The Dynamics of Statistical Discrimination*

Lawrence E. Blume

Department of Economics, Cornell University

and

The Santa Fe Institute

Email: lb19@cornell.edu

June 2006

Abstract

To Come.

Correspondent:

Professor Lawrence Blume

Department of Economics

Uris Hall

Cornell University

Ithaca NY 14850

Email: LB19@CORNELL.EDU

Fax: 607-255-2818

Running Head: Statistical Discrimination

Social scientists have identified self-fulfilling expectations as an important source of discrimination in labour markets and other social exchanges. Various theories highlight the role of negative racial attitudes in creating social outcomes that in turn help perpetuate the initial beliefs. The discriminatory outcomes are seen, then, as an equilibrium of beliefs and behaviors. Among sociologists, Merton (1948) is one source for this view.[1] A similar view of discrimination was put forward by Myrdal (1944), who referred to an equilibrium of beliefs and behaviors as 'the principle of cumulation' and, less abstractly, as 'the vicious circle'. Myrdal observes that social systems may have multiple equilibria, and that different equilibria may be more or less robust to perturbations. Remedial policies, he notes, might work by destabilizing one equilibrium in favor of another.

The contemporary economic version of this reasoning is best represented by the statistical discrimination models of Arrow (1972) and (1973, section 4), which explicitly model labour market outcomes as a rational expectations equilibrium. Employers' beliefs about workers' skill levels determine their willingness to hire, which in turn determines the rate of return on human capital investment, which determines workers' actual skill levels. The model is closed by assuming that expectations are rational. Arrow observes that this system can have multiple equilibria with distinct employment and wage levels, each with correct expectations.

Statistical discrimination models explain labour market outcomes which differ spatially or across groups by asserting that the different instances lie in different equilibria. So East Asians and Blacks, for instance, are in different equilibria. Such explanations beg the question because they say nothing about how this comes to be. The only role of policy is to eliminate or ameliorate a bad equilibrium. This defect is not specific to statistical discrimination models; it is a feature of all the many models across different fields of economics in which multiple equilibria arise from strategic

complementarities. The explanation for how a particular equilibrium comes to be is a vague appeal to history. Krugman (1991) writes, 'As economists grow more willing to make use of models in which there are important multiple equilibria, they will have to take a position on what determines the choice of equilibrium. Most economists who have thought about it at all have assumed that history dictates the choice;...'[2] Underlying the assertion that history chooses the equilibrium is the implicit claim that the equilibrating process lies entirely outside the domain of theoretical analysis.

This claim is wrong. Evolutionary game theory has taught us that the natural random fluctuations in any dynamic process of a population response to its endogenous environment in fact select among equilibria.[3] The population process tends to hover around one equilibria at the expense of all other states. Furthermore, the identity of the equilibrium which is 'selected' is not generally determined by the perturbation process, but by the fundamentals of the deterministic part of the system. By studying the dynamics a large population and a random determination of agents' characteristics, we identify which equilibria are robust, or *stochastically stable* to small individual-level randomness. Typically there is a unique stochastically stable equilibrium.

Considerations of stochastic stability are not just an elegant analytical flourish. They have important theoretical implications for the traditional comparative statics analysis of the effects of parameter changes, including those induced by policy, and also for the empirical study of labour market data with statistical discrimination models. At the theoretical level, equilibrium selection provides a second channel through which parameter changes can have an effect. Section 3.4 contains examples of policy changes which have no effect on the set of (static) equilibria but which have nonetheless a profound observable effect because they change which equilibrium is selected. Theoretically and econometrically, taking dynamics seriously provides a model not just of equilibrium but of equilibrium

transition, which is important to have if the point of a policy innovation is to change equilibrium. At the econometric level, the dynamics of belief adjustment provide alternative sources of identification strategies for statistical discrimination and other multiple equilibrium models.

This paper highlights the role of learning in belief formation. In order to illustrate how learning drives equilibrium selection, this paper presents the simplest possible fixed-wage market equilibrium model. There is an outgroup, which is the potential victim of statistical discrimination. Firms have beliefs about the productivity of the outgroup. These beliefs are revised for each new cohort based upon firms' collective experience with the previous cohort. Workers must make a decision about investing in skills before entering the market, and in making this decision they rely upon their beliefs about labour market outcomes, which also arise from learning. The analysis describes the long-run behavior of employment outcomes and beliefs. The novelty of this paper lies in demonstrating how the stochastic fluctuations inherent to learning 'select' equilibria in the long run; how policy can change the selected equilibrium even when it has no effect at all on the existence of equilibria; and how different kinds of learning can lead to very different kinds of dynamic behavior, even in the long run.

1   The Model

The following model is a discrete-type version of the Coate and Loury (1993) model of statistical discrimination. Most of the analysis will be concerned with a single outgroup, although section 3 contains a brief discussion of intergroup competition for jobs. The workers participate in a market for skilled jobs. They can only carry out the work successfully if they have acquired the necessary skills.

Firms cannot observe the skill levels of job candidates, and must infer them.

1.1 Assumptions

There are M workers and N firms. It is important that these populations are *finite*, although ultimately we will examine large M and N limits. The worker population consists of three types of workers, classified according to the cost a worker pays to acquire skills. The most common type of worker, type c, can acquire skills necessary for work at a cost $c > 0$. A second type of worker, type 0, is naturally endowed with the skills or can acquire them for free. Under the latter interpretation, not acquiring skills at 0 cost is weakly dominated in the worker's decision problem, so we shall assume that type 0 workers are always skilled. The final group of workers, type ∞, is unteachable. That is, the cost for them of acquiring the necessary skills for the skilled labour market is infinite.

The assignment of type to a worker is random, and the assignment of each worker to a type is independent of the assignment of others. A given worker is type 0 with probability, ρ_0, type ∞ with probability ρ_∞, and type c with probability $1 - \rho_0 - \rho_\infty$. Although all types occur with positive probability, we expect most workers to be type c; that is, ρ_0 and ρ_∞ are small. To summarise, the total number of workers is fixed at M, but the numbers of each type are random. The skill level of workers of the no-cost and infinite-cost types is fixed *ex ante*. The skill level of those who can acquire skills at cost is endogenous to the model. In making the skill-acquisition decision, workers will hold some beliefs about employment possibilities. The probability of getting a skilled job is ν, which is held in common by all workers.

Each of the N firms wants one skilled worker. A worker's skill set is not observable at the time

of hire. Workers must be put through an apprenticeship or training program or observed on the job before their skill level is observed. The cost to the firm of hiring an unskilled worker is $\eta > 0$. Firms are of two types. Both types assign 0 value to unskilled workers. Type θ values a skilled worker at $\theta > 0$. Type σ values a skilled worker at $\sigma > 0$. The assignment of types to firms is iid, and the probability of type σ is ϵ. We presume that ϵ is positive, but small.

