

Contractual Incompleteness, Unemployment, and Labour Market Segmentation

STEFFEN ALTMANN

Institute for the Study of Labor (IZA) and University of Bonn

ARMIN FALK

University of Bonn and IZA

ANDREAS GRUNEWALD

University of Bonn

and

DAVID HUFFMAN

University of Oxford and IZA

First version received August 2011; final version accepted June 2013 (Eds.)

This article provides evidence that involuntary unemployment, and the segmentation of labour markets into firms offering “good” and “bad” jobs, may both arise as a consequence of contractual incompleteness. We provide a simple model that illustrates how unemployment and market segmentation may jointly emerge as part of a market equilibrium in environments where work effort is not third-party verifiable. Using experimental labour markets that differ only in the verifiability of effort, we demonstrate empirically that contractual incompleteness can cause unemployment and segmentation. Our data are also consistent with the key channels through which the model explains the emergence of both phenomena.

Key words: Contractual incompleteness, Incentives, Unemployment, Dual Labour Markets, Laboratory experiment.

JEL Codes: C91, J41, J64, M52, M55

1. INTRODUCTION

This article provides evidence that two important features of labour markets—the emergence of involuntary unemployment, and the segmentation of markets into firms offering “good” and “bad” jobs to apparently similar workers—may have a common underlying cause. In particular, both phenomena may arise jointly when employment relationships are characterized by contractual incompleteness, in the sense that work effort is not verifiable to third parties. We also provide evidence supporting a specific set of mechanisms for how contractual incompleteness can cause these two phenomena, in which there is a key role for the implicit incentive strategies adopted by firms.

Intuitively, when effort is not verifiable, firms may adopt an implicit incentive strategy for eliciting high work effort that involves paying relatively high wages, and conditionally renewing workers' contracts based on their performance. Under decreasing returns to scale, however, high wage payments can make it profitable for firms to hire fewer workers than technologically feasible, since the gains from higher overall production might be more than offset by higher wage costs. The result of such job rationing is endogenous involuntary unemployment. If a critical mass of firms rations jobs, however, a secondary employment sector may emerge simultaneously, where firms profitably fill all vacancies and pay relatively low wages. Such firms are able to pay lower rents, and elicit relatively lower but non-minimal effort, because of the unemployment pressure in the market. This might give rise to a segmented labour market in which the strategies of offering "good" high-rent jobs and "bad" low-rent jobs are, in equilibrium, equally profitable for firms. We show that the qualitative features of this intuition can be captured in a simple formal model, which builds on the classic efficiency wage frameworks by Shapiro and Stiglitz (1984) and Akerlof and Yellen (1990).

In order to provide empirical evidence on the causal impact of contractual incompleteness on unemployment and market segmentation, we study behaviour of firms and workers in competitive experimental labour markets. All firms in our markets share the same production technology which exhibits decreasing returns to scale from labour. In our main treatment, the Incomplete Contracts treatment (*IC treatment*), a worker's effort is observable to firms but not verifiable to third parties. Firms may, however, use implicit incentives to elicit non-minimal work effort in this treatment. In a control treatment, the Complete Contracts treatment (*C treatment*), work effort is verifiable and contracts are explicitly enforced. In all other respects, such as production technology or the number of market participants, the two treatment conditions are identical.

Our first main empirical result is that contractual incompleteness causes a strong increase in the level of unemployment. Our data also reveal important differences in how labour markets function in the presence or absence of explicit contract enforcement. In line with the hypothesized mechanisms, we find that firms in the IC treatment use implicit incentives involving conditional contract renewal and paying strictly positive worker rents. At the same time that they pay high wages, however, some firms choose to offer fewer jobs than possible. Endogenous unemployment in the IC treatment arises as a byproduct of this job rationing decision. Given that employed workers earn substantial rents, unemployment in the IC treatment is involuntary. In the C treatment where effort is explicitly enforced, labour market outcomes differ substantially along all these dimensions. Firms pay much lower wages and reap the major share of production surplus. Employment relations are shorter than in the IC treatment, and the overwhelming majority of firms does not ration jobs. As a result, endogenous unemployment in this treatment is very low and mostly voluntary, being caused by workers who do not accept existing contract offers. We also find support for the underlying mechanism that is hypothesized to drive these treatment differences, namely an impact of contractual incompleteness on the profitability of different contractual instruments: paying positive rents and using contingent contract renewal increases firm profits in the IC treatment, while being detrimental or irrelevant for firms in the C treatment.

Our second main empirical finding is that contractual incompleteness leads to a stable coexistence of different job types. After an initial phase in which we observe a trend towards job rationing in the IC treatment, a plateau is reached such that unemployment stabilizes at a high level, and a relatively constant fraction of firms continues to operate without rationing job offers. Whereas in the initial phase job rationing is the more profitable strategy for firms, in this later phase firms earn similar profits regardless of whether or not they ration jobs. Workers, however, earn substantially lower rents and exert lower effort in firms that do not ration jobs. In the long run, the situation in the IC treatment thus resembles a segmented labour market in

which some workers are employed in “primary-sector” jobs characterized by high worker rents, relatively stable employment relationships, and job rationing, while other workers work under less favourable conditions in “secondary-sector” jobs (Doeringer and Piore, 1971; Saint-Paul, 1996).¹ Given the equal profitability of these alternative firm strategies, segmentation has the character of a market equilibrium. By contrast, market segmentation is not observed in the C treatment where firms’ strategy of not rationing jobs and paying very low worker rents pervades the market. This indicates that the emergence of market segmentation is caused by contractual incompleteness. We also find support for a key mechanism hypothesized to sustain segmentation, which is an impact of unemployment pressure on worker behaviour: workers in the IC treatment are significantly less likely to shirk when reduced market activity indicates lower job finding chances.

The first important contribution of our paper lies in empirically identifying a direct causal link between contractual incompleteness and involuntary unemployment. While efficiency-wage theory has long hypothesized that this link may exist (Shapiro and Stiglitz, 1984; MacLeod and Malcolmson, 1989; Akerlof and Yellen, 1990), establishing this key claim of the theory is difficult if not impossible using field data. The empirical literature has made important contributions on other aspects of the efficiency-wage hypothesis, particularly the relationship between rents and worker performance (for a survey see Katz, 1986). However, evidence on key variables like work effort has necessarily been indirect, because effort is inherently difficult to measure in settings where efficiency wages would be relevant. As a solution, previous field studies have related indirect proxies for effort, such as worker discipline problems or survey measures of workplace performance, to wage premiums (*e.g.* Cappelli and Chauvin, 1991; Campbell III and Kamlani, 1997), or to measures of dismissal barriers and firing threat (*e.g.* Ichino and Riphahn, 2005). An experimental approach is complementary to these studies because of the possibility to exogenously vary the degree of contractual incompleteness and to directly measure its impact on involuntary unemployment.² In our setup we can also induce or accurately measure key variables such as worker effort and ability, or firms’ production technology. Thus, we are able to precisely assess whether contractual incompleteness influences worker rents or decisions on job rationing and contract acceptance, as hypothesized in the theory.

Our second main contribution concerns understanding the foundations of dual labour markets. The theoretical literature on dual labour markets has traditionally argued that market segmentation can result from contract enforcement problems and efficiency wages, if monitoring technologies differ exogenously across segments (Bulow and Summers, 1986). Some theoretical approaches have made segmentation endogenous, arising due to non-linearities in monitoring technology, differences in setup and adjustment costs, or on-the-job search (*e.g.* Albrecht and Vroman, 1992; Saint-Paul, 1996; Board and Meyer-ter-Vehn, 2013). We provide further theoretical insights on how market segmentation can emerge endogenously despite homogeneous technology across firms, when the production technology exhibits decreasing returns to scale. More importantly, we provide the first empirical evidence on the endogenous emergence of market segmentation due to contractual incompleteness. Some earlier experimental papers have observed firms offering different types of jobs which are more and less attractive from workers’ perspective (Brown *et al.*,

1. Note that secondary-sector workers in our setup also earn efficiency wages, but the wages and corresponding worker rents are lower than those in primary-sector jobs. This differs from some of the traditional models of dual labour markets (*e.g.* Bulow and Summers, 1986), in which secondary-sector jobs are—by assumption—associated with explicit contract enforcement and thus do not involve efficiency wages. Our results show how, holding contractual incompleteness constant, there can emerge “good” and “bad” jobs in terms of similar workers earning systematically higher and lower rents, respectively.

2. For a general discussion of the role of lab experiments in labour economics see, *e.g.* Charness and Kuhn (2011).

2004; Bartling *et al.*, 2012). Importantly, however, in these studies jobs that are good and bad for workers also exhibit strong and systematic differences in firm profits. In contrast, we find that firms who use the good-job and bad-job strategy are equally profitable, in line with the notion that market segmentation can be supported as a stable equilibrium outcome.

A number of other papers have used experimental techniques to study the consequences of contractual incompleteness in labour market settings (*e.g.* Fehr *et al.*, 1993; Brown *et al.*, 2004; Brown *et al.*, 2012; Linardi and Camerer, 2012). Our study differs from this literature in that it incorporates an analysis of endogenously arising unemployment. By contrast, unemployment in earlier papers was exogenously given, ruled out by the design of the experiment, or determined by exogenous stochastic shocks. Previous studies have shown that contractual incompleteness can lead to an adoption of implicit incentive strategies that involve rent payments and contingent contract renewal (*e.g.* Brown *et al.*, 2004).³ We add a missing dimension to this literature, showing how contractual incompleteness can also affect aggregate-level market outcomes. Our findings thus provide a missing empirical link, illustrating how contractual incompleteness, the use of implicit incentives, unemployment, and market segmentation can all be intimately related.

In the remainder of this article, we first present the setup and procedures of the experiment, before discussing theoretical hypotheses in Section 3. In Section 4, we present our empirical results, and Section 5 concludes.

2. DESIGN AND PROCEDURES OF THE EXPERIMENT

To study the impact of contractual incompleteness on unemployment and market segmentation, we implemented experimental labour markets where we exogenously varied the verifiability of work effort. As our workhorse, we used a variant of the gift-exchange game (Fehr *et al.*, 1993). Firms and workers in a given market interacted for 18 periods. Each period consisted of a market phase in which firms offered employment contracts and hired workers, and a work phase in which work effort was determined. In our main treatment, the IC treatment, effort was not verifiable and workers thus could depart from the contractually stipulated effort level. By contrast, the effort level specified in the employment contract was explicitly enforced in our control treatment, the C treatment. Varying the verifiability of work effort while keeping everything else identical (production technology, supply and demand of labour, etc.) allows us to causally identify the effects of contractual incompleteness on labour market outcomes.

2.1. *The market phase*

Firms were the contract makers in the market phase. To offer a contract, firms stipulated a non-contingent (upfront) wage payment, w , and a desired level of effort, \hat{e} . Firms could make two types of contract offers: public offers that were available to all workers and could also be observed by all other firms, or private offers that were only available to one specific worker. Public offers allowed firms to reach the entire market if they wanted to fill a vacancy regardless of a particular trading partner. Private offers made it possible for firms to target specific workers. This is a necessary feature if firms want to apply a strategy that involves systematic rehiring of workers based on their previous performance. If an employer wanted to (re)hire a specific worker via a private contract offer, she had to specify the ID of the worker in the contract offer. In this case,

3. For recent surveys on lab and field experiments that study contractual and non-contractual solutions to alleviate moral hazard in the labour market see Charness and Kuhn (2011) and Bandiera *et al.* (2011). See Brown and Zehnder (2007) and Brown and Serra-Garcia (2012) for applications in credit markets.

only the selected worker was informed about the contract offer, and only this worker could accept the offer.

In a given market period, each firm could hire up to two workers. As long as none of her contract offers had been accepted, a firm could make as many private and public offers as she wanted. A worker could accept any public contract offer available in the market, and any private offer he had received. Workers were not informed about the number of private offers in the market as a whole, but they could infer labour market conditions and the tightness of the market from the number of publicly posted vacancies in a given period. Once a worker accepted a contract offer, he was not allowed to accept further offers in this period. Additionally, all other outstanding offers of the respective firm were removed from the list of available contracts. The firm could then decide to open a second vacancy and hire another worker, by entering new contract offers. This feature of opening first and second vacancies sequentially was implemented to prevent “accidental hiring”, such that a firm who wanted to employ only one worker but entered multiple contract offers could not have a second offer accepted before being able to withdraw her remaining contract offers. Note that while firms and workers could endogenously build up long-term employment relationships by repeatedly offering and agreeing on (private) contract offers, it was not possible for market participants to directly announce or sign a multi-period employment contract.

