Labour Demand Research: Towards a Better Match between Better Theory and Better Data

John T. Addison
Department of Economics, Moore School of Business, University of South Carolina, and
Department of Economics and Finance, Durham University Business School

Pedro Portugal∗
Banco de Portugal, and Faculdade de Economia Universidade Nova de Lisboa

José Varejão
Faculdade de Economia da Universidade do Porto and CEF.UP†

June 6, 2014

Abstract

At first blush, most advances in labour demand were achieved by the late 1980s. Since then progress might appear to have stalled. We argue to the contrary that significant progress has been made in understanding labour market frictions and imperfections, and in modelling search behaviour and heterogeneous preferences. Perhaps most notable have been the improvements in data, in the form of longitudinal matched employer-employee data, and in techniques and algorithms (e.g. for solving heterogeneous parameter models). In short, the Cinderella status of the field is frankly overdrawn. Nevertheless, a chief lacuna remains the need for a better match between theory and data. This paper provides a critical albeit eclectic assessment of these developments, along the dimensions of the static and dynamic theory of labour demand, wage formation, and estimation, noting advances and limitations. As is conventional, somewhat greater emphasis is placed on the latter.

JEL codes: J23, J3, J4, J5, D4, F66
Keywords: labour demand, input heterogeneity, labour adjustment costs, wage and employment determination, product and labour market imperfections, multiple fixed effects, exogenous wages, establishment-level functions

∗Corresponding author: Pedro Portugal, Banco de Portugal, Av. Almirante Reis, 71, 1150-165 Lisboa, Portugal; E-mail: pportugal@bportugal.pt
†José Varejão acknowledges funding from Fundação para a Ciência e Tecnologia in the framework of the project PEst-OE/EGE/UI4105/2014.
1 Introduction

For the general reader, research on labour demand may seem to have progressed little in the past 25 years especially when compared with other areas of research within labour economics. Thus, the static theory of labour demand has long been established; reliable estimates of the most relevant parameters have long been available; and the essential features of labour adjustment and the underlying micro-mechanisms were identified in the late 1980s. Yet we shall argue that significant progress has nonetheless been achieved in this research area.

On the theoretical front, the most important developments have stemmed from the recognition of the labour market frictions stemming from union presence and intervention, together with other imperfections that impart positive slope to the labour supply curve. To this mix one would need to add search frictions and heterogeneous preferences.

Furthermore, the importance of good data – especially longitudinal matched employer-employee data – to progress in labour demand studies cannot be overemphasized. In fact, benefiting from increasingly detailed data, most of the progress achieved in recent years has come from empirical studies that have allowed for a finer characterization of the labour input and renewed interest in the study of interrelated factor demand, the interaction between the labour and product markets, and a richer description of the dynamics of labour demand.

Significant progress on the empirical front has involved new methods applied to better data, thereby advancing our knowledge of labour demand and our ability to predict the effects of policies that impact wages and employment. These new approaches include IV methods combined with panel data, quasi-experimental methods and other micro-evaluation techniques for policy interventions, and high-dimensional fixed-effects models or heterogeneous-parameter models. All became common practice in labour demand research.

In this article we provide a critical assessment of the major developments achieved in the past quarter-century and identify the most promising such developments, those
that we consider to be most likely to ensure the vitality of labour demand research in the years ahead. The article is organised as follows. Section 2 reviews the results achieved by the empirical literature on the statics and the dynamics of labour demand. Section 3 reviews progress in studies of wages and employment determination. Section 4 discusses empirical methods and corresponding estimation issues. Section 5 concludes.

2 The Theory of Labour Demand

2.1 The Static Theory of Labour Demand

In its standard form, static labour demand theory focuses on employers’ decisions regarding the quantity of labour to be used in production and on how these desired quantities change in response to marginal changes in product demand and factor prices. It is essentially a branch of production theory, especially concerned with studying the transmission mechanism running from shocks to the product market and to prices set in input markets to employment and wages, the structure of production being an essential component of the transmission mechanism. Labour demand is typically described as a derived demand even if occasionally we are reminded of the specific nature of the labour input. Unsurprisingly, most progress achieved in this domain simply mirrors progress made in production theory.

In its simplest version, the starting point of labour demand theory is a representative profit maximizing (cost minimizing) firm that is able to adjust the quantity of the labour input used in production at no cost. In the general case of n-inputs, and allowing all quantities to vary (long-run analysis), conditional and unconditional demand functions are derived by solving the firm optimization problem which requires the specification of the production function, assumed to be strictly increasing and strictly concave. From the labour demand curve the parameters of interest – labour demand elasticity and the elasticities of substitution between different inputs (including different types of labour) – are derived.

Studies of static labour demand have produced a substantial amount of knowledge that is now instrumental for predicting the effects of policies that can alter wage and
employment outcomes, either because they change the relative prices of the inputs (P-policies) or their quantities (Q-policies).¹ Differently put, static labour demand studies are not only important per se, but also because they are an essential component of the policy advisor’s toolkit.

