Resolving the Public-Sector Wage Premium Puzzle by Indirect Inference[1]

Patrick Minford[2], Yi Wang[3] and Peng Zhou[4]

Abstract

This paper investigates the public-sector wage premium in the UK using a microfounded economic model and indirect inference. To answer the question whether there is public-sector wage premium, we ask an equivalent question—whether a model assuming perfect competition can explain the data. The neoclassical labour economic model is tested and estimated without introducing any ad hoc gap between the theoretical and empirical models. Popular econometric models are used as auxiliary models to summarise the data features, based on which we evaluate the distance between the observed data and the model-simulated data. We show that it is not the non-market factors, but the total costs and benefits of working in different sectors and so simple market forces, that create the public-sector wage premium. In other words, there is no inefficiency or unfairness in the labour market to justify government intervention. In addition, selection bias test can be incorporated into the indirect inference procedures in a straightforward way, and we find no evidence for it in the data.

Key Words: Public-Sector Wage Premium, Selection Bias, Indirect Inference, Monte Carlo

JEL Classification: C21, C35, J31, J45

---

[2] E26, Aberconway Building, Cardiff Business School, Cardiff University, UK. Email: minfordp@cardiff.ac.uk
[3] Cardiff School of Management, Cardiff Metropolitan University, UK. Email: ywang@cardiffmet.ac.uk
[4] Corresponding Author: D47, Aberconway Building, Cardiff Business School, Cardiff University, UK. Email: zhoup1@cardiff.ac.uk

There has been a long discussion in many countries as to whether public sector workers are paid too much (Smith, 1976; Robinson and Tomes, 1984; Disney and Gosling, 1998; Melly, 2005; Gomes, 2015; Jones et al, 2018). The global financial crisis and the Great Recession revived the debate over the need to restructure the public sector. The wage premium in the public sector lies at the centre of this debate in the mass media and the literature. Though most studies agree that the public-sector wage premium[5] (PSWP) has gone up since the 2008 financial crisis and that females in the public sector tend to enjoy a higher wage premium than their male counterparts (Blackaby, 2012), the empirical literature in labour economics mainly adopts the Mincer-type econometric models to estimate the wage premium but the identification issue is not fully addressed. Existing empirical techniques of estimating PSWP fall into one of the four types (Wang and Zhou, 2017; 2019): single-equation regression method (e.g. Blackaby et al, 2012; Chatterji et al, 2011), decomposition-based method (e.g. Morikawa, 2016; Anton et al, 2015), matching-based method (e.g. Gibson, 2009; Ramoni-Perazzi and Bellante, 2006), and multiple-equation regression method (e.g. Afonso and Gomes, 2014; Maczulskij, 2013). However, due to the missing theoretical microfoundation for the estimated PSWP, it is not clear whether the public sector wage should be changed to improve the efficiency or fairness of the labour market. We call this the public-sector wage premium puzzle. The former part of the puzzle is a matter of positive analysis, and the latter is a normative issue. We note that almost all the existing empirical methods use econometrics on reduced-form models. Very few attempts have been made to confront the microdata with the structural economic models *per se*.

The usual method in applied microeconomics for testing for the presence of such an effect is to estimate a reduced-form linear model on the data in which the personal characteristics of working householders are entered as controls and an additional dummy is entered to test for the presence of the effect. In this paper we address the question by reversing the order of the analysis: instead of fitting an elaborate reduced-form model and then asking whether a structural model of interest to us is consistent with it, we begin by specifying the structural model of interest and then asking whether the reduced-form we find in the data is consistent with it. We argue here for a different empirical strategy based on the underlying structural model. There are several reasons for this.

First, the structural model is nonlinear and theoretically identified. In traditional applied microeconomics or microeconometrics, the linearity (or log linearity) of regression models greatly reduces the accuracy of the predictions of highly nonlinear theoretical economic model. With these deficiencies, there is a considerable risk that what is tested or estimated in the

---

[5] In this paper, the wage premium is defined as only including the pecuniary wage. We are aware that there are other non- pecuniary benefits of working in the public sector, such as job security and better pension scheme. These factors are, however, not observable and partially absorbed by other observed factors included in the analysis (e.g. gender, marital status, number of children, etc.).

linearised reduced form is not what the corresponding economic model actually implies. More-over, there is the problem of identification: the econometric model may be consistent with another economic model altogether (for example by reverse causation or causation by omitted factors as occurs with selection bias). While efforts are made to create identifying restrictions (such as on the signs of coefficients), these are rarely sufficient to rule out all rival structural models to the one the researcher hopes to be testing.