The market phase ended when all firms had filled both vacancies, or when all firms had indicated that they did not want to post further vacancies.⁴ During the subsequent work phase, workers received a summary of their own contract terms, and information on whether and under which conditions their firm had employed a second worker. This information was mainly provided to ensure common knowledge within a firm on whether the firm operated as a one-worker or two-worker firm. While complete information regarding co-workers’ wages may not be fully realistic, some degree of transparency is likely to be present in many work settings: even with a firm policy encouraging wage secrecy, workers may have a reasonably accurate idea about co-workers’ earnings.⁵

2.2. *The work phase*

After the end of the market phase, workers who had accepted a contract entered the work phase in which actual work effort, e , was determined. Since effort was contractible in the C treatment, effort levels corresponding to the contractually stipulated ones were exogenously implemented by the experimenter ($e = \hat{e}$). This corresponds to an environment in which the penalties for breaching a contract are sufficiently high to make it optimal for workers to comply with the contractually stipulated effort levels. By contrast, work effort was observable by the firm, but not verifiable to third parties in the IC treatment. Therefore, a worker could exert equal, less, or more effort than stipulated in his employment contract. Workers’ effort choices, together with firms’ wage payments, determined material payoffs of firms and workers. Before the next period started, a firm and its worker(s) were informed about work efforts and the resulting payoffs for the firm and the workers employed by this firm.

4. We also had a maximum trading time of 200 seconds for each market phase. This constraint was, however, only binding in few occasions (mostly in the C treatment). The impact of the time limit on unemployment and other market outcomes reported below is thus limited, and confined to the control treatment.

5. Empirically, co-worker wages have no significant impact on workers’ effort choices in our setup. Efforts in the IC treatment strongly depend on a worker’s own contract terms, but they are not significantly related to the wage or the desired effort level of the co-worker (results can be found in Table A.1 in the Supplementary Material). The finding that wage inequalities per se might not affect behaviour is in line with recent evidence on social comparison processes in similar setups (Charness and Kuhn, 2007; Abeler *et al.*, 2010; Gächter *et al.*, 2012).

TABLE 1
Schedule of effort costs

Effort level e	1	2	3	4	5	6	7	8	9	10
Cost of effort $c(e)$	0	1	2	4	6	8	10	12	15	18

2.3. Parameters and procedures

Participants' roles were randomly assigned at the beginning of the experiment and kept constant throughout all market periods. In every market, we had 17 workers and 7 firms. Since firms could employ at most two workers, this implies that three workers were "exogenously" unemployed in each period.

A worker's payoff in a given period, π_W , was given by

$$\pi_W = \begin{cases} w - c(e) & \text{if worker accepted a contract } [w, \hat{e}] \\ 0 & \text{if unemployed} \end{cases}$$

A worker who remained unemployed in a given period received a payoff of 0 points. An employed worker received the wage w specified in his contract and had to bear the cost of the work effort he provided, $c(e)$. The set of feasible efforts and wages was given by $e \in \{1, 2, \dots, 10\}$ and $w \in \{0, 1, 2, \dots, 100\}$. As illustrated in Table 1, we induced a convex (monetary) effort cost schedule in the experiment.

A firm's profit depended on the number of workers hired, the wage(s) paid, and the effort exerted by the worker(s). Firms' production technology was characterized by decreasing returns to scale. Decreasing returns are often argued to arise with increases in firm size, for instance due to higher bureaucratic or coordination costs in larger organizations. We conceptualized this as a reduction in workers' productivity if a firm hired two workers. Specifically, each unit of effort by a worker increased output (and the firm's payoff) by 10 points if only one worker was employed by the firm. If two workers were employed, each unit of effort increased the firm's payoff by 7 points. This corresponds, for instance, to a work environment where workers in larger firms need to spend some of their time doing administrative tasks that are not directly productive. The payoff of a firm, π_F , can thus be summarized as follows:

$$\pi_F = \begin{cases} 10e_1 - w_1 & \text{if one worker employed} \\ 7(e_1 + e_2) - w_1 - w_2 & \text{if two workers employed} \\ 0 & \text{else} \end{cases}$$

e_1 (e_2) denotes the effort provided by the first (second) worker, and w_1 (w_2) is the wage paid to the first (second) worker employed by the firm. Note that this specification of the production technology implies that overall surplus is maximized when two workers are employed and maximum effort is exerted. Payoff functions π_F and π_W , workers' cost schedule $c(e)$, and the number of firms and workers in the market were common knowledge.

The experiment was carried out in the BonnEconLab at the University of Bonn. A total of 240 subjects, mainly university students from all majors, took part in the experiments.⁶ Every subject participated only in one session, and we conducted five independent market sessions for each treatment. At the beginning of a session, participants received a short verbal summary and

6. For a methodological discussion on the use of student subjects in economics research see Gächter (2010).

detailed written information on the rules of the experiment.⁷ The experiment started only after all participants had answered several control questions correctly. Before the start of the experiment, the key features of the experiment rules were summarized again by the experimenter, and subjects played one trial period of the market phase to ensure that they understood how to use the computer program. At the end of a session, all earnings from the experiment were paid out in cash at an exchange rate of 25 points = 1 euro. Sessions lasted about 110 minutes, and subjects earned on average 25.49 euros (about 35 USD at the time of the experiment), including a showup fee of 8 euros. The experiments were computerized using the software “z-Tree” (Fischbacher, 2007); subjects were recruited with the online recruitment system by Greiner (2003).

3. BEHAVIOURAL PREDICTIONS

The treatments described in the previous section allow us to identify the causal impact of contractual incompleteness on unemployment and labour market segmentation in our setting. Furthermore, using the data from the experiment we can investigate whether both phenomena arise in a way that is consistent with a certain theoretical equilibrium. A simple model, which is derived in the theoretical appendix, informs our hypotheses. In the model, we show how unemployment and segmentation can be part of a market equilibrium when effort is non-verifiable, and how these aggregate-level outcomes arise along with a very specific set of strategic behaviours by workers and firms. These individual-level mechanisms become additional qualitative predictions, which should be satisfied empirically if unemployment and segmentation are to be explained by the type of equilibrium formalized in the model.

Our model builds on two important strands of efficiency-wage theory. One is the “shirking version” of Shapiro and Stiglitz (1984), in which materially selfish agents are motivated to work by the prospect of earning future rents, and a threat of being fired in case of shirking. The other is the “gift-exchange version” of Akerlof and Yellen (1990), in which workers are fair-minded, in the sense of experiencing a psychological benefit or cost of fulfilling a contract, depending on the generosity of the rents offered in their current contract. Our model incorporates both motivations: a fraction of agents is assumed to have fairness concerns, while the rest is materially selfish.⁸ This assumption is in line with abundant previous evidence from the lab and field, showing that a mix of selfish and fair types is typically present in a given population (see, *e.g.* Fehr *et al.*, 1993; Bewley, 1999; Fehr and Gächter, 2000; Cohn *et al.*, 2012). To match the experimental setting, our model features a finite horizon.⁹

3.1. *Non-verifiable effort*

The presence of both fair and selfish types has important implications for the type of market equilibria that can emerge in finite-horizon settings where effort is non-verifiable, as is the case

7. A translation of the instructions can be found in the Supplementary Material. To rule out that differences in participants’ experiences from their employment relationships outside the lab could influence our results, instructions were framed in a neutral goods-market language.

8. We adopt a simple approach to incorporating such motivations: The more “kind” or generous is the current contract in terms of the level of offered rents, the greater is the psychological motivation for a fair agent to fulfil the contract. This captures the key feature of all reciprocity-based fairness models, that fair types reward kind actions and punish unkind actions (*e.g.* Rabin, 1993; Falk and Fischbacher, 2006).

9. For a more abstract model that is tied less directly to the specifics of the experimental design, see the working paper version of our article, in which we show in an infinite-horizon setting how unemployment and segmentation can be part of a stationary labour market equilibrium under contractual incompleteness (see Altmann *et al.*, 2011).

in our IC treatment. Intuitively, the presence of some fair types may lead firms to pay non-minimal wages in the final period, anticipating that fair workers voluntarily provide non-minimal effort in response to fair wages. This generates a rent from being employed in the final period. The prospect of earning this final-period rent, in turn, opens up possibilities for equilibria in which firms use implicit incentives to motivate agents in the pre-final period(s). There can thus exist equilibria in the IC treatment that involve firms paying rents and conditioning contract renewal on workers' previous performance; in pre-final periods of such equilibria, selfish and fair types pool and fulfill their contracts in order to qualify for reemployment and earning future rents.

Our model shows how—in an environment where effort is non-verifiable—there exist equilibria in which involuntary unemployment can arise as a byproduct of the implicit-incentive strategies adopted by firms. Intuitively, starting from a situation where unemployment is low, the rents needed to deter shirking may be rather high because shirkers who are fired can relatively easily find another job. If a firm adopts a strategy of paying high wages and worker rents, however, decreasing returns to scale can make it profitable to “stay small”. If wages are high, the absolute increase in wage costs can be larger than the gain in terms of extra output that results from hiring more workers.¹⁰ It can thus be profitable to hire fewer workers than technologically feasible. Endogenous unemployment arises as a byproduct of such job rationing, and since firms pay positive rents, this unemployment is involuntary. The model identifies several key components that should be observed in the IC treatment if such an equilibrium arises. First, the profit-maximizing strategy of firms should involve paying strictly positive worker rents and conditional contract renewal, such that workers who shirk have a higher likelihood of being dismissed than workers who provide the contractually stipulated effort level. Second, from a worker perspective, the long-run costs of shirking in terms of forgone future rents should more than offset the short-run gains due to lower effort costs. Third, unemployment should emerge due to firms deciding to hire fewer workers than possible, rather than workers rejecting available job offers.

The model also shows how an equilibrium with non-verifiable effort can involve market segmentation, emerging jointly with involuntary unemployment. Intuitively, while some firms may adopt the strategy of eliciting high effort with high rents and rationing jobs (“good jobs”), the resulting unemployment reduces the job-finding chances for unemployed workers, and therefore the attractiveness of shirking for those employed. If the unemployment pressure in the market is strong enough, a fraction of firms might be able to operate simultaneously and equally profitably by hiring two workers, and offering jobs involving lower worker rents, and somewhat lower effort levels (“bad jobs”). Importantly, without sufficient unemployment pressure, the lower rents in such secondary-sector jobs might not be able to prevent shirking, and offering bad jobs would thus not be profitable for firms. In the Appendix, we derive sufficient conditions for segmentation to be part of a stable market equilibrium, such that the strategies of offering good jobs and bad jobs are equally profitable for firms, and neither type of firm has an incentive to deviate. If a segmentation equilibrium arises in our experiment, the model thus predicts that one-worker firms and two-worker should coexist in the market, and that firms in both segments are equally profitable. Furthermore, the firms who ration jobs should be the ones offering high wages and worker rents, and workers in such firms should exert high effort levels. In firms who hire two workers, wages and offered worker rents are predicted to be lower, as are worker effort levels.

10. For example, with the production function used in our experiment, a firm that wants to elicit an effort level of 10 is better off hiring only one worker if the wage needed to elicit this effort level is higher than 40 ($10 \cdot 10 - 41 = 59 > 7 \cdot 10 + 7 \cdot 10 - 41 - 41 = 58$).

3.2. *Verifiable effort*

The mechanisms we have described as potential sources of unemployment and market segmentation do not apply in our control treatment where effort is verifiable. In the Appendix we discuss equilibrium characteristics for an environment with explicit contract enforcement, as is the case in the C treatment. Holding constant other assumptions besides the verifiability of work effort, the equilibrium is characterized by all firms hiring two workers and thus zero endogenous unemployment. Firms elicit maximum effort from workers, while paying wages just slightly above worker effort costs. Because effort is explicitly enforced, there is also no need to engage in repeated contract renewal with specific workers. Since all firms use a homogenous strategy of offering jobs with close to zero rents, there is no market segmentation.