Given the close ties with production theory, it is not surprising that most theoretical developments in static labour demand theory stem from production studies. These have contributed with increasingly complex and flexible representations of the production technology, evolving from the standard Cobb-Douglas production function, through CES, to generalized Leontief or translog functions, all of which have been used to derive labour demand equations readily usable in empirical work.

But important as these developments may have been, the truth is that they have added little to our knowledge of labour demand, with most estimates of the elasticity of labour demand falling within the interval 0.4-0.6 independently of the functional form adopted. Similar estimates were obtained with non-structural specifications of the labour demand equation.²

Increasingly available micro (firm-level) data have however made it possible to incorporate labour heterogeneity into the study of labour demand. By allowing for labour heterogeneity along occupational (notably, skilled versus non-skilled labour) and nationality (immigrants versus natives) lines, the empirical literature was able to produce evidence on the substitutability between skilled and non-skilled workers and between native and immigrant workers, thereby shedding light on some of the most researched theoretical and policy topics (respectively, skill-biased technological change and the debate over immigration policy) in recent years.

It is now widely accepted that technology change is the motive force behind the increasing demand for skilled workers. Nevertheless, Hamermesh (1993a, p.113) cautions

---

¹For a workable typology of static demand policies that includes P and Q-type policies, see Hamermesh (1993a), chapter 5.

²The mean estimate of the labour demand elasticity in Lichter’s et al. (2014) sample of 105 different studies is 0.508. These authors conclude from the meta-regression analysis they conduct that the elasticity of labour demand is bracketed by the interval [0.072, 0.446] with their preferred estimate being equal to 0.246, close to Hamermesh’s (1993) ‘best guess’ of 0.3. Admittedly, the ‘3 for 10’ rule underestimates the employment response to an increase in labour costs that is likely much bigger if capital is allowed to substitute for labor (Hamermesh, 2014).

3
on the basis of his extended review of the literature that “there may be” capital-skill complementarity and argues for the need to produce estimates of the parameters of interest based on data that allow for the effective disaggregation of labour into meaningful groups (avoiding in particular the overlapping of relevant skills across groups), as well as appropriate measures of the capital input.

Arguably, the most significant progress in tying labour demand shifts to technology change has been achieved by Autor et al. (2008) and Autor and Dorn (2009). In their re-evaluation of alternative explanations for changes in wage inequality, Autor et al. (2008) find that computerization sharply changed the mix of job tasks in production, increasing the demand for cognitive and interpersonal skills (required to perform abstract tasks), reducing the demand for routine analytical and mechanical skills (used in routine tasks), while having little direct impact on the demand for non-routine manual skills (used in manual tasks). These results call for a richer version of the skill-biased technical change hypothesis in which new technologies and highly-skilled (highly-educated) workers are complements for one another but substitutes for moderately-skilled (moderately-educated) workers, and have no significant relationship to low-skilled work. For their part, Autor and Dorn (2009) offer an integrated explanation for the polarization of employment and wages and, testing for it with a spatial general equilibrium model, confirm its predictions with U.S. data. From a labour demand perspective, these results are remarkably consistent with Hamermesh’s above-mentioned call for more meaningful disaggregation of the labour input, the implication being that the simple skilled versus non-skilled distinction may be too broad. The same implication also applies to capital, as hinted at by Bergström and Panas (1992) who find that estimates of total factor productivity are sensitive to the choice of the disaggregation of inputs. Progress along this dimension has yet to be achieved.

The intensification of migration flows in recent years has renewed the debate over the labour market consequences of immigration, that is, on how immigrants affect the economic opportunities of natives, their employment and their wages. The elasticity of substitution between immigrants and the native-born population is a critical parameter of interest here, not only for the study of the effects of inflows of migrants on native em-
ployment but also in the case of comparable migrant and native groups on the wages of the latter. Although early studies of these topics adopted a structural approach (Grossman, 1982; Borjas, 1987), specifying a production function of some form (translog or generalized Leontief), to derive labour demand equations from which the elasticities of interest were obtained and estimated, this approach was soon to be displaced by the increasingly popular spatial correlation approach. Although by no means unanimous in its findings, a fair reading of this literature indicates a small mostly negative effect of immigration on the employment and wages of native-born workers. Key methodological issues in this regard include the failure to account for natives’ outmigration from receiving areas, local demand shocks, and long-run effects of intercity trade (Card, 2001). Progress in this area is likely to be achieved in the near future by a return to structural approaches such as the one adopted by Ottaviano and Peri (2012). In a production function framework, these authors are able to combine own-group effects with cross-group effects to estimate wage effects for different indigenous groups.

As was noted earlier in respect of the technical change literature, a crucial issue in this approach is how to deal with worker heterogeneity. Borjas (2003) convincingly argues that the worker groups commonly considered in the past (schooling categories) are too broad (too heterogeneous). Further disaggregation is necessary as immigration is not evenly balanced across experience cells, even within schooling groups. In similar vein, Borjas (1987) has also noted that aggregating different immigrant groups (as regards national origin and/or ethnic background) is not neutral. Ottaviano et al. (2013) extend the study of the labour market impact of immigration and combine it with the study of the relocation abroad of jobs by multinationals (offshoring). They show that neither issue should be analysed in isolation, as they compete with each other and with the hiring of native workers by domestic employers. But they also conclude that to understand the economic effects of immigration and offshoring it is essential to account for the inherent heterogeneity of jobs (tasks) that require distinct abilities (skills).

