Imagine a group of $N$ identical workers, all of whom are “permanently” attached to a firm. The workers are risk averse; the firm is risk neutral. The future value of the workers’ labor to the firm, $\theta$, is unknown today, but the workers and firm would like to come up with a Pareto-efficient arrangement today that stipulates, for every possible $\theta$, how many workers will be hired and the incomes of every one of them “tomorrow” (when $\theta$ is realized). The reason firms and workers might want to do this in advance is to take advantage of some gains from trade (involving the provision of insurance by the risk-neutral firm to the risk averse workers) that could not be realized in an ex-post spot market for labor services.

Once this arrangement (or “implicit contract”) is in place, the covariation of employment and wages we observe will simply consist of the “playing out” of the rules of the contract in response to the realizations of $\theta$. As we demonstrate below, this covariation can be very different from that generated by “spot” labor markets which operate after $\theta$ is realized (and can make both workers and firms better off).

1. The base-case model (“$B$” unconstrained)

In our initial treatment, we will imagine that there is the possibility of private UI (payments by the firm to its unemployed workers), optimally chosen to be part of an efficient contract. We will change this assumption later.

Throughout our treatment here, we will assume that employment adjusts only on the “extensive” margin (i.e. the number of people employed), so there is no variation in hours per worker. As a result, we can say that each worker supplies one unit of labor to the firm if employed and zero units otherwise. This assumption is relaxed in the Burdett-Wright and van Audenrode papers on the reading list, which consider the effects of short-time working and short-time-compensation. We will also assume that, if some workers are not employed in a given state of nature ($\theta$), these workers are chosen at random from the group. Since all workers (by assumption) are permanently attached to the firm, we interpret these episodes of nonwork as temporary layoffs. The phenomenon of layoffs by inverse seniority, plus a more continuous distinction between insiders and outsiders, is explored in my own paper on the reading list. Finally, we will assume that the realized value of the state, $\theta$, is sufficiently observable to firms, workers and outside authorities that binding, contingent contracts can be based on it. The cases of asymmetric information regarding $\theta$, and of “nonverifiable” $\theta$’s, are considered, among others, by O. Hart, “Optimal Labor Contracts Under Asymmetric Information: An Introduction”, ReStud January 1983.

In more detail, let the value of the firm’s output in state $\theta$ be given by $\theta F(L)$ (so the “demand” shock $\theta$ acts just like an output price shock: you can think of it this way if you like). Denote the density of $\theta$ by the continuous function $g(\theta)$; this density function is known to both workers and firms when they are designing the optimal labor contract.

Let a worker’s utility if he/she is employed be given by $U(w)$, where $w$ is wage income received if working, $U'>0$, and $U''<0$.

Let a worker’s utility if he/she is not employed be given by $U(\bar{w}+B)$, where $B$ is a payment made by the firm to its unemployed workers. (Think of $B$ as a kind of “private” UI chosen optimally in the labor contract. In unionized auto workers’ contracts this is called SUB: “supplementary unemployment benefits”). The parameter $\bar{w}$ captures the total value (in dollar-equivalent terms) of (a) the extra leisure you get when you’re not working; (b) any government-funded UI benefits you receive; plus (c) any income (net of leisure costs) you might earn from alternative employment while on temporary layoff from your “main” job. Note that our formulation of utility assumes that consumption and leisure are perfect substitutes in the sense that $U(\text{Consump, Leisure})=U(\text{Consump}+\text{Leisure})$. (It is easiest to conceptualize risk aversion in this

---

1 Also, the practice problem for this topic considers the case where adjustment is all on the intensive margin (hours per worker).
Note that there are three endogenous variables here: \( L, w \) and \( B \). In general, a contract in this world consists of an ordered triplet, \((L, w, B)\) for every possible state of nature, \( \theta \). Since there is a continuum of possible states of nature, \( \theta \), a contract consists of three continuous functions, \( L(\theta), w(\theta), \) and \( B(\theta) \), defined on the domain of \( g(\theta) \).

What is an optimal contract? We use the concept of Pareto-optimality, i.e. a contract is optimal if neither party can be made better off without making the other worse off. The two parties in this case are (a) the firm, and (b) a representative worker. One justification for arguing that real-world contracts are Pareto-optimal is that, whatever the ultimate distribution of the gains from trade, sensible negotiators in a long-run relationship ought not to waste opportunities for mutually beneficial exchange. Alternatively, one can imagine a labor market that is ex ante strongly competitive, in the sense that identical workers are freely mobile among firms and there is free entry by firms. Such a competitive market for contracts would imply that no firm can attract workers unless it offers the contract that maximizes workers’ expected utility subject to a zero profit constraint. In this interpretation, labor markets are competitive ex ante, but ex post workers are tied to firms in long-run relationships. The question of the feasibility of insurance contracts when workers are free to quit firms ex post (and when firms may want to renege on their commitments to workers ex post) has been addressed by many authors, including Harris and Holmstrom, ReStud 1982).