Intuitively, firms can elicit high effort without paying high rents or using conditional contract renewal because contractibility of effort eliminates a source of worker bargaining power: not only are firms on the short side of the market, but workers have no option to shirk once they are employed. The presence of fair types has relatively minor implications for behaviour in an environment with explicit contract enforcement, as workers cannot shirk on the job.¹¹ Firms strictly prefer hiring two workers, because this is profitable given that wage and rent payments are low. We thus hypothesize that firms in the C treatment use a strategy of paying worker rents close to the minimal possible level, and filling both vacancies. Furthermore, repeated rehiring of specific workers should be irrelevant for firm profits, and firms are thus predicted to rely more heavily on public contract offers in the C treatment. When a firm wants to fill a vacancy without concern for who is hired, public offers have the advantage that they reach the entire market.

Our discussion has illustrated how involuntary unemployment and market segmentation can be part of a market equilibrium when effort is not verifiable. It is well known, however, that there are typically multiple possible equilibria in repeated games (see, *e.g.* Fudenberg and Maskin, 1986), and this is true as well in our setup. For instance, there is a range of different equilibria for our IC setting, some of which involve endogenous unemployment and segmentation, but others do not. Alternative equilibria may take forms such as all firms hiring two workers, resulting in no endogenous unemployment, or all firms hiring only one worker; both of these cases exhibit no segmentation, since all jobs are homogenous. Our model illustrates how the equilibrium characteristics depend on the job separation rate between firms and non-shirking workers on which players endogenously coordinate in equilibrium.¹² According to our model, segmentation equilibria in environments with excess supply of labour always involve positive equilibrium separation rates, whereas segmentation is not possible with a separation rate of zero.¹³ While equilibrium multiplicity limits the predictive power

11. Fairness could play a role in the decision of whether or not to reject a contract offer. Theories of fairness, and ample empirical evidence, however, indicate that a key motive underlying fairness is a desire to punish unfair actions by an opponent. In a setting like the C treatment, where receivers of offers face competition, rejection of contract offers has limited effectiveness as a punishment strategy, and even offers involving low worker rents are thus likely to be accepted (see, *e.g.* Fehr and Schmidt, 1999).

12. Note that the equilibrium can involve a positive firing probability for non-shirking workers; a no-shirking equilibrium requires that firing rates for shirkers are higher than for workers who fulfill their contract, but these probabilities need not be as extreme as 1 and 0, respectively. Intuitively, in a no-shirking equilibrium with pooling, workers provide the contractually stipulated effort level, and firms may thus as well hire a new good-performing worker as rehire the previous good-performing worker. Furthermore, as we discuss in more detail in the Appendix, an individual firm has no incentive to deviate from the equilibrium separation rate.

13. As discussed in the theoretical Appendix, in the (extreme) case that the no-shirking equilibrium involves zero job separations, new jobs never open up and there are no labour-market flows. With no flows, shirkers who are fired can never find a new job again, and the value of unemployment is always zero. In this case the value of unemployment does

of repeated game models, in our view they are still very useful for heuristic purposes. The specification of “candidate” equilibria gives rise to additional testable hypotheses for behaviour, which must be satisfied if the data are to be explained by a particular type of equilibrium. For instance, equal profitability of both firm types and positive separation rates between firms and non-shirking workers are both necessary features, and thus testable qualitative implications of an equilibrium involving market segmentation. The presence of multiple equilibria also makes it even more important to study actual behaviour in a tightly controlled environment, as it is ultimately an empirical question which type of equilibrium emerges.

4. RESULTS

We organize our discussion of results as follows. In Section 4.1 we analyse the impact of contractual incompleteness on the level of unemployment, and examine in more detail whether the channels through which unemployment emerges are in line with the mechanisms featured above. A similar structure applies to Section 4.2, where we study the influence of contractual incompleteness on labour market segmentation.

4.1. Contractual incompleteness and unemployment

Figure 1 depicts the average unemployment rates for the C treatment (black) and the IC treatment (grey), as well as the fraction of workers who are exogenously unemployed due to excess supply of labour (dashed grey line). We observe strong and significant differences in unemployment between treatments ($p < 0.01$).¹⁴ In the C treatment where contracts are explicitly enforced, unemployment remains close to the minimum possible level: in most periods, the unemployment level lies only 1–5 percentage points above the baseline level that is due to excess supply of labour. The unemployment pattern looks markedly different, by contrast, when effort is not verifiable. In particular, we observe a sharp increase in unemployment over the first seven market periods, before unemployment stabilizes and remains high for the rest of the game.¹⁵ Contractual incompleteness thus causes a strong increase in the overall level of unemployment.

Result 1. *We observe strong differences in unemployment between treatment conditions. Under explicit contract enforcement (C treatment), unemployment levels are close to the minimal possible level. When effort is not verifiable (IC treatment), unemployment rises strongly before stabilizing at a relatively high level.*

not depend on the number of one-worker firms, eliminating a key mechanism for sustaining segmentation. Zero labour market flows, and a zero value of unemployment, also work against the equilibrium featuring endogenous unemployment when there is excess supply of labour, because it facilitates firms being able to elicit effort with relatively low rents, and it may thus be profitable to fill all vacancies.

14. This test is based on a panel estimation with session-level random effects. Unless otherwise noted, this procedure is also applied for all other tests on aggregate-level outcomes reported in this section. Test statistics for individual-level data (e.g. wages in Table 2) are based on estimations with individual-level random effects and standard errors accounting for potential clustering at the individual level. Results are robust to applying non-parametric tests based on session-level averages for analyzing the main treatment differences, or to using alternative estimation strategies, such as allowing for multi-level random effects at the session and individual level, or clustering of standard errors at the session level. Reported p -values are always two-sided.

15. Regressing per-period changes in unemployment on a constant, we estimate an average change of $\Delta_t = 2.9$ percentage points for the first seven period in the IC treatment ($p = 0.044$). If we do the analog for the remaining periods, the average change is much smaller and insignificant ($\Delta_t = 0.5$; $p = 0.517$).

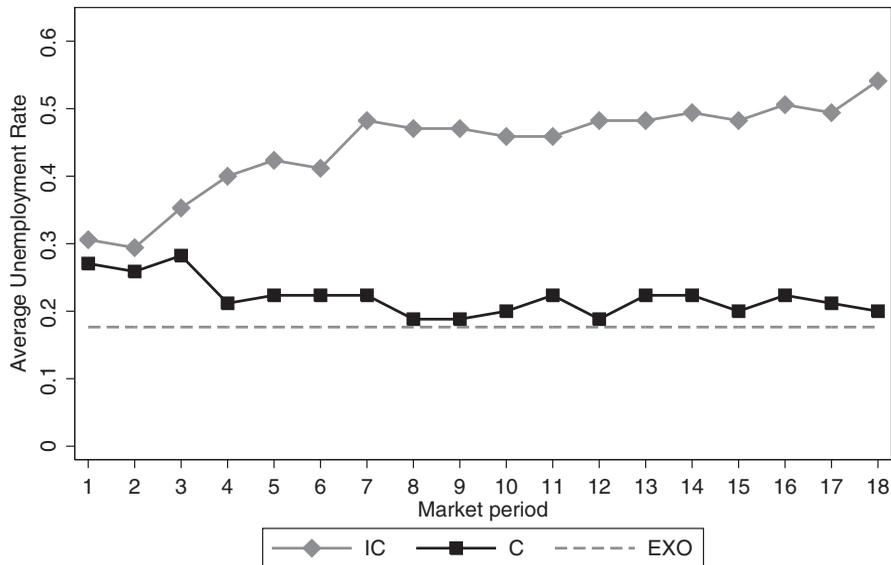


FIGURE 1

Unemployment over time. Average unemployment rate in the IC treatment (grey) and the C treatment (black). Fraction of exogenously unemployed workers (dashed grey).

Rows (1) and (2) of Table 2 shed further light on the driving forces behind the differences in unemployment. As could be inferred from the low levels of endogenous unemployment in the C treatment, we find that firms in this treatment almost always try to fill two vacancies (in 96.7% of cases) and, with an acceptance rate of 97.9%, workers essentially always accept. In the IC treatment, by contrast, firms offer only 67.5% of the possible jobs. At the same time, available contract offers are accepted in 99.8% of cases.¹⁶ Firms' decision to hire fewer workers than technologically possible—*i.e.* to ration jobs—is thus the main source of endogenous unemployment in the IC treatment.

Rows (3)–(9) of Table 2 summarize further differences in the characteristics of employment relationships, in terms of the contract terms being offered, worker behaviour, and employment duration. In the C treatment, firms pay worker rents of only 5.7 points, hire workers via public offers in 84% of cases, and elicit effort close to the maximum possible level ($e=9.7$). In contrast, wages and offered worker rents are substantially higher in the IC treatment, reaching levels of 34.6 and 21.1 points, respectively. In addition, 71.7% of concluded contracts in this treatment are initiated via private offers, and long-term employment relationships are frequently observed. For instance, 60% of firms rehire the same worker for at least 9 consecutive market periods. Effort levels are lower than in the C treatment, but at an average level of 7, they lie far above the minimum. Although effort is not verifiable, workers in the IC treatment on average deviate from the contractually stipulated level by only 1.2 points.¹⁷ These findings provide first indications that

16. In the few cases in which we observe available contracts offers being rejected, these exhibit very low or even negative offered worker rents. On average, the rejected contract offers stipulate worker rents of only 0.9 points (1.9 points in the C treatment, and -6.6 points in the IC treatment). As discussed above, rejections of such offers with extremely low worker rents might be an indication of workers' fairness concerns.

17. Note that, strictly speaking, any deviations from the contractually stipulated effort level in pre-final market periods are inconsistent with the notion of a no-shirking equilibrium in our model.

TABLE 2
Market characteristics

	C treatment	IC treatment	
(1) Fraction of possible jobs offered	0.967	0.675	$p < 0.01$
(2) Fraction of posted vacancies accepted	0.979	0.998	$p < 0.01$
(3) Wages	22.8	34.6	$p < 0.01$
(4) Rents offered by firms ($w - c(\hat{e})$)	5.7	21.1	$p < 0.01$
(5) Realized worker rents ($w - c(e)$)	5.7	23.6	$p < 0.01$
(6) Fraction of private contracts	0.160	0.717	$p < 0.01$
(7) Fraction of firms with employment relationships ≥ 9 market periods	0.029	0.600	$p < 0.01$
(8) Effort	9.7	7.0	$p < 0.01$
(9) Desired effort	9.7	8.2	$p < 0.01$

Notes: Mean values across treatments. The p-value for row (7) is derived from a linear-probability model where we regress a dummy variable equal to 1 if a given firm has at least one employment relationship for ≥ 9 consecutive market periods during the experiment on a treatment dummy. Reported p-values for all other tests are derived using the procedures described in Footnote 14.

firms in the IC treatment successfully use contract renewal and worker rents to establish implicit performance incentives. The co-existence of high worker rents and job rationing through firms also demonstrates that unemployment in the IC treatment is involuntary. At the same time, the observation of low rents and a lack of repeated contract renewal in the C treatment indicate that firms manage to exploit their high bargaining power when contracts are explicitly enforced.

In Table 3 we investigate two additional hypotheses about differences in the functioning of employment relationships, which cannot be directly inferred from the level differences in Table 2. Columns (1) and (2) report linear-probability estimates on the determinants of firms' contract renewal decisions. This allows us to examine whether the greater prevalence of long-term relationships in the IC treatment indeed reflects a policy of *conditional* contract renewal, with rehiring and firing decisions depending on a worker's previous performance. Comparing rehiring rates across treatments, the positive coefficient for the IC treatment in Column (1) confirms that contract renewal is overall more prevalent when effort is not verifiable. The negative coefficient of the shirking indicator, however, demonstrates that firms in the IC treatment do indeed strongly condition contract renewal decisions on workers' behaviour. The average likelihood that a contract is renewed drops from 70.5% in the case of contract fulfilment to only 26.2% if a worker deviates from the contractually stipulated effort level. Note that while separations occur less frequently in the case of contract fulfilment, the estimates show that the separation rate after contract fulfilment is still positive in the IC treatment. This is important since, according to our model, equilibria involving endogenous unemployment and market segmentation exhibit non-zero separation rates in equilibrium. Our data further reveal that 85.3% of the separations between firms and non-shirking workers reflect firms not offering a new contract to their previous worker, in line with the mechanism in our model. In the remaining 14.7% of cases, workers quit and refuse an available offer for being rehired by their previous firm.¹⁸

Column (2) adds controls for other important characteristics of employment relations, including the rents offered in the current period's contract, and an indicator for past contract renewal, which equals 1 if a firm has already renewed a worker's employment contract in the current period (via a private contract offer). In this specification shirking continues to be a crucial

18. In our model, we abstract away from such voluntary quitting and on-the-job search. Incorporating this additional mechanism for equilibrium flows would be an interesting extension to our model.

