1 Overview

Overtime working implies that actual hours of work per period are in excess of standard contractual hours.[1] An example of standard time is a 35-hour workweek consisting of five 7-hour weekdays. If an individual works a 50-hour week then this would entail 15 weekly hours of overtime. This may be spread over weekdays, or confined to specific weekdays or it may involve, additionally or solely, weekend work.

For the vast majority of workers, overtime hours are remunerated at a different rate of pay from standard hours. Typically they command a premium that consists of a fixed multiple of standard hourly pay. A premium of time and a half applies to most overtime workers in the United States. This means that a worker receives 50 per cent more for working an overtime hour than a standard hour equivalent. Double time, time and a quarter, time and a third are other common multiples found elsewhere. Over the week, overtime hours may not be paid at a constant premium. It is not uncommon, for example, for workers to receive a higher premium for weekend compared to weekday overtime activity. Overtime is not necessarily paid at higher rates than standard time, however. For a minority of workers, overtime hours are remunerated at the same rate or even at a reduced rate. Further, some workers are rewarded indirectly for undertaking overtime. The most common practice in this latter respect is for the firm to offer days off in lieu.

In some countries, the rules governing pay and hours of overtime working – i.e. the level of the premium and the maximum number of per-period standard hours after which overtime pay applies – are imposed through government legislation. At one end of the spectrum, the government sets the overtime premium at such a high level that few firms exceed it. At the other end, all aspects of pay and hours are set at firm level within a much more *laissez faire* labour market climate. In between these extremes, less severe constraints apply with respect to coverage and/or the level of the premium. The outside regulator need not be the government. It is not uncommon for a national-level union or a federation of employers to formulate overtime rules that are then adopted by member firms.

---

[1] Alternative expressions for standard hours include 'normal hours', 'straight time hours' and 'regular hours'. This book adopts standard hours throughout. Despite its (Oxford) dictionary status, 'regular hours' is especially misleading. As we will see at several later stages, many workers and employers view overtime working itself as a regular work activity.

Whether or not overtime rules are imposed by outside legislation, there is still the question of whether employees within a given firm can opt voluntarily to work overtime if such work is available. While many employees can exercise such a choice, there are also many cases in which the overtime decision is taken unilaterally by the employer and that it is a condition of employment that reasonable amounts of overtime be undertaken when requested. Given that varying degrees of freedom do exist across firms then we might expect that employees who face relatively significant within-firm mandated overtime hours may be compensated for the related disutility (Ehrenberg and Schumann, 1984).

It has recently emerged that many individuals claim to work unpaid overtime. This is especially true of managerial and professional workers. This may well represent a short-run state of affairs, with payment taking the form of deferred compensation. For example, individuals may choose to work longer hours than stipulated in their contracts in order to achieve recognition from their peers that they hope will lead to future accelerated promotion and/or pecuniary reward. There are several other possible explanations for the unusual claim of offering labour services in the workplace for no pay, and this topic will be visited at several stages in the text. To simplify matters, the following convention will be followed. When referring to 'overtime' or 'overtime working', it is always implicit that it is *paid overtime* that is under consideration. Otherwise, the term 'unpaid overtime' will be used.

Overtime working might be regarded as a peripheral activity, representing at best a small fraction of a firm's labour input and payroll. In fact, although there is a wide international variation in incidence, overtime is a quantitatively important labour market phenomenon in respect both of its contributions to total labour input and to take-home pay. It is a particularly significant component of work time in several major economies. As examples, average weekly overtime in all industries in Japan and the United Kingdom is roughly 3 hours per week and between 3 and 4 hours among non-supervisory production workers in the United States. Moreover, the proportions of males working overtime reached as high as 0.4 in UK and US manufacturing in the late-1990s cyclical peak. An individual working average overtime hours in these countries would receive in excess of 20 per cent of total direct remuneration in the form of overtime pay.