Mathematically, Pareto-optimal contracts can be found by maximizing one party’s utility, holding the other’s utility fixed at some predetermined level. Although one could do it either way, we’ll proceed by maximizing the expected utility of a representative worker subject to a fixed level of expected profits. (the latter would be zero if a free-entry condition applied to firms, but the results derived below don’t depend on the level at which we fix firms’ profits).

Thus the Pareto-optimal contract chooses three functions, \( L(\theta), w(\theta), \) and \( B(\theta) \), to maximize:

\[
E_{\theta} \left\{ \frac{L}{N} U(w) + \frac{N-L}{N} U(\overline{w} + B) \right\}
\]

subject to:

\[
E_{\theta} \left\{ \theta F(L) - wL - B(N - L) \right\} \geq \overline{\pi}
\]

(Strictly speaking, one should also impose the inequality constraint \( L \leq N \), but we won’t bother with this detail here.) Equation (1) uses the fact that a representative worker’s employment probability in any state, given random allocation of layoffs, is just \( \frac{L}{N} \).

To solve this, first simplify notation by multiplying the maximand (1) by the exogenous constant \( N \). (This won’t change the solution). Then, associating the Lagrange multiplier \( \lambda \) with the expected profit constraint (2) allows us to derive the first-order conditions:

w.r.t. \( L \):

\[
U(w) - U(\overline{w} + B) + \lambda \theta F'(L) - \lambda w + \lambda B = 0, \forall \theta
\]

w.r.t. \( w \):

\[
LU'(w) - \lambda L = 0, \forall \theta
\]

w.r.t. \( B \):

\[
(N - L)U'(\overline{w} + B) - \lambda (N - L) = 0, \forall \theta
\]

(mini)-Exercise: why does \( g(\theta) \) not appear in these conditions? Explicitly write out the Lagrangian and show why it disappears.
Note that there is a continuum of conditions for each of \( L, w, \) and \( B, \) with a different condition associated with each possible realization of \( \theta. \) Aside: this technique of optimally choosing an entire function by differentiating the objective function with respect to the value of the function at a continuum of points in its domain is called “pointwise optimization”. It just extends standard multivariable optimization to a continuum of choice variables. You could derive the same solution for the case of a finite number of possible states of nature, \( \theta, \) the usual way, but it seems more natural to think of the technology or output prices represented by \( \theta \) as continuous.

Simplifying (4) yields:

\[
U'(w) = \lambda, \ \forall \ \theta
\]  

(6)

In words, the marginal utility (and therefore the income, \( w \)) of employed workers should be the same regardless of the state of nature, \( \theta. \) Since workers are risk averse and firms risk neutral, it is efficient for firms to fully insure workers in this sense.

Simplifying (5) yields:

\[
U' (\bar{w} + B) = \lambda, \ \forall \ \theta
\]  

(7)

Thus, firms should structure private unemployment benefits, \( B, \) so that the marginal utilities of all unemployed workers (and therefore each worker’s “effective income”, \( \bar{w} + B \)) are unaffected by \( \theta \) as well. Since workers are risk averse and firms risk neutral, it is efficient for firms to fully insure workers in this sense as well. Assuming \( \bar{w} \) is independent of \( \theta, \) this means that optimal supplementary unemployment benefits paid to each unemployed worker, \( B, \) should be independent of \( \theta \) as well.

Combining (6) and (7) (and using the fact that \( U' \) is a strictly monotone function) yields:

\[
w = \bar{w} + B, \ \forall \ \theta
\]  

(8)

Thus, “effective” income should be the same, whether the worker has a job or not. (This does not necessarily imply that money income is the same, as \( \bar{w} \) can include the value of unemployed workers’ leisure). Since workers are risk averse and firms risk neutral, it is efficient for firms to fully insure workers in this third sense as well. Note the conceptual and mathematical parallels between the optimal income smoothing across states of nature derived here and the optimal income smoothing across time periods in the life cycle model of labor supply.

Finally, substituting (8) into (3) and simplifying yields:

\[
\theta F'(L) = \bar{w}, \ \forall \ \theta
\]  

(9)

Thus, the firm uses a different amount of labor in each state, \( \theta. \) It hires labor until its VMP just equals \( \bar{w}, \) which is the social opportunity cost of labor from the point of view of this worker-firm coalition.

Note that in this world firms (randomly) assign workers to layoff or work status. (Even though labor markets –i.e. markets for contracts-- may be ex ante competitive, what happens after a worker signs up with a firm is not a “neoclassical” case of workers or firms choosing what to do based on market prices, i.e. the wage). In this particular case of the model, workers will be indifferent to which status they are assigned. It is noteworthy that this result does NOT generalize to utility functions in which leisure and consumption are not perfect substitutes. In fact, in the world where \( U_{\text{Leisure}} > 0, \) it will be efficient to have both income and leisure be higher among the unemployed: all workers will prefer to be laid off, but only some will be allowed to go. When \( U_{\text{Leisure}} < 0 \) but Consumption and Leisure are not perfect substitutes, it is possible for layoffs to be (ex post) dispreferred by workers, even though it is optimal ex ante from workers’ point of
view to agree to a contract in which involuntary layoffs are a possibility. (you will demonstrate closely related results in the practice problem; also see Rosen 1985).