TABLE 3
Rehiring and long-run incentives

	Dependent variable:			
	1 if worker is re-hired in $t+1$		Future rents	
	(1)	(2)	(3)	(4)
IC treatment	0.639*** (0.061)	0.232*** (0.072)	155.560*** (15.443)	211.128*** (20.969)
Shirking (t)	-0.427*** (0.057)	-0.205*** (0.058)	-66.335*** (13.811)	-74.598*** (11.671)
Offered rents (t)		0.003 (0.002)		0.795* (0.449)
IC \times Offered rents (t)		0.002 (0.003)		0.677 (0.714)
Contract renewed (t)		0.416*** (0.083)		5.215 (5.621)
IC \times Contract renewed (t)		0.045 (0.098)		50.940*** (12.394)
Constant	0.065*** (0.020)	-0.007 (0.022)	30.214*** (2.075)	56.387*** (6.248)
Market period	no	yes	no	yes
R^2	0.373	0.551	0.353	0.611
N	1,935	1,935	1,935	1,935

Notes: Random-effects models; the reported standard errors (in parentheses) account for clustering at the individual level. Columns (1)–(2): the dependent variable equals 1 if a firm renews the contract of a worker through a private contract in period $t+1$. Columns (3)–(4): the dependent variable “future rents” is the sum of a worker’s earnings from period $t+1$ until period 18. “Shirking” is an indicator equal to 1 if the worker deviates from the contractually stipulated effort level in the current period (*i.e.* $e_t < \bar{e}_t$). “Contract renewed” is an indicator equal to 1 if a firm has interacted with the same worker in period $t-1$, and rehired the worker through a private contract offer in period t . “IC \times Offered rents” and “IC \times Contract renewed” are interaction terms of the respective measure with the treatment dummy. Columns (2) and (4) additionally control for a linear time trend, as well as an interaction term of the time trend with the treatment dummy. ***, **, and * indicate significance on the 1-percent, 5-percent, and 10-percent level, respectively.

factor in firms’ contract renewal decisions in the IC treatment.¹⁹ This indicates that firms in the IC treatment systematically engage in conditional contract renewal rather than, *e.g.* just having a stronger taste for repeatedly interacting with a given worker. Further estimates (available upon request) show that the results reported in Table 3 are also robust to using alternative measures of worker slacking, such as the degree of deviation from the contractually stipulated effort level.

Columns (3) and (4) of Table 3 show that the observed differences in rent payments and contract renewal strategies have important consequences for workers’ long-term earnings prospects. Confirming the observations from Table 2, the treatment dummy in Column (3) illustrates that workers in the IC treatment generally earn higher future rents than those in the C treatment. More importantly, the coefficient for shirking indicates that—within the IC treatment—the long-run benefits of contract fulfilment in a given period considerably outweigh the short-run gains

19. By design, shirking is not possible in the C treatment, and thus we do not include an interaction term between shirking and the treatment dummy. Interestingly, the coefficient of “Contract renewed” in the C treatment is significant and positive (Column 2). This only reflects a small number of firms, however, who engage intensely in contract renewal in this treatment; as illustrated in Column (1), the overall likelihood that a firm in the C treatment rehires her worker is very low. Our estimates from Table 4, discussed below, further indicate that this strategy does not increase the respective firms’ profits.

TABLE 4
Profitability of contractual instruments

	Dependent variable: firm profit from a given contract		
	(1)	(2)	(3)
Offered rents	-0.863*** (0.115)		-0.869*** (0.112)
(Offered rents) ²	-0.004 (0.003)		-0.004 (0.002)
IC × Offered rents	1.767*** (0.196)		1.529*** (0.193)
IC × (Offered rents) ²	-0.007* (0.004)		-0.006 (0.004)
Contract renewed		-3.178 (3.249)	0.865 (1.297)
IC × Contract renewed		16.837*** (3.669)	9.802*** (2.268)
IC treatment	-43.983*** (2.762)	-24.411*** (2.216)	-41.695*** (2.701)
Constant	49.556*** (1.678)	39.027*** (1.725)	49.600*** (1.659)
Market period	yes	yes	yes
Final period	yes	yes	yes
R ²	0.515	0.454	0.532
N	2,042	2,042	2,042

Notes: Random-effects models; the reported standard errors (in parentheses) account for clustering at the individual level. The dependent variable is the level of a firm's earnings from a given employment contract. See Table 3 for definitions of the remaining variables. The estimations additionally control for a linear time trend, an indicator equal to 1 in period 18 to capture a possible endgame effect, as well as interaction terms of the respective variables with the treatment dummy. ***, **, and * indicate significance on the 1-percent, 5-percent, and 10-percent level, respectively.

from shirking.²⁰ Controlling for other aspects of contract terms yields similar results (see Column (4) of Table 3). This underlines that workers in the IC treatment face strong implicit performance incentives. The relatively high effort levels in this treatment are thus understandable from a long-term incentive perspective.

The data from the IC treatment also illustrate the relevance of both (implicit) material incentives and fairness concerns. This can best be seen in the final market period, in which the potential for future interactions and the prospect of qualifying for future rents vanish. Nevertheless, we find that 46.2% of workers in the IC treatment provide above-minimal efforts in this period, on average choosing an effort level of 6.0.²¹ This underlines the relevance of voluntary gift-exchange in our setting, paralleling what has commonly been observed in similar labour-market settings in the lab and field (see, e.g. Fehr *et al.*, 1993; Brown *et al.*, 2004; Cohn *et al.*, 2012). At the same time, our data on effort provision also illustrate the importance of future rents and contingent contract renewal for motivating workers in the IC treatment: those workers who only exert minimal effort in the final period on average provide an effort of 7.7 in period 17 in which the potential for future interaction is still intact.

20. For instance, a worker who shirks in period 10 earns on average 88.4 points less during periods 11–18 compared to a worker who provides the contractually stipulated effort in the same period. These losses in future rents are higher than the maximally possible short-run gains of shirking due to reduced effort costs.

21. Note that this number might be a lower bound for the fraction of fair-minded workers: another 12.8% of workers provide minimal efforts in the final period, but do so in response to receiving a wage of only 1. Such very low-final-period wage offers might reflect some employers exhibiting “betrayal aversion” (Bohnet *et al.*, 2008; Fehr *et al.*, 2007) or holding particularly pessimistic beliefs about the degree of workers' cooperativeness.

Result 2. *Firms in the IC treatment pay higher workers rents, engage more heavily in repeated contract renewal, and offer fewer jobs than firms in the C treatment. The combination of job rationing and high worker rents implies that endogenous unemployment in the IC treatment is involuntary. The strategies of firms in the IC treatment generate strong implicit performance incentives for workers.*

So far, our findings on treatment differences in worker rents, the prevalence of conditional contract renewal, and other key characteristics of employment relationships are all consistent with our theoretical hypotheses. However, our model also suggests a specific mechanism for *why* these differences emerge, namely that the absence of explicit contract enforcement makes specific contractual instruments profitable for firms. Table 4 examines this key underlying mechanism. In the estimations, we focus on the types of contractual instruments studied in Table 2, related to worker rent levels and contingent contract renewal. Column (1) shows that higher rent payments decrease the profitability of a contract for firms in the C treatment. In the IC treatment, by contrast, we see a qualitatively opposite relationship. Firm profits are increasing in the level of rents that a firm offers to her worker; this holds up to a point, after which further increases in wage costs dominate the profit increases due to higher work efforts (for further illustration, see also Figure A.1 in the Supplementary Material). Column (2) reports estimates on the profitability of repeated contract renewal, using the “renewed contract” indicator from Table 3. We find that firm profits in the IC treatment are strictly higher in employment relationships that involve repeated rehiring. In contrast, firm profits in the C treatment are not systematically affected by whether firms and workers interact repeatedly. Allowing for a simultaneous influence of both contractual instruments does not affect our main findings (Columns (3) of Table 4). In sum, we find important qualitative differences in how contractual instruments affect firm profits, and we see that the level differences in the variables shown in Table 2 are in line with firms’ profit incentives. We discuss the profitability of differences in job rationing in the next section, where we study the impact of contractual incompleteness on market segmentation.

Result 3. *Contractual incompleteness changes the profitability of specific contractual instruments. Paying high worker rents and using repeated contract renewal has a positive impact on firm profits when effort is not verifiable. Under explicit contract enforcement, both instruments are irrelevant for or detrimental to firm profits.*

4.2. *Contractual incompleteness and labour market segmentation*

In this section, we turn to investigating the impact of contractual incompleteness on labour market segmentation, and we analyse whether individual behaviour is consistent with our theoretical hypotheses for how segmentation can be an equilibrium phenomenon. We have already seen in Figure 1 that, after a strong initial increase, the level of unemployment in the IC treatment stabilizes during the later phase of the experiment. Given that endogenous unemployment is almost exclusively driven by job rationing, this also implies that the proportion of firms who ration jobs is increasing initially, but then reaches a plateau. Indeed, a relatively stable fraction of about 25% of firms continues to employ two workers during the later phase of IC treatment. This provides a first indication that there might be market segmentation under contractual incompleteness, with two different firm types coexisting in the long run.

For market segmentation to be an equilibrium phenomenon according to our model, however, it is necessary that the firm strategies characterizing the different market segments are equally profitable. This can occur if the unemployment pressure arising due to job rationing by some firms allows the non-rationing strategy to be viable for other firms. This is exactly what we observe in

our data. Job rationing is strictly profitable for firms in the early periods of the IC treatment, in which unemployment pressure is relatively low, and in which we observe strong dynamics towards adopting the one-worker strategy. On average, one-worker firms earn roughly 50% more than two-worker firms during the first seven market periods (average per-period profits are 37.2 and 24.7 points, respectively; $p < 0.01$).²² In contrast, in the remaining periods where unemployment has reached a plateau and the fraction of two-worker firms stabilizes, profits between one-worker firms and two-worker firms do not differ significantly anymore. The average per-period profits for periods 8–18 are 36.7 and 35.3 points in one-worker and two-worker firms, respectively ($p = 0.361$).²³ The finding that, in the long run, profits of one-worker firms and two-worker firms are similar is consistent with the emergence of a segmentation equilibrium.

While both strategies yield similar profits for firms during the later market phase, workers in two-worker firms face much less favorable contract terms than their counterparts in one-worker firms (see the rightmost columns of Table 5). On average, firms who employ two workers offer about 40% lower worker rents than one-worker firms. At the same time, workers in one-worker firms work somewhat harder: average effort is approximately 20% higher, although this difference turns out to be statistically insignificant. Overall, however, the higher wages in one-worker firms result in much higher realized earnings for workers in those firms.²⁴ Thus, firms using the one-worker strategy and two-worker strategy in the IC treatment offer “good” jobs and “bad” jobs, respectively, consistent with the endogenous emergence of a segmented labour market.²⁵ Interestingly, the two segments in the IC treatment also seem to differ in the general stability of employment relationships: one-worker firms are somewhat more likely to hire their workers through private contract offers, and the average job tenure of workers—*i.e.* the average number of consecutive periods that a firm has rehired her current worker through private contract offers—is also longer in one-worker firms (see rows (7) and (8) of Table 5).

In contrast to the situation in the IC treatment, we do not see any indication of market segmentation under explicit contract enforcement (see the leftmost part of Table 5). In the few occasions in which firms do hire only one worker in the C treatment, wages and worker rents are only slightly above the values for two-worker firms (the differences are 1.5 and 1.2 points, respectively). Similarly, efforts are about 1% higher, and the fraction of private contracts is 2.2 percentage points higher in one-worker firms. While the latter effect turns out to be weakly significant, the differences between segments are generally much smaller than in the IC treatment. Most importantly, the data on firm profits demonstrate that employing a one-worker strategy in the C treatment is clearly suboptimal from a firm’s perspective. Firms who hire only one worker in the C treatment earn almost 20 points less than the ones employing two workers. This underlines

22. In line with the idea that job rationing goes hand in hand with paying high worker rents to elicit high work effort, we observe that 67.4% of firms who switch to a one-worker strategy in the IC treatment at the same time increase their wage payments.

23. In the Supplementary Material, we illustrate in more detail how the profitability of the two hiring strategies adapts between the early and late phase of the experiment. We show that two-worker firms initially tend to make substantially lower profits, irrespective of the rents they pay to their workers. In the late phase, profits do not depend on firm size anymore if firms choose the respective profit-maximizing rent levels.