Second, we can test it with substantial power through our proposed Indirect Inference proce-dure, never before used in such a microeconomic context, though familiar now in much mac-roeconomic work.

Third, we can deal directly with concerns about possible selection bias, that could well affect the reduced-form estimate of the PSWP. In the rest of this paper we set out our proposed method and our detailed results. We show that the usual conclusion that there is a PSWP is convincingly over-turned.

This structural model, if it passes this test, has clear welfare implications: it implies that the labour market is by and large efficient, and no government intervention is needed. If it holds, the PSWP estimated in reduced form models is only a statistical phenomenon; it is part of the data description but only an implication of a model in which there is no such premium: the premium comes about purely as an incidental implication of supply and demand characteristics of the people surveyed in the data.

Of course, we could have chosen to test other structural models in the labour economics liter-ature (e.g. imperfect competition or search). It remains open to other researchers to do so. However, we reason first that if our model can pass the powerful empirical test we use and will explain in full shortly, then it is at least a major candidate to be considered in policy analysis. We also know (from both Le et al, 2016 and Meenagh et al, 2019) that under this test mis-specified models are rejected with high, close to 100%, probability; therefore, if the model we have tested were to be true, then other models would be rejected. Arguably, we cannot rule out the possibility that our model is somewhat false and hence that another model, also somewhat false, could also pass our test, but we leave the search for such a model for future research.

To summarise, we go behind the statistical reduced form models that describe the data and ask what structural model could be generating these descriptions; and specifically, whether a model derived from standard optimising theory under perfect competition could be that model. To answer this question we use the powerful testing procedure of indirect inference.

In the remainder of this paper, we first discuss the method of indirect inference. Then in section 2, we set our underlying structural model. In section 3, we set out the auxiliary model we will

use to describe the data behaviour. Section 4 shows the results of our test and section 5 concludes.

1   Indirect Inference

1.1   Indirect Inference Test

In this section we explain the method of Indirect Inference that we use for comparing the above models to the data as we report below. Indirect Inference originated as a systems estimator with Smith (1993), Gregory and Smith (1991) and Gourieroux, et al. (1993) for getting around the difficulties in estimating models with complex likelihood functions. Minford, et al. (2009) and Le, et al. (2011) then developed the method as a testing procedure based on comparing the simulated behaviour of a structural model with the behaviour found in the data. The idea is simple enough: the data relationships are described by some model- known as the 'auxiliary model'-, implied by the reduced form, and whose role is to describe the data behaviour. Then the structural model being tested, if it is the data-generating process will, when simulated with random model shocks generate data which on average has the same behaviour; the joint distribution of the auxiliary model parameters implied by these data samples details the probability of each parameter set. From it we can compute the probability of the parameter set found in the actual data sample. If that probability falls below a lower threshold, say 5%, the model is rejected.

Thus, first we construct an auxiliary (empirical) model which is fitted to the actual data to give descriptors of the data. Such descriptors can be coefficients of the auxiliary model or functions of them (such as moments or impulse response functions); they are a parsimonious description of the facts, implied by the reduced form, i.e. the solved-out data relationships from the model. As in our application here, the data descriptors can be the parameters of the reduced form model itself, which is therefore treated as the auxiliary model. The method then turns to the model to be tested (the structural model) and uses it to generate a large number of simulated data samples on each of which the same auxiliary model is estimated. This establishes an empirical distribution of the data descriptors conditional on the hypothesis that the structural model is true. The method then evaluates the distance between the actual data descriptors and this distribution of them, to decide whether the structural model is rejected at a selected level of significance; this testing process is ultimately carried out on the model once it has been estimated, also by indirect inference (which chooses model parameters to minimise the above distance).