But, of course, high incidence alone is insufficient reason to devote a monograph to the topic. Until relatively recently, the main interest has derived from economic analysis that has treated overtime as an important element of active labour market policy at national, industrial and firm levels. On the demand side, overtime has been viewed as representing a relative speedy form of short-term factor adjustment on the firm's intensive margin that involves costs and returns that distinguish it from alternative factors, such as stocks of employment and capital. A substantial policy oriented literature has grown around the idea of exogenously manipulating these costs, usually through government legislative intervention, in order to achieve employment policy goals. Work sharing in the form of more employees working fewer hours has been the central interest. On the supply side, it is common to treat an individual worker as supplying hours that maximise utility subject to a budget constraint. If optimal hours are higher than (by assumption) a standard workday constraint the individual would wish to work overtime hours. Suppose, however, that an individual wants to work no more than maximum standard hours while the firm – perhaps due to

technical or organisational constraints – requires longer hours. The firm may attempt to 'encourage' a supply of overtime by paying a premium for additional hours. In other words, overtime pay may be seen as a means by which the firm establishes patterns of working time that would otherwise not have been forthcoming. As we will see later, one application of this possibility is the use of overtime pay to discourage absenteeism.

Since the early 1990s, there has been a serious challenge to the notion that there is significant scope for policy intervention in respect of overtime pay and hours. This has arisen from an alternative approach to modelling hours' determination. Both firms and workers have preferences over wage-hours combinations. *Ceteris paribus*, job attributes are viewed in terms of their associated per-period lengths of scheduled working hours. The wage acts as a compensating differential for jobs of different lengths.[2] Contrary to traditional demand and supply analyses of working time, hours are treated as indivisible blocks of time, such as the required length of the working day. Suppose that the parties reach agreement over an average hourly wage and the number of daily hours. Assume that maximum daily standard hours and the overtime premium are set by legislation. The terms of the wage-hours agreement can be maintained in the face of a mandatory change in the overtime regulations. For example, an increase in the mandated premium can be exactly offset by a reduction in standard hourly rate so as to leave the average hourly wage and daily hours unaltered.

It is worth spelling out in a little more detail the ideas behind the two most dominant approaches, the workers-hours demand model and the wage-hours compensating differential model.

1.1   The two core models

The two leading workers-hours models that motivate the core debates on overtime working were introduced in the 1960s.

Brechling (1965) was the first to develop in detail the workers-hours demand model. Output, capital, technology and factor prices are treated as exogenous to the cost minimising firm. The firm produces output at the lowest cost combination of workers and average hours per worker. The study contains many of the results that, perhaps, are more immediately associated with later work. Among several long-lasting features, it distinguishes between equilibrium outcomes that represent short-time working or maximum standard time working or overtime working. It shows how changes in factor prices and in standard hours move us away from these equilibria. Substitution between workers and hours is central to its analysis. It explores what were later to become the most important working time topics in United States and European policy debates. For the United States, an economy with mandatory overtime rules, this concerned the employment effects of a rise in the overtime premium.

---

[2] Interestingly, the logic of this approach is most easily seen against the background of a labour demand set-up in its strictest form. Suppose we lived in a world where all employment contracts imposed purely employer-determined fixed hours and fixed hourly pay. Supply-side considerations cannot be suppressed under these conditions. Thus, 'we can think of individuals as shifting amongst employers until, in equilibrium, each individual is working in a job which offers him the [fixed] hours he would have liked to work, given the pay and tax parameters' (Ashworth, McGlone and Ulph, 1977).

4   **The Economics of Overtime Working**

Suppose we fix the capital stock and the level of technology. If the firm employs overtime hours in equilibrium, a rise in the premium produces a workers-hours substitution – i.e. substitution away from the margin that experiences the price rise. As for Europe, governments and unions have long been fixated on policies designed to induce work sharing through cuts in the standard workweek. Brechling shows that, for cost minimising firms working overtime in equilibrium, a reduction in standard hours produces a *fall* in employment and a rise in average overtime hours. This perverse result has stubbornly recurred over a large number of model variants on Brechling's original theme. It has been a dominant argument among economists who have advised against such policy interventions.