Although the skill of a given worker is unobservable *ex ante* employment, firms have beliefs about the likelihood that a typical worker is skilled. Expectations will be held in common across firms. Let π denote the probability any firm assigns to the event that a given worker is skilled. Beliefs π, which should in some way relate to the distribution of types, are determined in equilibrium in the static model and through learning in the dynamic models. Workers have no opportunity to signal their skill. The only marker they exhibit is their outgroup membership. In a multiple-group model we would assume that π is group specific: one value for whites, another for blacks, and so forth.

The market model is very simple, in order to facilitate the computations to come. Workers are assigned to firms in such a way that each firm sees no more than one worker. Let $q = \min\{N/M, 1\}$ denote the probability that a worker is matched with any firm. The firm with a worker at the door must decide whether to hire the worker or not. The wage rate for skilled workers is fixed at w, and $c < w < \theta$. (Were this assumption not met, either workers would never acquire skills or firms would never hire workers. Only when these three model parameters are ordered in this way can anything interesting happen.) For workers, the labour market offers two possibilities: A skilled worker is matched with a firm, and earns the wage w. A worker who is not offered a job or who is fired from a skilled job goes immediately to the unskilled market, where the return is normalised to 0.[4]

1.2   Static Equilibrium

Firms' hiring decisions and the skill acquisition decision of those workers who can acquire skills at cost c are determined in equilibrium. In any equilibrium, workers must maximise their expected returns in making the skill-acquisition decision, firms must maximise their expected profits in making hiring decisions, and expectations must be correct.

A worker shows up at the door, and the firm must decide whether or not to hire her. The profit from not hiring is 0. The type-θ profit-maximizing firm will hire the worker if and only if

$$\pi\theta - \big(\pi w + (1-\pi)\eta\big) \geq 0,$$

that is, that the expected net profits from hiring are non-negative. The firm's *reservation belief* π^* is that belief about the probability of a worker being skilled which makes the firm just indifferent between hiring and not; that is, the belief at which the expected profits of hiring are exactly 0. The reservation belief is uniquely determined by the model parameters w, θ and η: $\pi^* = \eta/(\theta + \eta - w)$. Assume that $\big((1-\rho_0)\eta + \rho_0 w\big)/\rho_0 > \theta$. Then $\pi^* > 0$. Type σ firms undertake a similar calculation. Assume $\sigma > \big((1-\rho_0)\eta + \rho_0 w\big)/\rho_0$. This inequality guarantees that a type σ firm will always hire an outgroup worker. In summary, a type σ firm will always hire an outgroup worker. A type θ firm may or may not, depending upon its beliefs.

A worker must make a decision whether or not to invest in skill acquisition. For worker types 0 and ∞ this decision is trivial: Always invest and never invest, respectively. Type c workers believe with probability ν that they will be offered a job, and so the expected return to skill acquisitiion is $\nu w - c$. The return to having no skill is 0. The type c workers' *reservation belief* ν^* is that belief about

employment at which those worker's who can acquire skills at cost c are indifferent about whether or not to do so. That is, ν^* solves

$$\nu w - c = 0.$$

When $w > c > 0$, reservation beliefs ν^* lie strictly between 0 and 1.

In an equilibrium of the static model, firms maximise profit, workers make a return-maximizing skill-acquisition decision, and all beliefs are correct. An equilibrium can be described by two variables: ρ_f, the probability that a type θ firm offers a worker a job, and ρ_w the probability that a type c worker acquires skills.

Definition 1. *An equilibrium is a pair (ρ_f, ρ_w) of action probabilities such that*

1. *ρ_f maximises $\rho_f\left(\pi\theta - \pi w - (1 - \pi)\eta\right)$,*

2. *ρ_w maximises $\rho_w(\nu w - c)$,*

3. *$\pi = \rho_0 + (1 - \rho_\infty - \rho_0)\rho_w$,*

4. *$\nu = (1 - \epsilon)q\rho_f + \epsilon q$.*

This definition allows for randomisation in the event that firms or workers are indifferent over their choices, but generically it will be the case that equilibria are pure, that ρ_f and ρ_w take on only the values 0 or 1. There are two possible types of pure equilibria.

Definition 2. *A* full-employment equilibrium *is an equilibrium in which those workers who can acquire skills, and all who are matched are offered jobs; $\rho_f = 1$, $\rho_w = 1$, $\pi = 1 - \rho_\infty$ and $\nu = q$. An* underemployment equilibrium *is an equilibrium in which workers who can acquire skills at cost c choose not to, and only type σ firms offer jobs; $\rho_f = 0$, $\rho_w = 0$, $\pi = \rho_0$ and $\nu = q\epsilon$.*

The analysis of this model is straightforward. Rather than go through all possible parameter combinations, the following theorem (and the remainder of this paper) treat only the most important cases. Proofs of all results are discussed in the appendix.

Theorem 1. *Assume that $\theta > w > c$, that $v^* < q$ and that $\rho_0 < \pi^* < 1 - \rho_\infty$. If $q\epsilon < v^*$, then both a full-employment and an underemployment equilibrium exist, and these are the only pure equilibria. If $v^* < q\epsilon$, then the only pure equilibrium is a full-employment equilibrium.*

Both ρ_0 and ρ_∞ should be small, so $\rho_0 < \pi^* < 1 - \rho_\infty$ is the 'typical case'. This will be the maintained assumption for the remainder of the paper. With parameter values in this range, and $q\epsilon < v^*$, the model exhibits multiple equilibria. Both full- and underemployment equilibria exist. Most statistical discrimination models stop here, having noted the possibility of different social configurations, and explaining the observed configuration by reference to forces which lie entirely outside the model.

1.3 Dynamics

This subsection introduces the basic market dynamics. In the next section a learning model will be employed to generate the beliefs which drive the market. The empirical distribution learning model takes current beliefs to be the empirical distribution of some finite past history of market outcomes. For the sake of simplicity, the history length will be 1 in this paper. Firms and workers have beliefs. Based upon these beliefs they take actions. These actions determine market activity. Beliefs are then updated from the market outcomes.

This paper presents a model of labour market entry. Workers make a human capital acquisition decision and then enter the labour market. Either they are hired and screened, or not hired. The

skilled hired workers keep their jobs, while the rest of the workers go off to the unskilled market, and a new cohort arrives. The timing of the dynamics is described in the following figure. Workers are dated

| I | M | I | M | I | M | I | M |

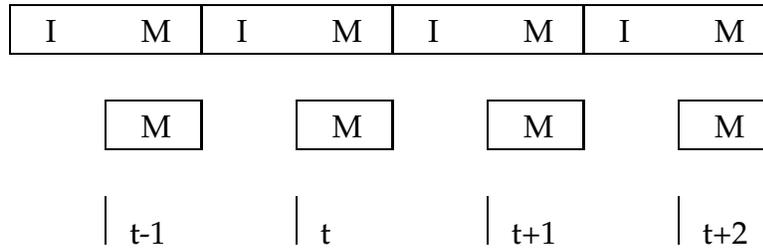| | M | | M | | M | | M |

| t-1 | t | t+1 | t+2 |

Fig. 1: Timing.

not by the year of their birth but the year in which they come to the job market. A worker's life has two parts: A skill-acquisition period, denoted I (for *investment* in Figure 1) and a subsequent market period, denoted M in the diagram. At the beginning of period t, 'old' workers arrive in the market with any skills they have acquired, and are matched with firms. Firms will make job offers based on their beliefs π_t. Of the M workers in the market at date t, K_t will receive jobs, and J_t of those with jobs will in fact have skills. It will be more convenient to normalise by the number of workers so as to work with population fractions rather than body counts. Let $k_t = K_t/M$ and $j_t = J_t/M$. The pair of numbers (j_t, k_t) is the market outcome at date t. The goal of this paper is to describe the stochastic process $\{(j_t, k_t)\}_{t=0}^{\infty}$. From j_t and k_t firms and workers will update their beliefs to π_{t+1} and ν_{t+1}, respectively. At this point, in the second half of date t, 'young' date $t+1$ workers make their skill acquisition decisions based on their beliefs ν_{t+1}. All data is public, and known to both workers and firms.