The common denominator of the intersection between the labour demand literature on the one hand with the skill-biased technical change and immigration literatures on
the other, is the need for better data that allow for a deeper characterization of all the inputs used in production, especially different types of labour. Increasingly rich linked employer-employee data (LEED) offering the possibility of controlling for multiple fixed effects (not simply firm and worker fixed effects, but also job fixed effects) emerge as the most promising way forward for research in the field of static labour demand studies. Considering that LEED data are more developed in Europe, it is likely that the main progress in this field will come from non-U.S. studies, as anticipated/encouraged by Hamermesh’s (2002) call for a truly International Labour Economics discipline. The use of such data in these areas is clearly in its infancy (see Malchow-Møller, 2009). New sources of data or the increasing possibility of combining different data sets covering different areas of firms’ operations is most likely to pay dividends here. As these data become available, the road will open for linking labour demand theory not only with production theory but also with the economic theory of the firm. Recent or renewed interest in topics such as the links between labour markets and product markets, international trade, financial markets, and corporate finance open the opportunity for a revival of interest in labour demand.

The distinction between employment and hours has long been recognized as key to the study of labour demand. However, this is a topic that remains open to research. The difficulty here is to disentangle in the data changes in the price of hours that do not change the price of workers (or change it differently), and vice-versa. Hence, no evidence is available on cross-elasticities between employment and hours, even if a few studies do address the related topic of the employment effect of mandated or bargained changes in the standard workweek (e.g. Hunt, 1999).

Implicit in the entire literature on labour demand is the fact that both production functions and the demand for labour are defined over some fixed (typically unspecified) interval of time. Very few studies have recognized that hours of the day and days of the week are not all the same to workers (or to employers) and should be treated differently (the major exceptions are Winston, 1982, and Hamermesh, 1996). That said, Cardoso et al. (2012), using data from a representative sample of Portuguese firms, describe how the number of employees at work varies across the 168-hour week, and examine
the role of pay penalties for work at nonstandard times in affecting the variation of the number of those at work in different hourly time slots. They find that the wage premia attaching to night or weekend work significantly alter the timing of work. Better data, meaning data for longer periods that include changes in the penalties for working at different times, may open the way for a completely new and worthwhile research topic.

Job creation and job destruction generated by the entry and exit of establishments (viz. the extensive margin of employment adjustment) are discrete events often ignored in the estimation of labor demand functions. Nonetheless, establishment births and deaths account for a large share of job turnover, particularly in countries with high firing costs. Hamermesh (1989) proposed the notion of quasi-elasticities of labor demand through entry and exit, and suggested that these might be quite large. Unfortunately, studies seeking to estimate the impact of wages upon the entry and exit of firms are sparse.3 Once again, however, with the recent availability of increasingly rich matched employer-employee datasets containing accurate identifiers of entries and exits, studies pursuing this line of inquiry research are clearly in the offing and to be welcomed.

2.2 The Dynamic Theory of Labour Demand

The dynamic theory of labour demand is driven by two major goals: first, to explain the cyclical behaviour of the productivity of labour; and, second, to understand the workings and the effects of job security policies (Hamermesh, 1993a, p. 205). Consequently, the focus of the research effort is upon the costs of adjusting the quantities of the labour input, with two separate questions - issues of magnitude and structure - deserving attention.

Until today, very little direct evidence exists as to the magnitude of labour adjustment costs. In his influential article on the quasi-fixed nature of the labour input, Oi (1962) reported evidence from a study on the magnitude of hiring and training costs for one North-American company in 1951. For this company, the total cost of annual turnover represented 5.4 percent of its annual wage payments. He also reported that the incidence of these costs was higher for high-wage, high-skill jobs. The few other

3But see Abowd et al. (2001).
studies that also attempt to measure adjustment costs (reviewed in Hamermesh, 1993a) converge in the finding that these costs are large, even if they all ignore, as indeed does Oi, the internal component of adjustment costs, namely, costs arising from the disruption of production as the quantity of labour (net costs) or the identity of workers (gross costs) are changed.

In recent years, indexes of labour market rigidity have become increasingly popular. However, our knowledge of the true magnitude of labour adjustment costs is pretty much the same as it was when Oi (1962) first demonstrated their importance for labour demand studies. Needless to say, we continue to lack evidence on how these costs vary across industries/firms using different production technologies, and across countries characterized by different labour market institutions. Progress in this area would certainly be welcome.