Figure 1 Joint Distribution of Estimates of Auxiliary Regression Parameters
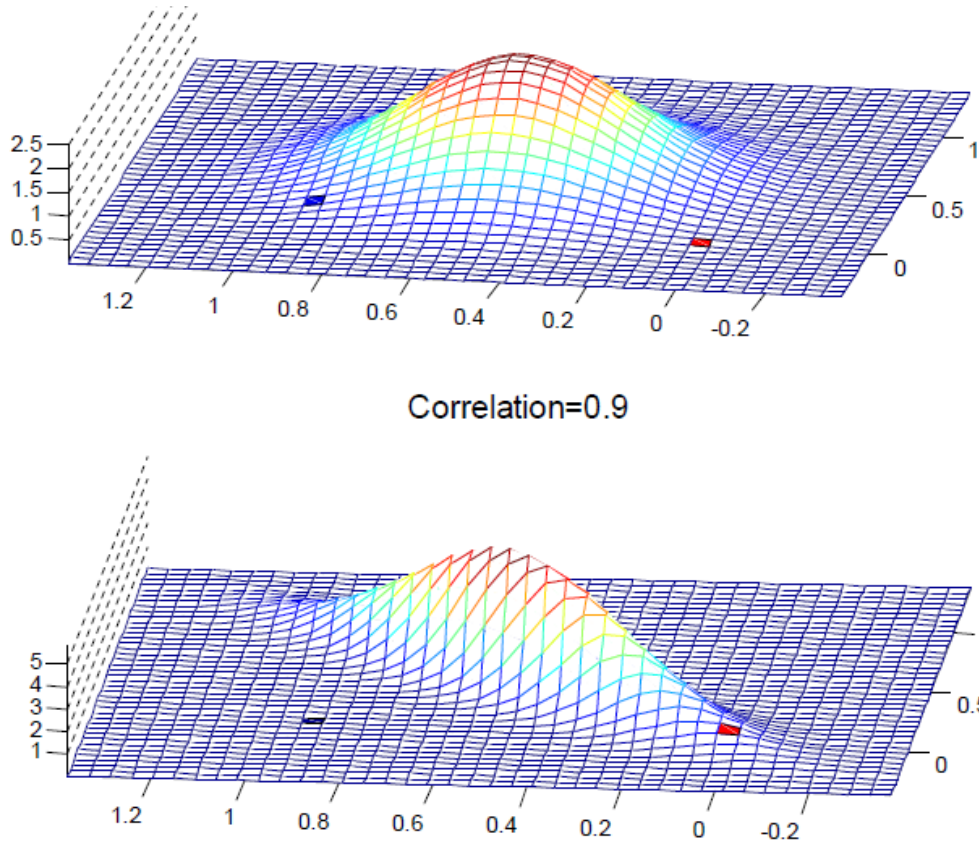
Notes: The top is the case where estimates of auxiliary model coefficients uncorrelated. The bottom is the case where they covary, i.e. the covariance matrix $\boldsymbol{\Omega}$ has non-zero off-diagonal elements. The PDF of k simulated coefficients $\boldsymbol{\vartheta}$ (which depends on the structural parameters) is given by $\frac{1}{(2\pi)^k|\Omega|} \exp\left(\frac{1}{2}\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^{(n)}\right)'\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^{(n)}\right)\right)$.

The illustrations in Figure 1 show the joint distribution implied by the structural model over two selected descriptor coefficients where the structural model implies either (top) no covariance between the two or (bottom) a high covariance; in general structural models imply non-zero covariances between the descriptor coefficients, as found in the covariance matrix, Ω, The blue and red dots represent possible descriptor coefficient values found in the data sample (thus the red dot would cause the model to be rejected on the top panel, accepted on the bottom one; and vice versa for the blue dot). The notes on the Figure 1 show the probability distribution function; notice that its exponent is the Wald statistic which is thus the key indicator we use in evaluating the model.

We now go through the practical details of the II test procedure in our application here. The structural economic model is a system of equations derived from worker/firm optimisation behaviour and the market clearing condition. It consists of some endogenous variables ($\mathbf{y}$) to be explained (wage and working hours in our case) and residuals ($\mathbf{z}$) that the structural model cannot explain (i.e. the discrepancies between the observed and model-predicted wages), given

the structural parameters ($\boldsymbol{\theta}$). $\mathbf{z}$ differ across workers and jobs, whose characteristics are the exogenous conditioning variables ($\mathbf{x}$). The part of $\mathbf{z}$ that $\mathbf{x}$ cannot capture are idiosyncratic shocks or "innovations" ($\boldsymbol{\varepsilon}$).

