Self-fulfilling Early Contracting Rush

Li, Hao

University of Toronto

Wing Suen

The University of Hong Kong

April 9, 2003

ABSTRACT: In labor markets for entry-level professionals and in other related markets, job applicants' concern for availability of positions and employers' concern for availability of qualified applicants can drive some participants on the two sides to sign early job contracts. The rush to early contracting can be self-fulfilling, as both its effect on expectations about demand-supply balance in the subsequent spot market and the effect on it from changes in the demand-supply balance can be non-monotone. Matching markets with more risk-averse participants, a greater uncertainty regarding relative supply of positions, or a more polarized distribution of applicant qualities can be more vulnerable to self-fulfilling early contracting rushes. Employers can have a collective interest in preventing early offers to a few promising applicants from starting the rushes.

ACKNOWLEDGEMENTS: An earlier version was circulated with the title "Early Contracting Rush". We thank the Editor and two anonymous referees for helpful comments that improved the paper. Li's work is supported by a grant from Social Sciences and Humanities Research Council of Canada.

1. Introduction

Some markets, especially entry-level labor markets for professionals, have experienced difficulties in controlling the timing of interview and appointment dates. Participants on both sides of such markets tend to arrange interviews and make offers ahead of an agreed upon starting date, or in the absence of such a date, before important information about ability of applicants and desirability of positions becomes available.[1] But not all markets are so vulnerable to rolling back of the appointment date. The market for freshly minted Ph.D. economics graduates has followed the same recruitment routine of interviews at the American Economic Association Meetings and subsequent campus visits year after year. Even for markets that have had timing problems, some are more successful than others in enforcing the policy of a uniform starting date. Why do these differences exist across markets? We believe that a model of self-enforcing multiple equilibria in early contracting can offer some insights.

There is a good deal of evidence that supports the existence of self-enforcing multiple equilibria in the unraveling of appointment dates. Roth and Xing (1994) refer to Wald's (1990) description of the experience of a failed attempt to enforce a uniform appointment date in the market for federal judicial law clerks. According to Wald, in the spring of 1989, the District of Columbia Judicial Council adopted a resolution that committed itself to the practice of not making offers to law clerk applicants prior to May 1 of the applicant's second year in law school. This resolution was also adopted by the First, Second, Third, Fourth, Sixth, Eighth, and Tenth circuits, but was rejected by the Fifth, Seventh, and the Eleventh circuits. There were some variations to the adopted resolution: some made compliance with the May 1 deadline contingent upon the compliance of other circuits; some agreed unilaterally. Again according to Wald, as May 1, 1990 approached, "a few judges weakened at the end and made calls ahead of the deadline. This, in turn, provoked

[1] Studies of such markets have been pioneered in a series of papers by Roth and his co-authors (Roth, 1984, 1991, Mongell and Roth, 1991, Roth and Xing, 1994). One of the most recent examples of the rush to contract early occurred the 2001 draft season of the National Basketball Association, which has gained some negative publicity with the dominance of top draft picks by high-school graduates who skip college basketball entirely.

the students to call other judges they preferred before the noon deadline, so there was a destabilizing flurry of pre-deadline transactions."

In a recent paper, Avery, Jolls, Posner, and Roth (2001) argue that there are two related difficulties in enforcing a uniform deadline for offers. The first one is the congestion of proposals and decisions at the starting time of the deadline. This occurs because market participations have too little time to consider more than a few choices, and the fear of losing candidates or positions to competitors drives them to a frenzy in which offers have to be made and accepted. The second problem in enforcing a uniform deadline is cheating by applicants and employers who contact their favorite choices before the deadline. Avery et al. (2001) argue that part of the reason for cheating is the anticipated congestion at the beginning of the deadline. Since the turnaround time is short, it can be critical for applicants and employers to know how committed their top choices are. But even if the congestion problem is non-existent, the cheating problem can arise because of the incentives to use early appointments to insure against risks from match outcomes in the spot market. These incentives can be self-fulfilling, as illustrated by the results of two surveys conducted by Avery et al. (2001) with the federal judges. When the judges were asked whether they believed that their colleagues would adhere to a start-date for interviews of September 1 of the third year of law school, if the date was established by the Judicial Conference, more than seventy percent of the responding judges stated that they did not believe all or virtually all of their colleagues would adhere. The same surveys showed that "most judges say they are willing to comply if others are, but the problem is that they do not believe that most others will comply."

Incentives to sign early contracts in a competitive market can be understood in terms of the trade-off between the insurance benefits and the sorting inefficiencies generated by early contracts. Li and Rosen (1998) consider such a model in which an aggregate uncertainty about market conditions prompts risk-averse market participants to engage in early contracting before their productive characteristics are completely known. A unique equilibrium is derived, in which some participants enter early matches, while others match in the spot market after the aggregate uncertainty is resolved and their productive characteristics become known. However, the model of Li and Rosen (1998) does not completely capture the self-fulfilling property of this process. The critical assumption in Li and Rosen (1998) responsible for the uniqueness of early contracting equilibrium turns out to be that there is no uncertainty about firms' hiring needs. When a job applicant contracts early with an employer, the expected number of remaining applicants in the spot market falls by less than one, because not all workers will turn out to be productive, while the number of job positions in the spot market falls by one. Therefore an increase in early contracting beyond the equilibrium level would make jobs more scarce in the spot market. Since firms do not face uncertainty about their own hiring needs, as jobs become more scarce, the scope for mutually beneficial early contracting would decline because firms would face little uncertainty in their spot market payoff. Equilibrium in the Li and Rosen model is therefore "self-correcting."

When there are uncertainties both about quality of applicants and about hiring needs of employers, however, multiple equilibria can occur. There are two reasons why equilibria need not be "self-correcting" in the presence of two-sided uncertainties. First, since expected vacancies in the early market may not materialize in the spot market, an increase in the extent of early contracting does not always make jobs more scarce in the spot market. Indeed, as early contracting spreads from applicants with high expected abilities to those who are not so promising, the residual demand for positions in the spot market may first fall before rising relative to supply. This means that different degrees of early contracting can be consistent with the same demand-supply balance in the spot market. Second, even when early contracting is so extensive that a further increase does make jobs more scarce in the spot market, the result may be a greater instead of a smaller scope of mutually beneficial early contracting. This is because when employers also face uncertainty regarding their hiring needs, incentives to contract early are the greatest if the spot market is perceived to be balanced, as an imbalance in either direction reduces the chance that an applicant and an employer can strike a deal in the early market for insurance purpose. Non-monotonicities in these two relations can give rise to multiple early contracting equilibria. As a result, market sentiment is important in understanding whether early contracting rushes occur. If the market is calm and few participant are anticipated to "jump the gun" by offering contracts early, then no one will have the incentive to make early offers. But if the market is hot and a significant fraction of market participants are making early contacts, others will want to follow suit and another equilibrium arises with early matches by some participants.

Our study of multiple equilibria in early contracting rushes is potentially useful in assessing viability of new reforms and regulations. Given that it is close to impossible for a uniform-date policy to plug all "leaks" of early offers, some understanding of how leaks feed themselves is important for identifying potential weak spots of the reforms and increasing the effectiveness of preventive measures. Moreover, the existence of multiple equilibria in our model facilitates stability analysis. An equilibrium in the early market may be thought of as unstable if "small shocks" to market sentiment begin a self-fulfilling process that leads to a new equilibrium with more wide-spread early contracting. Matching markets with more risk-averse participants, a greater uncertainty regarding relative supply of positions, or a more polarized distribution of applicant qualities can be more vulnerable to self-fulfilling early contracting rushes. Instability of an equilibrium with a limited extent of early contracting also helps to explain why reforms in some markets (e.g., the judicial law clerks market) were initially successful in containing early offers before breaking down entirely.

This paper is organized as follows. The model of early contracting is presented in the next section. In Section 3 we show how multiple equilibria of early contracting arise, either because participants' expectation of the balance of demand and supply in the spot market has non-monotone effects on their decisions to contract early, or because individual decisions to sign early contracts have non-monotone effects on the balance of demand and supply in the spot market. Section 4 discusses stability and welfare implications of multiple equilibria. We address the issue of when early contracting rushes are likely to occur due to the vulnerability of the equilibrium with no early contracting, show that multiple early contracting equilibria cannot be Pareto-ranked, and compare the welfare of different groups of market participants across the early contracting equilibria. Section 5 extends the analysis in the paper by allowing fixed-wage early contracts and heterogeneity on both sides of the market. The final section summarizes the results and concludes the paper.

2. The Model

The setup of the model generally follows Li and Rosen (1998), with important differences pointed out along the way. There are two periods when pairwise contracts can be agreed upon. In period one, both workers and firms face individual uncertainty about their productivity. An output of 1 is produced in period two if and only if a productive worker is matched with a productive firm; otherwise, the output is zero. In addition to individual uncertainty, there is also some aggregate uncertainty that affects market demand and supply in period two.