Modelling the dynamics of labour demand requires the specification of the adjustment cost function, hence the structure of these costs. Early studies of the dynamics of labour demand assumed that labour quantities smoothly adjust to their long-run desired levels. Hence, partial adjustment-type models became standard and proved well-fitted to aggregate data, even before they were rationalized with an underlying technology of convex (quadratic) adjustment costs at the micro level. The popularity of these models was not accompanied by convincing evidence that, at the employer level, adjustment costs should be best seen as convex in structure as opposed to being non-convex (i.e. fixed or linear). Hamermesh (1989) provided evidence on monthly plant level data showing that adjustment is discontinuous in the sense that large jumps follow prolonged periods of inaction. Subsequent studies (e.g. Caballero et al., 1997; Varejão and Portugal, 2007; Kramarz and Michaud, 2010) provided further evidence confirming that the adjustment technology at the micro level is inherently non-convex. This is now undisputed.

The standard approach to deriving the optimal path of employment adjustment simply solves a dynamic optimization problem where firms maximize the present discounted value of future profits and face costs of adjusting the labour input. For reasons of tractability, simplifying assumptions as to the production technology (e.g. Cobb-
Douglas), nature of the product market (e.g. perfect competition) and the expectation generation process (e.g. adaptive expectations) are typically, although by no means universally, imposed. Labour adjustment costs are commonly assumed to arise out of the net adjustment of labour quantities and to be quadratic in structure. More recently, studies incorporating non-convex adjustment cost structures have been developed and brought to the data under some additional simplifying assumptions (Nilsen et al., 2007).

The vast majority of studies of the dynamics of labour demand, including those previously cited, consider separate adjustment of a single factor: labour. However, it is clear that lumpy adjustment of one input may be due not only to non-convexities in the technology of adjustment of that input but also in those of other inputs as well. Full understanding of the dynamics of input (labour) demand requires the specification of models that incorporate all inputs used in production while allowing for interactions between them. The few articles that focus on interrelated factor demand (Letterie et al., 2004; Nilsen et al., 2009; Asphjell et al., 2010) show that the adjustment of one factor input cannot be understood without considering adjustment of the other inputs, especially when the latter are large (Letterie et al., 2004), and further that the cost advantage of simultaneous adjustment of capital and labour is large in the case of labour (Asphjell et al., 2010). The implication of the above is that more severe biases are to be expected when labour demand parameters are estimated separately from those of other inputs.

For reasons detailed in the previous subsection, the study of the dynamic interaction of adjustment of different inputs is especially valuable as regards the distinction between employment and hours, discussion of which is either absent from the literature or is only considered assuming hours can be adjusted freely.

Empirical studies of the dynamics of labour demand are very demanding in terms of the quantity and quality of the data they require. This is especially true regarding the frequency of the data. Typically, yearly data are used. Frankly, this is not appropriate. Low-frequency data can be expected to bias the results against non-convex adjustment costs as temporal aggregation, no less than spatial aggregation, smoothes away any signal of discontinuous adjustment that would be observed at the appropriate frequency.
(Hamermesh, 1993b). Hamermesh (1993a, p. 261) recommends that dynamic labour demand topics should be addressed with at least quarterly data. This is another area where progress has yet to be realized.

An important and fast-developing literature on temporary employment contracts emerged in recent years at the intersection of job protection studies and dynamic labour demand theory. From earlier studies that simply modelled temporary contracts as low (zero) firing cost contracts, to more recent approaches that incorporate very specific but arguably important features of this form of employment, the focus of researchers’ efforts has been upon the employment, unemployment and productivity consequences of job protection. As it turns out, these consequences hinge on the purposes for which low firing cost contracts are used. Hence, research is increasingly directed towards explaining the employer’s choice between the two types of contracts. The standard explanation draws on a standard adjustment cost framework: temporary contracts are preferred because they allow employers to save on future dismissal costs. However, Cahuc et al. (2012) emphasize the importance of accounting for specific details of the norms regulating the use of these contracts and show that the fact that temporary contracts cannot be terminated before the ending date stipulated \textit{ab initio} is key in explaining the demand for these contracts.

The cost of adjustment approach to the explanation of the incidence of temporary employment in high and low firing cost labour markets is not undisputed. A burgeoning alternative literature emphasizes the role of temporary contracts as mechanisms for screening workers for permanent positions and their effects on productivity via match quality (Faccini, 2008; Bucher, 2010). The race between the two approaches has yet to produce convincing theoretical or empirical results.

Under standard theory, the labour demand decisions of the employer, and in particular the characterization of the employment-wage relationship, have been considered in isolation from the notion that firms and employees bargain over wages, working conditions, work schedules, or even employment. Taking account of the bargaining power of the employees (the power of unions) and the bargaining power of the firm (monopsony
power) has decisive implications for the specification of the employment wage relationship, calling for the need to consider the role of union power in driving (up) negotiated wages (see Booth, 2014) together with the (upward) impact of firm-level labour supply elasticities on wage formation (see Webber, 2013).

3 Wage Formation: On- and Off the Demand Curve Models and a Wider Classification

In the competitive model of wage and employment determination the firm determines employment subject to an exogenously-given market determined wage. One important branch of the literature has sought to introduce market imperfections in the form of unions. Latterly, more attention has of course been devoted to labour market frictions that impart positive slope to the supply curve (Manning, 2003). Studies in both areas, however, have tended to exhibit separate development. Recently, this apartheid has been eroded with new approaches that link these two labour market imperfections, while also introducing product market distortions.

