.1831	0.1354	88	0.6097	-0.1911	-3.1910	-0.0981
17	0.1922	-0.2720	-0.7068	-0.2627	89	0.4182	0.1773	2.3591	0.2210
18	0.6625	-0.5602	-1.1826	-0.4505	90	1.0625	-0.3254	-3.2654	-0.0432
19	0.9839	-0.6832	-1.4403	-0.4411	91	1.0829	-0.2370	-4.5701	0.0562
20	0.3872	-0.0913	-4.2410	-0.0538	92	1.2904	-0.5075	-2.5429	-0.0912
21	0.6575	-0.1834	-3.5845	-0.0754	93	1.1816	-0.3239	-3.6480	0.0251
22	0.2662	0.2240	1.1883	0.2417	94	0.5297	-0.0217	-24.3997	0.0484
23	0.4417	0.0051	86.0516	0.0539	95	1.3417	-0.4746	-2.8271	-0.0246
24	0.8583	-0.2466	-3.4802	-0.0625	96	0.9287	-0.1717	-5.4090	0.0439
25	0.9397	-0.2744	-3.4251	-0.0536	97	0.2445	-0.1564	-1.5631	-0.1415
26	0.8621	-0.0944	-9.1312	0.0914	98	0.8552	-0.3989	-2.1437	-0.2161
27	1.2667	-0.8495	-1.4911	-0.4484	99	1.0006	-0.4995	-2.0030	-0.2493
28	1.1900	-0.4753	-2.5035	-0.1213	100	0.2555	0.1109	2.3041	0.1272
29	1.3097	-0.5248	-2.4955	-0.0960	101	0.8856	-0.3612	-2.4517	-0.1652
30	1.2706	-0.3625	-3.5051	0.0411	102	0.3132	0.0100	31.3040	0.0345
31	-0.0595	-0.3742	0.1590	-0.3733	103	0.7115	-0.1562	-4.5548	-0.0296
32	0.9969	-0.4278	-2.3302	-0.1794	104	0.0643	-0.1021	-0.6302	-0.1010
33	0.6777	-0.2286	-2.9642	-0.1138	105	0.4719	-0.0951	-4.9644	-0.0394
34	1.0368	-0.5784	-1.7926	-0.3096	106	0.4973	-0.2059	-2.4147	-0.1441
35	0.7675	-0.3449	-2.2252	-0.1977	107	0.6755	0.0846	7.9839	0.1987
36	1.1889	-0.3783	-3.1426	-0.0249	108	0.6115	-0.3674	-1.6645	-0.2739
37	0.1980	-0.3227	-0.6138	-0.3128	109	1.3271	-0.6388	-2.0776	-0.1985
38	0.9082	-0.3991	-2.2753	-0.1929	110	0.9640	-0.5045	-1.9105	-0.2722
39	0.2332	-0.2347	-0.9940	-0.2211	111	1.2711	-0.4968	-2.5586	-0.0929
40	0.6685	-0.3560	-1.8776	-0.2443	112	0.8454	-0.1589	-5.3205	0.0198
41	0.6618	-0.3121	-2.1208	-0.2026	113	1.3292	-0.5646	-2.3543	-0.1229
42	0.8237	-0.2662	-3.0945	-0.0966	114	1.0573	-0.6047	-1.7486	-0.3252
43	0.1437	-0.0717	-2.0050	-0.0665	115	0.3067	-0.1817	-1.6880	-0.1582
44	0.9529	-0.4895	-1.9467	-0.2625	116	0.6625	-0.4602	-1.4397	-0.3504
45	0.4308	-0.0385	-11.1854	0.0079	117	0.4815	-0.2442	-1.9721	-0.1862
46	1.6086	-0.8281	-1.9425	-0.1812	118	0.9013	-0.3626	-2.4856	-0.1595
47	1.4733	-0.6813	-2.1624	-0.1387	119	0.7876	-0.5031	-1.5655	-0.3480
48	0.7072	-0.0415	-17.0338	0.0835	120	0.6440	-0.4462	-1.4432	-0.3425
49	0.3966	0.2939	1.3497	0.3332	121	0.4609	-0.5663	-0.8139	-0.5132
50	1.2067	-0.3578	-3.3726	0.0062	122	0.6578	-0.3741	-1.7586	-0.2659
51	0.8591	-0.3650	-2.3539	-0.1805	123	0.3886	-0.3685	-1.0546	-0.3307
52	0.6872	-0.1926	-3.5690	-0.0745	124	1.3149	-0.6295	-2.0887	-0.1973
53	0.7694	-0.3221	-2.3884	-0.1741	125	0.4065	-0.0500	-8.1303	-0.0087
54	0.0865	-0.0829	-1.0441	-0.0810	126	0.5678	-0.2438	-2.3288	-0.1632
55	0.2615	-0.0454	-5.7582	-0.0283	127	0.9439	-0.5881	-1.6049	-0.3654
56	0.5985	-0.2024	-2.9575	-0.1128	128	0.5932	-0.3795	-1.5632	-0.2915
57	0.6093	-0.0791	-7.7060	0.0137	129	0.1527	-0.1230	-1.2415	-0.1172
58	0.4313	0.0113	38.1864	0.0578	130	0.9427	-0.1675	-5.6285	0.0547
59	0.5385	-0.1523	-3.5367	-0.0798	131	0.6557	-0.1657	-3.9557	-0.0583
60	0.9281	-0.4294	-2.1614	-0.2141	132	0.4023	-0.0634	-6.3459	-0.0229
61	1.3071	-0.6911	-1.8913	-0.2640	133	1.0817	-0.3743	-2.8902	-0.0818
62	0.5215	0.0259	20.1202	0.0939	134	0.0649	-0.1817	-0.3571	-0.1806
63	1.3430	-0.7524	-1.7850	-0.3015	135	0.7306	-0.4554	-1.6045	-0.3219
64	1.2269	-0.4939	-2.4842	-0.1176	136	0.3443	0.1009	3.4129	0.1305
65	0.9493	-0.2674	-3.5498	-0.0421	137	0.1911	-0.0832	-2.2967	-0.0741
66	1.1808	-0.6173	-1.9128	-0.2687	138	0.1449	0.0302	4.7948	0.0355
67	0.8596	0.0799	10.7651	0.2646	139	0.4444	0.1551	2.8644	0.2045
68	-0.0949	-0.0307	3.0906	-0.0284	140	0.2546	0.1243	2.0480	0.1405
69	0.5286	0.0250	21.1191	0.0949	141	1.0664	-0.4082	-2.6121	-0.1240
70	0.4675	-0.3332	-1.4032	-0.2785	142	0.9282	-0.3067	-3.0263	-0.0913
71	0.9171	-0.2619	-3.5023	-0.0516	143	-0.2902	-0.0992	2.9243	-0.0782
72	-0.1596	-0.0553	2.8863	-0.0489					