Individual uncertainty features prominently in discussions of early contracting, because it generates both insurance incentives to contract early and the cost of sorting inefficiency.[2] We model individual uncertainty as follows. Workers are characterized by their types λ. A type-λ worker has probability λ of becoming productive in period two. Worker type λ is assumed to be continuously distributed on the support $[\lambda_{\min}, \lambda_{\max}]$, with distribution function F and density function f. The assumption of continuous type distribution avoids dealing with discrete distributions as in Li and Rosen (1998), and is more realistic in markets with large number of participants on both sides. Let $\overline{\lambda}$ be the mean type of workers. On the other side of the market, we assume that all firms are of the same type: each firm has probability $\mu < 1$ of becoming productive in period two. The assumption of homogeneous firms simplifies the exposition; the multiple equilibria result does not depend on this assumption (see Section 5). Further, we assume

$$\lambda_{\max} > \mu > \lambda_{\min}.$$

That is, some workers in the market have higher probabilities than firms to be productive, while others are less likely to be productive. This is impossible in the model of Li and Rosen (1998), who assume one-sided individual uncertainty, with $\mu = 1 \geq \lambda_{\max}$, and derive

[2] In some markets, unraveling of the appointment date has been pushed so far back that the uncertainty about abilities of the applicants becomes substantial. For example, in the market for federal law clerks, with the appointment date unraveled to the middle of the second year of the three-year law program, there is significant uncertainty about the ability of the candidate at the time of early appointment (Avery et al., 2001).

a unique equilibrium. We will show later that the assumption of two-sided uncertainties here is responsible for the multiple equilibria result in the present paper.

Let the measure of workers be 1 and the measure of firms be n, which can be either greater than or smaller than 1. Since there is a continuum of workers and firms, in the absence of any aggregate uncertainty about the second period spot market, when all wait for the spot market, either for certain there will be a shortage of positions (if $\overline{\lambda} > n\mu$), or for certain there will be a shortage of applicants (if $\overline{\lambda} < n\mu$). It will become clear later that early contracts are then impossible, because either all workers, or all firms will refuse to match early. As in Li and Rosen (1998), aggregate uncertainty is necessary for early contracting to occur in a model with binary productivity.[3] Unlike Li and Rosen (1998), where aggregate uncertainty is created by discreteness of type distribution, here we introduce it through exogenous shocks to the spot market. We assume that in period two, before firms and workers are matched, an additional net measure x of productive firms comes into existence. This shock x is a random variable distributed continuously on $[x_{\min}, x_{\max}]$, with distribution function H and density function h. We allow x to be positive or negative.[4] We assume that

$$x_{\min} < \overline{\lambda} - n\mu < x_{\max}.$$

This assumption means that starting from a situation where all participants wait for the spot market, both workers and firms have positive probabilities of being on the short side of the market.

Wages are assumed to be flexible in both the first period market and the second period market. In the spot market of period two, unproductive workers and firms cannot produce and receive 0. Due to our binary assumption, productive workers and firms receive either nothing or all of the output, depending on the market condition. All productive workers

[3] See Li and Suen (2000) and Suen (2000) for models where realized productivity is a continuous variable. These models generate early contracting equilibrium without aggregate uncertainty.

[4] We interpret a positive value of x as more new firms than new applicants entering the spot market, but it also can result from some applicants changing their minds about applying for a position between the first and the second period. Similarly, a negative value of x can result from firms withdrawing from the spot market due to unexpected economic downturns or even bankruptcy. See, for example, Roth and Xing (1994). The results in the present paper do not depend on the interpretations.

receive 1 and productive firms receive 0 if workers are on the short side—there are fewer productive workers than productive firms. The opposite is true when productive firms are on the short side. In the first period early market, an early contract between a firm and a worker is a promise by a firm to pay $r \in [0, 1]$ to the worker in period two if both turn out to be productive, and 0 otherwise. The assumption of flexible wages in both the first period market and the second period market suits some markets, such as the labor market of American law firms, where salary wars have been reported in the rush to make early offers (Roth and Xing, 1994), but is less appropriate for markets such as the one for federal law clerks where salaries are non-negotiable. In Section 5, we adapt the analysis to markets where wages are fixed.

Finally, let u and v represent the von Neumann-Morgenstern utility functions of workers and firms, respectively. Both u and v are assumed to be weakly concave, with strict concavity for at least one of them. For convenience, we normalize by assuming that $u(0) = v(0) = 0$ and $u(1) = v(1) = 1$. It will be seen from the ensuing analysis that early contracting can never occur if participants on both sides of the market are risk-neutral. This is because in our model early contracting is the equilibrium outcome where participants trade off the insurance benefits against the sorting inefficiencies. With both sides risk-neutral, there is no insurance gain and in equilibrium all participants wait for the spot market. However, early contracting can occur if only one side of the market is risk-averse. Thus our analysis applies to markets where workers are averse to the risks in the job market outcome but firms are neutral to the risks in filling their positions.

3. The Analysis

A road map for the following analysis is perhaps helpful. First, for any perceived market condition in period two, we define the ask price, which gives the minimum wage offer in an early contract for workers to sign up, and the bid price, which gives the maximum offer that firms are prepared to give to each type of worker. Second, we show how the bid and ask prices determine a non-monotone relationship between the perceived market condition in period two and the extent of early contracting in period one: incentives to contract early are the greatest when the period two market is perceived to be more or less balanced. Third, we show that the relationship from the extent of early contracting to the market condition can also be non-monotone: the prospect that productive workers will be on the short side of the spot market first rises as high types of workers form early contracts, and then declines as early contracting spreads to lower types of workers. After defining an equilibrium as a pair of market condition in period two and extent of early contracting in period one that satisfy the two relationships, we show how the non-monotonicity of either relationship can lead to multiple self-fulfilling early contracting equilibria.

3.1. Bid and ask prices

Incentives of the participants to engage in early contracting depend on their expectations of the spot market condition of demand and supply. In our model with binary productivities in the spot market, expectations are summarized by the probability that workers are on the short side of the spot market. We denote this probability by π.

A type-λ worker prefers early contracting to waiting if he receives r in the early contracting market such that $\lambda\mu u(r) \geq \lambda\pi$. Define the "ask price" by $r^w(\pi)$, that is,

$$\mu u(r^w(\pi)) \equiv \pi.$$

Note that this price is independent of worker's type λ. It is straightforward to verify the following intuitive properties of the ask price function: (i) $r^w(0) = 0$; (ii) $r^w(1) > 1$; (iii) $r^w(\pi)$ is increasing and convex in π; and (iv) $r^w(\pi)$ is decreasing in μ for any π. The first property follows from the normalization that $u(0) = 0$. The second property follows from our assumption that $\mu < 1$. If productive workers are short for sure in the spot market, workers of all types will demand more than the entire output for them to sign up early with firms, to compensate for the fact that the firm's promise in an early contract is fulfilled only when it turns out to be productive. A greater prospect of shortages of productive workers in period two means that workers have to be compensated more to sign up in period one, so r^w is increasing in π. Convexity of r^w follows from risk-aversion of workers. Finally, if the firms' prospect is better, then workers of any type can be satisfied with lower wages in early contracts, so $r^w(\pi)$ is decreasing in μ for any π.

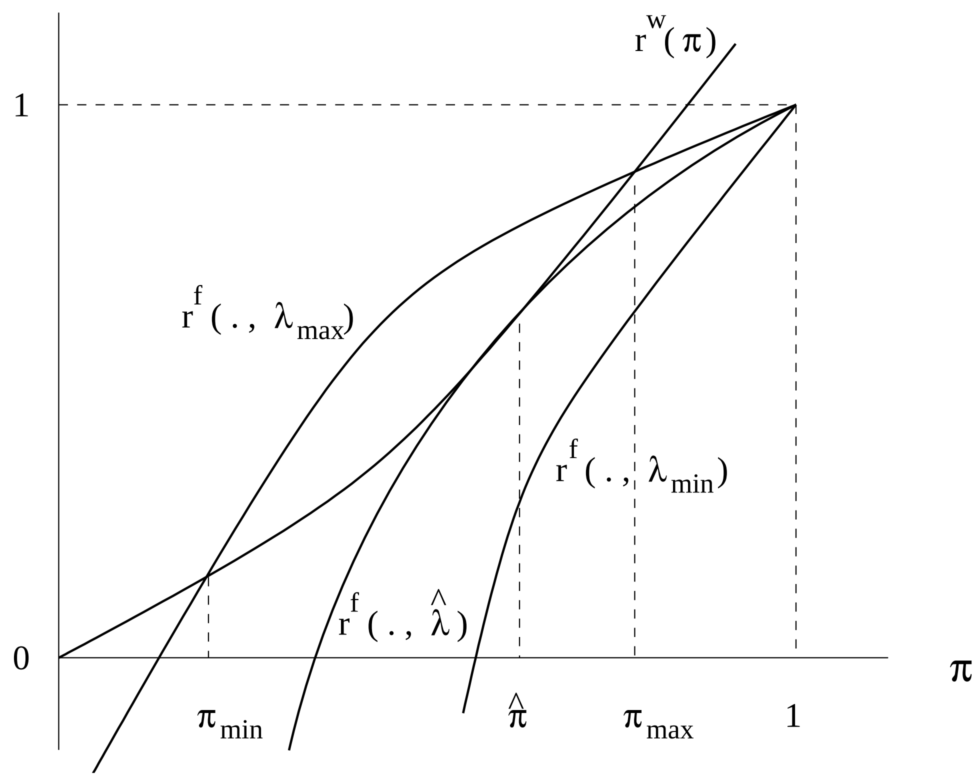

Figure 1

On the other side of the market, a firm prefers early contracting with a type-λ worker to waiting if the price r it pays satisfies $\lambda\mu v(1-r) \geq \mu(1-\pi)$. Denote the "bid price" for a λ-type worker by $r^f(\pi, \lambda)$. Then