$$f\left(\mathbf{y}, \mathbf{z}\left(\mathbf{x}, \boldsymbol{\varepsilon}\right), \boldsymbol{\theta}\right) \equiv f\left(\mathbf{y}, \mathbf{x}, \boldsymbol{\varepsilon}, \boldsymbol{\theta}\right) = 0.$$

Assume this model can be solved in a reduced form:

$$\mathbf{y} = g\left(\mathbf{z}\left(\mathbf{x}, \boldsymbol{\varepsilon}\right), \boldsymbol{\theta}\right) = g\left(\mathbf{x}, \boldsymbol{\varepsilon}, \boldsymbol{\theta}\right).$$

Given some calibrated parameter values $\boldsymbol{\theta}_0$, the observed endogenous variables ($\mathbf{y}^{(a)}$) and the conditioning variables ($\mathbf{x}^{(a)}$), we will be able to compute all the actual innovations termed as $\boldsymbol{\varepsilon}^{(a)}$ based on the structural form $f\left(\mathbf{y}^{(a)}, \mathbf{x}^{(a)}, \boldsymbol{\varepsilon}^{(a)}, \boldsymbol{\theta}_0\right) = 0$. The number of shocks must be equal to the number of observed endogenous variables; otherwise, we will have "stochastic singularity", which would (absurdly) imply that some endogenous variables are deterministically related to the rest.

Then the actual conditioning variables ($\mathbf{x}^{(a)}$) and the actual innovations ($\boldsymbol{\varepsilon}^{(a)}$) are bootstrapped $S = 1000$ times, resulting in S sets of exogenous variable realisations $\mathbf{z}^{(n)}$. Using these S sets of exogenous variables, we simulate S sets of endogenous variables $\mathbf{y}^{(n)}$ by substituting the bootstrapped exogenous variables and calibrated parameters into the reduced form:

$$\mathbf{y}^{(n)} = g\left(\mathbf{z}^{(n)}\left(\mathbf{x}^{(n)}, \boldsymbol{\varepsilon}^{(n)}\right), \boldsymbol{\theta}_0\right).$$

Then, we can use the appropriate auxiliary model (here the GOLS) to summarise the features of both the actual and the simulated data of the endogenous variables. The parameters of the auxiliary model are denoted as $\boldsymbol{\vartheta}$, so there will be a $\boldsymbol{\vartheta}^{(a)}$ based on the actual data $\mathbf{x}^{(a)}$ and N sets of $\boldsymbol{\vartheta}^{(n)}$ based on the simulated data $\mathbf{x}^{(n)}$. A standard Wald test can be implemented by computing the Wald statistic: $\text{Wald}(\boldsymbol{\theta}_0) \equiv \left(\boldsymbol{\vartheta}^{(a)} - \overline{\boldsymbol{\vartheta}}^{(n)}\right)'\left(Var\left[\boldsymbol{\vartheta}^{(n)}\right]\right)^{-1}\left(\boldsymbol{\vartheta}^{(a)} - \overline{\boldsymbol{\vartheta}}^{(n)}\right)$.

The Wald statistic has a χ^2 distribution with degrees of freedom equal to the number of parameters in the vector $\boldsymbol{\vartheta}$. If the Wald statistic lies within the 95% confidence interval, then the original model $f(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}_0) = 0$ is said to be able to generate the actual data, i.e. the model is true. Otherwise, the model is rejected. The flowchart in Figure 2 illustrates the workings of the II test procedures.