Might it be more realistic to concentrate on fixed work schedules consisting of indivisible daily hours rather than on the optimal fine-tuning of average hours in the face of relative factor price effects? Technological, organisational and customer considerations may lead firms to select the most cost effective daily lengths of production and business trading hours. Different attitudes to work and leisure – prompted by such factors as personal characteristics, family circumstances and peer group influences – may induce individual suppliers of labour to search for the job length/wage earnings combination that optimises their utility.[3] Market equilibrium may then be thought of as reflecting the satisfaction of joint preferences across employers and individual workers with labour demand equal to labour supply at all job lengths. In effect, the wage acts as a compensating differential for jobs of different length. In essence, this is the core motivation behind the influential work of Lewis (1969a) into the determination of daily hours. It brings together the hours preferences of firms and workers by determining the mutually agreed daily hours and wage earnings combinations.

Why does the Lewis approach carry radical implications for the standard demand model? Let the firm and its employees agree, via the above process, to an average hourly wage of w_0 for a 10-hour working day. Suppose that maximum standard hours and the overtime premium are set by legislation. Trejo (1991) nailed down the critical point. At the time that wages and hours are set, the parties will encompass these exogenously imposed hours and pay rules, adjusting the standard hourly rate in order to achieve w_0. For example, suppose that the mandated overtime premium is 1.5 times the standard hourly rate and that daily standard hours cannot exceed 8 hours per day. Then, w_0 would be the average of 8 hours of standard pay plus 2 hours of overtime 'weighted' at 1.5 times the standard rate. Now, suppose that the government increases the overtime premium to 2 times the standard rate. The average hourly wage would rise above w_0 because the 2 hours of overtime would now receive an increased weight. But the parties would wish to stick to their original wage agreement because this jointly satisfies their preferences with regard to a job length of 10 hours per day. To maintain a given equilibrium, they could simply agree to retain w_0 by a suitable reduction in the standard hourly rate.[4] Alternatively, if the government reduced

---

[3] There is evidence that workers match with their preferred lengths of jobs. For example, based on the Canadian Survey of Work Reduction, Kahn and Lang (1995) show that, in general, workers who state that they prefer to work fewer (longer) hours do in fact work short (long) hours. Increased job tenure appears to lead to more satisfaction with length of hours suggesting that hours matching improves with seniority.

[4] Actually, Ehrenberg and Schumann (1982a, p.135), without reference to Lewis' paper, realised this point: 'One plausible response to a legislated increase in the premium is for . . . [firms and employees] . . . to voluntarily

maximum standard daily hours, the average hourly wage – for given daily hours – would again rise because fewer hours would be weighted by the lower standard rate and more by the higher overtime rate. Again, the parties would be happy with a downward adjustment of the standard rate that served exactly to offset the increase. Contrary to the Brechling model, the initial hours equilibrium is unaffected by these interventions. This implies that there is no incentive to alter the workforce size, or other input factors, given such a government intervention.

An interesting issue surrounds the question of why the Lewis model, and not standard neoclassical supply-side analysis, has offered the main challenge to the workers-hours demand model. There is a well-established supply-side literature that deals with the time allocations of individuals facing differing wage schedules. Unfortunately, it has little or nothing to say about the demand side of the problem. But overtime working is essentially a firm-level activity and there is an undeniably important element of employer input into most overtime decisions.[5] However, in most instances overtime is a voluntary activity and so the employer cannot unilaterally impose overtime work on individual workers. Herein lies a major problem. The supply-side reactions of employees to changes in overtime rules are predicted to differ radically from those emanating from demand considerations. For example, a rise in the overtime premium is likely to induce the supply response of a greater willingness to work overtime – or to offer more overtime hours per period – while the demand reaction would be to substitute out of overtime and to employ more workers. Models that have attempted to integrate traditional demand and supply analyses of pay and hours have proved to be intractable and largely unhelpful. Yet, standing alone, labour supply models cannot offer a serious alternative to workers-hours demand because cutting out employer-based decision making is simply too far removed from reality. The Lewis approach side steps these problems by regarding the length of per-period working hours as defining a job attribute over which employers and employees register preferences (Trejo, 2003).