$$\lambda v(1 - r^f(\pi, \lambda)) \equiv 1 - \pi.$$

This bid price function has the following properties: (i) $r^f(0, \lambda) < 0$; (ii) $r^f(1, \lambda) = 1$; (iii) $r^f(\pi, \lambda)$ is increasing and concave in π; and (iv) $r^f(\pi, \lambda)$ is increasing in λ. The first two properties follow from our normalization that $v(1) = 1$ and $v(0) = 0$. A greater prospect of shortages of productive workers in period two means that firms are willing to offer higher wages in early contracts, so r^f is increasing in π for any λ. Concavity of r^f follows from risk-aversion of firms. Finally, for any fixed π, firms are willing to offer higher wages in period one to more promising workers, so r^f is increasing in λ for any π.

We sketch the ask price function and a family of the bid price functions in Figure 1. In our model the insurance incentives to contract early are the greatest when the period two market is expected to be more or less balanced. As can be seen from Figure 1, concavity of r^f and convexity of r^w in π imply that for any type λ the difference between the bid and the ask prices are greater for intermediate values of π than for extreme values. If workers are desperate for early matches because there is an expected over-supply of productive workers in the period two market (i.e., if π is close to 0), firms' best early offers fall short of workers' demand due to the uncertainty about workers' productivity (i.e., $r^f(\pi,\lambda) < r^w(\pi)$ because $\lambda < 1$). Conversely, if firms are desperate because they expect to have a hard time finding productive workers from the spot market (i.e., if π is close to 1), workers demand more than what firms can offer in early contracts due to the uncertainty about firms' prospect (i.e., $r^w(\pi) > r^f(\pi,\lambda)$ for any λ because $\mu < 1$).

Since r^f is increasing in λ for any π, the bid price functions are ordered by worker type. In Figure 1, if workers and firms are risk-neutral, both r^w and r^f would be linear. In this case, since $r^w(0) = 0 > r^f(0,\lambda)$ and $r^w(1) > 1 = r^f(1,\lambda)$, the ask price function r^w would lie above even the highest bid price function r^f for any π. Early contracting is impossible if both sides of the market are risk-neutral. If at least one side is risk-averse, the bid price function $r^f(\pi,\lambda)$ can rise above the ask price function $r^w(\pi)$ for some intermediate values of π. As long as at least one side of the participants are sufficiently risk-averse there is a unique worker type $\hat{\lambda}$ (not necessarily between $\lambda_{\min}$ and $\lambda_{\max}$) such that $r^f(\pi,\hat{\lambda})$ is tangent to $r^w(\pi)$.[5] We assume that $\hat{\lambda} < \lambda_{\max}$; otherwise, early contracting can never occur because the ask price is higher than the bid price for any type. Then, for any $\lambda > \hat{\lambda}$, the bid price function $r^f(\pi,\lambda)$ crosses the ask price function $r^w(\pi)$ exactly twice. Let $\pi_{\min} < \pi_{\max}$ be the two solutions to the equation

$$r^f(\pi,\lambda_{\max}) = r^w(\pi),$$

[5] To see the existence and uniqueness of $\hat{\lambda}$, for each λ let $\hat{\pi}(\lambda)$ be the unique value of π at which r^w and r^f have the same slope: $dr^w(\hat{\pi})/d\pi \equiv \partial r^f(\hat{\pi},\lambda)/\partial\pi$. Let $\hat{\pi}(\lambda) = 0$ if $dr^w(\pi)/d\pi > \partial r^f(\pi,\lambda)/\partial\pi$ for all π, and $\hat{\pi}(\lambda) = 1$ if $dr^w(\pi)/d\pi < \partial r^f(\pi,\lambda)/\partial\pi$ for all π. Define $\delta(\lambda)$ as the distance between r^w and r^f at $\hat{\pi}$, that is, $\delta(\lambda) \equiv r^w(\hat{\pi}(\lambda)) - r^f(\hat{\pi}(\lambda),\lambda)$. By construction $d\delta(\lambda)/d\lambda = -\partial r^f(\hat{\pi}(\lambda),\lambda)/\partial\lambda < 0$. Then, there is a unique type $\hat{\lambda}$ such that $r^f(\pi,\hat{\lambda})$ is tangent to $r^w(\pi)$ at $\hat{\pi}(\hat{\lambda})$, if (i) $\delta(\lambda) > 0$ for λ close to 0; and (ii) $\delta(\lambda) < 0$ for λ close to 1. The first of the two conditions is always satisfied, because $r^f(\pi,\lambda)$ falls entirely below $r^w(\pi)$ if λ is sufficiently small. For any given μ, the second condition is satisfied, if either u is sufficiently concave so that r^w increases slowly with π, or v is sufficiently concave so that r^f increases quickly with π.

and $\hat{\pi}$ be the tangency point of $r^f(\pi, \hat{\lambda})$ and $r^w(\pi)$, i.e. the unique solution to the equation

$$r^f(\pi, \hat{\lambda}) = r^w(\pi).$$

As can be seen from Figure 1, $\pi_{\min} < \hat{\pi} < \pi_{\max}$.

3.2. Two non-monotone relationships

Early contracting is mutually beneficial to a type-λ worker and a firm if the bid price $r^f(\pi, \lambda)$ for type λ exceeds the ask price $r^w(\pi)$. Since $r^f(\pi, \lambda)$ is increasing in λ, the "ordering property" holds: If firms are willing to bid for workers of type λ, they are also willing to bid for workers of types higher than λ. Because workers' willingness to accept early contracts is independent of their types, the ordering property implies a critical worker type λ_0, for whom the bid price is no lower than the ask price. Workers with $\lambda \geq \lambda_0$ will contract early while those with $\lambda < \lambda_0$ will wait.

From Figure 1, the critical type λ_0 depends on perceived market conditions in period two. We can write $\lambda_0 = l(\pi)$. We define the l function without regard to the constraint that $\lambda_0 \geq \lambda_{\min}$, in order to focus on the general shape of the function. The l function is sketched in each of the two panels in Figure 2. See the lower panel for the labels corresponding to the variables we have defined. For $\pi < \pi_{\min}$ or $\pi > \pi_{\max}$, the bid price is lower than the ask price for all λ, and so $l(\pi) = \lambda_{\max}$. For any π such that $l(\pi) < \lambda_{\max}$, by definition $r^w(\pi) = r^f(\pi, l(\pi))$. For $\pi = \hat{\pi}$, the bid price is not lower than the ask price for all $\lambda \geq \hat{\lambda}$, and so $l(\hat{\pi}) = \hat{\lambda}$. That is, $\hat{\lambda}$ is the lowest type that can sign early contracts, regardless of the market condition. As the value of π deviates from $\hat{\pi}$ in either direction, the critical type λ_0 rises. Hence, the l function is U-shaped, attaining a minimum of $\hat{\lambda}$ at $\pi = \hat{\pi}$.[6]

Because the bid prices are ordered by worker type, intermediate values of π imply not only that insurance incentives for early contracting are greater for any fixed worker type, but also that insurance incentives exist for more types. This translates into the

[6] The l function is differentiable at $\hat{\pi}$, implying that l is indeed U-shaped. To see this, note that $l(\pi)$ is implicitly defined by $r^f(\pi, l(\pi)) = r^w(\pi)$ for both $\pi \geq \hat{\pi}$ and $\pi \leq \hat{\pi}$. The right-derivative of l at $\hat{\pi}$ takes the same form as the left-derivative: both are given by the ratio of $dr^w(\hat{\pi})/d\pi - \partial r^f(\hat{\pi}, l(\hat{\pi}))/\partial\pi$ to $\partial r^f(\hat{\pi}, l(\hat{\pi}))/\partial\lambda$. By definition, $r^w(\pi)$ and $r^f(\pi, l(\hat{\pi}))$ are tangent at $\hat{\pi}$, and so $dl(\hat{\pi})/d\pi$ exists and is equal to zero.