Figure 2 Flow Chart of Indirect Inference

Figure 2 Flow Chart of Indirect Inference
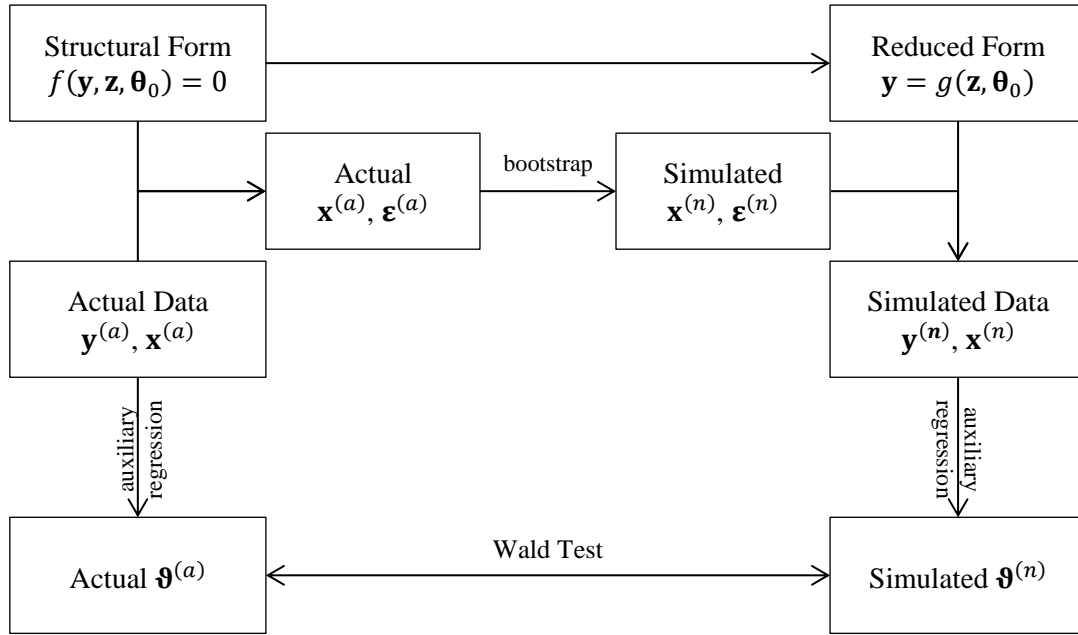
Note that the conclusion of the test does not depend on the likelihood of the data, but the likelihood of a specific feature of the data—the chosen auxiliary model or auxiliary function of the data. That is why it is called indirect inference, in contrast to direct inference based directly on the data. Why use II instead of direct FIML and the likelihood tests based on it, such as the Likelihood Ratio (LR) test? Asymptotically—i.e. with very large samples—there would be no difference: both tests would have infinite power. However, in practice economists are faced with small samples: some micro panel samples are large but once one has controlled for myriad special factors their residual sample variation effectively shrinks to a small size too. Hence, we really need to know how powerful our tests are in small samples. The evidence on this we have so far about II (see Le et al, 2016, and Meenagh et al, 2019, for recent surveys) is that it is considerably more powerful than the LR test. Furthermore, by increasing the number of "data features" to be matched its power can be increased steadily until the data features exhaust the differential implications from the model: for example, in a large macro model such as Smets and Wouters (2007) the VAR reduced form extends to some 200 coefficients and as the VAR used to describe the data is increased in size so does the power of the II Wald test. However, in practice the investigator requires a power that is appropriate to the problem: namely such that there is the possibility of finding a tractable model that passes the test while also giving strong reassurance that the model cannot be badly false. As we will see below in a Monte Carlo experiment, some of the auxiliary models are far too powerful on this criterion while others are not powerful enough. Under II we can choose the power of the test flexibly to meet our purposes as investigators or policymakers; the trouble with LR is that in general it provides just

one all-purpose level of power that cannot be varied, and this level is generally far too weak. II can give us a Goldilocks power that is "just right".

1.2 Indirect Inference Estimation

We implement the II test for an initial calibration $\boldsymbol{\theta}_0$. As a starting point, the model may be rejected because this initial calibration may not be the best according to the auxiliary model criterion. An optimisation procedure can then be carried out to search for the optimal calibration $\hat{\boldsymbol{\theta}}$, which minimises the objective function—the Wald statistic. The procedure will raise the probability of accepting the model to the maximum possible. The resulting optimal calibration $\hat{\boldsymbol{\theta}}$ is therefore the II estimation of the model parameters:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} \text{Wald}(\boldsymbol{\theta}).$$

Note that the estimation here is a multivariate global optimisation problem, which has a stochastic and non-smooth objective function. It is usually impossible to derive the analytical solution for $\hat{\boldsymbol{\theta}}$. Instead, a numerical algorithm is typically used to search for the optimal calibration within the parameter space. Various global optimisation algorithms are available for this purpose, such as the simulated annealing and genetic algorithms.

The simulated annealing algorithm (for example, Le et al, 2010, 2011) has the disadvantage that the optimum may still depend on the starting point (despite the name of "global" optimisation algorithm). The genetic algorithm provides a more thorough search in the parameter space using a population-based iteration (simulated annealing is point-based iteration), and it is not dependent on the starting point.[6] We will use this more robust algorithm to undertake the II estimation.