3.1 Models of Wage and Employment Determination

A logical starting point is the right-to-manage model (Nickell and Andrews, 1983) since this is a generalization of the monopoly union model (Dunlop, 1944). In the former model, the two sides bargain over wages and the firm retains the prerogative to determine employment. In the monopoly union model, the union sets the wage, and the firm sets employment as before. Denoting union bargaining power by $\gamma$, the perfectly competitive outcome occurs where $\gamma = 0$ and the monopoly union outcome where $\gamma = 1$. The-right-to manage outcome is delineated by these special cases, and is thus given by $0 \leq \gamma \leq 1$. In terms of testable predictions, however, the right to manage and the monopoly union models are the same, with the proviso that the variables proxying union power affect wages directly but not employment (Booth, 1995, p. 127).

But wage and employment outcomes on the demand curve are not Pareto efficient (McDonald and Solow, 1981). In the efficient contracts model, the firm and the union
are, subject to the condition that each does as well as if they disagree, assumed to achieve a settlement on a contract curve, which is the locus of efficient outcomes where no mutually beneficial alternatives remain unexploited. In contrast to the right-to-manage and monopoly union models, the firm under efficient contracting operates off the demand curve at a level of employment in excess of the profit maximizing level for that wage. For the vertical contract curve, the level of employment will be determined by the competitive wage alone and the contractual wage drops out of the employment equation.

Clearly, at points along the contract curve the marginal rates of substitution of employment for wages are the same for both sides. That is, the slopes $dw/dn$ of the firm isoprofit and union indifference curves are equal. The former is given by $pq'(n) - w$, where $p$ is product price, $n$ is employment, $q(n)$ is the production function, and $w$ is the negotiated wage. The latter is given by $-[u(w)u(b)/u'(w)]$, where $u(\cdot)$ is the underlying expected utility function of the union and $b$ is the alternative wage. The equilibrium of the contract curve model thus differs from the right-to-manage model (where $pq'(n) = w$) in the addition of the term containing the alternative wage.

The generalized Nash bargaining solution under efficient contracts is described by Booth (1995, pp. 131-134, 154-156), and occurs where the surplus or rent sharing schedule intersects the contract curve. It is equivalently described in broad strategic bargaining terms by Addison and Chilton (1997, pp.169-170, 190), who also point out that strategic bargaining may not lead to outcomes on the contract curve in situations of imperfect information and also that it may be efficient to delegate to the employer the choice of employment if some relevant information only becomes available after the contract is negotiated.

The empirical literature has largely focused on testing which of the two models more accurately characterizes wage and employment determination. The test is whether the actual wage and the alternative wage, or both under weak efficiency, play some role in wage determination. Using data for the International Typographical Worker Union (ITU), 1948-65, and a variety of alternative wage measures, Brown and Ashenfelter (1986) find that the contract wage is negatively and significantly related to their em-
ployment measure, while for its part the alternative wage often turns out to be positively related to employment.

Disputation has centred on the use of the efficient contracts model as the null. In particular, MaCurdy and Pencavel (1986) contend that there are no inclusion or exclusion restrictions enabling the researcher to determine whether contracts are efficient, and no compelling reasons or empirical evidence pointing to a vertical contract curve (see also Pencavel’s (1991, p. 114) criticism of Abowd’s (1989) empirical finding of the latter). Accordingly, the ‘absence’ of the alternative does not enable the researcher to reject efficiency. On the other hand, MaCurdy and Pencavel argue that the exclusion restrictions lacking from the efficient contracts model with respect to the alternative wage do apply in the case of the on-the-demand curve model, where the acceptable null hypothesis therefore is that the marginal product of labour equals its cost.

Using the same data as Brown and Ashenfelter (1986), MaCurdy and Pencavel report that variations in marginal product are not only explained by the current wage but also by other variables such as the alternative wage and the level of employment. As a result of such variables entering the union objective function, it is concluded that the bargaining solution is not located on the demand curve - and that the outcome is at least consistent with efficiency.5

The baseline literature admits of no consensus as between the union models of wage and employment determination with some studies pointing to on-the-demand curve outcomes and others to efficient contract solutions.6 But one common appeal to the facts that employers do not in practice bargain over employment will not suffice, since it can be shown that there are alternative enforcement mechanisms to contractually specified employment that dominate outcomes on the demand curve even if the firm

4Another criticism is that, in a model with quadratic adjustment costs (see section 1.2) all variables influencing profitability, utility, and union bargaining power will enter the employment relation (see Booth, 1995, p. 139).

5Note that the strength of the exclusion restriction for the demand curve model hinges in part on the quality of the estimate of marginal productivity and one important problem here is that variables modifying the utility of the union can have an impact on the behaviour of individuals, affecting their productivity and (hence) shifting labour demand (Cahuc and Zylberberg, 2001, p. 428).