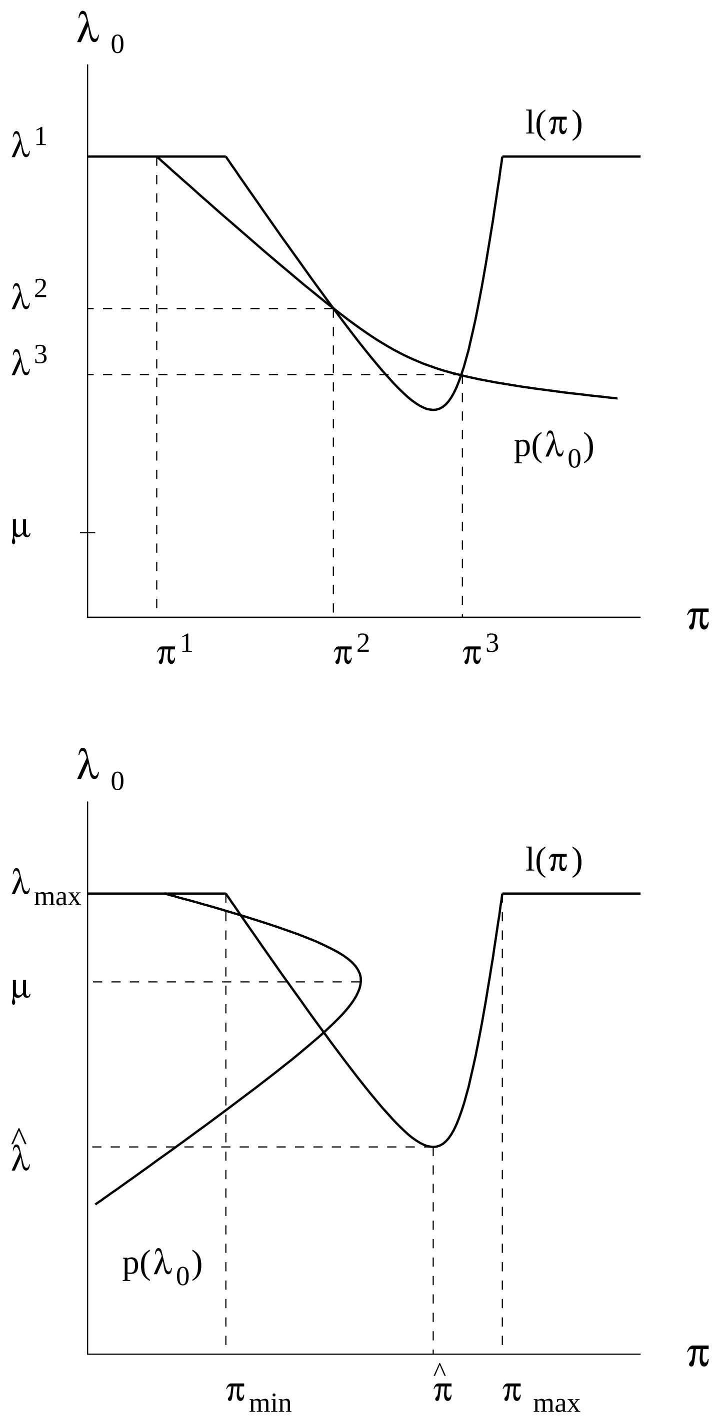

Figure 2

non-monotone relationship from π to λ_0 through the l function. The non-monotonicity of the l function contrasts with the model of Li and Rosen (1998). In their model, individual uncertainty is one-sided, with $\mu = 1$. Consequently, $r^w(1) = r^f(1, \lambda) = 1$ for any λ. In this case the bid price function $r^f(\pi, \lambda)$ for any type λ always intersects the ask price function

$r^w(\pi)$ at $\pi = 1$. This implies that the two functions can be tangent to each other only at $\pi = 1$, and so $\hat{\pi} = 1$. Thus, even though insurance gains from early contracting become smaller for all types as π becomes closer to 1, as in the present model, the gains become available to more types at the same time. As a result, $l(\pi)$ is monotonically decreasing in π: the greater is π, the more worker types that can strike an early deal with firms.

In an early contracting equilibrium, both π and λ_0 are endogenously determined. Having considered how π affects λ_0, we now characterize how λ_0 affects π in the period two market. Denote this function as $p(\lambda_0)$. Since all workers with $\lambda < \lambda_0$ stay in the period two market, the measure of productive workers in period two is $\int_{\lambda_{\min}}^{\lambda_0} \lambda f(\lambda) d\lambda$. The measure of productive firms, on the other hand, is $(n - (1 - F(\lambda_0)))\mu + x$. Define the excess supply of workers before the shock x is realized as

$$e(\lambda_0) = \int_{\lambda_{\min}}^{\lambda_0} \lambda f(\lambda) d\lambda - (n - (1 - F(\lambda_0)))\mu.$$

Then the probability that workers are on the short side of the market is equal to

$$p(\lambda_0) = \begin{cases} 0, & \text{if } e(\lambda_0) \geq x_{\max}; \\ 1 - H(e(\lambda_0)), & \text{if } x_{\min} < e(\lambda_0) < x_{\max}; \\ 1, & \text{if } e(\lambda_0) \leq x_{\min}. \end{cases}$$

The assumption that $x_{\min} < \overline{\lambda} - n\mu < x_{\max}$ implies that $p(\lambda_0)$ is strictly between 0 and 1 at least for λ_0 close to $\lambda_{\max}$. For any such λ_0, the derivative of the function p with respect to λ_0 is

$$-h(e(\lambda_0))f(\lambda_0)(\lambda_0 - \mu).$$

Since early contracting satisfies the ordering property, our assumption that $\lambda_{\max} < \mu < \lambda_{\min}$ implies that initially when workers who sign early contracts have higher probabilities of becoming productive than do firms, the signing of more workers increases the chance that productive workers will be short in the spot market. That is, $p(\lambda_0)$ increases as λ_0 decreases from $\lambda_{\max}$. See the upper panel of Figure 2. However, if aggregate uncertainty still exists when the prospects of the threshold worker type λ_0 drop to the level of firms (i.e., if $x_{\min} < e(\mu) < x_{\max}$), then $p(\lambda_0)$ starts to decrease as λ_0 falls below μ. This is the case depicted in the lower panel of Figure 2. In this case μ is relatively high, so eventually as the last workers who sign early contracts have lower probabilities of becoming productive than do firms, the signing of more workers reduces the chance that productive workers will be on the short side.

The potentially non-monotone property of the p function depends critically on the assumption of $\mu < \lambda_{\max}$. If $\mu = 1$, as in Li and Rosen (1998), the above derivation of p shows that the function is monotonically increasing for any λ_0. Intuitively, in this case the first-order impact of more early contracts is that firms become more scarce in the spot market, so that the probability π of workers being short in period two monotonically declines as more early matches are made.

3.3. Incomplete and complete early contracting

In a rational expectations equilibrium, the probability π that productive workers are short in period two and the threshold λ_0 of worker types that enter early matches are determined endogenously and are consistent with each other.[7] An early contracting equilibrium can be naturally defined by a pair of variables π and λ_0 that satisfy the two relationships: $\lambda_0 = l(\pi)$ and $\pi = p(\lambda_0)$. In such an equilibrium, the extent of early contracting is endogenously limited by the insurance gains from early contracting. However, it is possible that the extent of early contracting is exogenously limited by the period one market size. For example, it can happen that the insurance gains from early contracting still exist when all firms have entered early matches. We distinguish two types of equilibria according to whether the extent of early contracting is limited by insurance gains or by the market size.

Definition 3.1. An incomplete early contracting equilibrium is a pair (π^*, λ_0^*), with $\lambda_0^* > \lambda_{\min}$ and $1 - F(\lambda_0^*) < n$, such that $\pi^* = p(\lambda_0^*)$, and $\lambda_0^* = l(\pi^*)$.

An incomplete equilibrium is an intersection of the two functions $l(\pi)$ and $p(\lambda_0)$, provided that the extent of early contracting does not exceed the size of the early market. Given π^*, worker type λ_0^* is the last one for whom the bid price exceeds the ask price, so all insurance gains from early contracting are exhausted. Early contracting is incomplete in

[7] For discussions of formal definition of early contracting equilibrium, see Li and Rosen (1998) and Li and Suen (2000).

this type of equilibria in that not all firms and not all workers contract in the early market. Since firms are identical, an incomplete equilibrium with $\lambda_0^* < \lambda_{\max}$ is associated with a schedule of early wage offers $r^*(\lambda)$, for $\lambda \in [\lambda_0^*, \lambda_{\max}]$, such that all firms are indifferent between waiting for the spot market and signing early contracts with any worker of type $\lambda \in [\lambda_0^*, \lambda_{\max}]$. This wage offer schedule is then given by:

$$r^*(\lambda) = r^f(\pi^*, \lambda),$$

for all $\lambda \in [\lambda_0^*, \lambda_{\max}]$. In an incomplete equilibrium with identical firms, all insurance benefits from early matches are captured by workers. Among workers who enter early matches, higher types benefit more from early contracting than lower types do.