**Trivial exercise:** illustrate the FOC in (9) for two different values of \( \theta \) (states of nature/product demand). Show diagrammatically that an increase in federally-funded unemployment insurance benefits (i.e. an exogenous rise in \( \overline{w} \)) (a) raises the fraction of workers on temporary layoff in both states of nature, and (b) raises the difference in employment levels between the two states, i.e. it increases the cyclical amplitude of employment fluctuations.

2. The \( B=0 \) case.

The above model has at least two troubling features. First, very few firms provide supplementary unemployment benefits. Second, one reason these models were initially developed was to provide an explanation for downward wage rigidity. This was seen as a way to help explain the involuntary unemployment that occurs during recessions. Unfortunately, while the model does generate wage rigidity, condition (9) means that it generates exactly same amount and pattern of employment as a spot labor market. Further, as noted we can only get involuntary unemployment when \( UCL < 0 \), (otherwise workers prefer unemployment). It seems hard to imagine that the existence of involuntary unemployment would rest on such a fine data point on which we have little evidence or intuition.

A (somewhat less than satisfactory) solution to both the above problems, adopted by quite a few authors including Feldstein, can be had by simply forcing \( B \) to equal zero. While one could justify a \( B=0 \) constraint with a number of arguments (for example a positive \( B \) might inefficiently reduce search for a new job among the temporarily laid-off workers), this justification would have to come from outside the model, since it is not generated within the model itself.

When \( B=0 \), the FOC in (5) no longer apply, because \( B \) is no longer a choice variable. The FOC for \( L \) and \( w \) in (3) and (4) now become:

\[
w.r.t. \ L:\quad U(w) - U(\overline{w}) + \lambda \{F'(L) - w\} = 0, \ \forall \ \theta \quad \text{(10)}
\]

\[
w.r.t. \ w:\quad U'(w) = \lambda, \ \forall \ \theta \quad \text{(11)}
\]

Solving (11) for \( \lambda \) and substituting into (10) yields:

\[
\theta F'(L) = w - \frac{U(w) - U(\overline{w})}{U'(w)}, \ \forall \ \theta \quad \text{(12)}
\]

**Exercise:** Show that risk aversion (i.e. concavity of \( U \)), implies that \( \frac{U(w) - U(\overline{w})}{U'(w)} > w - \overline{w} \).

Condition (12) can therefore be written:

\[
\theta F'(L) = a, \ \forall \ \theta \quad \text{(13)}
\]

where \( a < \overline{w} \).

In choosing employment, the firm should now behave as if labor were cheaper than its true opportunity cost, \( \overline{w} \). Therefore it will hire more labor in every state of nature than in the base case. (It also follows – from the same diagrammatic argument as in the previous page’s exercise-- that cyclical variation in
employment will be reduced relative to the base case). The reason is that (when the firm can’t just give cash to unemployed workers) the only way to insure workers is to keep them employed.

Thus, while this modification to the model does generate rigid wages and involuntary unemployment (because $w > \bar{w}$), it actually generates less unemployment than one would see in a spot market world. That is however not necessarily an empirical problem since we have no way of knowing how much unemployment there would be in a pure spot market world.

The biggest problem with this model is explaining why $B=0$. As noted, one reason might be that it’s usually socially more efficient for workers on temporary layoff to search for (and take) other jobs than to stay on the old firm’s payroll. The only way to induce efficient search behavior among the laid off may be to have a low or zero level of $B$.

On the other hand, if workers have reasonable direct (on their own, not through their employers) access to capital markets (certainly workers can save to tide themselves over low- $\theta$ periods), one might question the empirical importance of worker risk aversion in determining contract structure. In that case one would have to search for other explanations of wage rigidity and/or involuntary unemployment, for example those discussed in Truman Bewley’s book, or in Shapiro-Stiglitz model of efficiency wages. That said, recall the evidence in Carrington et al., which seems to demonstrate considerable wage smoothing among employed workers, relative to the self employed.

Other models of wage rigidity include:


Herweg, Fabian, Daniel Muller, and Philipp Weinschenk. 2010. “Binary Payment Schemes: Moral Hazard and Loss Aversion.” *American Economic Review* 100 (5): 2451–77  (If the agent’s reference point is her rational expectations about the wage (as in Koszegi and Rabin 2006, 2007), then the optimal contract often has a “bonus structure” with only two possible wage levels). See also Koszegi’s JEL 2014 review paper, section 2.1.

Macera, Rosario. 2012. “Intertemporal Incentives under Loss Aversion.” (extends this to a dynamic context, where under some circumstances the current wage is completely unresponsive to performance, with incentives provided by future opportunities.


Models of wage compression:

When the primary source of uncertainty is individual ability, and ability is revealed over time, wage rigidity is closely related to wage compression.