1.2 Key developments and issues

Why did the workers-hours labour demand model dominate the study of overtime working during the 1970s and 1980s? Arguably, and especially in a European context, job creation policy was a dominant labour market theme of this era and so economists wanted to work within a framework that offered the possibilities of influencing employment outcomes through working time policy interventions. Workers-hours substitution by optimising firms in the face of factor price changes provided a simple means of arguing the pros and cons

agree to a reduction in the level of straight-time wages, or fringes, or both, leaving total compensation for the initial number of hours unchanged.' Owen (1989) also raises this issue but emphasises that legislation introduced to change the premium may well also attempt to prevent employers from offsetting downward adjustments of standard rates. Owen also points out that there are potential difficulties of reducing standard rates under actual union contracts.

[5] Lewis (1969b) emphasises that a motivation of his approach to working time is to overcome the inadequacy of received theory that stems from 'its exclusive reliance on the supply side to explain hours data. The theory assumes that hours of work per employee per period of time is a matter of no consequence to employers.'

of such policy initiatives. The subsequent rise of the Lewis model might then be seen as an *ex post* rationalisation of the general failure of government and union policy to influence work sharing outcomes. A simpler, but not necessarily unconnected, explanation of early demand-side dominance lies in the fact that two distinguished labour market economists produced seminal pieces of work along the lines pursued by Brechling. The first emphasised the role of fixed, or hours-independent, labour costs in the demand model and contained an empirical application related to US railroads (Rosen, 1968). A PhD student of Brechling undertook the second, and arguably most important study of all. Thus, Ehrenberg (1971a) rigorously developed static and dynamic versions of the cost minimising model and provided detailed empirical work. Economists went down the labour demand route because this work stimulated a broad range of theoretical, empirical and policy issues that were not only confined to the working time domain.[6]

In common with a number of contemporary studies, Brechling's 1965 paper is also concerned with the analysis of time-wise adjustment between actual and desired employment within the context of the labour demand model.[7] Nadiri and Rosen (1969) extended this area of demand analysis to embrace both labour and capital dimensions. Suppose the firm envisages a long-term expansion of production. It may undertake this by a combination of (a) a larger workforce, (b) a greater utilisation of the existing workforce, (c) an expansion of its stock of capital, (d) a greater utilisation of its capital equipment.[8] The extent to which it will have recourse to one or other action is likely to embrace considerations of relative prices. As already intimated, the overtime wage premium is likely to be a key consideration in (b). *Ceteris paribus,* a rise in the overtime premium, in relation to other marginal input costs, is now far more difficult to evaluate. Outcomes depend on degrees of complementarity and substitutability among factor inputs.

In the short term, the adjustment potential of each input factor is conditioned by its own adjustment speed and the relative adjustment speeds of the other factors. If we were to concentrate on this time frame – and rule out adjustment of capital – then the workers-hours adjustment problem may still be extended to include the stock of inventories. As investigated by Topel (1982), the firm may face the choice – involving an evaluation of relative prices – of increasing overtime hours or reducing inventories below planned levels as its response to an unanticipated short-run increase in sales. Nakamura (1993) brings overtime hours directly into dynamic factor demand analysis of the Japanese labour market. It is found that an output shock stimulates an overshooting of overtime to compensate for a slow adjustment of employment to a new long-run level.

---

[6] While cost minimisation dominated the early workers-hours studies, it was later expanded to embrace profit maximisation. The best contribution is that of Calmfors and Hoel (1988). In several notable instances, this served to complicate predicted outcomes. For example, a rise in the premium produced workers-hours substitution (the substitution effect) but the rise in total labour costs for given output caused a reduction in sales that impacted negatively on the workforce size (the scale effect). Net outcomes were rendered ambiguous.

[7] Perhaps developments of the workers-hours model in this direction is most associated with the work of Fair (1969; see also Fair, 1985).

[8] Of course, there are other ways of increasing production. For example, on the intensive margin, the firm might try to encourage more effort per unit of time rather than change the length of the time unit itself.