2 The Structural Model

The model for testing the PWSP is based on a simple neoclassical labour market model with the assumption of perfect competition. The representative worker maximises utility subject to a budget constraint and a time constraint (the supply side of the labour market), while the representative firm maximises profit subject to a technology constraint (the demand side of the labour market). The labour market clears with a market-agreed wage (price of labour) and working hours (quantity of labour). We could have used a more sophisticated model such as multi-sector model or search-and-matching model, but what we ultimately want is to check if a model with competitive labour market can pass the test. If so, then the policy implication is that we do not need further policy intervention according to the first fundamental theorem of

---

[6] The genetic algorithm was initially developed by John Holland in the 1960s inspired by the evolution concept in the biological literature. It has been widely used in engineering, economics and finance recently (e.g. Foreman-Peck and Zhou, 2018).

welfare economics. Therefore, the choice of model depends on the question at hand. We are not trying to find a structural model to maximise the goodness of fit, but to answer a policy-related question here. The neoclassical model is exactly the null hypothesis we want to challenge, just like assuming a regression coefficient equal to 0 in microeconometrics.

2.1 The Supply Side

The representative worker faces the following standard optimisation problem:

$$\max_{C,X,L} U\left(C, X\right) = \left[C^{\frac{s-1}{s}} + \alpha X^{\frac{s-1}{s}} \right]^{\frac{s}{s-1}}, \text{ subject to:}$$

$$\text{Budget Constraint: } C = wL;$$

$$\text{Time Constraint: } X + L = T.$$

For simplicity, the utility function is assumed to have a constant elasticity of substitution (CES) with the elasticity equal to s. There are two utility inputs, consumption C and leisure X, and the relative utility weight on leisure is α. The budget constraint is expressed in real terms, so wL is real wage income. The time endowment T is allocated between leisure X and labour L.

The first order condition is obtained by taking the derivative with respect to L, leading to the intratemporal condition—the marginal rate of substitution between leisure and consumption is equal to the real wage:

$$w = \alpha \left(\frac{wL}{T-L} \right)^{\frac{1}{s}}$$

This is the marginal condition for the representative worker, so it is satisfied by all observations only when the workers are homogeneous. In reality, individual characteristics, such as age, gender, race and education, are all different across individual workers. It is assumed that the wages and hours we observed among the individuals are all market-agreed amounts taking into account of these individual characteristics. Therefore, for each particular individual, the marginal condition is:

$$w_i = \alpha_i \left(\frac{w_i L_i}{T-L_i} \right)^{\frac{1}{s_i}}$$

The difference in the structural parameters on the supply side (α_i, s_i) is derived from the individual characteristics ($\mathbf{ind}_i$). Therefore, if we extract the difference in these parameters and express them as an "error terms" (S_i), we can rewrite the individual marginal condition as:

$$w_i = \alpha \left(\frac{w_i L_i}{T - L_i} \right)^{\frac{1}{s}} S_i \qquad\qquad ...(1)$$

The residual S_i can be interpreted as an "exogenous shock". We can break this exogenous sup-ply-side "shock" S_i into a deterministic component capturing the differences in individual char-acteristics and a stochastic component ε_i^S, supposedly to be IID:

$$S_i = \overline{S} \times \exp\left(\boldsymbol{\eta}_S' \mathbf{ind}_i \right) \times \exp\left(\varepsilon_i^S \right)$$

The specification is chosen to be exponential so that the coefficients can be interpreted as elas-ticities. Take natural logarithms on both hand sides of this equation:

$$\ln S_i = \ln \overline{S} + \boldsymbol{\eta}_S' \mathbf{ind}_i + \varepsilon_i^S, \text{ where } \varepsilon_i^S \sim IID\left(0, \sigma_S^2\right) \qquad\qquad ...(2)$$

Here, $\mathbf{ind}_i$ is a vector of individual characteristics as used in the econometric modelling method, such as age, gender, race and education, and $\boldsymbol{\eta}_S$ is the coefficient vector of each term inside $\mathbf{ind}_i$. The innovation term ε_i^S is supposed to be an IID random variable under the null hypothesis (there is no selection bias, or equivalently there is no endogeneity bias), so ε_i^S is uncorrelated with the terms of $\mathbf{ind}_i$.