The learning model considered here has firms forecasting the date $t+1$ skill frequency out-

come based on the previous date's data. Since this data is publicly available, workers can accurately predict the firms' forecast, π_{t+1}, and then use the firms' decision rule to construct their own forecast ν_{t+1} of the likelihood of iemployment:

$$
\nu_{t+1} = \begin{cases} q & \text{if } \pi_{t+1} \geq \pi^*, \\ q\epsilon & \text{otherwise.} \end{cases}
$$

There are two other ways to do this: Workers could update from market data and firms could predict worker's expectations, or both could predict from market data. The details change, but the outcome of the analysis remains the same in all three cases.

If $\pi_t < \pi^*$, beliefs are said to be in the *low regime*, wherein the only workers who will have jobs are those who are offered a job by the type-σ firms. Among these workers, the type assignment is independent and the probability that any one worker is skilled is ρ_0, the probability that she is naturally gifted. When $\pi_t \geq \pi^*$, beliefs are said to be in the *high regime*. All matched workers will get jobs, and the only workers who will not be skilled are those with infinite acquisition costs. When beliefs are in the low regime, type c workers can respond in one of two ways, depending on the relationship between ν^* and $q\epsilon$. If $\nu^* > q\epsilon$, employment only by type σ firms is not enough to induce these workers to acquire skills, while if $\nu^* < q\epsilon$, these workers will acquire skills. Notice that when $\nu^* > q\epsilon$ and $\pi_t < \pi^*$, the mean of the distributions for j_t and k_t correspond to the underemployment equilibrium, and when $\pi_t \geq \pi^*$ the means correspond to the full-employment equilibrum. This description of market behavior is summarised in the following table, in which $b(n, p)$ is a random variable distributed binomially from a population of size n with success probability p:

Low regime: $\pi_t < \pi^*$	High regime: $\pi_t \geq \pi^*$
$\nu^* > q\epsilon$ $k_t \sim b(qM, \epsilon)$ $j_t \sim b(k_t, \rho_0)$	
$\nu^* < q\epsilon$ $k_t \sim b(qM, \epsilon)$ $j_t \sim b(k_t, 1 - \rho_\infty)$	$k_t \sim b(qM, 1)$ $j_t \sim b(qM, 1 - \rho_\infty)$

Table 1: Conditional Distributions of k_t, j_t.

2 Empirical Distribution Learning

Suppose that firms predict the probability of a given worker's being skilled according to the empirical distribution of skilled data over some fixed past number of market periods. For clarity we will take the number of periods to be one, but any fixed finite horizon would do. From the date t data both workers and firms estimate the likelihood that a date $t + 1$ worker will be skilled according to the rule

$$\pi_{t+1} = j_t / k_t.$$

This empirical rule determines firms' expectations.

Empirical learning makes the stochastic process $\{(j_t, k_t)\}_{t=0}^\infty$ Markov. The distribution of the date $t + 1$ variables is determined entirely by the ratio j_t / k_t. The Markov process has two transition regimes. If $j_t / k_t < \pi^*$, then the joint distribution of (j_{t+1}, k_{t+1}) is described by one of the two leftmost columns of the table, depending on the relationship of the parameters ν^* and $q\epsilon$, each of which gives

a distribution for k_{t+1}, and a conditional distribution for j_{t+1} given k_{t+1}. If $j_t/k_t \geq \pi^*$, the right column of the table describes the transition probabilities in a similar fashion. But the market process can be described even more simply. The probability that π_{t+1} is in, say, the high regime, depends only upon the joint distribution of j_t and k_t. There are only two possible joint distributions, and which one obtains depends only upon the regime that π_t belongs to. Thus the evolution of regimes is a Markov process which has only two states, H (high) and L (low). Conditional on regime of beliefs at date t, the actual distribution of market outcomes at date t is given by the table.

2.1   The Short Run

In the short run, the market in a given regime converges to mean behavior, and if the regime contains a static equilibrium, the equilibrium employment and skill rates will be the means. This is to be expected since the static equilibrium values in each regime are the mean values of the distributions in that regime, and the weak law of large numbers implies that the frequency data will converge to its mean. The fact is recorded in the following theorem.

Theorem 2. *For all $0 < q \leq 1$ and $\delta > 0$, if $\pi < \pi^*$ and $v^* > q\epsilon$, so that low regime contains an equilibrium, then as $M, N \to \infty$, $\max\{N/M, 1\} \to q$,*

$$\text{Prob}\left\{ \left| k_{t+1} - q\epsilon \right| > \delta \mid \frac{j_t}{k_t} = \pi \right\} \to 0 \quad \text{and} \quad \text{Prob}\left\{ \left| \frac{j_{t+1}}{k_{t+1}} - \rho_0 \right| > \delta \mid \frac{j_t}{k_t} = \pi \right\} \to 0;$$

that is, the fraction of workers who are employed and the fraction of employed workers with skills converge to their low regime equilibrium values in probability. If $\pi < \pi^*$ and $v^* < q\epsilon$, so that the low

regime does not contain an equilibrium, then

$$\text{Prob}\left\{\left|k_{t+1} - q\epsilon\right| > \delta \,\middle|\, \frac{j_t}{k_t} = \pi\right\} \to 0 \quad \text{and} \quad \text{Prob}\left\{\left|\frac{j_{t+1}}{k_{t+1}} - (1 - \rho_\infty)\right| > \delta \,\middle|\, \frac{j_t}{k_t} = \pi\right\} \to 0;$$

the fraction of workers who are employed converges in probability to the low regime market clearing level and the fraction of employed workers who are skilled converges to the high regime equilibrium level. Finally, if $\pi > \pi^*$, so that the market is in the high regime, then

$$\text{Prob}\left\{\left|k_{t+1} - q\right| > \delta \,\middle|\, \frac{j_t}{k_t} = \pi\right\} \to 0 \quad \text{and} \quad \text{Prob}\left\{\left|\frac{j_{t+1}}{k_{t+1}} - (1 - \rho_\infty)\right| > \delta \,\middle|\, \frac{j_t}{k_t} = \pi\right\} \to 0;$$

that is, the fraction of workers who are employed and the fraction of employed workers with skills converge to their high regime equilibrium values in probability.