Overtime is treated as one of the firm's most important intensive margin responses in the interrelated factor demand literature. This margin has also featured prominently in the analysis of the relationship between marginal cost and price over the business cycle, with the work of Bils (1985) providing the leading example. Overtime hours have been found to respond more quickly than employment stock to fluctuations in business activity. One reason for this is that overtime decisions are more easily reversible than employment hiring decisions. This helps to avoid the risk of writing off sunk human capital investments if market predictions turn out to be inaccurate. So, changes in overtime can be used as a short-term employment reaction to cyclical change in the face of uncertainty over the length and potential extent of the perturbation. Even when relative factor responses have been fully assessed, overtime may temporarily diverge from planned levels due to adjustment impediments in capital and labour inputs. Overtime's relationship to marginal cost and wage earnings over the cycle stems from similar considerations. Some firms will respond to an upturn in the cycle by extending both the average overtime hours of existing overtime workers and the proportions of their total workforces working overtime. If overtime hours are paid at a higher marginal rate than standard hours then marginal labour costs to the firm will rise. These rises will be greater if longer weekly hours themselves involve a higher marginal overtime premium (e.g. weekend working).

Most of the best known time-related overtime-adjustment investigations derive from factor demand models. But how do studies of employment and hours adjustments relate to the Lewis compensating differential model? We have already noted that this model has provided a radical alternative to demand-side developments in the comparative static evaluation of workers-hours reactions to mandatory changes in pay and hours. At the heart of this challenge is the idea that firms and workers jointly express preferences over jobs of given lengths. This suggests that hours, including overtime hours, contain important job-match effects. One of the latest developments in the study of overtime working is to uncouple pure dynamic adjustment speeds of hours from the influences of fixed effects. In fact, if job matching is a quantitatively important phenomenon then a failure to account for it in dynamic time-series analysis of overtime would tend to produce underestimates of the true speed of hours adjustments.

Most theoretical developments relating to overtime working in the workplace describe a homogeneous workforce, with all workers either participating or not participating in overtime. While this convenient assumption has helped considerably with problems of modelling tractability, it clearly is not representative of the experience of many overtime firms. Overtime working is largely a voluntary activity and not all eligible workers elect to work extra hours. Moreover, the firm itself may not experience sufficient demand to justify extending all workers' hours and so may attempt to select what it believes to be those individuals most suited to working longer hours. For whatever reasons, partial overtime coverage is commonly observed. This raises several extensions and questions concerning existing models. We need to find explanations of how changes in factor prices and standard hours affect *both* average overtime hours *and* the decision over whether or not to participate in overtime. We might even need to query the premises on which the various modelling

approaches are based. Take the Lewis model as an example. Suppose that employees in a typical firm either do or do not engage in overtime and, if they do, that they work different numbers of weekly overtime hours. Does this mean that the firm specifies a number of jobs of different lengths and finds workers with different preferences by paying a range of compensating differential wage rates? This may cause industrial relations problems if workers with the same ability and hourly productivity are paid different average hourly rates.

Overtime hours changes entail production, labour cost and wage effects. With respect to the last of these, overtime plays a key role in the study of the cyclical behaviour of real hourly earnings. At the aggregate level of the firm or industry, hourly earnings decompose into average standard wages, the average overtime premium and the proportion of the workforce working overtime. At the micro (individual-based) level, researchers are increasing keen to distinguish between standard hourly pay and hourly earnings (i.e. including the effects of overtime pay). In fact, the use of hourly earnings in studies of wage cyclicality and cross-section/time-series wage curves involves an important issue concerning overtime. Even if the wage rate and the premium were to remain constant, earnings can still fluctuate if the length of weekly hours is itself changing over the cycle. In particular, if weekly hours vary procyclically then so can the earnings of overtime workers even if standard hourly rates and the premium were to remain unchanged. This is because the proportion of premium to total pay would vary procyclically. This raises a critical issue in relation to Phillips curve and wage curve studies. Hours effects may serve to obfuscate pure wage effects with estimated earnings/unemployment elasticities representing composites of wage curves and hours curves.

Suppose each of the main components of daily earnings – i.e. standard pay, overtime pay and the proportion of overtime to total hours – display cyclical time-series behaviour. A critical question is whether the components are responding primarily to the same cyclical indicator. For example, fluctuations in standard pay may vary positively with all phases of the business cycle. Economists typically proxy the business cycle by an output- or employment-based indicator. By contrast, overtime pay and hours may respond to shorter-term fluctuations in demand towards the peaks in economic activity. For example, if firms find it difficult to adjust employment and capital inputs in tight labour markets towards the peaks of cycles, they may opt to employ longer overtime hours or to rundown inventories at a faster than planned rate. For this type of reasoning, overtime fluctuations have been argued to link closely with inventory cycles. Gaining an understanding of the influences of various cyclical indicators on wage earnings is usefully aided by the adoption of frequency domain methods of decomposing (stationary) time series into harmonic waves of varying phases and amplitudes. Recent work on the behaviour of wage earnings over the cycle has moved in this direction.