In the second type of early contracting equilibria, called "complete" early contracting equilibrium, either all workers or all firms enter early matches in period one. Take the case of all-worker complete equilibrium. This can happen only if $n > 1$ so there are more firms than workers in the period one market. Since $\hat{\lambda}$ is the lowest type that can enter early matches, a complete early contracting equilibrium with all workers entering early matches can occur only if $\lambda_{\min} > \hat{\lambda}$. Similarly, an all-firm complete early contracting equilibrium can occur only if there are more workers with type higher than $\hat{\lambda}$ than firms in period one, i.e., if $\lambda_n > \hat{\lambda}$, where λ_n satisfies

$$n = 1 - F(\lambda_n).$$

We have the following definition.

Definition 3.2. An all-worker complete early contracting equilibrium is a pair $(\pi^*, \lambda_{\min})$ such that $\pi^* = 1 - H(e(\lambda_{\min}))$ and $l(\pi^*) < \lambda_{\min}$. An all-firm complete early contracting equilibrium is a pair (π^*, λ_n) such that $\pi^* = 1 - H(e(\lambda_n))$ and $l(\pi^*) < \lambda_n$.

A complete equilibrium corresponds to the point $(\pi^*, \lambda_{\min})$ or (π^*, λ_n) on the $p(\lambda_0)$ function, provided it lies "above" the point $(\pi^*, l(\pi^*)))$ on the $l(\pi)$ function. In such an equilibrium, insurance gains from early contracting still exist after the limit of the period one market is reached on the workers' side or on the firms' side. In an all-worker complete equilibrium, since not all firms enter early matches, the early wage schedule is determined in the same way as in an incomplete equilibrium, with all insurance benefits going to the workers. In an all-firm complete equilibrium, in contrast, because there is a shortage of firms in the period one market, all insurance benefits from early contracting with the critical type λ_n go to the firms they are matched with. Since all firms are identical, they are indifferent between signing with the critical type at the ask price $r^w(\pi^*)$ and signing with types higher than λ_n. Thus, the equilibrium early wage schedule $r^*(\lambda)$, for $\lambda \in [\lambda_n, \lambda_{\max}]$, is given by

$$\lambda v(1 - r^*(\lambda)) = \lambda_n v(1 - r^w(\pi^*)).$$

For all types higher than λ_n, the insurance benefits from early contracting are split between workers and the firms they are matched with.

3.4. Multiple equilibria

Early contracting equilibria are graphically displayed in Figure 2. The upper panel of Figure 2 shows the case in which $\hat{\lambda} > \mu$, and the lower panel shows the case for $\hat{\lambda} < \mu$. Consider first incomplete early contracting equilibria, which correspond to the intersections of $p(\lambda_0)$ and $l(\pi)$. In each panel of Figure 2, we illustrate a case of multiple equilibria. The diagram is not meant to include all possibilities of how the two functions l and p can intersect each other; a complete catalogue of the possibilities is tedious and not very illuminating. Instead, we use the two panels to illustrate two different reasons for multiple equilibria. In the upper panel, multiple equilibria arise because of the U-shaped $l(\pi)$ function. In the lower panel, multiple equilibria arise because the function $p(\lambda_0)$ is non-monotone.[8]

A complete equilibrium arises when the extent of early contracting reaches the limit of the early market before the insurance benefits are exhausted. For a given set of parameters, there can be at most one complete early contracting equilibrium, but it can coexist with incomplete equilibria. Consider the upper panel of Figure 2, for example. Let the three intersections of $l(\pi)$ and $p(\lambda_0)$ be (π^1, λ^1), (π^2, λ^2), and (π^3, λ^3), in the order of increasing values of π. When $n > 1$ it can happen that $\lambda_{\min}$ lies between λ^2 and λ^3. Then, at

[8] If $p(\lambda_0)$ is monotonically increasing, as in Li and Rosen (1998), there would be a single intersection with the $l(\pi)$ function.

$\lambda_0 = \lambda_{\min}$ and the corresponding $\pi = p(\lambda_{\min})$, after all workers have entered early matches in period one, there are still firms that are willing to contract early provided more workers were available, because $l(p(\lambda_{\min})) < \lambda_{\min}$. This is an all-worker complete early contracting equilibrium. On the other hand, when $n < 1$, λ_n may lie between λ^2 and λ^3. At the point $(p(\lambda_n), \lambda_n)$ on the $p(\lambda_0)$ curve, after all firms have entered early matches with workers of type λ_n and higher, there are still workers who are willing to enter early matches provided more firms were available, because $l(p(\lambda_n)) < \lambda_n$. This is an all-firm complete early contracting equilibrium. In either case, (π^3, λ^3) is no longer an equilibrium.[9] The following proposition summarizes our discussion.

PROPOSITION 3.3. *Multiple early contracting equilibria can arise because $l(\pi)$ is non-monotone, or because $p(\lambda_0)$ is non-monotone.*

In Li and Rosen (1998), early contracting equilibrium is unique. The present paper makes two different assumptions. First, firms as well as workers face individual uncertainty: $\mu < 1$ in the present paper, whereas $\mu = 1$ in Li and Rosen. Second, types of workers are distributed continuously and aggregate uncertainty about the spot market is introduced through newcomers in the spot market, rather than through discrete type distribution of workers. As we have explained earlier, the first difference alone is responsible for generating multiple equilibria in the early market. It renders both the l function and the p function non-monotone. If $\mu = 1$, as in Li and Rosen (1998), we would have a downward sloping function $l(\pi)$ and an upward sloping $p(\lambda_0)$ in Figure 2. In that case, if there is an intersection of the two functions, it will be unique.

Therefore, the assumption that firms also face individual uncertainty potentially generates multiple early contracting equilibria in two ways. The economic meanings of non-monotonicity of the two functions l and p are different. Non-monotonicity of the function l captures the idea that uncertainty about the spot market, and hence the insurance benefits from early contracting, is the greatest when the spot market is neither too tight nor too

[9] If $\lambda_{\min}$ or λ_n is between $\lambda_{\max}$ and λ^2, the point $(p(\lambda_{\min}), \lambda_{\min})$, or $(p(\lambda_n), \lambda_n)$ on the $p(\lambda_0)$ curve does not correspond to a complete equilibrium because it lies below the corresponding point on the l function. In this case, the only early contracting equilibrium is (π^1, λ^1).

slack. A U-shaped p function captures the idea that the feedback effect of early contracting on the spot market is not monotone: as more participants sign early, the probability that productive workers will be short in the spot market first goes up because those who sign early are on more likely to be more productive than firms, but eventually goes down as workers who are less promising also sign up early.

4. Stability and Welfare Implications

Existence of multiple early contracting equilibria is more than a theoretical possibility in the class of matching models where risk-sharing motivates participants to contract early without adequate information about each other. In this section we introduce stability analysis to further understanding of cross-market differences in terms of how vulnerable they are to early contracting rushes. We also provide welfare comparisons of different groups of market participants across the equilibria that can help explain observed efforts in some matching markets to regulate timing of offers.

4.1. Stability and vulnerability

Borrowing from the standard pseudo-dynamic stability analysis (Henderson and Quandt, 1980), stability of an equilibrium (π^*, λ_0^*) depends on the relative slopes of the functions l and p. If the l function is (locally) downward sloping and the p function is (locally) upward sloping (see Figure 2, for example), this could lead to "cobweb" type of dynamics as in demand-supply analysis. This kind of analysis implicitly assumes that out-of-equilibrium π and λ_0 take turns to adjust, and the resulting dynamics in terms of λ_0 is non-monotone. In our setup, however, workers are heterogeneous and their incentives to contract early are ordered by types. So it makes sense to consider the kind of pseudo-dynamics where market participants make sequential decisions in an orderly fashion and out-of-equilibrium adjustments are monotone in terms of λ_0.[10] We therefore adopt a stability analysis which

[10] The stability definition introduced below is based on our assumption of heterogeneous worker ordering property. If workers are homogeneous, equilibrium can still be defined by an intersection of the l and p functions, except that λ_0 now represents the fraction of workers that contract early. The l function is locally constant and the p function is always monotone, implying that there will be a unique equilibrium and that it is stable according to both the standard cobweb dynamics and our definition below. In this case there is no advantage in using our definition, but then the uniqueness of equilibrium makes stability an uninteresting issue.

amounts to assuming instantaneous adjustments of π: as λ_0 gradually moves from the starting point in the direction of $l(p(\lambda_0))$, π adjusts to keep pace with changes in λ_0 by staying on the p function. This assumption reduces a two-dimensional dynamic adjustment problem to a one-dimensional problem, and ensures that the direction of adjustment in terms of λ_0 is always monotone regardless of the slopes of l and p functions. Formally, we introduce the following definition.