6The two models will of course coincide where unions are indifferent to unemployment among their ranks but evidence in favour of such a corner solution is sparse (see Carruth et al., 1986, and the cautionary remarks of Pencavel, 1991, p. 71).
retains the right to determine employment (Espinosa and Rhee, 1989). The bigger issue is of course that these two alternatives are not nested in the standard treatments, thereby undermining tests claiming to have a basis in a nesting model. Testing for efficient bargaining hinges crucially on assumptions made about the firm’s revenue function and the union’s utility function, so that acceptance of a particular model may reflect invalid parameterization of these functions. Even if wage and employment outcomes do not lie on the demand curve there is currently no basis for arguing that off the demand curve outcomes are (a) efficient and (b) stable through time.

However, some progress has been recorded in treatments that seek more adequately to integrate the bargaining problem while constructing econometric models that nest the labour demand and efficient contract models. In particular, Doiron (1992) deploys a generalized Nash bargaining solution to identify a bargaining frontier or contract curve and to predict the position of the contract on that frontier. Each is also identified for the on-the-demand curve model. The differences between the models in terms of their first-order conditions are again shown to reside in residual terms. Although the two models are not naturally nested, Doiron uses this difference to construct an artificial compound nesting model as well as providing non-nested tests of the two models based on their likelihood values. Although unable to reject either model, she is able to identify the demand curve(s) and the contract curve with the positions of each determined by estimated bargaining power at the means of exogenous variables in the system.

Although there are some problems associated with the selection of empirical proxies for bargaining power and the disagreement point, Doiron’s unified approach is of considerable interest and further insights might be expected from fully specified bargaining models. That said, rather more can be expected of the next development identified here, characterized by the identification of a joint market imperfections parameter.

---

7See also Naylor (2003) who draws on industrial organization theory to offer a structure-conduct-performance application and who argues that the axiomatic approach of the generalized Nash bargaining solution pays insufficient attention to strategic considerations and underlying bargaining structure. Here, we have implicitly assumed that a game-theoretic solution is the same as the axiomatic approach in wage bargaining, under the assumptions set down by Binmore (1987).
3.2 Towards a Wider Classification of Outcomes

Thus far, no mention has been made of search frictions or heterogeneous worker preferences that impart a positive slope to labour supply curves (*viz.* monopsony, albeit not in the strict sense of a single employer in the labour market) or to product market distortions. Perhaps the major contemporary development in the spirit of the earlier right-to-manage versus efficient bargaining models has been the attempt to graft product market imperfections as well as additional labour market imperfections (monopsony) onto the standard models, reviewed earlier.

Of modern studies seeking to bridge the gap between product market distortions and labour market imperfections the most recent analyses have drawn on Hall’s (1988) product market framework of estimating price-cost margins. In particular, Dobbelare and Mairesse (2013) compare factor elasticities for labour and materials from firm production functions with their revenue shares to obtain an estimate of a parameter of joint labour market imperfections, $\psi$. In this framework, product and labour market imperfections drive a wedge between the output elasticities of labour and materials and their respective revenue shares. The sign and statistical significance of $\psi$ enables a distinction to be drawn between the efficient bargaining model and the monopsony model, and in the former case derive estimates of the price-cost mark up and extent of rent sharing parameters and in the latter case derive the price-cost mark-up and labour supply elasticity parameters.

The upshot of this procedure is that is the authors are able to classify industries into different regimes according to the type of competition in the product and labour market. As noted earlier, $\psi$ is expressed as the difference between the estimated output elasticities of labour and materials and their revenue shares, where $\psi = 0$ in the case of perfect competition in the labour market, $\psi > 0$ in the efficient bargaining case; and $\psi < 0$ in the monopoly case.\footnote{Under efficient bargaining (monopsony), the marginal employee receives a wage that is greater (less) than marginal revenue, so that the output-elasticity/revenue ratio for labor exceeds (falls below) the corresponding ratio for materials, yielding $\psi > 0$ ($\psi < 0$).} Note that the test for labour market imperfections takes the materials market as perfectly competitive and compares it to the labour market.
Accordingly, in a perfectly competitive labour market, which also encompasses the right-to-manage case, the only source of difference between the estimated output elasticity of labour and the share of labour costs in revenue is the firm price-cost mark-up, just as in the materials market. *Vulgo*: there will be no difference between the estimated elasticities of labour and materials and their revenue shares, and hence $\psi = 0$ as noted.

In principle, on the basis of the joint market imperfections parameter, the model admits of six possible regimes, as follows:

1. PC-PR, namely perfect competition in the product market and perfect competition again to include right-to-manage bargaining in the labour market (3%; 1%);
2. IC-PR, namely imperfect competition in the product market and perfect competition/right-to-manage bargaining in the labour market (26%; 21%);
3. PC-EB, namely perfect competition in the product market and (strongly) efficient bargaining in the labour market (0%; 0%);
4. PC-MO, namely perfect competition in the product market and monopsony in the labor market (21%; 10%);
5. IC-EB, namely imperfect competition in the product market and (strongly) efficient bargaining in the labour market (45%; 63%); and,
6. IC-MO, namely imperfect competition in the product market and monopsony in the labour market (5%; 5%).