As already indicated, in some countries pay and hours conditions applying to overtime working are largely worked out at the level of the firm while, in others, firms are subject to high degrees of exogenous constraints. However, even in the latter cases, overtime decisions are undertaken that lie outside the domain of regulation. For example, large numbers of firms in the United States pay overtime rates for hours that occur *before* maximum permitted

weekly standard hours have been reached.[9] The question then arises as to why, given wages and hours are determined within the workplace, firms opt to pay overtime premia for marginal hours. One simple, but potent, explanation involves custom and practice. Long-term efficient contracts necessitate agreements over both wages and working hours. Suppose the firm takes standard weekly hours as given, perhaps following an industry or local labour market norm. Also, suppose that the firm requires working hours to be in excess of standard hours. How is premium pay determined in these circumstances? There is an indeterminacy problem if the parties attempt to set the optimal length of overtime hours and two optimal wage rates, i.e. the standard and overtime rates. It is far simpler for the firm and its workforce to allow the premium to be based on an historic norm, subject to custom and practice, and then to concentrate bargaining on the standard wage. In this event, firms facing the custom of a relatively high premium can achieve a competitive average hourly wage by setting relatively low standard rates. Firms experiencing low premiums would set relatively high rates. This argument is very much in the spirit of the Lewis model and applies, especially, to unregulated labour markets.

Premium pay for overtime may also be adopted by the employer as a simple means of attempting to offset potential difficulties among workers with diverse preferences over per-period lengths of working hours. Overtime schedules may be designed with the aim of curbing dissatisfaction, and associated poor work performance, with respect to pure standard time arrangements. The use of overtime in this way may help to improve matters when all employees in the firm are required to work the same length of daily or weekly hours. Greater potential difficulties arise when equally productive employees in the same occupation and/or involved in interactive job tasks within the firm undertake different lengths of per-period hours. It is not easy in these instances to pay different average hourly wage rates for different hours provision while also avoiding adverse industrial relations repercussions.

In recent times, European statistical sources have reported on the phenomenon of unpaid overtime. Managers and professional workers, in particular, report working significant numbers of weekly hours beyond those stipulated in their contracts and for which they receive no additional payments. This is in contrast to paid overtime which is especially prevalent among blue-collar and lower-paid workers. Aggregated across all individuals, unpaid hours are roughly comparable in number to paid overtime hours. Of course, for most individuals, there is no such thing as working for nothing in a long-run context. Early work concentrated on studying the longer-term consequences of offering such hours. These include accelerated promotion, above-average pay rises for given work characteristics and the link of unpaid work to remuneration systems linked to company performance.

Unpaid work also has implications for the study of wage earnings. Unpaid hours serve to drive a wedge between paid-for and effective hourly wages. If, at any given time, company managers work more than contracted weekly hours while non-managers work only paid hours then their effective hourly wage differential will be less than the wage differential if only paid-for hours are measured. This may impact, for example, on the analysis of returns

---

[9] Trejo (1993) provides evidence that suggests that over 20 per cent of overtime receives a premium for marginal hours worked before the 40 weekly standard hour limit set under the US Fair Labor Standard Act (FLSA) regulations.

to education. Suppose that in their early careers, managers and professionals work more unpaid hours than other workers. More highly educated individuals are likely to work as managers and professionals. In the early work years at least, returns to education will be exaggerated if paid hours rather than effective hours are used to deflate gross wages. Of course, working unpaid hours may help managers and professionals to enjoy steeper wage profiles at later stages in their careers but this does not detract from the strong possibility that wage-tenure and wage-experience profiles will differ with and without the accommodation of unpaid overtime.

The remainder of the text attempts to explore in greater depth these broad areas of interest.