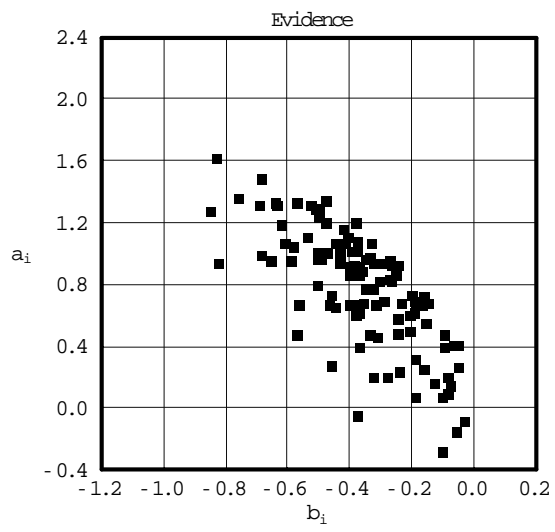
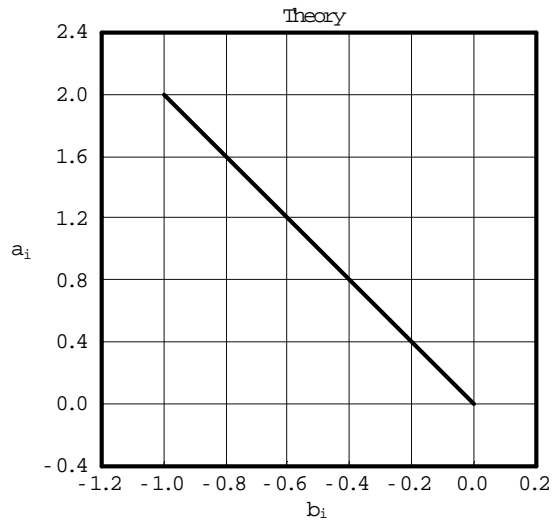