Dobbelaere and Mairesse provide production function estimates for French manufacturing industry of the joint market imperfections parameter, *inter al.*, for manufacturing industry as a whole, for 38 manufacturing industries, and finally at the firm level. The single $\psi$ estimate(s) for the most aggregative level of analysis indicate that imperfect competition in both the product and the labour markets is the prevalent regime. The less aggregative industry analysis classifies each industry into one of the six distinct product and labour market regimes. The figures in parentheses in the six-way classification above give firstly the share of the 38 industries in each category and secondly the share of firms they represent. Thus, the dominant regime is IC-EB (as was suggested by the aggregate analysis) followed by IC-PR and PC-MO, etc.

For each of the three predominant regimes the authors then offer an interesting
analysis of differences in the output elasticities of the production function and $\psi$ (as well as in the price mark-up and extent of rent sharing or labour supply elasticity). In addition, for the dominant IC-EB regime, they link the measures of industry-specific product and labour market imperfections – again, the price-cost mark-up and the extent of rent sharing – to varying levels of four industry specific observables (profitability, unionization, import penetration and intensity of technology). Positive correlations are anticipated for price cost margins and price-cost mark-ups and between union density and price-cost mark-ups (as well as extent of rent sharing), and negative correlations are predicted between import penetration and price-cost mark-ups (and rent sharing), and possibly between RD intensity and price-cost mark-ups. The results broadly match these priors and are intended to add to the plausibility of the main findings.

In a final set of iterations, this time at firm level, Dobbelaere and Mairesse examine firm heterogeneity in product and labour market imperfections within the (predominant) regimes to which they belong. Although, the results of the authors’ random coefficient regressions at firm level conform with median values of key parameters for industries in the three predominant market forms, there is nevertheless evidence of considerable within-regime firm dispersion in the joint market imperfections indicator and the corresponding product and labour market imperfection parameters. Accordingly, the jury is still out on the question of the importance of regime differences across firms within a given industry.

Other limitations of this far-reaching study - all of which are recognized by the authors - include the estimation of a revenue rather than an output production function, the omission of a measure of adjustment costs that likely influence the estimated output-elasticity share ratios, and perhaps most important of all the absence of a richer dataset with matched employer-employee data. That said, the contribution of this attempt to determine the relevance of labour and product market distortions is to be contrasted with the frankly thin gruel offered by much of the investigation of the two standard models considered under 3.1.
4 Estimation Issues

The successful story of the evolution of the empirical research on the estimation of labour demand functions is too rich and involved to be told in such a short article (but see Hamermesh, 1993a). Here, our attention will first focus on what we consider to be unsatisfying features of the conventional approaches to the estimation of labour demand functions before attempting to offer some promising avenues allowing us to overcome or circumvent such limitations.

A critical parameter in the estimation of labor demand functions is the elasticity of labour demand with respect to wages. The simultaneity between labour supply and labour demand necessarily raises the issue of endogeneity. With longitudinal data on firms, an important source of endogeneity is partially attenuated via controls for firm fixed effects. But changes in wages remain problematic. Practitioners have learned to rely upon the elegant solution offered by Arellano and Bond (1991) using past information on wage differences and wage levels to instrument for current wages. This is an adequate route where the analyst is able to test for the validity of the instruments. But the extraordinary success of the Arellano and Bond procedure has in an important sense distracted researchers from inquiring into other exogenous sources of wage variation. In some particular situations - at least for part of the workforce - it may be convincingly argued that employers confront exogenous changes in the price of labor. Indeed, a branch of the literature on minimum wages has pursued exactly this route, by identifying the workers affected by minimum wage legislation and relating the wage hikes to their employment outcomes (Abowd et al., 2000). In similar vein, the specific features of the wage setting system may usefully be explored to identify exogenous wage variations. Thus, for example, wage agreements between trade unions and employers associations in continental Europe are often extended to the whole industry. Here, detailed information on the timing of the extension and/or the impact of the extension on the wage bill of the firm can be used to identify the impact of wages on employment outcomes, including firm exits (see Catalán and Villamor, 2012; Martins, 2013; Carneiro et al., 2014). Furthermore, the presence of downward nominal wage rigidities...
imposed by country-specific legislation may be explored in analogous fashion to the minimum wage literature to link job creation and job destruction with nominal wage rigidities.\textsuperscript{9}

There are, of course, other types of labour cost besides wages. In some instances, one can directly observe exogenous changes in these costs. Cases in point include changes in social security taxes (as in Gruber, 1997) or in firing costs (as in Martins, 2009).

Unanticipated labour supply shocks can also be instrumental in identifying the parameters of the labour demand function. In particular, studies exploring massive labour immigrant shocks in the framework set by Card (1990), Carrington and Lima (1996) and Angrist (1996) are especially revealing and assuredly merit further exploration.

