Labour Economics 26 (2014) 34–42

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Collective versus decentralized wage bargaining and the efficient allocation of resources ☆

Xiaoming Cai [a], Pieter A. Gautier [a,*], Coen N. Teulings [b,c], Makoto Watanabe [a]

[a] VU University Amsterdam, Tinbergen Institute, The Netherlands
[b] University of Cambridge, United Kingdom
[c] University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 8 January 2013
Received in revised form 4 October 2013
Accepted 15 November 2013
Available online 28 November 2013

Keywords:
Collective bargaining
Efficiency
On-the-job search
Pay scales
Wage formation
Unions

ABSTRACT

In a search model with two sided heterogeneity and on-the-job search, we compare collective bargaining agreements (CBA) with a decentralized bargaining outcome case. Under CBA, a union chooses a pay-scale schedule and the firm can select a wage from this pay scale after observing match quality. An advantage of collective bargaining agreements is that search and business-stealing externalities can be internalized. A disadvantage is that it takes more time before an optimal allocation is reached. What the most desirable system is, depends on worker bargaining power (β) and the relative efficiency of on- versus off- the job search. We find both for the Netherlands and the US that as long as β lies between 0.1 and 0.7, CBA is less desirable.

1. Introduction

Under collective bargaining agreements (CBA), wages are determined at the industry level rather than at the individual level.[1] An advantage of collective wage bargaining is that potential externalities can be internalized while a disadvantage is that the allocative role of wages is reduced which leads to a sub optimal allocation of workers over jobs. To study this trade-off we need a model that allows for two sided heterogeneity and search frictions. Heterogeneity is important because we are interested in allocation. Search frictions are important because it takes time for workers to find the production units where they are needed most. Wages play in this setting a key role because they inform workers which firms need them most.

There exists a lot of cross-country variation in coverage rates of CBA, see OECD (2004, 2012). In the US, the fraction of workers who are covered by a collective agreement has been falling from 26% in 1980 to 13% in 2010. In the Netherlands, this fraction has been increasing over time from 70% in 1980 to over 80% in 2012. In Germany the CBA coverage rate fell from 80% in 1980 to 60% in 2010. The relation between coverage rates and wage dispersion has also been studied extensively. Card (1996) and Dinardo et al. (1996) give empirical evidence that unions compress wages and Blau and Kahn (1999), Hartog et al. (2002) and OECD (2004) show that in countries where coverage is high there is less wage dispersion.

To study the relation between CBA, wage dispersion and the allocation of workers we use a model similar to Marimon and Zilibotti (1999) and Gautier et al. (2010). The idea is that worker (s) and job types (c) are located on a circle and productivity is decreasing in the distance, x, between s and c.[2] In the simplest version of the centralized-wage setting case, all firm types offer the socially optimal wage under the constraint that it is the same for all job types while in the decentralized case, firm types are allowed to post different wages to different worker types. We assume that firms *cannot* ex ante commit to a wage schedule (if firms can commit, the decentralized outcome is more favorable).[3] The reason that firms pay positive wages that are increasing in productivity, even if they have all the bargaining power is that a higher wage reduces the quit rate.

Gautier et al. (2010) show that in this case firms engage in excessive vacancy creation due to a *business-stealing effect*. The idea is that firms do not internalize the output loss of firms they poach a worker from, in particular, they do not care whether they destroy relatively good or bad matches. Although each worker's transition is efficient, the

☆ We thank the editor, Chris Flinn, and two referees for useful comments. We also thank NWO and the Dutch Ministry of Social Affairs for a research grant.

* Corresponding author at: Department of Economics, VU Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, The Netherlands.
E-mail addresses: x.cai@vu.nl (X. Cai), p.a.gautier@vu.nl (P.A. Gautier), cnt23@cam.ac.uk (C.N. Teulings), makoto.wtnb@gmail.com (M. Watanabe).

[1] Even if unionization rates are low, CBA coverage can be substantial. In the Netherlands, the minister can for example make a CBA binding for an entire sector.

[2] The results of Gautier, Teulings and van Vuuren (2005) show that without on-the-job search, the circular model has the same characteristics as a Taylor expansion of the hierarchical model.

[3] Under continuous renegotiation, ex ante commitment becomes meaningless.

0927-5371/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.labeco.2013.11.003

expected marginal increase in aggregate output is too low to justify the entry cost. In the simplest CBA case where the union can set only one wage, the wage can be set at a level that generates the efficient level of entry. However, if all firms in a sector pay the same wage, workers do not know which firms need their services most and once they have a job they stop searching. We are interested in the trade-off between efficient entry and efficient allocation and which wage mechanism is most desirable from a social welfare point of view.

The desirability of CBA mainly depends on the efficiency of on-the-job search relative to off the job search (Ψ). If employed job seekers receive more than 20% of the number of offers as the unemployed workers (for the commitment case it is more than 10%), the cost of CBA exceeds the benefits. The reason for this is that the more efficient on-the-job search is, the faster workers flow to the jobs where they are needed most and the more costly it is if CBA prevents firms within an industry to pay a higher wage if a particular worker type is very valuable for them. One novelty relatively to Gautier et al. (2010) is that we also solve the model for general bargaining power. This allows us to derive a relationship between the desirability of CBA and a worker's bargaining power. For the estimated values of Ψ, the decentralized case performs better for most values of β. For intermediate values of β, the threshold value of Ψ above which decentralized wage setting performs better goes down because the business-stealing-externality is reduced. For very high values of β (above 0.8), the standard congestion externalities become more important and too few vacancies are created under decentralized bargaining.