Figure 1: Results II

that some of the remaining series include zero enrollment numbers in one year or other; because computing  $\log(n_{i,t}/n_{i,t}^*)$  from  $n_{i,t}$  is only possible when data are nonzero, we are forced to exclude these series from the empirical analysis as well. After processing the data in this way, we are left with a sample which is documented in tables 1 and 2: it consists of time series for 143 fields of study, with each series being composed of annual data ranging from 1972 to 1991.

When we transform the data for  $n_{i,t}$  included in our sample into logarithms  $\log(n_{i,t})$ , remove a linear least-squares trend  $\log(n_{i,t}^*)$ , and use the resulting data for  $\log(n_{i,t}/n_{i,t}^*) = \log(n_{i,t}) - \log(n_{i,t}^*)$  to estimate the autoregressive process specified in equation (6) for each field, we obtain the results documented in table 3. As the table shows, there are 108 out of 143 fields where  $a_i^2/4 + b_i$  is negative so that  $a_i/2 \pm \sqrt{a_i^2/4 + b_i}$  is complex and the time path of  $\log(n_{i,t}/n_{i,t}^*)$  is cyclical. In 104 and, thus, in 96 percent of these fields the lag-one coefficient  $a_i$  is situated in the theoretically predicted interval  $(0, 2)$ ; this is an empirical result which strongly supports the herd behavior model of labor supply cycles. In 108 and, thus, in 100 percent of the fields the lag-two coefficient  $b_i$  is element of the theoretically predicted interval  $(-1, 0)$ , a result that provides even stronger support for the herd behavior model. In 74 and, thus, in 69 percent of the fields the coefficient ratio  $a_i/b_i$  is found in the interval  $(-3, -1)$  and, hence, in some neighborhood of the theoretically predicted value  $-2$ , another result that supports the model to be evaluated. All in all, we find that the empirical results documented in table 3 confirm the theoretical predictions of the herd behavior model, a finding that becomes even more evident when we regard the two diagrams which are depicted in figure 1.

The upper diagram depicted in the figure shows those combinations of  $a_i$  and  $b_i$  which conform to the theoretical predictions of the herd behavior model: because  $a_i \in (0, 2)$ ,  $b_i \in (-1, 0)$ , and  $a_i/b_i = -2$ , they all lie on a straight line joining the points  $(0, 0)$  and  $(-1, 2)$ . The lower diagram shows the combinations of  $a_i$  and  $b_i$  which conform to the empirical results listed in table 3: since the herd behavior model can only be applied to fields where the motion of  $\log(n_{i,t}/n_{i,t}^*)$  is cyclical, we have omitted all combinations that refer to non-cyclical fields from the diagram. As we can see by comparing the diagrams, most of the combinations observed empirically are in the close neighborhood of the combinations predicted theoretically; while the theoretical combinations are located on the line  $a_i = -2 \cdot b_i$  (for  $-1 < b_i < 0$ ), the empirical combinations are located near the OLS regression line  $a_i = -1.9931 \cdot b_i$  (with  $R^2 = 0.4794$ ), which is almost identical. Given this coincidence of the empirical and theoretical results, we are confident to be on the right track when using the herd behavior model to explain the cyclical behavior of worker supply observed in many professional labor markets.