DEFINITION 4.1. An incomplete early contracting equilibrium with λ_0^* is stable if there is a neighborhood around λ_0^* such that for any λ_0 in the neighborhood, $\lambda_0 < \lambda_0^*$ implies that $l(p(\lambda_0)) > \lambda_0$ and $\lambda_0 > \lambda_0^*$ implies that $l(p(\lambda_0)) < \lambda_0$.

The above definition leads to the following characterization in terms of slopes of l and p functions around λ_0^*. Take the linear approximation of the combined function $l(p(\lambda_0))$ around λ_0^*, we have

$$l(p(\lambda_0)) = \lambda_0^* + l'(\pi^*)p'(\lambda_0^*)(\lambda_0 - \lambda_0^*).$$

Then our definition of stability amounts to the condition that

$$l'(\pi^*)p'(\lambda_0^*) < 1.$$

It follows that if one of the two functions l and p is downward sloping and the other is upward sloping, then the intersection λ_0^* is always stable. Furthermore, if both functions are downward sloping or upward sloping, then λ_0^* is stable if and only if p is steeper than l in the π-λ_0 diagram. In other words, an intersection is stable if and only if the $p(\lambda_0)$ function intersects the $l(\pi)$ function from above.[11] Note that under the standard cobweb-dynamics, λ_0^* is stable if and only if $|l'p'| < 1$. Thus our adopted stability concept is weaker. In particular, it imposes no restrictions on the slopes when one of the two functions is downward sloping and the other is upward sloping.

One implication of our definition of stability is that the unique equilibrium constructed by Li and Rosen (1998) is stable by our definition, because the l function is downward

[11] It is straightforward to show that in a generic situation, the number of equilibria is finite and odd, which implies existence. Moreover, if we rank the equilibria in increasing order of λ_0, then the equilibria are alternatively stable and unstable.

sloping and the p function is upward sloping. Note also that whenever no early contracting is an equilibrium, with $\lambda_0^* = \lambda_{\max}$, it is also stable by our definition because the l function is flat at such an equilibrium. See Figure 2.

Thus far, we have considered only incomplete early contracting equilibria. But the same definition of stability applies to complete early contracting equilibria as well. Since by definition a complete equilibrium contracting corresponds to a point on the $p(\lambda_0)$ function that is above the $l(\lambda_0)$ function, any such equilibrium is stable. A complete early contracting equilibrium is always reached monotonically as more and more workers enter early matches with firms. The result that any complete equilibrium is stable is reassuring and adds to the attraction to our concept of stability.

Our model of multiple early contracting equilibria shares some similarities with the bank runs model of Diamond and Dybvig (1983). Multiplicity of equilibria arises in situations where coordination is important such as in models of bank runs, because actions by agents can be self-fulfilling. In our model, workers have different characteristics in the first period. Multiple equilibria arise not from coordination but from the non-monotone effects of early contracting by some agents on the insurance benefits from early contracting for the remaining agents (non-monotonicity of the l function), or from the non-monotone feedback effects of early contracting on the spot market (non-monotonicity of the p function).[12] Despite the differences, our stability concept allows us to consider comparative statics issues in a similar spirit as in Diamond and Dybvig (1983). Consider again the case of $\mu < \hat{\lambda}$ shown in the upper panel of Figure 2, where (π^1, λ^1) and (π^3, λ^3) are stable, but (π^2, λ^2) is not. We are concerned with the transition from (π^1, λ^1) to (π^3, λ^3). In our model, we can tell the following "big push" story. Since (π^1, λ^1) is stable and (π^2, λ^2) is not, it takes a portion of the participants to sign early contracts for the equilibrium to switch from (π^1, λ^1) to (π^3, λ^3). The closer is λ^2 to λ^1, the more "vulnerable" is the

[12] There is a recent literature that attempts to reduce multiple equilibria to a unique equilibrium by introducing small individual heterogeneity. Examples of such works include Postlewaite and Xavier (1986), and Morris and Shin (1998). The idea of this literature is that when agents do not have common knowledge about the fundamental variables of the environment and instead choose their actions based on independent signals about these variables, the self-fulfilling property may fail. So far this literature has assumed that the agents are homogeneous except for the signals they receive.

market to early contracting rushes. Moderate levels of anxiety in the market can create self-sustaining momentum of early contracting.

What characteristics of the market make an early contracting rush more likely? In the upper panel of Figure 2, we can see that whether λ^2 is close to λ^1 depends on the position and shape of both the l and the p functions. If $\pi_{\min}$ is smaller so that the l function starts to decrease for small values of π, or if l decreases fast for small values of π, then λ^2 is closer to λ^1. From Figure 1, we find that these conditions obtain if the insurance gains from early contracting are large for promising worker types, which in turn occurs if workers and firms are highly risk-averse, if many workers are highly promising, or if the prospect of the firms is good. The position and shape of p also matter. If the p function shifts to the right, or if p increases fast for small values of λ_0 (as the critical worker type λ_0 decreases), then λ^2 is closer to λ^1. From the definition of p, an overall decrease in qualities of workers or an increase in firms' prospect μ will increase the probability that qualified workers are short in the spot market for any λ_0, and cause p to shift to the right. On the other hand, recall that the derivative of the function p with respect to λ_0 is $-h(e(\lambda_0))f(\lambda_0)(\lambda_0-\mu)$. So p increases fast as the critical worker type λ_0 decreases from λ^1, if the density h is great for values of shock x around $\overline{\lambda}-n\mu$, the density f is great for promising worker types, or the prospect of the firms μ is low. The effect of μ is therefore ambiguous, but the following factors unambiguously contribute to vulnerability of the market: highly risk-averse workers and firms, a polarized distribution of worker qualities (i.e., a great number of highly promising workers for a fixed average quality $\overline{\lambda}$), and significant aggregate uncertainty concentrated around the initial stages of early contracting.

The above analyses of stability and comparative statics may be used to understand cross-market differences in the extent of early contracting. Part of the reason that the market for Ph.D. economics graduates has been fairly immune to early offers may be the absence of a significant number of highly promising candidates that have established themselves early in the Ph.D. programs. A few early "superstars" in each recruitment season are not sufficient to generate the kind of self-fulfilling competitive process that spreads to any significant portion of the market. In contrast, if markets such as the one for legal clerks are characterized by relative homogeneity and concentration of applicants near the top of the ranking, our analysis suggests that the situation of no early offers can be an equilibrium but it can be vulnerable to the market sentiment because it is close to an unstable equilibrium. For the same reason, reforms in such markets may be initially successful in containing early offers if they lead to an equilibrium situation, but can unravel quickly if the equilibrium itself is vulnerable to early contracting rushes.

4.2. Welfare analysis

Reforms and other concerted efforts in controlling the practice of early contracting in some markets raise the theoretical issue concerning the welfare implications of early contracting. Questions about the welfare effects of banning the practice of early contracting have been addressed in Li and Rosen (1998) and Li and Suen (2000). The present model shares the basic welfare trade-off in the two earlier papers: early contracting increases the chance of mismatch, but provides insurance gains to risk-averse agents. Existence of multiple early contracting equilibria, however, poses new questions for welfare analysis: Can equilibria be Pareto-ranked? If not, how do the two sides of the market fare in the different equilibria?

Consider first incomplete early contracting equilibria (including equilibria where there is no early contracting). Compare two such equilibria with different spot market tightness π^*. Since firms are identical and not all firms can successfully enter early matches, early wage offers adjust to ensure that all firms are indifferent between waiting for the spot market and making early deals with workers above the threshold type. This implies that all firms are worse off in the incomplete early contracting equilibrium with a greater π^*. For workers who are below the critical type in both equilibria and who wait for the spot market, their equilibrium payoff is higher in the equilibrium with a greater π^*. For workers who are above the critical type in both equilibria and who sign early contracts, the equilibrium early wage offer schedule $r^*(\lambda)$ shifts up with π^*, so they are also better off in the equilibrium with a greater π^*. Finally, for workers who switch from waiting in one equilibrium to early contracting in the other, the welfare comparison depends on whether the equilibrium with more extensive early contracting (lower critical worker type λ_0) has a greater π^*. Suppose that the equilibrium with more early contracting has a higher π^* (see the upper panel of Figure 2). Then workers who switch from waiting to early contracting become better off, because their waiting option becomes more attractive and they capture the insurance benefits from early contracting. Note that in this case, all workers agree on the preference between the two equilibria and they all have the opposite preference as the firms. In the other case, the equilibrium with more early contracting may have a lower π^* (this can happen in the lower panel of Figure 2 if the p function moves sufficiently to the right.) Then, among the workers who switch from waiting to early contracting, the lower types tend to be worse off because their waiting option becomes less attractive and the insurance benefits from early contracting for these types are small. The higher types tend to be better off in spite of lower payoffs from their waiting option, because they are able to capture more insurance benefits from early contracting.