2.2 The Demand Side

A representative firm faces the following standard optimisation problem:

$$\max_{Y,L} \pi = Y - wL, \text{ subject to:}$$

Technology Constraint: $Y = AL^\gamma$.

Y is the output of the firm, π denotes the economic profit, and A is to capture the average total factor productivity level in the production function. Note that the output Y here is private goods and services for the private sector, but it can be public goods and services in the public sector. Like any economic model, the profit here is broadly defined to include any goal such as social welfare or votes. It is essentially a model of optimal efficiency. The profit/output is monetised to make the objective function commensurable, but there is no need for worrying about the measurement discrepancy between Y and GDP data, since we are not building a macroeco-nomic model. This paper focuses on the labour market, so capital is treated as given in the production function, and it is absorbed into A. The first order condition with respect to L is the standard marginal condition for a firm—marginal product of labour equals to marginal cost of labour:

$$w = \gamma A L^{\gamma - 1}$$

9

Again, this is the marginal condition for the representative firm or job, so it holds for all observations only when jobs are homogenous. In reality, job attributes, such as industry, sector, occupation, work mode and location, are all different. It is again assumed that the wages and hours we observed among the individuals are all market-agreed amounts taking into account of these job attributes. The marginal condition for a particular job is:

$$w_i = \gamma_i A_i L_i^{\gamma_i - 1}$$

To make the condition linked with the representative firm's marginal condition, a demand-side surplus term (D_i) is needed to account for the effects of job attributes:

$$w_i = \gamma A L_i^{\gamma - 1} D_i \qquad \qquad \text{...(3)}$$

Similar to the supply-side surplus, the exogenous error term D_i can also be further decomposed into a job attributes component and an IID innovation:

$$D_i = \bar{D} \times \exp\left(\mathbf{\eta}'_D \mathbf{job}_i\right) \times \exp\left(\varepsilon_i^D\right)$$

Here, $\mathbf{job}_i$ is a vector of job attributes, such as industry, sector, occupation, work mode and location, and $\mathbf{\eta}_D$ is the coefficient vector of each term of $\mathbf{job}_i$. In particular, we exclude from the microfounded model the public-sector dummy, i.e. whether the job is in public sector or private sector. We do this because this is an *ad hoc* addition to the microfounded model that is not consistent with the perfect competition assumption. The whole point of our investigation is to see whether our microfounded model can match the data features summarised by the auxiliary model (note that the auxiliary model does include the public-sector dummy as this model is intended to be a pure description of the behaviour found in the data, which we know needs to include this dummy) even without such an *ad hoc* factor.

Take natural logarithms to rewrite this equation into a regression-like model:

$$\ln D_i = \ln \bar{D} + \mathbf{\eta}'_D \mathbf{job}_i + \varepsilon_i^D \text{, where } \varepsilon_i^D \sim IID\left(0, \sigma_D^2\right) \qquad \qquad \text{...(4)}$$

2.3 Market Equilibrium

If the labour market clears, the supply of a particular sort of labour L_i is equal to the demand for it. To summarise, equation (1) and equation (3) describe the equilibrium[7].

---

[7] Note that this is a partial equilibrium in the labour market, not a general equilibrium of the whole macroeconomy, so it does not require the clearance of goods market.

$$\begin{cases} w_i = \alpha \left(\dfrac{w_i L_i}{T - L_i} \right)^{\frac{1}{s}} S_i \\ w_i = \gamma A L_i^{\gamma-1} D_i \end{cases}$$

There are two endogenous variables in this system, the real wage w_i and the working hours L_i, and there are two exogenous variables, S_i and D_i, which are further modelled by two generalised linear regressions (2) and (4).

$$\begin{cases} \ln S_i = \ln \bar{S} + \boldsymbol{\eta}'_S \mathbf{ind}_i + \varepsilon_i^S \\ \ln D_i = \ln \bar{D} + \boldsymbol{\eta}'_D \mathbf{job}_i + \varepsilon_i^D \end{cases}$$

Note that the $\boldsymbol{\eta}$'s in the two equations are not exactly the same as the regression coefficients. The strict interpretation of $\boldsymbol{\eta}^S$ is the "elasticities of supply-side surplus", and that of $\boldsymbol{\eta}^D$ is the "elasticities of demand-side surplus". In contrast, the $\boldsymbol{\beta}$s in the econometric models are the elasticities of wage. Accordingly, there are two innovations (regression error terms), ε_i^S and ε_i^D, respectively describing the idiosyncratic disturbances on the supply-side surplus and demand-side surplus. Again, they are different from the error terms in the regressions. In fact, the error term (of the reduced-form model) should be a function of the two innovations (of the structural model equations). The method of solving this nonlinear equation system is detailed in Appendix 1.