These statements describe the short-run behavior of the stochastic labour market process. If the process is in the full-employment regime, it is likely to remain in that regime, and within that regime, near the equilibrium, and increasingly so as the number of firms and workers becomes large. What happens in the underemployment regime depends upon the relationship between ν^* and $q\epsilon$. If ν^* is large, the process, once in the underemployment regime, will tend to remain there, and nearer the equilibrium as the market grows large. If ν^* is small relative to $q\epsilon$, the case where the low regime contains no equilibrium, the market will tend to jump to the full-employment regime.

2.2   Equilibrium Selection

Equilibrium selection is concerned with the asymptotic analysis of the regime process when the static model exhibits multiple equilibria. The high regime has full employment, with mean behavior characterised by the full employment equilibrium, and the low regime has underemployment, with mean behavior characterised by the underemployment equilibrium. What fraction of time will the regime process spend in each regime? The answer to this question is given by the stationary distribution for the regime process. The probability it assigns to a given regime equals the fraction of time the process will spend in that regime in the long run. The phrase *equilibrium selection* refers to the following startling fact: Except for a knife-edge case of parameter values, when the number of firms and workers is large, the invariant distribution assigns probability approximately 1 to one of the regimes. That is, for some parameter values, the process remains almost entirely in the high regime, and occurrences of the low regime are rare, while for most other parameter values, just the reverse is true. The process remains mostly in the low regime, the underemployment equilibrium, and occurrences of full employment are rare. Both equilibria exist, but only one describes the typical behavior of the system.

We begin the asymptotic analysis by considering the multiple equilibrium case of $\nu^* > q\epsilon$. Since ρ_∞ and ρ_0 exceed 0, the labour market process does not lock in to one regime forever. Although transitions between regions are rare, they will surely happen. The long run behavior of the process is describing the long-run frequency with which each region is visited. This frequency is given by the

invariant distribution for the s_t-process. Transition probabilities for the s_t-process are:

$$\text{Prob}\{s_{t+1} = H \,|\, s_t = L\} = \text{Prob}\left\{\frac{j_t}{k_t} \geq \pi^* \,\Big|\, \frac{j_{t-1}}{k_{t-1}} < \pi^*\right\},$$

$$\text{Prob}\{s_{t+1} = L \,|\, s_t = H\} = \text{Prob}\left\{\frac{j_t}{k_t} < \pi^* \,\Big|\, \frac{j_{t-1}}{k_{t-1}} \geq \pi^*\right\}.$$

For any numbers of firms and workers, the probability of observing data which moves the process from one region to the next is positive, and so the s_t-process is irreducible. Thus it has a unique invariant distribution μ. A calculation shows that

$$\frac{\mu(H)}{\mu(L)} = \frac{\text{Prob}\{s_{t+1} = H \,|\, s_t = L\}}{\text{Prob}\{s_{t+1} = L \,|\, s_t = H\}}. \tag{1}$$

Equilibrium selection means that for almost all parameter values, this ratio either converges to ∞ or to 0 the numbers of firms and workers become large.

Computing exact expressions for the right hand side of (1) appears to be intractable, but they can be approximated using large deviation theory. Approximating the denominator is straightforward. When $s_t = H$, all firms hire workers, so the number of employed workers is qN, and $k_t = q$. The number of unskilled workers is a binomial random variable. The mean fraction of unskilled workers is ρ_∞. A transition occurs if and only if the actual fraction of unskilled workers exceeds $1 - \pi^*$. Computation of the probability that this will happen when $1 - \pi^*$ exceeds ρ_∞ is the canonical problem of large deviation theory, which computes the rate at which this probability goes to 0 (as it must) as

the population size grows, and the answer is

$$\lim_{\substack{M,N\to\infty \\ N/M\to q}} \frac{1}{M} \log \text{Prob}\left\{s_{t+1} = L \middle| s_t = H\right\} = -qI(1 - \pi^*, \rho_\infty) \tag{2}$$

where $I(p,q)$ is the *relative entropy* of p with respect to q:

$$I(p,q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}.$$

The right hand side of this expression is called a *rate function*, because it tells the rate at which this probability goes to 0 as the population becomes large.

The numerator is more complicated to estimate, because in state L, k_t is random. Nonetheless there is a rate function. A calculation with the standard tools of large deviation theory proves that

$$\lim_{\substack{M,N\to\infty \\ N/M\to q}} \frac{1}{M} \log \text{Prob}\left\{s_{t+1} = H \middle| s_t = L\right\} = q\log(1 - \epsilon + \epsilon\exp{-I(\pi^*, \rho_0)}). \tag{3}$$

Notice that if $\epsilon = 1$, then the size of the employed worker population is non-random, and the rate function becomes a conventional relative entropy. When the number of employed workers is not random, the problem of estimating the transition probability is the same as that in going from the high regime to the low.

Combining these two estimates gives the central result about the long run behavior of the labour market with empirical learning.

Theorem 3. *Suppose that* $q\epsilon < v^*$ *and that* $\theta > w > c$. *Then*

$$\lim_{\substack{M,N \to \infty \\ N/M \to q}} \frac{1}{M} \log \frac{\mu(H)}{\mu(L)} = q \log(1 - \epsilon + \epsilon \exp -I(\pi^*, \rho_0)) + qI(1 - \pi^*, \rho_\infty). \tag{4}$$

If the right hand side of equation (4) is positive, then as M and M grow large as described under the limit, $\mu(H)$, the long run probability of the high regime, converges to 1. If the right hand side is negative, then $\mu(H)$ converges to 0 and so $\mu(L)$ converges to 1. So except when parameters are such that the right hand side is exactly 0, equilibrium selection occurs: For large enough numbers of firms and workers, the fraction of time spent in one of the regimes is nearly 1.

2.3 Policy Analysis and Comparative Dynamics

The relative entropy $I(p, q)$ is a measure of distance between the probabilities p and q. The function is non-negative, convex, and takes the value 0 only when $p = q$. The function $-\log(1 - \epsilon + \epsilon \exp -I(p, q))$, the rate function for the underemployment regime, is not convex, but it is non-negative, 0 only when $p = q$, and increasing in $|p - q|$. Figure 2 plots the 0-contour of the right hand side of equation (3) when $\rho_0 = \rho_\infty = 0.1$. For values of π^* and ϵ above the graph, $\mu(H)$ converges to 0 as the market becomes large; the underemployment equilibrium is selected. For values of π^* and ϵ below, $\mu(H)$ converges to 1 as the market becomes large; the full employment equilibrium is selected.