We have now reached the end of this section, but before finishing it, we will briefly return to a point already discussed at the end of section 2: the relationship between the intensity of herd behavior on the one hand and the cyclicity of labor supply on the other hand. Returning to this point is worthwhile because we

have numbers for the intensity of herding effects now; since  $h'_i(1) = a_i/2$  — this equation follows from equations (4) and (6) —, the results in table 3 imply that herding effects range between 0.0322 in Hebrew language and literature and 0.8043 in general medicine. When we insert these numbers into the formulas for the damping factor and the period of labor supply cycles,  $1/\sqrt{h'_i(1)}$  and  $2 \cdot \pi / \arccos(\sqrt{h'_i(1)})$ , we find that the cycles arising in Hebrew have a damping factor of 5.5762 and a period of 4.5187 years, whereas the damping factor and the period of the cycles arising in medicine are 1.1151 and 13.7106 years. In the field with the weakest herding effect, Hebrew, cycles are therefore extremely weak and should not give rise to any remarkable problems; in the field with the strongest herding effect, medicine, cycles are extraordinarily strong, however, and give rise to serious economic problems whenever wages are not perfectly flexible in the medical labor market. It is therefore quite important to find out why herding effects differ across fields or markets, a question which we will examine in our next paper.

## 4 Conclusion

In this paper we have examined the question why the supply of new workers exhibits periodic ups and downs in a large number of professional labor markets, a phenomenon which may have serious consequences for employment and economic growth. We have proceeded from the assumption that new workers have only poor information about occupational alternatives when making their choice of career, and this assumption distinguishes our approach from other ones where new workers are assumed to be well informed about present or even future wages. Drawing upon the observation that people who have to make a decision without having enough information tend to imitate each other, we have presented a model with herd behavior in occupational choice where herding, when combined with one other assumption, leads to a cyclical behavior of the supply of new workers. Apart from this model, we have also presented empirical evidence, which strongly supports the herd behavior explanation of labor supply cycles.

At the end of this paper let us briefly describe our next project, the explanation of differences in the intensity of herd behavior across professional labor markets, and sketch our preliminary theoretical and empirical results. As it stands by now, our theoretical model attributes differences in the intensity of herd behavior to differences in the size of labor markets: while in large markets a change in the number of new workers easily becomes public so that a herding effect can start, in small markets a change in the number of new workers may stay unnoticed and a herding effect cannot arise. The empirical evidence seems to confirm this relationship: using the same data we have used in this paper, we find that there is indeed a positive relation between the intensity of the herding effect on the one hand and the size of a labor market on the other hand. Of course,

these results are purely preliminary; whether they will bear closer examination will only be known when our coming project is finished.

## References

- Betts, Julian R. (1996), “What Do Students Know about Wages? Evidence from a Survey of Undergraduates”, *Journal of Human Resources* 31, 27–56.
- Drost, André (2002), “The Dynamics of Occupational Choice: Theory and Evidence”, *Labour* 16, 201–233.
- Freeman, Richard B. (1971), “The Labor Market for College-Trained Manpower”, Cambridge, Massachusetts: Harvard University Press.
- Freeman, Richard B. (1972), “Labor Market Adjustments in Psychology”, *American Psychologist* 27, 382–392.
- Freeman, Richard B. (1975a), “Legal Cobwebs: A Recursive Model of the Market for New Lawyers”, *Review of Economics and Statistics* 57, 171–179.
- Freeman, Richard B. (1975b), “Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948–1973”, *American Economic Review* 65, 27–39.
- Freeman, Richard B. (1976a), “A Cobweb Model of the Supply and Starting Salary of New Engineers”, *Industrial and Labor Relations Review* 29, 236–248.
- Freeman, Richard B. (1976b), “The Overeducated American”, New York: Academic Press.
- Freeman, Richard B., Leonard, Jonathan S. (1978), “Autoregressive Degree Patterns: Evidence of Endogenous Cycles in the Market?” *Proceedings of the 30th Annual Meetings of the Industrial Relations Research Association* 30, 10–19.
- Leonard, Jonathan S. (1982), “Wage Expectations in the Labor Market: Survey Evidence on Rationality”, *Review of Economics and Statistics* 64, 157–161.
- Romer, Paul M. (2000), “Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers?” NBER Working Paper 7723.
- Zarkin, Gary A. (1982), “Occupational Choice: An Application to the Market for Public School Teachers”, Ph.D. thesis, Chicago: University of Chicago.
- Zarkin, Gary A. (1983), “Cobweb versus Rational Expectations Models: Lessons from the Market for Public School Teachers”, *Economics Letters* 13, 87–95.
- Zarkin, Gary A. (1985), “Occupational Choice: An Application to the Market for Public School Teachers”, *Quarterly Journal of Economics* 100, 409–446.