In most countries, CBA takes the form of pay scales. Therefore, in Section 5, we allow the union to choose upon n different wages. Given this set of wages, a firm decides after observing the quality of the match, which of the n wages it offers to the worker. To our knowledge, this is the first paper that endogenizes union-based pay scales. We find that $n = 4$ performs almost as well as $n \to \infty$. CBA with sufficiently many pay scales ($n \geq 4$) is only socially more efficient for very low β's (≥ 0.1) and for very high β's (≥ 0.7) than the decentralized wage setting scheme. The main reason for this is that they set the lowest wage in the pay scale too high. If pay scales contain too few wages, there are again not enough transitions from bad to good matches.

There are a number of other papers that study the effect of centralized bargaining in frictional labor markets. Lindbeck and Snower (1986) consider insider-wage setting. Pissarides (1986) asks whether the standard search externalities will be internalized by a union. He finds that this is the case only if the union's policy is chosen by unemployed persons. If employed persons can influence the union's policy, unemployment and wages will be too high. We find that in a dynamic setting with modest discount rates, it matters very little whether unions maximize the expected welfare of the average employed worker or of unemployed workers. In both cases, the union realizes that at some point in time, employed workers may become unemployed and that too high wage demands are welfare reducing because it reduces vacancy creation. Therefore, in a dynamic setting, the negative welfare effects of insider-wage setting are a lot smaller than in a static model. Other models that have frictions and centralized bargaining include Delacroix (2006) and Bauer and Lingens (2011). None of those models look at the trade-off between an efficient allocation and internalizing externalities as we do. Teulings and Hartog (1998) argue that an advantage of CBA is that it reduces hold-up problems (because individual workers and firms cannot influence the wage) while at the same time it also allows wages to respond to aggregate shocks. This paper ignores this and only looks at wage dispersion across jobs. Finally, Krusell and Rudanko (2012) focus on union wage setting rather than CBA's. They assume that in the short run, unions raise current wages above the efficient level, in order to appropriate surpluses from firms with existing matches. We abstract from that here but since this makes the CBA perform worse, it will not change our main conclusions about the performance of decentralized wage setting and CBA.

The paper is organized as follows. Section 2 starts with the assumptions, derives the equilibrium conditions, and characterizes the equilibrium. Section 3 discusses the two wage mechanisms. Section 4 conducts welfare analysis and Section 5 introduces pay scales and unions maximizing the value of employment rather than unemployment. Finally, Section 6 concludes.

2. The model

2.1. Assumptions

For the decentralized case, we use the model of Gautier et al. (2010) which extends Marimon and Zilibotti (1999) to allow for on-the-job search. The model is briefly summarized below. Worker types (s) and job types (c) are locations on a circle with circumference 1. The production technology has constant-returns-to scale so it is easiest to think of firms as consisting of one worker. A matched firm–worker pair produces Y which depends on the "spherical distance" between s and c: $x(s,c) = min\{|s - c|, 1 - |s - c|\}$, which is common knowledge to both the worker and the firm. Note $0 \leq x(s,c) \leq 1/2$. Specifically, $Y(s,c) = Y(x)$. Since we interpret x as an indicator of mismatch between workers and jobs, $Y(x)$ has a maximum at 0, and the value of the maximum is normalized to unity: $Y(0) = 1$. We assume that $Y(x)$ is twice differentiable and strictly quasiconcave. If we think of $Y(x)$ as a second order Taylor approximation of a more general (differentiable) production function around the optimal assignment, the derivative of $Y(x)$ at 0 should be 0. The simplest functional form that meets those criteria is,

$$Y(x) = 1 - \frac{1}{2}\gamma x^2.$$

Low values of γ imply that a precise match is not very important. In the limit, $\gamma \to 0$, the model reduces to a standard Diamond–Mortensen–Pissarides type of matching model with identical workers and firms.

We assume that both labor and vacancy supply are uniformly distributed over the circle (the latter can be shown to be an equilibrium). Total labor supply in period t equals $L(t)$ and the total number of vacancies per unit of labor supply is given by $v(c) = v$. The flow cost of maintaining a vacancy is equal to K per period and the flow value of non-market time is B.

We assume that the discount rate ρ equals the population growth rate (golden growth) and that all new workers start out as unemployed. The implications of this assumption are the same as when we assume that the discount rate ρ is much smaller than the job-finding and separation rate, $\rho \gg \delta,\lambda$. This is a common assumption in the wage-posting literature (see for example, Burdett and Mortensen, 1998).

Next, we discuss the job search technology. Let m be the total number of contacts between job seekers and vacancies per unit of labor supply and u be the unemployment rate. We think it is reasonable to assume that two workers with an empty intersection of matching sets do not cause congestion on each other. Therefore, we take a quadratic contact technology,

$$m = \lambda[u + \psi(1-u)]v.$$

The parameter $\Psi, 0 \leq \Psi \leq 1$, measures the relative efficiency of on-the-job search versus search while unemployed. Marimon and Zilibotti (1999) consider the case, $\Psi = 0$, which is related to the stochastic matching model of Pissarides (2000). If off- and on-the-job search are equally efficient, $\Psi = 1$, the model is relatively simple and analytical results can be obtained. For the general Ψ case we rely on numerical simulations. The overall efficiency of the matching process is captured by λ. The Walrasian equilibrium is obtained for $\lambda \to \infty$. Finally, matches between workers and jobs are destroyed at an exogenous rate $\delta > 0$.

We focus on two wage-setting schemes. First, we add CBA to this framework. The simplest implementation is to interpret CBA as the