24. Notably, we find a negative correlation between wages and firm size in the IC treatment. It is important to bear in mind, however, that our design rules out some of the most important factors that are typically discussed as reasons for positive firm size-wage differentials in labour markets, *e.g.* larger firms hiring higher-quality workers in terms of observed and unobserved skills, having higher degrees of unionization, or facing stronger monitoring difficulties (*e.g.* Brown and Medoff, 1989; Abowd *et al.*, 1999).

25. This ranking of job qualities across firm types does not only hold for average worker rents in a given period, but also in terms of workers’ overall future earnings (calculated using the method from Table 3). In any given period, workers who are employed in one-worker firms have higher total earnings over the rest of the game than workers who are employed in two-worker firms in the corresponding period.

TABLE 5
Market segmentation in the IC and C treatment

	C treatment			IC treatment		
	1-worker		2-worker	1-worker		2-worker
(1) Firm profits	77.1	$p < .01$	95.5	36.7	$p = .361$	35.3
(2) Wages	22.9	$p = .115$	21.4	42.3	$p < .01$	27.7
(3) Rents offered by firms	4.9	$p = .313$	3.7	26.9	$p < .01$	16.0
(4) Realized worker rents	4.9	$p = .313$	3.7	29.2	$p < .01$	17.6
(5) Effort	10.0	$p = .161$	9.9	7.9	$p = .718$	6.5
(6) Fraction of private contracts	0.192	$p = .082$	0.170	0.900	$p = .012$	0.775
(7) Job tenure	1.0	NA	1.9	6.8	$p < .01$	4.5
(8) Fraction of firms	0.068		0.932	0.738		0.262

Notes: Mean values of market characteristics during the late phase of the experiment (periods 8–18). The reported p -values are derived from random-effects estimations in which the respective dependent variable is regressed on a dummy equal to 1 if a contract comes from a one-worker firm (standard errors account for clustering at the individual level).

why the latter strategy dominates the market, with 93% of firms using a two-worker strategy when contracts are explicitly enforced.

Result 4. *Contractual incompleteness leads to a segmentation of the labour market. In the long run, two types of firms coexist in the market when effort is not verifiable. These earn similar profits, but differ qualitatively with respect to wage payments, worker rents, and effort provision.*

As a final step we investigate more directly one of the main mechanisms underlying our theoretical explanation for segmentation, which is a feedback from unemployment pressure to behaviour of workers. The emergence of the secondary sector is possible in equilibrium, because the unemployment pressure arising as a byproduct of job rationing makes workers less likely to shirk and willing to put in higher effort for a given wage. If this is true, we should observe workers being more likely to shirk on the job when they receive information that signals low unemployment pressure and high chances to acquire a job. While workers in the experiment did not have precise information on the level of unemployment or the job acquisition rate in a given period, they could infer the tightness of the labour market from activity in the contracting stage. The most salient indicator of less favourable market conditions from a worker's perspective is the number of vacancies in a given period that are posted through public contract offers. Since unemployed workers disproportionately have to rely on public contract offers for finding a new job, a low number of publicly posted vacancies indicates high costs of unemployment and, consequently, higher unemployment pressure for those employed.

Columns (1)–(3) of Table 6 demonstrate that a decrease in the number of public job offers in the market is associated with a significant reduction in workers' propensity to shirk. This holds after controlling for the contract terms faced by a worker, an "endgame dummy" to account for the sharp increase in shirking in the final period of the experiment, and a general time trend. Paralleling the observations on shirking, Columns (4)–(6) of Table 6 show that information signalling lower job finding chances is associated with a general increase in workers' performance in terms of effort. This illustrates how changing market conditions feed back into workers' behaviour under contractual incompleteness.

Our data also allow us to rule out some alternative explanations for why the secondary-sector strategy might become profitable. In particular, we find no evidence that workers who are employed in two-worker firms during the late phase of the experiment are inherently more willing to provide higher efforts for a given wage. In other words, there is no support for a

TABLE 6
Shirking and effort provision in the IC treatment as a function of market conditions

	Dependent variable:					
	1 if $e < \hat{e}$			Effort		
	(1)	(2)	(3)	(4)	(5)	(6)
# public vacancies	0.021*** (0.006)	0.019*** (0.006)	0.019** (0.008)	-0.064** (0.027)	-0.053** (0.025)	-0.060* (0.032)
Wage	-0.019*** (0.001)	-0.019*** (0.002)	-0.019*** (0.002)	0.112*** (0.007)	0.110*** (0.008)	0.110*** (0.008)
Desired effort	0.111*** (0.009)	0.111*** (0.009)	0.111*** (0.009)	0.223*** (0.042)	0.224*** (0.042)	0.223*** (0.043)
Contract renewed		-0.045 (0.048)	-0.045 (0.048)		0.259 (0.204)	0.280 (0.205)
Market period			-0.000 (0.005)			-0.005 (0.023)
Final period	0.380*** (0.073)	0.383*** (0.073)	0.387*** (0.087)	-3.148*** (0.490)	-3.164*** (0.486)	-3.113*** (0.540)
Constant	0.122* (0.065)	0.132** (0.064)	0.139 (0.090)	1.386*** (0.227)	1.331*** (0.212)	1.391*** (0.378)
R^2	0.302	0.309	0.310	0.672	0.676	0.676
N	849	849	849	849	849	849

Notes: Random-effects models; the reported standard errors (in parentheses) account for clustering at the individual level. Column (1)–(3): the dependent variable is a dummy equal to 1 if the worker deviates from the contractually stipulated effort level. Column (4)–(6): the dependent variable is the level of effort provided in a given period. “# public vacancies” is the number of vacancies in a given market period that were posted via public contract offers. “Final period” is a dummy variable equal to 1 in the final period of the game. ***, **, and * indicate significance on the 1-percent, 5-percent, and 10-percent level, respectively.

sorting explanation in which two-worker firms become more profitable over time, on the basis of eventually finding especially diligent workers. Instead, our data indicate that a within-worker change in the willingness to provide effort under tighter labour market conditions accounts for the increased profitability of two-worker firms.²⁶

Result 5. *Workers’ behaviour in the IC treatment changes in accordance with tightening market conditions. This contributes to the increasing profitability of two-worker firms and the long-run segmentation of the labour market.*

5. CONCLUSIONS

In this paper, we provide evidence that involuntary unemployment and labour market segmentation may be caused by contractual incompleteness. A simple model shows how both phenomena can jointly arise in a labour market equilibrium when effort is non-verifiable, and our data are consistent with the key mechanisms of the equilibrium. The findings thus support a perspective in which two fundamentally important aspects of labour markets are intimately related.

26. This result comes from restricting the analysis to workers who are mainly employed in two-worker firms during the later phase in the IC treatment. We find that the willingness to provide effort for a given wage for these workers increases between the early and late phase. Thus, we observe a within-worker change, which is consistent with adaptation to market conditions, but not with an explanation where two-worker firms eventually found a particularly hard-working group of workers. Results can be found in the Supplementary Material.

Our analysis suggests several avenues for future research. First, it would be interesting to study how unemployment and segmentation are affected by the availability of other explicit contracting strategies (*e.g.* piece rates), as well as other contracting schemes that do not require verifiability of work effort, such as voluntary bonus payments or possibilities to sign multi-period employment contracts. Second, our setup could be used to study the impact of institutions such as unemployment insurance or employment protection, which might affect both workers' on-the-job behaviour and employee turnover. This would be especially interesting since our model predicts that the emergence and characteristics of market segmentation depend on the equilibrium job separation rates. Third, one could examine the impact of different information structures on market outcomes, for example by changing the information that workers have about labour market tightness. Fourth, while the number of market participants was held constant in our experiments, our setup could be modified in future work to study the implications of free market entry. Finally, our framework could be used to analyze how hiring decisions, rent payments, and other labour market outcomes interact with the available production technology. Different forms of technology, where productivities are unequal across jobs in a two-worker firm, could generate within-firm segmentation, in the sense of identical workers earning different rents within the same firm. In many labour markets, there are also additional factors that directly influence a firm's hiring incentives, such as market entry costs, heterogeneity in outside options for firms, or differences in the substitutability of capital for labour inputs. Studying how these factors affect market performance under contractual incompleteness is a potentially fruitful direction for future work.

APPENDIX

In this appendix, we provide a model that formalizes the intuition for how contractual incompleteness can cause involuntary unemployment and labour market segmentation.

A.1. Model setup

We denote by $N+U$ the mass of workers and by $\frac{N}{2}$ the mass of firms. Workers and firms interact for T periods in discrete time. All firms can hire up to two workers in a given period. Maximum employment is thus N , and excess supply of workers implies that the minimum level of unemployment is U .²⁷ In any period, each firm offers zero, one or two contracts stipulating a desired level of effort, \hat{e} and a wage, w . Firms may either rehire an employee from the previous period, or recruit from the pool of unemployed workers. In the latter case, the worker who receives the offer is randomly drawn from the pool of unemployed agents who have not received an offer yet. Subsequently, workers can either accept or decline the posted contract. If effort is verifiable, the contract determines both the (upfront) wage and the effort level. If efforts are only observable but not third-party verifiable, workers have discretion over their level of effort. We assume that workers are homogeneous in their ability and can choose between n different effort levels $e \in \{e_1, \dots, e_n\}$, where $e_i < e_{i+1}$ and $c(e_i) = c_i$ denotes the increasing cost of effort.

There are two types of workers: a fraction λ of the population is fair (f) and a fraction of $1-\lambda$ is selfish (s). The utility of an employed selfish worker in any period is the difference of the received wage and the effort cost:

$$u_s = w - c(e).$$

Fair workers, in contrast, face an additional psychological cost or benefit, $g(w - c(\hat{e}))$, if they fulfil the contract. Utility for fair workers is thus denoted as:

$$u_f = \begin{cases} w - c(e) + g(w - c(\hat{e})) & \text{if the worker fulfils the contract } (e = \hat{e}) \\ w - c(e) & \text{if the worker shirks } (e < \hat{e}) \end{cases}$$

We assume that $g(\cdot)$ is strictly increasing in the rent offered by the firm, and $g(\tilde{x}) = 0$ for some $\tilde{x} \geq 0$. Thus, more generous contract terms make shirking increasingly less attractive for a fair worker, and $g(\cdot)$ changes from negative

27. While a level of $U > 0$ is implemented in the experiment, all qualitative results also hold for $U = 0$.

(a cost of contract fulfilment) to positive (an additional benefit of contract fulfilment), when offered rents exceed a benchmark level of fairness, \bar{x} . This captures in a simple and tractable way the idea that fair workers may experience a psychological benefit (cost) of fulfilling (breaching) agreements (e.g. Ellingsen and Johannesson, 2004; MacLeod, 2007). It also incorporates a central theme of all reciprocity-based fairness models, that fair types are willing to behave kindly towards someone who has been kind or generous before (e.g. Rabin, 1993; Falk and Fischbacher, 2006), and the idea that the kindness of an opponent is evaluated relative to some social norm or fairness benchmark (e.g. Akerlof and Yellen, 1990; Fehr and Schmidt, 1999).²⁸ In our setting, kindness of the firm is captured by the size of offered rents, relative to the benchmark \bar{x} , which in turn affects a fair worker's psychological utility $g(\cdot)$ of fulfilling the terms of the contract to which they agreed. Denote by $g^{-1}(\cdot)$ the inverse function of $g(\cdot)$, which exists and is well defined due to the monotonicity of $g(\cdot)$.

Firms are characterized by the following production technology, with output increasing in the level of effort:

$$0 \leq f(e_1) = z_1 \leq \dots \leq f(e_n) = z_n, 0 \leq f(2e_1) = z_{n+1} \leq \dots \leq f(2e_n) = z_{2n},$$

z_1 to z_n denote the output of a firm that employs one worker who exerts e_1 to e_n , and z_{n+1} to z_{2n} are the corresponding output levels of two-worker firms.²⁹ Furthermore, we require the production technology to exhibit a weak form of decreasing returns to scale (Part 1 of Assumption 1) and to be efficient (Part 2 of Assumption 1). The latter means that in the one-shot version of the game the wage needed to induce an extra unit of effort by a fair agent is smaller than the induced gain in output. Hence, in a market with only fair workers and no repeated interaction, firms and workers find it most beneficial if maximum effort is implemented.