Changes in parameter values work through two channels in this model. First, they determine whether the underemployment equilibrium exists, and the location of the equilibria, the equilibrium
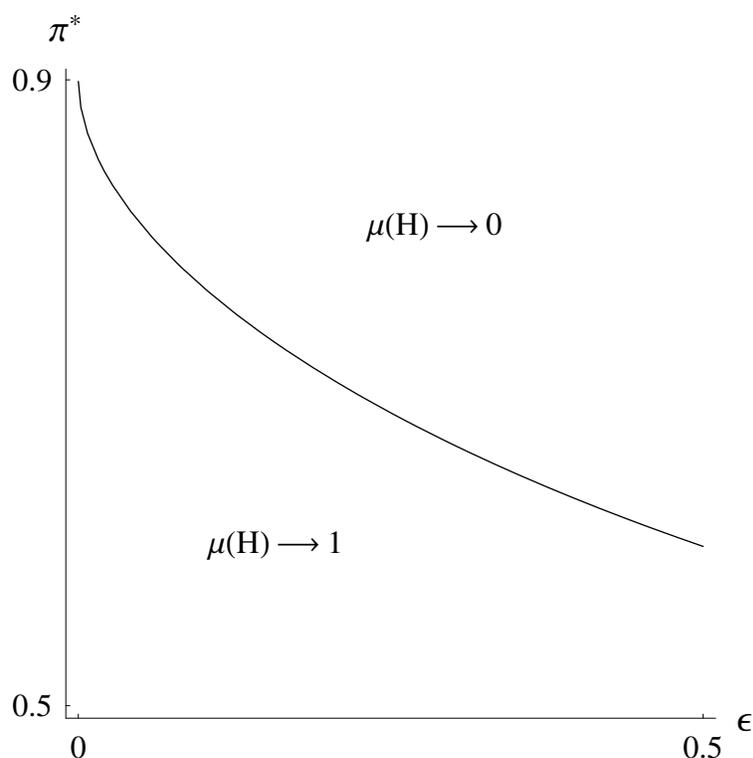
Fig. 2: $\log \mu(H)/\mu(L) = 0$.

employment rates. The impact of parameter changes through this channel can be observed in the static model as well as in the dynamic model. Second, they determine which equilibrium is selected, which is most likely to be observed in a given market. The effects of this channel are unobservable in the static model. They are a consequence of the stochastic perturbations which arise from the learning dynamics.

Figure 2 demonstrates several properties of the model. In this model, the traditional comparative statics are completely uninteresting. So long as $\pi^* < 1 - \rho_\infty$ and $\nu^* > q\epsilon$, the full employment equilibrium and the underemployment equilibria both exist. The threshold belief π^* has no other ef-

fect on the location of equilibrium, and the only effect of ϵ is to determine the unemployment rate $(q\epsilon)$ in the underemployment equilibrium. Suppose that $q = 1$ and $\nu^* = 0.5$. Then changes in π^* within the range plotted in figure 2 has no effect on the comparative statics of equilibrium, and changes in ϵ determine only the low regime unemployment rate. But both parameters have a huge effect on selection. As the parameter values cross the graph from right to left, the probability of equilibrium under-employment shifts from near 1 to near 0. This is the effect of equilibrium selection.

Observe that transitions are very sharp. In the infinite population limit, the transition is discontinuous. The probability of full employment is 0 everywhere to the right of the graph and 1 everywhere to the left. Policy changes that move π^* have no effect if they fail to cross the graph. Policy changes that move ϵ have a continuous effect until they hit the graph, where the effect on the low regime unemployment rate is arbitrarily steep. This dramatic 'phase transition' effect of policy is common to equilibrium selection in multiple equilibrium models. See Brock and Durlauf (2001) and Blume and Durlauf (2003).

The selection channel for parameter effects has important policy implications that are masked by considering only parametric variation of the equilibrium set. Three policies illustrate these selection effects: Hiring quotas, hiring subsidies and job training programs.

Hiring Quotas: A conceivable affirmative action response to underemployment is the institution of a quota on new hires. Firms are required to make a certain number of new hires from the out group. This has the notional effect of raising ϵ, the probability of being hired in the low expectation regime. If the change in ϵ is large enough, the low-regime equilibrium will disappear. The probability of getting a job will be high enough that skill investment becomes worthwhile even when firms hold unfavorable beliefs about workers. If the new ϵ is not large enough to make the equilibrium

disappear, it will have no comparative static effect. With the parameters of the figure 2, making the underemployment equilibrium disappear would require an $\epsilon \geq 0.5$. Nonetheless, smaller ϵ could have a selection effect, and the selection effect works in the opposite direction. If ϵ is located to the left of the graph, and if ϵ increases enough to cross the graph, then the long run probability of being in the high regime falls from near 1 to near 0. The current underemployment equilibrium that was the anomaly now becomes the usual condition. The intuition is clear. With a higher ϵ, more workers are going to be observed. If more of them are observed, firms will have a more accurate idea of their low skill level, and it will be harder to transit out of the low regime. The conclusion is the same as that of the Coate and Loury (1993) analysis, but for a different reason. There, hiring programs reduce workers incentives to acquire skills. Here, it reduces firms ability to learn that the workforce is skilled.

Firm Subsidies: Another affirmative action possibility is to subsidise the hiring of new minority workers, perhaps through tax incentives. The effect of a hiring subsidy is to lower π^*, the threshold fraction of skilled minority workers in the population that would make it just worthwhile to hire. This policy has no comparative static effect so long as the full-employment equilibrium previously existed. Yet it can have a positive selection effect if the parameter pair moves down across the graph in figure 2. Subsidies and quotas work in different ways. A quota changes system dynamics only in the low state, by forcing more workers on firms. A subsidy affects system dynamics by changing the hiring firms calculus.

This policy exercise is slightly misleading, because it presumes that the subsidy remains in place when the economy transits to the high regime. Suppose instead that the government commits to a subsidy every time the economy falls into the low regime, but removes the subsidy when high

employment is reestablished. If the threshold belief with the subsidy is π^{**}, then

$$\lim_{\substack{M,N\to\infty \\ N/M\to q}} \frac{1}{M} \log \frac{\mu(H)}{\mu(L)} = q \log(1 - \epsilon + \epsilon \exp -I(\pi^{**}, \rho_0)) + qI(1 - \pi^*, \rho_\infty)$$

Again there will be a threshold selection effect with the low-only subsidy, but the magnitude of the policy necessary to create the selection switch will be different Comparing the subsidy in both regimes versus the subsidy in the low regime only, at the same subsidy rate,

$$\lim_{\substack{M,N\to\infty \\ N/M\to q}} \frac{1}{M} \left\{ \log \frac{\mu(H)}{\mu(L)} \bigg|_{\text{both}} - \log \frac{\mu(H)}{\mu(L)} \bigg|_{\text{low}} \right\} = q\big(I(1 - \pi^{**}, \rho_\infty) - I(1 - \pi^*, \rho_\infty)\big) > 0.$$

A given reduction of the threshold firm belief from π^* to π^{**} which would cause a selection switch when employed in both regimes may not be enough to accomplish a switch of equilibrium when employed only in the low regime. While this result is only common sense, it is useful to understand why. In when the subsidy is employed in both regimes, lowering the threshold belief makes it easier to observe a sample of workers above the threshold in the low regime, and harder to observe a sample of workers below the threshold in the high regime. Thus it is easier to transit from low to high, and harder to transit from high to low. When the subsidy is in place only in the low regime, only the first effect is present. In terms of equation (1), the subsidy in both regimes both increases the numerator and decreases the dominator, while the subsidy only in the low regime only decreases the denominator.