Welfare implications of complete early contracting equilibria can be similarly addressed, depending on whether complete early contracting increases or decreases the spot market tightness π^*. If more early contracting increases π^* (the function p decreases for the relevant values of λ_0), then it hurts firms and benefits all workers. This is the case when workers have promising prospects in the early market ($\overline{\lambda}$ is high relative to μ), so that more early contracting means that many promising workers enter early matches, which reduces the relative demand for jobs in the spot market. If more early contracting decreases π^* (the function p increases for the relevant values of λ_0), then it benefits all firms and possibly some of the workers but hurts most of the latter. This is the case when firms are relatively more promising (μ is high relative to $\overline{\lambda}$) or when early contracting is extensive. When more workers of relatively low types enter early matches, the demand for jobs relative to the supply of positions becomes greater in the subsequent period. This bids down the early wage offers, hurting most of the workers while benefiting all firms.

The above welfare analysis implies that firms have a collective interest in limiting the extent of early contracting when early offers are mainly going to the applicants that show greater promises than firms. This is consistent with the evidence documented in the works by Alvin Roth and his co-authors (Mongell and Roth, 1991; Roth, 1991; Roth and Xing, 1994) that in many market it is often the firms that initiate reforms to control the timing of offers. In this sense, our result resolves the puzzling conclusion in Li and Rosen (1998) that banning early contracting should hurt all firms. Li and Rosen's (1998) conclusion arises from the assumption of one-sided individual uncertainty: with μ equal to 1 and greater than any λ, a smaller λ_0 always leads to a smaller π and therefore makes firms better off. This counter factual conclusion is avoided in our model. The assumption that $\mu < \lambda_{\max}$ implies that, at least initially, more extensive early contracting increases the chance that productive workers will be on the short side of the spot market, and therefore is not beneficial to firms.

5. Extensions

In this section we extend the multiple equilibria result in two directions. First, we allow wages to be rigid in either the early market of period one or the spot market of period two. Second, we allow firms as well as workers to be heterogeneous in having different prospects of becoming productive. These two extensions demonstrate that our multiple equilibria result is largely robust, and that our theoretical framework is quite versatile.

5.1. Fixed-wage contracts

We have assumed that wages are competitively determined in both the early market and the spot market. In this section we discuss two situations of fixed-wage contracts: wage rigidity in the early market, and wage rigidity in the spot market. The first situation may arise when participants attempt to regulate the market, as it can be easier to control the terms of contracts than to control the timing of offers. The second situation may arise when firms try to use incentives such as signing bonuses in early offers in a market where wages in the spot market are fixed. For simplicity, we consider only incomplete early contracting equilibria.

With wage rigidity in the early market only, equilibrium in the spot market continues to be determined by which side of the market turns out to be on the short side. With probability π productive workers are short and receive 1, and with probability $1 - \pi$ they are long and receive 0. Unproductive workers and firms always receive 0. Let s be the fixed share received by workers in early contracts, regardless of type. The firms' share is

$1 - s$. Given any π, early contracting is beneficial between a firm and a type-λ if

$$\mu u(s) \geq \pi;$$
$$\lambda v(1-s) \geq 1-\pi.$$

With fixed wage s in the early contracts, the ordering property continues to hold: if early contracting is beneficial for type λ workers, then it is also beneficial for workers of higher types.

As before, let $l(\pi)$ be the function that defines the cutoff worker type as a function of π. This function differs from what we defined in Section 3 because the wage offer s is now fixed in the early market. To derive $l(\pi)$, let $\underline{\pi}_s$ and $\overline{\pi}_s$ be the points of π where the constant function $r = s$ intersects the highest bid price function $r^f(\pi, \lambda_{\max})$ and the ask price function $r^w(\pi)$, respectively. That is,

$$r^f(\underline{\pi}_s, \lambda_{\max}) = s;$$
$$r^w(\overline{\pi}_s) = s.$$

Refer to Figure 1. In words, $\underline{\pi}_s$ and $\overline{\pi}_s$ are, respectively, the lowest and the highest value of π that allows early contracts at s between firms and some type of workers. We assume that

$$r^w(\pi_{\min}) < s < r^w(\pi_{\max}),$$

so that $\underline{\pi}_s < \overline{\pi}_s$; otherwise, early contracting at s is impossible. Then, the l function is given by:

$$l(\pi) \begin{cases} = \lambda_{\max}, & \text{if } \pi < \underline{\pi}_s \text{ or } \pi > \overline{\pi}_s; \\ \text{satisfies } r^f(\pi, l(\pi)) = s, & \text{if } \underline{\pi}_s \leq \pi \leq \overline{\pi}_s. \end{cases}$$

Thus, wage rigidity in the early market reduces the potential extent of early contracting. Compared to the l function defined in Section 3, we see that $[\underline{\pi}_s, \overline{\pi}_s] \subset [\pi_{\min}, \pi_{\max}]$. Even though gains from insurance exist for any $\pi \in [\pi_{\min}, \pi_{\max}]$, the fixed wage s can be either higher than the bid price for type $\lambda_{\max}$ worker (when $\pi < \underline{\pi}_s$), or lower than the ask price of all workers (when $\pi > \overline{\pi}_s$). Moreover, for $\pi \in [\underline{\pi}_s, \overline{\pi}_s]$, because r^f is increasing in both of its two arguments, the cutoff worker type $l(\pi)$ is a decreasing function of π in this range: a higher π implies more participants signing early contracts. Finally, there is a discontinuity of the l function at the point of $\pi = \overline{\pi}_s$: the extent of early contracting drops from the maximum to zero.

With the function l redefined as above, early contracting equilibria can be identified in the same way as before, as intersections between the l function and the p function. As we have seen, the l function is monotonic non-increasing (except for a point of upward discontinuity at $\pi = \overline{\pi}_s$). Nevertheless the p function remains the same as in Section 3 and can be non-monotone in λ_0. This means that multiple early contracting equilibria can still occur because of the feedback effect of early contracting on the subsequent spot market condition.

Now suppose that wage rigidity exists only in the spot market. Let s be the fixed share received by a productive worker, and $1-s$ the share received by the worker's productive matching partner. Unproductive workers and firms get nothing. Since wages are fixed, market-clearing in the spot market requires a rationing mechanism. Consider the following random matching mechanism. If productive workers are short, then all these workers receive s with probability 1 and productive firms receive $1-s$ with a probability equal to the ratio of the measure of productive workers to productive firms. If productive workers are long, then all productive firms receive $1-s$ with probability 1 and productive workers receive s with a probability equal to the ratio of the measure of productive firms to productive workers. Let α be the probability in period one that productive workers receive utility $u(s)$ in the spot market, and β the probability that productive firms receive utility $v(1-s)$. Given α and β, early contracting is beneficial for a firm and a type λ worker if there exists an early contract share r such that

$$\mu u(r) \geq \alpha u(s);$$

$$\lambda v(1-r) \geq \beta v(1-s).$$

With fixed wage s in the spot contracts, the ordering property continues to hold. If early contracting with type λ workers is mutually beneficial, so is it with workers of higher types as well. Therefore, if λ_0 represents the cutoff worker type, we have

$$\alpha(\lambda_0) = p(\lambda_0) + \int_{x_{\min}}^{e(\lambda_0)} \frac{(n-(1-F(\lambda_0))\mu + x}{\int_{\lambda_{\min}}^{\lambda_0} \lambda f(\lambda) d\lambda} dH(x);$$

$$\beta(\lambda_0) = 1 - p(\lambda_0) + \int_{e(\lambda_0)}^{x_{\max}} \frac{\int_{\lambda_{\min}}^{\lambda_0} \lambda f(\lambda) d\lambda}{(n-(1-F(\lambda_0))\mu + x} dH(x),$$

where $e(\lambda_0)$ is the excess supply of workers before the shock x is realized, and $p(\lambda_0)$ is the probability that productive workers will be short in period two, as defined in Section 3.

An early contracting equilibrium with wage rigidity in the spot market is then given by a cutoff worker type λ_0 and the corresponding early wage offer r_0 such that

$$\mu u(r_0) = \alpha(\lambda_0)u(s);$$

$$\lambda_0 v(1 - r_0) = \beta(\lambda_0)v(1 - s).$$

Compare the above two equations with the case of flexible wage contracts in Section 3. For any cutoff worker type λ_0, we have $\alpha(\lambda_0) > p(\lambda_0)$ and $\beta(\lambda_0) > 1 - p(\lambda_0)$. On one hand, wage rigidity in the spot market means that the downside of waiting to find a match in the spot market becomes less grim for both workers and firms, because being on the long side of the spot market is no longer associated with a zero payoff. This tends to reduce the insurance benefits of early contracts and make early offers less likely. On the other hand, since $u(s) < 1$ and $v(1 - s) < 1$, the upside of waiting to find a match in the spot market becomes less attractive as well. This has the opposite effect of encouraging participants to sign early contracts. Which of the two effects dominates depends on the fixed wage offer s. For extreme wage shares (s is close to 0 or 1), wage rigidity in the spot market discourages early offers and makes multiple equilibria less likely by reducing the bid prices too low or by raising the ask price too high. For more even shares ($u(s)$ is about the same as $v(1 - s)$), wage rigidity in the spot market can encourage early contract, and this is more likely to occur if both workers and firms are highly risk-averse.