Assumption 1. Let $z_i < z_j$ be output levels with corresponding effort input $e_i < e_j$. Then,

1. $\frac{z_j - z_i}{e_j - e_i} > \frac{z_{n+j} - z_{n+i}}{2(e_j - e_i)}$
2. $z_{n+j} - 2(g^{-1}(c_j) + c_j) > z_{n+i} - 2(g^{-1}(c_i) + c_i) > z_i - (g^{-1}(c_i) + c_i) > 0$

We further assume that output translates directly into firm revenue, all firms have access to the same production technology, and maximize total profits (i.e. revenue minus overall wage costs). At the end of each period, firms decide whether to renew the contract with their worker(s). As a simplification we assume that a firm which renews the contract of its worker(s) also keeps its size constant. We denote by $b_k^t \in [0, 1]$ with $k \in \{1w, 2w\}$ the probability that a one-worker firm (1w) or two-worker firm (2w) dismisses a worker who has exerted the desired effort level in period t . Purely for expositional reasons, we assume that firms always separate from workers who deviate from the contractually stipulated effort level ($a_k^t = 1$).³⁰ Workers and firms discount the future at a rate $r \in [0, 1]$.³¹

A.2. Equilibria with non-verifiable effort

If effort is non-verifiable, the type of equilibrium depends on the shape of firms' production technology and the psychological cost function for fair workers. Since our main interest is to illustrate how contractual incompleteness can give rise to equilibria involving endogenous unemployment and market segmentation, we concentrate on the existence of these segmentation equilibria. As a first step we show that the endgame of the model features positive continuation values. Second, we characterize the no-shirking conditions for both worker types in pre-final periods. Third, assuming positive continuation values for the last period, we pin down the conditions leading to a segmentation equilibrium in pre-final periods.

28. There is substantial empirical support for the notion that individuals exhibit a preference for fulfilling promises and agreements (e.g. Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006; Vanberg, 2008), are willing to reward kind actions and punish unkind ones (e.g. Fehr *et al.*, 1993; Fehr and Gächter, 2000; Kube *et al.*, 2013), and base their judgments on norms and fairness standards (e.g. Bewley, 1999; Cohn *et al.*, 2012). For simplicity, we do not model the determination of \bar{x} , and abstract away from some more nuanced notions of fairness, for example ones in which the psychological benefit/cost of a fair worker additionally depends on the distance between the chosen effort level and the contractually stipulated one.

29. Note that we rule out the possibility that firms hire two workers and elicit different effort levels from each worker. This corresponds to assuming that a single firm does not offer both bad and good jobs. However, allowing for this possibility would not change the existence of the segmentation equilibrium that is our focus, although existence would be for a more restricted range of parameters.

30. Endogenizing a_k^t would not change the characteristics of feasible equilibria. In particular, no-shirking conditions for separation rates $0 < a_k^t < 1$ can be derived analogously to conditions (A.1) and (A.3) below.

31. In the specific setting of our experiment in which the time horizon is short, the assumption of zero time discounting is plausible. All of the following results hold for the case of $r = 0$.

A.2.1. Endgame and continuation value. Selfish workers always choose to shirk in the final period. In contrast, fair workers fulfil their contract if their effort cost is smaller than their psychological return from fulfilling the contract. Anticipating this behaviour of workers, firms either stay out of the market or offer wages such that fair workers are indifferent between working and shirking. The latter is profitable if there exists at least one effort level e_i such that either a one- or a two-worker firm expects positive payoffs from inducing e_i , given its belief that it faces a fair worker.

Assumption 2.

$$\exists e_i: \lambda z_i - (g^{-1}(c_i) + c_i) > 0 \text{ or } \lambda^2 z_{n+i} + 2(1-\lambda)\lambda z_i - 2(g^{-1}(c_i) + c_i) > 0.$$

Assumption 2 ensures that there are enough fair types such that firms are willing to offer contracts with positive worker rents in the one-shot version of the game.³² This generates positive continuation values for workers in the pre-final period, if there has not been any screening in previous periods. Note that all firms use homogeneous contracting terms that involve positive rents in the final period.³³ Hence, unemployment is involuntary from a worker's perspective. We denote the value of a job for a worker of type j who is employed by a firm of type k in period t by V_{kj}^t . The value of unemployment in period t for a worker of type j is denoted by V_{uj}^t . We set all continuation values to zero in period $T+1$ and denote by L_{1w}^t, L_{2w}^t the number of jobs in one-worker and two-worker firms in period t .

A.2.2. No-shirking conditions for workers in pre-final periods. We start the analysis of pre-final periods by characterizing worker behaviour. Workers trade off the short-run gains of low effort costs due to shirking against potential long-run costs due to higher risk of dismissal and unemployment. Let w_k^{t-1} and e_k^{t-1} denote the wage and desired effort level, offered by a firm of type k in period $t-1$. For a fair worker, the no-shirking condition in period $t-1$ is then:

$$w_k^{t-1} - c_k^{t-1} + g(w_k^{t-1} - c_k^{t-1}) + (1-r) \left[(1-b_k^{t-1})V_{kf}^t + b_k^{t-1}V_{uf}^t \right] \geq w_k^{t-1} + (1-r)V_{uf}^t \quad (\text{A.1})$$

$$\Rightarrow w_k^{t-1} \geq g^{-1} \left[c_k^{t-1} + (1-r)(b_k^{t-1} - 1)(V_{kf}^t - V_{uf}^t) \right] + c_k^{t-1}. \quad (\text{A.2})$$

A fair worker's utility in case of contract fulfillment (left hand side of equation (A.1)) consists of four components. The worker earns the current period's wage w_k^{t-1} , bears the cost of effort c_k^{t-1} , experiences psychological utility $g(\cdot)$, and receives the continuation value conditional on contract fulfillment. In case of shirking (right-hand side of (A.1)), the worker saves the effort cost and experiences no psychological utility. Furthermore, the current firm does not renew the worker's contract in the next period. Hence, the worker only receives the value of unemployment V_{uf}^t in the next period, which compromises the likelihood of finding a job of either type, and the likelihood of remaining unemployed in that period.

Selfish workers are not subject to the psychological cost, and since wages are paid before efforts are revealed, their effort choice is independent of the current period's wage. Selfish workers thus exert effort in period $t-1$ if:

$$w_k^{t-1} - c_k^{t-1} + (1-r) \left[(1-b_k^{t-1})V_{ks}^t + b_k^{t-1}V_{us}^t \right] \geq w_k^{t-1} + (1-r)V_{us}^t \quad (\text{A.3})$$

$$\Leftrightarrow V_{ks}^t - V_{us}^t \geq \frac{c_k^{t-1}}{(1-r)(1-b_k^{t-1})}.$$

Denote by $B^t(e_i, V_{us}^t)$ the set of all vectors of separation rates $(b_{1w}^{t-1}, b_{2w}^{t-1})$ such that selfish types are willing to exert effort e_i in both types of firms, for a given value of unemployment V_{us}^t . Note that B^t may be empty for some effort levels, if future rents cannot compensate selfish types for the respective effort costs in the given period, as it is for instance the case in the final period of the game. Moreover, define the set $\bar{B}^t = B^t(e_n, \max_{L_t} V_{us}^t)$, which constitutes the set of separation rates for which the highest effort level is implementable even in the case of minimal unemployment threat.

32. Strictly speaking, fair types are not necessary for a final-period rent if workers can generate positive output without incurring effort costs (*i.e.* if $c_1 = 0$ and $z_1, z_{n+1} > 0$, as was the case in our experimental setting). In this case, firms could profitably offer a minimal, but strictly positive worker rent in the final period, which in turn opens up the possibility for "reputation equilibria", even when all agents are selfish. Empirically, final-period rents are substantially above the minimal possible level, and many workers exert non-minimal effort in the final period; thus, an equilibrium based on fair types is better supported by the data (see our discussion in Section 4.1).

33. Since firms have homogenous contracting terms in the last period, we require that either firm type may rehire a worker from the pre-final period, using the final-period contract terms.

In what follows, we derive a necessary and sufficient condition for separation rates to be in \bar{B}^t . For a worker of type j in period t the value of unemployment is given by:

$$V_{ij}^t = \frac{(b_{1w}^{t-1}-1)L_{1w}^{t-1}+L_{1w}^t}{U+(1+b_{1w}^{t-1})L_{1w}^{t-1}+b_{2w}^{t-1}L_{2w}^{t-1}} V_{1w,j}^t + \frac{(b_{2w}^{t-1}-1)L_{2w}^{t-1}+L_{2w}^t}{U+(1+b_{1w}^{t-1})L_{1w}^{t-1}+b_{2w}^{t-1}L_{2w}^{t-1}} V_{2w,j}^t \quad (\text{A.4})$$

$$+ \left[1 - \frac{(b_{1w}^{t-1}-1)L_{1w}^{t-1}+L_{1w}^t}{U+(1+b_{1w}^{t-1})L_{1w}^{t-1}+b_{2w}^{t-1}L_{2w}^{t-1}} - \frac{(b_{2w}^{t-1}-1)L_{2w}^{t-1}+L_{2w}^t}{U+(1+b_{1w}^{t-1})L_{1w}^{t-1}+b_{2w}^{t-1}L_{2w}^{t-1}} \right] V_{ij}^{t+1}(1-r).$$

The equation illustrates the three components of the value of unemployment in period t : finding a new job in either firm type, or remaining unemployed in that period. V_{ij}^t is endogenously determined in equilibrium. Depending on the contract renewal strategies of either firm type in period $t-1$, it is more or less likely to get hired in a corresponding job in period t . The total derivative with respect to the number of one-worker jobs L_{1w}^{t-1} , shows that V_{ij}^t is monotone in L_{1w}^{t-1} . Hence, V_{ij}^t is smaller than the maximum of its boundary values:

$$\lim_{L_{1w}^{t-1} \rightarrow \frac{N}{2}} V_{ij}^t = \underbrace{\frac{(b_{1w}^{t-1}-1)\frac{N}{2}+L_{1w}^t}{U+(1+b_{1w}^{t-1})\frac{N}{2}} V_{1w,j}^t + \frac{L_{2w}^t}{U+(1+b_{1w}^{t-1})\frac{N}{2}} V_{2w,j}^t + \frac{U+L_{1w}^t}{U+(1+b_{1w}^{t-1})\frac{N}{2}} V_{ij}^{t+1}(1-r)}_{=H_{1,j}(b_{1w}^{t-1}, b_{2w}^{t-1})},$$

and

$$\lim_{L_{1w}^{t-1} \rightarrow 0} V_{ij}^t = \underbrace{\frac{L_{1w}^t}{U+b_{2w}^{t-1}N} V_{1w,j}^t + \frac{(b_{2w}^{t-1}-1)N+L_{2w}^t}{U+b_{2w}^{t-1}N} V_{2w,j}^t + \frac{U+L_{1w}^t}{U+b_{2w}^{t-1}N} V_{ij}^{t+1}(1-r)}_{=H_{2,j}(b_{1w}^{t-1}, b_{2w}^{t-1})}.$$

H_1 and H_2 represent the value of unemployment for maximal and minimal level of unemployment, respectively. Since the maximal value of unemployment is, thus, well defined for each combination of separation rates, so is the set \bar{B}^t . A necessary and sufficient condition for a vector of separation rates to be in \bar{B}^t is thus:

$$V_{ks}^t - \max \left\{ H_{1,s}(b_{1w}^{t-1}, b_{2w}^{t-1}), H_{2,s}(b_{1w}^{t-1}, b_{2w}^{t-1}) \right\} \geq \frac{c_n}{(1-r)(1-b_k^{t-1})}, \quad k \in \{1w, 2w\}.$$

In the following, we derive conditions for segmentation equilibria to arise, involving separation rates that induce selfish workers to work. By definition of \bar{B}^t , all separation rates in \bar{B}^t are lower than whatever is the separation rate between firms and shirkers, a_k^t , because a lower separation rate for non-shirkers is a key component of incentives. As discussed in the text (see p. 19), the data indicate that selfish types are indeed willing to work in pre-final periods of the experiment, before shirking in the final period, consistent with such an equilibrium.

A.2.3. Sufficient conditions for a segmentation equilibrium in pre-final periods. We begin by stating a condition that will be the key sufficient condition for the existence of a segmentation equilibrium.