Job Training Programs: The effect of job training programs is to increase the fraction of the population who acquire skills for free; that is, to increase ρ_0. (Alternatively, one could consider programs of wide scope whose notional effect is to lower c.) Increasing ρ_0 has no effect on equilibrium

market outcomes until it becomes large enough that the underemployment equilibrium disappears, when $\rho_0 > \pi^*$. However, increasing ρ_0 has a selection effect: The right hand side of equation (4) increases in ρ_0, and the equilibrium selection switch can take place for parameter values in the range where the underemployment equilibrium still exists. For instance, suppose that $\epsilon = 0.1$, $\pi^* = 0.5$ and $\rho_\infty = 0.1$, Full- and underemployment equilibria will coexist for $\rho_0 < 0.5$, but Mathematica suggests that the full employment equilibrium is selected when $\rho_0 > 0.28$.

Discrimination: Although economists have been creative about coming up with structural reasons for differential labour market outcomes for ingroup and outgroup participants, discrimination cannot be ruled out. Hu and Taber (2005), for instance, infer from layoff and plant closing data that simple discrimination by firms still has explanatory power against the more modern informational explanations. Suppose that some fraction of firms simply refuses to hire outgroup members. This affects the parameter q, the probability of matching with a firm. This parameter change has implications for comparative statics, because if $q\epsilon > v^* > q'\epsilon$, then a parameter shift from q to q' makes the underemployment equilibrium disappear. Beyond this, q has no effect on equilibrium selection. Here is an instance of a parameter change for which dynamic equilibrium selection has nothing to say which cannot be observed in the static model.

2.4 Dynamics with One Equilibrium

When $v^* < q\epsilon$, there is only one equilibrium. Should firm expectations be in the low regime, workers become skilled anyways, so such beliefs cannot be self-fulfilling. In terms of the dynamic model, the low regime becomes a transient state. The transtion from the underemployment regime to full

employment is no longer a rare event, but a transition from full employment to underemployment is. Calculations identical to those for the previous Theorem show:

Theorem 4. *Suppose that* $q\epsilon > v^*$ *and* $\theta > w > c$. *Then*

$$\lim_{\substack{M,N\to\infty \\ N/M\to q}} \log \frac{1}{M} \frac{\mu(H)}{\mu(L)} = qI(1 - \pi^*, \rho_\infty) \tag{5}$$

This number is always positive for $\rho_0 < \pi^* < 1 - \rho_\infty$. The invariant distribution always piles up in the full-employment regime as the market becomes large, because leaving the underemployment regime is typical while leaving the full-employment regime is rare.

This case is illustrative of a phenomenon which would be more interesting were there more than two worker types. Suppose there were $m + 1$ worker types with skill acquisition costs $0 = c_1 < \cdots < c_m < c_{m+1} = \infty$. Suppose too that there are multiple firm types with different skilled worker valuations, $\eta = \theta_n > \cdots > \theta_1 > 0$. The range of firm expectations can be divided up into $n + 1$ regimes, where in regime k all workers of type k or less acquire skills. Some of these regions will contain an equilibrium: The probability of observing a type k or lower worker, assuming they and only they are skilled, is such that firms with type l or higher will want to hire, and the fraction of these types in the population is such that the probability of being matched with one of these firms is within the range of worker beliefs that will make exactly worker types 0 through k acquire skills. Other regimes contain no equilibria. This extension of the model is analysed just as was the $m = n = 2$ case. For every regime containing an equilibrium there will be a set of parameters (possibly entry) for which that regime is selected. For the remaining regimes, the probability of remaining in them for any positive fraction of time in the long run goes to 0 with the market size, just as was the case here with the low

regime. The set of parameter values for which there will be ties in the rate function comparison is negligible.

2.5 Multiple Groups

Another extension of the model is to allow for two or more competing groups. This theory would not be about ingroup/outgroup competition. There are all kinds of structural reasons not captured in statistical discrimination models which determine why we do not expect the relative wage positions of the ingroup and outgroup to switch. But we can imagine two outgroups competing in the same labour market. Firms have beliefs about each group, each group has beliefs about the probability of getting a skilled job, and the model proceeds just as does the single group model. However, in a fixed wage market, not much interesting happens. For interesting parameter values there are four states, corresponding to a a high and low regime for each group, each containing an equilibrium. Again one constructs the transition matrix, estimating its entries with the tools of large deviation theory.

The small adjustment required by the multigroup model is that, when both groups are in the low regime and workers are matched with those firms willing to hire them, the appropriate model to describe match outcomes is sampling without replacement. A group A worker matched with 1 firm increases the odds that a group B worker is matched with firm 2. However, the literature contains a rate function for sampling without replacement (Dembo and Zeitouni 1992), so the analysis is straightforward, and the story is very much the same as that presented here. If the number of available jobs in the low-low regime is small relative to the size of each group, sampling with and without replacement are asymptotically very much the same thing. In this case the two-group model

is just two disconnected one-group models running side by side. In either case, however, the random perturbations of the model select an equilibrium regime, and that regime could either be one of the symmetric states or one of the non-symmetric states.

A more interesting version of this model endogenises wages. The additional effect this creates is that one group can crowd out the other. If group A is in its high regime, then the wage rate that would result if group B were also high is falls far enough that group B and/or group A would no longer be willing to invest in skill acquisition. Such a model requires a compelling wage determination story. See Blume (2003) for the details of such a model, but the fundamental principle of equilibrium selection remains intact.

3   Conclusion

This paper investigates equilibrium selection in a statistical discrimination model. We have recently come to understand how the dynamics of behavior select among equilibrium in simple games. This paper brings that understanding to the study of learning dynamics in a simple labour market. Its purpose is to see how learning dynamics reinforces or significantly modifies conclusions drawn from the analysis of static statistical discrimination models. Accordingly, the simplest possible statistical discrimination model is constructed in order to clearly illustrate the different effects learning can have. The paper develops a framework which in its static version admits the possibility of multiple equilibria with different employment levels. When learning dynamics are introduced, it is seen that the effects of noisy data can influence equilibrium selection in significant ways, which depend upon the precise nature of how firms and workers revise beliefs.

Learning here is accomplished by constructing empirical distributions from the last period's market outcomes. Thus the stochastic process describing labour market outcomes is Markov. Although the process bounces back and forth across different equilibrium regimes, one regime occurs most frequently, and its frequency of occurrence converges to 1 as the size of the market grows. Such states are said to be *stochastically stable*. Although the static model has multiple equilibria, the remaining equilibria appear as *metastable states*. The market process tends to remain near an equilibrium regime, but will from time to time wander to another equilibrium regime. The time it remains near any given equilibrium regime before drifting away grows to infinity with the market size, but the time spent near the stochastically stable grows large at an exponential rate compared to the time spent near the other equilibria. Were there other non-equilibrium regimes, as there is when $v^* < q\epsilon$, the time spent in these regimes does not grow large with market size.

Readers of the stochastic evolutionary game theory literature will recognise the analysis of empirical learning as a 'basin-hopping' exercise like those of Blume (1993), Kandori, Mailath, and Rob (1993) and Young (1993). These papers examine stochastically perturbed best- or better-reply adjustment dynamics in finite games, and relate the invariant distribution (the long run outcomes) to the Nash equilibria of the static game. Two criticisms of this literature have emerged: How sensitive are the results to the specification of the stochastic perturbation, and how long is the long run?