In the estimation of labour demand functions one typically assumes the homogeneity of labour inputs and the micro-homogeneity of coefficients. Both assumptions are of course questionable and researchers have duly attempted to distinguish among a limited set of heterogeneous labor inputs (\textit{e.g.} skilled/unskilled, educated/non-educated, blue collar/white-collar workers) and/or among a broad industries. Given the increasing richness of contemporary datasets and current computer capacities, one can visualize how a detailed classification of occupations or job tasks may be sensibly incorporated in the estimation of production and labour demand functions.\textsuperscript{10} That is, by properly distinguishing the skills of workers necessary to fill job tasks, one can expect a better revelation of the substitutabilities and complementarities among inputs and their evolution over time.

The other issue is what can be learned from relaxing the notion that production units behave identically (\textit{i.e.}, have common parameters). A useful benchmark, given sufficiently long panels and high-frequency data, would be to estimate distinct labor demand functions at the establishment level.\textsuperscript{11} In a sense, this corresponds to a generalization of the approach of Hamermesh (1989) to the whole economy (or, more modestly,

\textsuperscript{9}See Kramarz (2001) for a review of micro-econometric studies of nominal wage rigidities.

\textsuperscript{10}See, for example, Torres \textit{et al.} (2013) for estimation of a Cobb-Douglas production function with close to 30,000 labour inputs in the context of an assortative matching model.

\textsuperscript{11}A convenient alternative would be to impose a subset of common coefficients (\textit{e.g.} time dummies) and allow the coefficients to vary, as in Cerejeira and Guimarães (2012) and Krueger and Mueller (2013).
to the industry).\textsuperscript{12} In Dobbelaere and Mairesse (2013) it is shown that the (firm) heterogeneity of the production function coefficients contains relevant information allowing the researcher to characterize the nature of the product and the labour market. An identical empirical strategy should prove fruitful in understanding why it is that firms adjust at different speeds and with different intensities to product and labour market shocks.\textsuperscript{13} In the same vein, the estimation of labour supply elasticities at the firm level provided by Webber (2013) for the U.S. offers convincing evidence that monopsony power is critical to understanding wage determination, as forcefully hinted at in wage models that stress the role of firm heterogeneity in setting wage policies (Abowd \textit{et al.}, 1999; Torres \textit{et al.}, 2013).

A promising alternative estimation procedure, albeit one that has yet to achieve full traction is the structural approach to the dynamics of labor demand developed by Cooper \textit{et al.} (2004). Starting from a neoclassical dynamic programming problem of profit maximization, these authors distinguish between different structures of labour adjustment costs, having recourse to simulation methods that match the relevant empirical moments of the employment adjustment distribution. The main tension in this approach is, of course, the selection of the empirical moments. But this choice is also its strength. Being able to observe interesting (or more revealing) moments may help in obtaining richer dynamics and more convincing estimates of the structural parameters. Despite the complexity of the approach, a number of possible extensions recommend themselves. They include: distinguishing between different types of firing costs (or types of labor contracts); matching the moments of the distribution of the duration of labor adjustment inaction (Ejarque and Portugal, 2007); modeling explicitly the entry and exit of firms (Schott, 2013); incorporating the notion of labor input indivisibilities; and replicating the moments of distributions of job flows and worker flows and, by implication, the distribution worker churning (Nilsen and Ejarque, 2008).

\textsuperscript{12}See Lee \textit{et al.} (1990) for a demonstration of the gains to be had from disaggregate analysis, and the dangers involved in aggregation.

\textsuperscript{13}In this set-up, the researcher still faces the thorny problem of distinguishing between the distribution of the true micro-heterogeneity and that of the sampling error. But see Dobbelaere and Mairesse (2013) for a way to distinguish between the two under a random coefficients approach.
5 Conclusions

Stricto sensu, research on labour demand is not amongst the liveliest and most successful areas of research within labour economics over the past 25 years. Yet, we have argued that significant progress has taken place. First, theory has been successful in bringing product and labour market imperfections into the study of labour demand. As part and parcel of this process, the rather sterile debate as to whether wage-employment outcomes are located on or off the demand curve has been enlivened not only by developments integrating the bargaining problem (while constructing econometric models that nest the main models) but also by the attempt to fuse product market imperfections and additional labour market imperfections onto the basic models. Second, better data and in particular rich longitudinal matched employer-employee datasets and improved estimation techniques have yielded sound estimates of the parameters of interest. Equipped in this manner, and benefiting from a myriad of policy interventions that impact labour demand, analysts are now able to provide competent, if not always uncontroversial, policy advice. Third, bridge-building with other fields of study has occurred. The initial results of an integrative approach involving international trade and corporate finance are now beginning to surface.

We also argue that there are still a number of promising research avenues within labour demand theory that will guarantee the field a long and active future life. Such areas include, but are not restricted to, the incorporation of labour (and other input) heterogeneity and the detection of exogenous sources of wage variation on which to base estimates of the relevant elasticities. Given that structural modelling of labour demand is still in its infancy, sizable payoffs from model extension should be anticipated. In sum, while the current outlook is positive, the future is yet more compellingly so.
References