Condition 1. *There exist $(b_{1w}, b_{2w}) \in \bar{B}^t$ such that for all $c_i \leq c_j$:*

- (a) $\Gamma_1^{t-1}(b_{1w}, b_{2w}, c_j, c_i) < z_{n+i} - z_j < \Gamma_2^{t-1}(b_{1w}, b_{2w}, c_j, c_i)$
- (b) $(1-b_k)N \leq L_k^t$

Γ_1^{t-1} and Γ_2^{t-1} will be defined in the course of the proof and depend on the fairness considerations of workers. They represent the difference in wages between one- and two-worker firms for the tightest labour market (Γ_1^{t-1}) and the least tight labour market (Γ_2^{t-1}). The essence of part (a) of Condition 1 is that the production function is (i) ‘‘sufficiently concave’’ such that the output differential between one-worker and two-worker firms cannot become too large and (ii) steep enough such that a one-worker strategy inducing high effort does not dominate a two-worker strategy with lower effort levels for all possible labour market conditions. Part (b) of Condition 1 is purely technical: it guarantees that the number of jobs in period t is larger or equal to the number of workers who have their contract renewed.

In the next step, we prove a lemma (Lemma 1) that states that a segmentation equilibrium exists in an arbitrary pre-final period $t-1$. The lemma assumes that Condition 1 holds for the arbitrary period, along with two additional assumptions: the first assumption is about continuation values and ensures that firms participate in the market. The second assumption is that firms believe that an arbitrary unemployed worker is equally likely to be a fair type as the currently employed worker ($\hat{\lambda}_U = \hat{\lambda}_E$), which rules out screening in previous periods. In a final step, we will prove a

proposition (Proposition 1), that says if Condition 1 holds for *all* pre-final periods, then there exists an equilibrium with segmentation in every pre-final period, without needing either of the two additional assumptions involved in the lemma.

Lemma 1. *Suppose Condition 1 is fulfilled, the continuation values satisfy $(1-r)V_{uf}^{t+1} < V_{2w,f}^t, V_{1w,f}^t$, and $\hat{\lambda}_U = \hat{\lambda}_E$ for all firms. Then there exists an equilibrium with a segmented labour market in period $t-1$ that exhibits the following properties:*

1. *Effort levels $e_{1w}^{t-1} \geq e_{2w}^{t-1}$ are realized in one- and two-worker firms, respectively, with wages $w_{1w}^{t-1} \geq w_{2w}^{t-1}$ such that fair workers are indifferent between working and shirking.*
2. *Fair and selfish workers exert effort for the given wages and shirk if they get paid less.*
3. *There is a number $L_{1w}^{t-1} > 0$ of one-worker firms, a number $L_{2w}^{t-1} > 0$ of two-worker firms, and a number $L_{1w}^{t-1} + U$ of unemployed agents.*
4. *Workers who do not exert the stipulated effort level or who are known to be selfish are fired with certainty, those who exert the contractually stipulated effort have separation rates of $(b_{1w}^{t-1}, b_{2w}^{t-1}) \in B$.*
5. *Unemployment is involuntary for both fair and selfish workers.*
6. *For both types of workers, jobs in one-worker firms yield higher rents than jobs in two-worker firms: $w_{1w}^{t-1} - c_{1w}^{t-1} + g(w_{1w}^{t-1} - c_{1w}^{t-1}) \geq w_{2w}^{t-1} - c_{2w}^{t-1} + g(w_{2w}^{t-1} - c_{2w}^{t-1})$ and $w_{1w}^{t-1} - c_{1w}^{t-1} \geq w_{2w}^{t-1} - c_{2w}^{t-1}$.*

To prove the lemma, we first characterize firms' optimal wage-effort schedules for given behaviour of workers and given separation rates. In a second step, we show that there is an intermediate number of one-worker firms and corresponding separation rates, such that the derived wage-effort schedules for one-worker and two-worker firms are equally profitable for firms. This gives rise to a segmentation equilibrium, if the offered wage-effort schedules are incentive compatible for workers, which we show in the last step.

In any period, firms decide first on the wage and stipulated effort level. For any level of effort, it is optimal for firms to pay wages that make fair workers indifferent between working and shirking: if a firm offered a contract with a lower wage, all workers would shirk, thereby decreasing firm profits. An offer of a higher wage inducing the same effort clearly also diminishes firm profits. We denote by $e_{1w}^{t-1}, e_{2w}^{t-1}$ the profit-maximizing levels of effort given the value of unemployment and implied wage payments.³⁴ The effort level in one-worker firms needs to be higher than in two-worker firms. Otherwise firms could profitably deviate, due to the decreasing returns to scale production function. To see this, suppose instead that $e_{1w}^{t-1} < e_{2w}^{t-1}$. From the optimal behaviour of the firms we know:

$$\begin{aligned} f(e_{1w}^{t-1}) - w_{1w}^{t-1} &\geq f(e_{2w}^{t-1}) - w_{2w}^{t-1} \quad \text{and} \quad f(2e_{2w}^{t-1}) - 2w_{2w}^{t-1} \geq f(2e_{1w}^{t-1}) - 2w_{1w}^{t-1} \\ &\Rightarrow 2(f(e_{2w}^{t-1}) - f(e_{1w}^{t-1})) \leq f(2e_{2w}^{t-1}) - f(2e_{1w}^{t-1}). \end{aligned}$$

This is a contradiction to Assumption 1.

We next turn to the participation decision of firms. Note that $(1-r)V_{uf}^{t+1} < V_{2w,f}^t, V_{1w,f}^t$ together with (A.4) implies that, for at least one firm type, future rents of employed fair workers exceed those of unemployed fair workers. The fair workers' no-shirking condition (A.2) thus implies that there is an incentive compatible wage-effort schedule such that the wage is below $g^{-1}(c_i) + c_i$. Since the production technology is efficient (Assumption 1), firms will offer contracts in the market. The decision of firms then boils down to deciding whether to employ one worker or two workers. A one-worker strategy is more profitable if:

$$\begin{aligned} z_{1w}^{t-1} - w_{1w}^{t-1} &> z_{2w}^{t-1} - 2w_{2w}^{t-1} \\ \Leftrightarrow 0 &< z_{1w}^{t-1} - (g^{-1}(c_{1w}^{t-1}) + (1-r)(b_{1w}^{t-1} - 1)(V_{1w,f}^t - V_{uf}^t) + c_{1w}^{t-1}) \\ &\quad - z_{2w}^{t-1} + 2(g^{-1}(c_{2w}^{t-1}) + (1-r)(b_{2w}^{t-1} - 1)(V_{2w,f}^t - V_{uf}^t) + c_{2w}^{t-1}) \equiv \Delta \end{aligned}$$

Whether this is the case depends on the difference in outputs between one-worker firms (z_{1w}^{t-1}) and two-worker firms (z_{2w}^{t-1}), and the tightness of the labour market, which determines V_{uf}^t . Δ is continuous in V_{uf}^t , and therefore a shift in the sign of this inequality leads to at least one level for the value of unemployment such that firms are indifferent between

34. If firms are indifferent between two or more levels of induced effort, we assume that there is a tie breaking rule that is homogeneous across firms.

both strategies.³⁵ This reversal in the sign of Δ is given if:³⁶

$$\begin{aligned} & \Gamma_1^{t-1}(b_{1w}^{t-1}, b_{2w}^{t-1}, c_{1w}^{t-1}, c_{2w}^{t-1}) \equiv 2g^{-1}\left(c_{2w}^{t-1} + (1-r)(b_{2w}^{t-1} - 1)(V_{2w,f}^t - H_{1,f})\right) + 2c_{2w}^{t-1} \\ & - g^{-1}\left(c_{1w}^{t-1} + (1-r)(b_{1w}^{t-1} - 1)(V_{1w,f}^t - H_{1,f})\right) - c_{1w}^{t-1} \\ & < z_{2w}^{t-1} - z_{1w}^{t-1} \\ & < 2g^{-1}\left(c_{2w}^{t-1} + (1-r)(b_{2w}^{t-1} - 1)(V_{2w,f}^t - H_{2,f})\right) + 2c_{2w}^{t-1} \\ & - g^{-1}\left(c_{1w}^{t-1} + (1-r)(b_{1w}^{t-1} - 1)(V_{1w,f}^t - H_{2,f})\right) - c_{1w}^{t-1} \equiv \Gamma_2^{t-1}(b_{1w}^{t-1}, b_{2w}^{t-1}, c_{1w}^{t-1}, c_{2w}^{t-1}) \end{aligned}$$

Condition 1 ensures that a combination of separation rates from the set \bar{B}^t exists such that this is fulfilled. Hence, there is a number L_{1w}^{t-1} such that firms are indifferent between both strategies. This gives rise to a segmentation equilibrium in which L_{1w}^{t-1} firms employ only one worker at a high level of effort. These firms do not fill their second vacancy and thus the equilibrium features endogenous unemployment. Note that the separation rates in any segmentation equilibrium must be larger than zero. If, by contrast, separation rates were zero, firms would renew all their contracts and there would be no vacancies in the next period, independently of the fraction of one-worker firms. The value of unemployment would then be independent of the current fraction of firms that ration jobs ($H_{1,j} = H_{2,j}$). Hence, the profitability of firm strategies would not depend on the labour market conditions ($\Gamma_1^{t-1} = \Gamma_2^{t-1}$), and one firm size would dominate the alternative strategy throughout.

After firms have observed work effort, they decide whether to renew a worker's contract. In Lemma 1 we assume that firms have no additional information about the type of agent they employ compared to unemployed agents ($\hat{\lambda}_U = \hat{\lambda}_E$), which is a characteristic of fair and selfish workers pooling. Hence, there is always an "equally good" worker available, implying that firms are indifferent between dismissing the worker and renewing her contract. All separation rates from the set \bar{B}^t are thus incentive compatible, and the equilibrium can involve strictly positive separation rates between firms and non-shirking workers (with Proposition 1 below, we no longer assume $\hat{\lambda}_U = \hat{\lambda}_E$; rather this follows from the proposition).³⁷

Turning to the workers, fair types by construction choose to exert the stipulated effort level. Furthermore, selfish workers comply with the contract, since the separation rates are in \bar{B}^t . Moreover, the no-shirking conditions (A.1) and (A.3) directly show that unemployment is involuntary. In particular, workers who fulfil their contract receive at least the same utility as shirking workers, who in turn receive the value of unemployment *plus* the current wage.

Part 6 of Lemma 1 implies that, from the workers' perspective, there are two systematically different types of jobs: "primary-sector" jobs that pay high rents for high efforts, and "secondary-sector" jobs with lower rent payments and lower efforts. This difference arises since firms induce the profit-maximizing effort in either firm type. To see this, first suppose that rents for selfish workers are higher in the low-effort jobs of two-worker firms:

$$\begin{aligned} w_{1w}^{t-1} - c_{1w}^{t-1} \leq w_{2w}^{t-1} - c_{2w}^{t-1} & \Rightarrow 2w_{1w}^{t-1} - 2w_{2w}^{t-1} < f(2e_{1w}^{t-1}) - f(2e_{2w}^{t-1}) \\ & \Leftrightarrow f(2e_{2w}^{t-1}) - 2w_{2w}^{t-1} < f(2e_{1w}^{t-1}) - 2w_{1w}^{t-1} \end{aligned}$$

The second inequality follows from $e_{1w}^{t-1} \geq e_{2w}^{t-1}$ (Part 1 of Lemma 1) and the efficiency of the production function (Assumption 1). The last inequality is a contradiction to the profit maximization of firms. Hence, rents need to be higher

35. If the optimal induced effort choice changes in the course of varying the number of one-worker firms from zero to $\frac{N}{2}$, the segmentation equilibria may feature three different levels of effort.

36. There is also a segmentation equilibrium if $\lim_{L_{1w}^{t-1} \rightarrow \frac{N}{2}} \Delta > 0 > \lim_{L_{1w}^{t-1} \rightarrow 0} \Delta$. The following analysis is also valid for this case.

37. Note that in a no-shirking equilibrium with pooling, an individual firm has no incentive to deviate from the equilibrium separation rate. Firms cannot commit *ex ante*, at the time of hiring, to a particular contract renewal rate, because if effort is non-verifiable, effort-contingent contract renewal is not verifiable either. After being hired, workers thus choose their effort based on the equilibrium separation rate. When firms subsequently make re-hiring decisions, effort is already determined and firms have no incentive to deviate from the prevailing equilibrium separation rate. While firms might prefer an equilibrium with a different separation rate, this is the usual coordination problem with multiple equilibria.

for selfish types in high-effort jobs than in low-effort jobs. The monotonicity of g implies that the same ranking also holds for fair types.³⁸ This concludes the proof of Lemma 1.