The first question asks, if different kinds of learning rules are introduced, or if the data process is changed in some significant way, will the identity of the stochastically stable state change. While this question has not been carefully investigated, the answer to this question is almost certainly yes, so careful modelling of the learning process is required to determine the identity of the stochastically stable states. Suppose, for instance, that firms see noisy data rather than accurate market-generated

data. Then the rate functions computed in the two lemmas will be different — they will reflect the noise, and so the right hand side of equation (4) will change. The fact that equilibrium selection goes on is still true, but sets of parameters that select for the high and low regime, respectively, will be different.[5]

The answer to the second question is that, in this simple model, transitions from one equilibrium regime to the next can take a long time, ever longer as the market grows large. This is not a limit for the relevance of 'long run' analysis so much as it is an indictment of the overly simple structure of this particular model. For instance, there is ample evidence of the importance of a network structure to the organisation of labour markets (Topa 2001). A natural model for this kind of structure is a local interaction model such as Calvó-Armengol and Jackson (forthcoming). In stochastic population games, the mean transition time in 'global interaction' models such as Kandori, Mailath, and Rob (1993) and Young (1993) take a time which increases without bound in the size of the population of players, but in local interaction models such as Blume (1993), the transition times remain bounded as the population size grows. See, for instance, Young (2005). Should one take a learning model to a Calvó-Armengol and Jackson-type model, there is no reason to believe that a similar result would fail to hold. Furthermore, for analytical convenience our model steps through time; that is, period by period, data is revealed, expectations are formed and individuals and firms choose. In the real labour market, there is a continual flow of data, jobs and new workers. It is not at all clear how the notional time of the model translates into real-world time. But in any case a model with sufficient attention to realistic detail to seriously address this fundamental issue is beyond the scope of this paper.

Stochastic stability and the resultant equilibrium selection is not simply an analytical nicety. It has important implications for the theoretical and empirical analysis of statistical discrimination. At the

theoretical level, the effect of model parameters on the equilibrium selection provides a channel for parametric changes in model outcomes that is simply absent from the deterministic models heretofore studied. At a basic level this is a radical extension of conventional comparative statics analysis. We ask the traditional question, 'how does that equilibrium change with model parameters?' But it is important to understand, as the examples of section 3.4 show, that the intuition and the results can be quite different from those observed in the static model.

Learning models have empirical implications as well. In general, the problem of taking models with multiple equilibria to data are quite difficult.[6] The modelling strategy introduced here provides an alternative. In cross section, it provides a probability distribution over the different equilibria. Moro (2003) has used a static multiple equilibrium model estimated at three different time periods to test if equilibrium switching can explain reductions in the black-white wage gap. A serious econometric implementation of the toy model developed here would give much more leverage for the examination of this hypothesis.

The simple model explored here can be extended in a number of interesting ways. Having seen how selection works in the simple model studied here, one could imagine working out an analysis of the kind presented here for more complicated statistical discrimination settings such as that Moro and Norman (2004). Even in the model presented here, analysis is still feasible when different groups compete for jobs. It is also possible to extend the model to allow for endogenous wage determination. Yet another issue is the effect of environmental randomness. Suppose that wage rates fluctuated with macroeconomic conditions. Does the stochastic effect of environmental fluctuations help or hinder the stability of the full-employment regime? All these questions, even unanswered, serve to make the point that the analysis of multiple equilibria models should not end with a static

analysis and a vague appeal to history.

The theme of this paper extends beyond labor market models to other statistical discrimination models; in particular, to those of profiling, which is the central concern of this Feature. Knowles, Persico, and Todd (2001) provide the leading model of profiling in screening for criminal behavior and terrorist activity. The KPT model and its descendents have a unique equilibrium because there is no complementarity between the actions of profilers and profilees. If bad behaviors are considered more likely by the profilers, they take stronger actions of a kind which reduces the payoff to the bad behaviors to the profilees. In this paper the actions are complementary. The behavior that is being screened for is not criminal activity, but acquiring human capital. If in the view of firms, the profilers, this behavior becomes less likely, they take an action, not hiring, which decreases the relative payoff of skill acquisition to the profilees, the workers.

Although KPT models do not exhibit the complementarities that would generate multiple equilibria, such complementarities are in fact important in considering profiling policies. One complaint about a policy of profiling minority groups in the search for terrorists is that, for a certain segment of the minority population, there is a strategic complementarity. The feelings of resentment and alienation the policy engenders makes terrorist acts, attacking society, more attractive. One consequence of this argument is the possibility of multiple equilibria and of flipping from a benign to a malignent equilibrium. The relevant dynamics here are not learning, but the dynamics of alienation. The considerations raised here are nonetheless relevant.

Appendix: Proofs

Proof of Theorem 1: The proof of Theorem 1 is just a simple check. Suppose that $v^* > q\epsilon$. To see that full employment is an equilibrium, suppose that $\rho_f = 1$, and that firms as workers both have correct expectations. Then $v = q$. Since $v^* < q$, all type c workers will acquire skills, so $\pi = 1 - \rho_\infty$. Since $\pi^* < 1 - \rho_\infty$, $\rho_f = 1$ maximises profits.

For the underemployment equilibrium, suppose that $\rho_f = 0$. Then correct worker expectations has $v = q\epsilon < v^*$. So type c workers will choose not to acquire skills; that is, $\rho_w = 0$. Correct firm expectations are $\pi = \rho_0 < \pi^*$, so it is rational for firms not to employ workers, $\rho_f = 0$.

If $v^* < q\epsilon$, then type c workers will choose to acquire skills no matter what firms do, $\rho_w = 1$. Firm beliefs must be $\pi = 1 - \rho_0 > \pi^*$, and so firms will hire workers, $\rho_w = 1$. $\qquad\square$

Proof of Theorem 2: This theorem is just the immediate application of the weak law of large numbers to the relevant distributions. $\qquad\square$

Proof of Theorems 3 and 4: Proving Theorem 3 requires verifying equations (2) and (3). This is done using results from large deviation theory. A standard reference is (Dembo and Zeitouni 1992). The probability to be estimated in (2) is the probability that the sample mean X_N / N is less than p when X_N is distributed $b(q, N)$ with $q > p$. This is exactly the problem solved by Cramer's Theorem, and equation (2) supplies the answer.

Equation (3) is the main technical hurdle for this paper. Its proof is mostly a long pair of calculations which teaches little, and so I will provide here only a sketch. However, there is easy

intuition to see why it should be true. If the sample size in the low regime were not random, we could use the rate function from Cramer's theorem. An elementary fact about the rate function is that it gives a probability upper bound as well as an asymptotic equality:

$$\mathrm{Prob}\Big\{\frac{j_t}{k_t} > \pi^* \mid k_t = k\Big\} \leq \exp -kI(\rho_0, \pi^*)$$

Expecting over the sample size k gives an unconditional probability upper bound. Logging that bound and dividing by the population size M is thus a reasonable conjecture for the rate function of equation (3), and it turns out to be correct.