5.2. Heterogeneous firms

To allow different types of firms, suppose that μ is distributed on $[\mu_{\min}, \mu_{\max}]$, with distribution function G and density g. We consider only incomplete early contracting equilibria, where some workers and firms wait for the spot market in the second period. The ask price r^w of a worker is now a function of the type of the matched firm. Let us write it as $r^w(\pi, \mu)$. Clearly, r^w decreases with μ. The bid price function $r^f(\pi, \lambda)$ is defined as before, which is an increasing function of λ for any π. It follows that for any π, and any $\lambda < \lambda'$ and $\mu < \mu'$, if $r^f(\pi, \lambda) \geq r^w(\pi, \mu)$ then $r^f(\pi, \lambda') \geq r^w(\pi, \mu')$. Thus, the ordering property continues to hold with heterogeneous firms.

Given any probability π that productive workers are short in the spot market, let λ_0 be the lowest type of workers and μ_0 be the lowest type of firms which sign early contracts. The ordering property implies that for any π, there can be at most one pair, λ_0 and μ_0, which satisfies

$$r^f(\pi, \lambda_0) = r^w(\pi, \mu_0);$$

$$1 - F(\lambda_0) = n(1 - G(\mu_0)).$$

The first condition says that given π, the ask price of type-λ_0 workers and the bid price of type-μ_0 firms are equal, and the second condition says that the measure of worker types higher than λ_0 and the measure of firm types higher than μ_0 are equal.[13] Define the function $\lambda_0 = l(\pi)$ such that the pair $(l(\pi), \mu_0)$ solves the above two equations for any π.

We now show that the two reasons for having multiple early contracting equilibria are preserved in the case of heterogeneous firms. First we characterize the dependence of the critical types λ_0 and μ_0 on π. From the two above conditions that define λ_0 and μ_0, it is straightforward to show that $dl/d\pi$ has the same sign as

$$\frac{\partial r^w(\pi, \mu_0)}{\partial \pi} - \frac{\partial r^f(\pi, \lambda_0)}{\partial \pi}.$$

Let a unique pair of $\hat{\lambda}$ and $\hat{\mu}$ be such that $r^w(\pi, \hat{\mu})$ is tangent to $r^f(\pi, \hat{\lambda})$, and $1 - F(\hat{\lambda}) = n(1 - G(\hat{\mu}))$, and denote $\hat{\pi}$ as the unique value of π at which the tangency occurs. Since r^w is convex in π and r^f is concave in π, we have $dl/d\pi > 0$ if $\pi > \hat{\pi}$, and $dl/d\pi < 0$ if $\pi < \hat{\pi}$. Therefore, as in the case of homogeneous firms, the function $l(\pi)$ has the same shape as in Figure 2 ($\pi_{\min}$ and $\pi_{\max}$ are now the two intersections of $r^w(\pi, \mu_{\max})$ and $r^f(\pi, \lambda_{\max})$.) We obtain non-monotonicity of the $l(\pi)$ function for the same reason as in the case of homogeneous firms. The uncertainty about the spot market, and hence the insurance benefits from early contracting, is the greatest when the spot market is neither too tight nor too slack.

[13] The ordering property implies not only that early contracting involves only workers and firms above their respective critical types, but also that these workers and firms are matched "positive assortatively," i.e., highest type worker with highest type firm, and so on (Becker, 1981). Using positive assortative matching, we can determine the equilibrium schedule $r(\lambda)$ of early wage offers, which is not necessary for our purpose of demonstrating the existence of multiple equilibria. The ordering property distinguishes the present model from Li and Suen (2000), and Suen (2000), where both firms and workers are heterogeneous and early contracting is not necessarily positive-assortative. This difference arises because in the present paper the market participants will be either productive or unproductive, so the trade-off between early contracting and waiting does not depend on their characteristics.

Next, we characterize the feedback effect of early contracting on π. Given any critical worker type λ_0 and corresponding type μ_0 of firms, the measure of productive workers in period two is $\int_{\lambda_{\min}}^{\lambda_0} \lambda f(\lambda) d\lambda$, while the measure of productive firms is $n \int_{\mu_{\min}}^{\mu_0} \mu g(\mu) d\mu + x$. Therefore, the probability that productive workers are on the short side of the market is equal to $p(\lambda_0) = 1 - H(e(\lambda_0))$, where

$$e(\lambda_0) = \left(\int_{\lambda_{\min}}^{\lambda_0} \lambda f(\lambda) d\lambda - n \int_{\mu_{\min}}^{\mu_0} \mu g(\mu) d\mu \right).$$

Using $1 - F(\lambda_0) = n(1 - G(\mu_0))$, we find that

$$\frac{dp(\lambda_0)}{d\lambda_0} = -h(e(\lambda_0)) f(\lambda_0)(\lambda_0 - \mu_0).$$

Thus, $p(\lambda_0)$ is decreasing for $\lambda_0 > \mu_0$ and is increasing for $\lambda_0 < \mu_0$. The feedback effect of early contracting on π is similar to the case of homogeneous firms, with μ_0 replacing the constant μ. Non-monotonicity of the $p(\lambda_0)$ function can arise under a variety of assumptions on the distributions F and G. The intuition is the same as before. If workers who sign early contracts have high probabilities of becoming productive relative to the firms they are matched with, then more early matches imply a greater chance that productive workers will be short in the spot market. Conversely, if the last workers in early matching have low probabilities of becoming productive relative to the firms they are matched with, then more early contracting reduces the chance that productive workers will be on the short side.

6. Conclusion

In labor markets for entry-level professionals and in other related markets, job applicants' concern for availability of positions and employers' concern for availability of qualified applicants can drive some participants on the two sides to sign early job contracts before qualifications of applicants are ascertained and employers' hiring needs are confirmed. In a two-period model with these individual uncertainties and an aggregate uncertainty about the balance of demand and supply in the second period market, we show that multiple equilibria of early contracting can arise, because participants' expectation of the balance of demand and supply in the spot market has non-monotone effects on their decisions to contract early, or because individual decisions to sign early contracts have non-monotone effects on the balance of demand and supply. Comparative statics analysis shows that early contracting rushes are more likely to occur if workers and firms are highly risk-averse, if the distribution of worker qualities is polarized, or if there is significant aggregate uncertainty concentrated around the initial stages of early contracting. We show that early contracting equilibria cannot be Pareto-ranked. More early contracting hurts firms and benefits workers when early contracting is not extensive and the workers who sign up early are relatively promising candidates; the opposite is true when early contracting is so extensive that even workers with not so good prospects sign up early. Much of our analysis is shown to extend to the case of fixed wage contracts in either the early market or in the spot market, and to the case of heterogeneous firms.

References

Avery, Christopher, Christine Jolls, Richard Posner, and Alvin Roth. "The Market for Federal Judicial Law Clerks." *University of Chicago Law Review* 68 (Summer 2001): 793–902.

Becker, Gary. *A Treatise on the Family.* Cambridge, Mass.: Harvard University Press, 1981.

Diamond, Douglas W. and Philip H. Dybvig. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (June 1983): 401–19.

Henderson, James M. and Richard E. Quandt. *Microeconomic Theory: A Mathematical Approach*, 3d. ed. New York: McGraw Hill, 1980.

Li, Hao and Sherwin Rosen. "Unraveling in Matching Markets." *American Economic Review* 88 (June 1998): 371–87.

Li, Hao and Wing Suen. "Risk-sharing, Sorting, and Early Contracting." *Journal of Political Economy* 108 (October 2000): 1058–91.

Mongell, Susan and Alvin E. Roth. "Sorority Rush as a Two-sided Matching Mechanism." *American Economic Review* 81 (June 1991): 441–64.

Morris, Stephen and Hyun Song Shin. "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks." *American Economic Review* 88 (June 1998): 587–97.

Postlewaite, Andrew and Xavier Vives. "Bank Runs as an Equilibrium Phenomenon." *Journal of Political Economy* 95 (June 1987): 485–91.

Roth, Alvin. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy* 92 (December 1984): 991–1016.

Roth, Alvin. "A Natural Experiment in the Organization of Entry-Level Labor Markets: Regional Markets for New Physicians and Surgeons in the United Kingdom." *American Economic Review* 81 (June 1991): 415–40.

Roth, Alvin and Xiaolin Xing. "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions." *American Economic Review* 84 (September 1994): 992–1044.

Suen, Wing. "A Competitive Theory of Equilibrium and Disequilibrium Unraveling in Two-Sided Matching." *Rand Journal of Economics* 31 (Spring, 2000): 101–20.

Wald, Patricia. "Selecting Law Clerks." *Michigan Law Review* 89 (October 1990): 152–63.