Proposition 1. *If Condition 1 is fulfilled for all pre-final periods, there exists an equilibrium with a segmented labour market and involuntary unemployment in all pre-final periods.*

Given that Condition 1 is satisfied for all pre-final periods, there is an equilibrium with a segmented labour market in all pre-final periods, if (i) $(1-r)V_{uf}^{t+1} < V_{2w,f}^t, V_{1w,f}^t$ and (ii) $\hat{\lambda}_U = \hat{\lambda}_E$ hold for any arbitrary period $t-1$. For the second-to-last period, (i) holds trivially, and maintaining (ii), there is a segmentation equilibrium in that period. But then, the no-shirking condition (A.2) implies that $(1-r)V_{uf}^{t+1} < V_{2w,f}^t, V_{1w,f}^t$ holds from the perspective of the third-to-last period as well, and there is again segmentation. Using backward induction, and maintaining (ii), (i) is thus satisfied for all periods. Given (i) holds in all periods, workers behave homogeneously in every period, *i.e.* there is pooling of fair and selfish workers. In the first period firms have the prior $\hat{\lambda}_U = \hat{\lambda}_E = \lambda$, and due to pooling, this does not change in subsequent periods until the end of the final period, implying that (ii) holds for all periods. Hence, there exists a segmentation equilibrium with involuntary unemployment in all periods.

A.3. Equilibrium with verifiable effort

We now turn to the case of contractible effort. Consider first the final period of the game. In this period, selfish workers accept a contract offer if the wage at least covers the stipulated effort costs. Fair workers will accept a contract if $w - c(e) + g(w - c(\hat{e})) \geq 0$.

Given our assumptions about \tilde{x} , firms need to pay positive (but potentially low) monetary rents to induce fair workers to accept an offer. Since the production function is efficient, firms always offer a contract rather than staying out of the market. The optimal terms of the contract depend on the subjective probability of facing a fair worker, $\hat{\lambda}$. If, for instance, $\hat{\lambda} = 1$, a firm expects to face a fair worker with certainty and thus pays wages equal to the sum of effort and the psychological costs. The efficiency of the production technology implies that employment of two workers with maximal effort exertion is optimal in this case. Thus, there exists a cutoff $\tilde{\lambda}$ such that for all beliefs $\hat{\lambda} \geq \tilde{\lambda}$, firms use a homogeneous contracting strategy of hiring two workers, that involves paying wages to cover both types of costs, and workers exert the maximum effort level.

Maintaining $\hat{\lambda} > \tilde{\lambda}$ for all firms, in the second-to-last period backward induction implies that fair workers accept the same offer as before, since they do not expect any future utility rents. In contrast, selfish workers may anticipate a rent in the next period. Nevertheless, they do not accept any offer involving negative current rents, because this would reveal their type. If they did, the firm would have an incentive to renegotiate in the next period, and offer rents equivalent to the value of unemployment, which is zero in the last period. However, selfish workers always accept the contract needed to employ fair types, which ensures positive current and future rents. For any stipulated level of effort, the rent needed to hire selfish workers is therefore weakly larger than in the last period, while the rent for fair types is the same. This makes it even less attractive than in the last period for firms to offer a contract attracting only selfish workers, thus firms offer the same contract terms as in the last period.

Assuming that there are enough fair types in the population such that $\lambda \geq \tilde{\lambda}$, this argument holds for all previous periods. This implies for all periods that (i) firms offer a wage-effort schedule that induces maximal effort provision, (ii) firms always hire two workers and the level of unemployment is thus minimal, and (iii) all workers accept posted contracts.

Acknowledgments. We thank the editor, Imran Rasul, and four anonymous referees for their thoughtful comments. We also thank Johannes Abeler, Thomas Dohmen, Ernst Fehr, Simon Gächter, Paul Heidhues, Peter Kuhn, Daniel Müller, Moritz Ritter, Joel Sobel, Konstantinos Tatsiramos, and Matthias Wibrall for helpful discussions, and conference and seminar participants at Alicante, Bonn, the ESA conference at NYU, ESSLE, Graz, Maastricht, Milan, Nottingham, Oxford, Pittsburgh, Rotterdam, Rutgers, Swarthmore College, Temple, the University of Tennessee, Wharton, and Zurich for valuable comments. Franziska Tausch and Nicolas Kaufung provided excellent research assistance. Financial support from the Deutsche Forschungsgesellschaft through SFB/TR 15 is gratefully acknowledged.

38. Another way to assess the job quality in one- and two-worker firms is by comparing workers' expected lifetime utilities. It follows directly from the no-shirking conditions that, also from this perspective, fair workers prefer jobs in one-worker firms over those in two-worker firms. For selfish workers the same follows necessarily if one-worker firms offer more stable employment relationships than two-worker firms, *i.e.* if rehiring rates are higher in one-worker firms. Empirically, we find that this is the case, with rehiring rates of 79.0% in one-worker firms and 61.6% in two-worker firms.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

- ABELER, J., ALTMANN, S., KUBE, S. and WIBRAL, M. (2010), "Gift Exchange and Workers' Fairness Concerns: when Equality is Unfair", *Journal of the European Economic Association*, **8**, 1299–1324.
- ABOWD, J. M., KRAMARZ, F. and MARGOLIS, D. N. (1999), "High Wage Workers and High Wage Firms", *Econometrica*, **67**, 251–333.
- AKERLOF, G. A. and YELLEN, J. L. (1990), "The Fair Wage-effort Hypothesis and Unemployment", *Quarterly Journal of Economics*, **105**, 255–283.
- ALBRECHT, J. and VROMAN, S. (1992), "Dual Labor Markets, Efficiency Wages, and Search", *Journal of Labor Economics*, **10**, 438–461.
- ALTMANN, S., FALK, A., GRUNEWALD, A. and HUFFMAN, D. (2011), "Implicit Contracts, Unemployment, and Labor Market Segmentation" (Working Paper, University of Bonn).
- BANDIERA, O., BARANKAY, I. and RASUL, I. (2011), "Field Experiments with Firms", *Journal of Economic Perspectives*, **25**, 63–82.
- BARTLING, B., FEHR, E. and SCHMIDT, K. M. (2012), "Screening, Competition, and Job Design: Economic Origins of Good Jobs", *American Economic Review*, **102**, 834–864.
- BEWLEY, T. (1999), *Why Wages Don't Fall During a Recession* (Cambridge, MA: Harvard University Press).
- BOARD, S. and MEYER-TER-VEHN, M. (2013), "Relational Contracts in Competitive Labor Markets" (Working Paper, UCLA).
- BOHNET, I., GREIG, F., HERRMANN, B. and ZECKHAUSER, R. (2008), "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States", *American Economic Review*, **98**, 294–310.
- BROWN, C. and MEDOFF, J. (1989), "The Employer Size – Wage Effect", *Journal of Political Economy*, **97**, 1027–1059.
- BROWN, M., FALK, A. and FEHR, E. (2004), "Relational Contracts and the Nature of Market Interactions", *Econometrica*, **72**, 747–780.
- BROWN, M., FALK, A. and FEHR, E. (2012), "Competition and Relational Contracts: the Role of Unemployment as a Disciplinary Device", *Journal of the European Economic Association*, **10**, 887–907.
- BROWN, M. and SERRA-GARCIA, M. (2012), "Debt Enforcement and Relational Contracting" (Working Paper, University of Munich).
- BROWN, M. and ZEHNDER, C. (2007), "Credit Reporting, Relationship Banking, and Loan Repayment", *Journal of Money, Credit and Banking*, **39**, 1883–1918.
- BULOW, J. I. and SUMMERS, L. H. (1986), "A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment", *Journal of Labor Economics*, **4**, 376–414.
- CAMPBELL III, C. M. and KAMLANI, K. S. (1997), "The Reasons for Wage Rigidity: Evidence from a Survey of Firms", *Quarterly Journal of Economics*, **112**, 759–789.
- CAPPELLI, P. and CHAUVIN, K. (1991), "An Interplant Test of the Efficiency Wage Hypothesis", *Quarterly Journal of Economics*, **106**, 769–787.
- CHARNESS, G. and DUFWENBERG, M. (2006), "Promises and Partnerships", *Econometrica*, **74**, 1579–1601.
- CHARNESS, G. and KUHN, P. J. (2007), "Does Pay Inequality Affect Worker Effort? Experimental Evidence", *Journal of Labor Economics*, **25**, 693–723.
- CHARNESS, G. and KUHN, P. J. (2011), "Lab Labor: What can Labor Economists Learn from the Lab?", in Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics*, Vol. 4. (Amsterdam: Elsevier) 229–330.
- COHN, A., FEHR, E. and GÖTTE, L. (2012), "Fair Wages and Effort: Evidence from a Field Experiment" (Working Paper, University of Zurich).
- DOERINGER, P. B. and PIORE, M. J. (1971), *Internal Labor Markets and Manpower Analysis* (Lexington, MA: Heath Lexington Books).
- ELLINGSEN, T. and JOHANNESSON, M. (2004), "Promises, Threats and Fairness", *The Economic Journal*, **114**, 397–420.
- FALK, A. and FISCHBACHER, U. (2006), "A Theory of Reciprocity", *Games and Economic Behavior*, **54**, 293–315.
- FEHR, E. and GÄCHTER, S. (2000), "Fairness and Retaliation: The Economics of Reciprocity", *Journal of Economic Perspectives*, **14**, 159–181.
- FEHR, E., KIRCHSTEIGER, G. and RIEDL, A. (1993), "Does Fairness Prevent Market Clearing? An Experimental Investigation", *Quarterly Journal of Economics*, **108**, 437–460.
- FEHR, E., KLEIN, A. and SCHMIDT, K. M. (2007), "Fairness and Contract Design", *Econometrica*, **75**, 121–154.
- FEHR, E. and SCHMIDT, K. M. (1999), "A Theory of Fairness, Competition and Cooperation", *Quarterly Journal of Economics*, **114**, 817–868.
- FISCHBACHER, U. (2007), "Z-tree: Zurich Toolbox for Ready-made Economic Experiments", *Experimental Economics*, **10**, 171–178.
- FUDENBERG, D. and MASKIN, E. (1986), "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information", *Econometrica*, **54**, 533–554.

- GÄCHTER, S. (2010), "(Dis) advantages of Student Subjects: What is Your Research Question?", *Behavioral and Brain Sciences*, **33**, 92–93.
- GÄCHTER, S., NOSENZO, D. and SEFTON, M. (2012), "The Impact of Social Comparisons on Reciprocity", *Scandinavian Journal of Economics*, **114**, 1346–1367.
- GREINER, B. (2003), "An Online Recruitment System for Economic Experiments", in Kremer, K. and Macho, V. (eds), *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, (Göttingen: Ges. für Wiss. Datenverarbeitung) 79–93.
- ICHINO, A. and RIPHAHN, R. (2005), "The Effect of Employment Protection on Worker Effort: Absenteeism during and after Probation", *Journal of the European Economic Association* **3**(1), 120–143.
- KATZ, L. F. (1986), "Efficiency Wage Theories: A Partial Evaluation", in Fischer, S. (ed), *NBER Macroeconomic Annual* (Cambridge, MA: MIT Press).
- KUBE, S., MARÉCHAL, M. and PUPPE, C. (2013), "Do Wage Cuts Damage Work Morale? Evidence from a Natural Field Experiment", *Journal of the European Economic Association*.
- LINARDI, S. and CAMERER, C. (2012), "Can Relational Contracts Survive Stochastic Interruptions? Experimental Evidence", (Working Paper, California Institute of Technology).
- MACLEOD, W. B. (2007), "Can Contract Theory Explain Social Preferences?", *American Economic Review*, **97**, 187–192.
- MACLEOD, W. B. and MALCOMSON, J. M. (1989), "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment", *Econometrica*, **57**, 447–480.
- RABIN, M. (1993), "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, **83**, 1281–1302.
- SAINT-PAUL, G. (1996), *Dual Labor Markets – A Macroeconomic Perspective*, 1st edn (Cambridge, MA: MIT Press).
- SHAPIRO, C. and STIGLITZ, J. E. (1984), "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review*, **74**, 433–444.
- VANBERG, C. (2008), "Why do People Keep their Promises? An Experimental Test of Two Explanations", *Econometrica*, **76**, 1467–1480.