Now for the adumbration of the proof. The Gärtner-Ellis Theorem describes a calculation, taking the Legendre transformation of the scaled cumulant generating function, whose end result is a rate function $\Lambda(p, q)$ for the sequence $\{K_N, J_N\}_{N=1}^{\infty}$ (where N indexes the number of firms). The Contraction Principle says that the rate function for any real random variable sequence $\{f(K_N, J_N)\}_{N=1}^{\infty}$ is $\mathcal{I}_f(r) = \inf\{\Lambda(p, q) : f(p, q) = r\}$. In our case, $\mathcal{I}(r) = \inf_p \Lambda(p, rp)$. Computing this object verifies the conjecture.

With these results, equation (1) can now be used to compute the scaled log invariant odds ratios of Theorems 3 and 4. $\qquad\square$

The intermediate product $\Lambda(p, q)$ is a hideous-looking rate function but it has economic interest. Empirical distribution learning says, hire a worker at date t and run the risk of a loss if and only if $J_{t-1}/K_{t-1} \geq \pi^*$. This criterion, that the sample mean be above the threshold, is independent of the size K_{T-1} of the worker sample. One may object that it is much easier to achieve the threshold in small samples than in large, and so it would be appropriate to make the criterion tougher when

K_{t-1} is small. So for instance, let $\phi(x)$ be a decreasing function of x such that $\lim_{x\to\infty} \phi(x) = 1$. Consider instead the decision rule which says to hire workers if and only if $\phi(K_{t-1})J_{t-1}/K_{t-1} \geq \pi^*$. This, of course, is completely ad hoc, but no more so than empirical distribution learning itself. For any such rule we can apply the contraction principle to get a rate function for the probability of a low regime to high regime transition, and carry on the analysis parallel to that of sections 2.2 and 2.4.

Notes

[1]In Merton (1967, p. 477) he illustrates a self-fulfilling prophecy, 'In the beginning, a *false* definition of the situation evoking a new behaviour which makes the original false conception come *true*,' with a fable of a bank run.

[2]See, for instance, Loury (2002) on statistical discrimination, and David (1985) on adoption externalities. Cooper (1994) tries to make a virtue of excessive historical determinism. Other social scientists can reach the same point without relying on a formal model. Putnam (1993), writing on social capital in Italy, attributes the different levels of corruption in the north and south to 'the mists of the dark ages'.

[3]See Blume (1993), Kandori, Mailath, and Rob (1993) and Young (1993).

[4]Implicitly, the value to an unskilled worker of being hired and then fired is assumed to be the same as that of having never been offered a job. One might imagine that part of the cost η is a payment to the worker. This gives the unskilled worker an incentive to participate in the job market. This would change the calculation of the reservation belief ν^* defined below, but otherwise has no effect on the analysis.

[5]Earlier versions of this paper also explored a Bayesian updating scheme, in which the market process locks into one of the two equilibrium regimes, high or low employment, with probability 1.

[6]See, for instance, Tamer (2003).

References

Arrow, K. J. (1972). 'Some mathematical models of race in the labor market', in (A. Pascal, ed.), *Racial Discrimination in Economic Life*, pp. 187–204. Lexington MA: Lexington Books.

Arrow, K. J. (1973). 'The theory of discrimination', in (O. Ashenfelter and A. Rees, eds.), *Discrimination in Labor Markets*, pp. 3–33, Princeton NJ: Princeton University Press.

Blume, L. and Durlauf, S. (2003). 'Equilibrium concepts for social interactions models', *International Game Theory Review*, vol. 5(3) (September), pp. 193–210.

Blume, L. E. (1993). 'The statistical mechanics of strategic interaction', *Games and Economic Behavior*, vol. 5(3) (July), pp. 387–424.

Blume, L. E. (2003). 'Group Competition and Statistical Discrimination', unpublished, Cornell University.

Brock, W. A., and Durlauf, S. (2001). 'Discrete choice with social interactions', *Review of Economic Studies*, vol. 68(2) (April), pp. 235–260.

Calvó-Armengol, A. and Jackson, M. O. (forthcoming). 'Networks in labor markets: Wage and employment dynamics and inequality', *Journal of Economic Theory*.

Coate, S. and Loury, G. C.. (1993). 'Will affirmative-action policies eliminate negative stereotypes?' *American Economic Review*, vol. 83(5) (December), pp. 1220–1240.

Cooper, R. (1994). 'Equilibrium selection in imperfectly competitive markets with multiple equilibria', ECONOMIC JOURNAL, vol. 104(426) (September), pp. 1106–22.

David, P. (1985). 'Clio and the economics of QWERTY', *American Economic Review*, vol. 75(2) (May), pp. 332–337.

Dembo, A. and Zeitouni, O. (1992). *Large Deviations Techniques and Applications*, Boston MA: Jones and Bartlett.

Hu, L. and Taber, C. (2005). 'Layoffs, lemons, race and gender', unpublished, NBER Working Paper No. 11481.

Kandori, M., G. Mailath, and Rob, R. (1993). 'Learning, mutation and long run equilibrium in games', *Econometrica*, vol. 61(1) (January), pp. 29–56.

Knowles, J., Persico, N., and Todd, P. (2001). 'Racial bias in motor vehicle searches: Theory and evidence', *Journal of Political Economy*, vol. 104(1) (February), pp. 203–229.

Krugman, P. (1991). 'History versus expectations', *Quarterly Journal of Economics*, vol. 106(2) (May), pp. 651–667.

Loury, G. C. (2002). *The Anatomy of Racial Inequality*. Cambridge MA: Harvard University Press.

Merton, R. K. (1948). 'The self-fulfilling prophecy', *The Antioch Review*, vol. 8 (Summer), pp. 193–210.

Merton, R. K. (1967). *Social Theory and Social Structure*. New York: The Free Press.

Moro, A. (2003). 'The effect of statistical discrimination on black-white wage inequality: Estimating a model with multiple equilibria', *International Economic Review*, vol. 44(2) (May), pp. 467–500.

Moro, A., and Norman, P. (2004). 'A general equilibrium model of statistical discrimination', *Journal of Economic Theory*, vol. 114(1) (January), pp. 1–30.

Myrdal, G. (1944). *An American Dilemma: The Negro Problem and Modern Democracy*. New York: Harper & Row.

Putnam, R. D. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton: Princeton University Press.

Tamer, E. (2003). 'Incomplete simultaneous discrete response model with multiple equilibria', *Review of Economic Studies*, vol. 70(1) (January), pp. 147–165.

Topa, G. (2001). 'Social interactions, local spillovers and unemployment', *Review of Economic Studies*, vol. 68(2) (April), pp. 261–295.

Young, H. P. (1993). 'The evolution of conventions', *Econometrica*, vol. 61(1) (January), pp. 57–84.

Young, H. P. (2005). 'The diffusion of innovations in social networks', in (L. E. Blume and S. N. Durlauf, eds.), *The Economy as a Complex Evolving System III*, Oxford: Oxford University Press.