In the present model the focus is on the "intensive margin", i.e. the working hours rather than the "extensive margin", the participation decision—whether to work at all (Hansen, 1985). This is desirable because it matches the microdata we use, the Labour Force Surveys (LFS) collected by the Office for National Statistics (ONS) in the UK in 2011. The original dataset accounts for a 25% random sample of individuals aged 20-64 years. Full-time students, unpaid family workers, and people on government training schemes are excluded. There are 6,216 observations finally included in the analysis. The average wage in the public sector is 9.38% higher than the private sector (23.67% for females and 6.62% for males). This crude data feature is in line with the stylised facts identified by the previous econometric literature. II will establish whether the positive PSWP found in the data can be accounted for by a plain structural model of perfect competition in which there is individual optimisation behaviour and possibly selection bias. We will exploit this last possibility by explicitly modelling both a no-selection-bias structural model and a structural model with explicit selection bias. In this way we can generalise our test of the basic structural model so that it is not imperilled by selection bias.

3 Which Reduced Form Should be Chosen as the Auxiliary Model?

In the traditional empirical literature, as noted above, reduced form models are usually adopted to estimate the PSWP. We will use these as our potential auxiliary models, there to describe the data behaviour. Five main reduced forms can be distinguished in the PSWP literature:

- **Type 1**: Single-Equation-Regression Model. This directly estimates in a wage determination equation the coefficient of the dummy variable describing whether or not an individual is working in the public sector. The simplest way is OLS as in Blackaby et al (2012) and quantile regression is also commonly used to correct for outliers.
- **Type 2**: Decomposition-Based Model. Based on two separate regressions on the sub-samples, it allows for sectoral heterogeneities in all regressors (slopes) in addition to the sector average (intercept). This type of method includes the Blinder-Oaxaca decomposition (BOD) adopted in the early literature (Smith, 1976; Gunderson, 1979) and the later extensions by Juhn et al (1993) and Melly (2005).
- **Type 3**: Matching-Based Model. Based on a sector choice regression, it calculates the wage premium by finding the counterpart individuals in the two sectors in terms of a certain matching criterion. The most popular matching-based methods are Propensity Score Matching (PSM) and Nearest Neighbour Matching (NNM), as used in Ramoni-Perazzi and Bellante (2006) and Gibson (2009).
- **Type 4**: Multiple-Equation-Regression Model. The fourth type includes the Heckman selection model (HSM), the treatment effect models, simultaneous equation models as well as the 2SLS estimator. They address the problem of selection bias by using an explicit selection equation or excluded instruments to account for the sector choice, so that the estimated coefficients in the wage equation are unbiased.
- **Type 5**: GOLS "Grouped OLS" (GOLS) is in fact a variant of Type 1. This is done by grouping the 35 coefficients of the OLS regression into 8 categories, one of which is the PSWP. The details of the grouping are shown in Table 4 in Appendix 2. The grouped auxiliary parameters are basically the arithmetic average of the underlying coefficients of the OLS regression (Type 1). Since the estimated coefficients of the original regressors are normally distributed asymptotically, the average (a linear combination) of them is also normally distributed. By doing this grouping, the dimensionality of the auxiliary parameter vector has been reduced to obtain the appropriate level of power.

These models vary in the amount of information (i.e. the number of parameters in the model) they provide about the data behaviour. This turns out to be of great importance in determining the power of the II test. The more information is included in the auxiliary model the harder it is for the structural economic model to pass the test of matching this behaviour. At one extreme, only the "real world" structural model would match a perfect and total description of the data

behaviour. Yet a test so powerful would reject all models, giving us no insight into the causal processes underlying behaviour. At the other extreme an auxiliary model with little information will discriminate poorly and have very low power. We need a test power that allows good discrimination while accepting models that are reasonably accurate. To establish this, we begin our discussion with a Monte Carlo experiment to explain our choice of auxiliary model. In this experiment we gauge the power of the indirect inference test for different auxiliary models, each containing more or less information about the data behaviour.

In this experiment we follow Le et al (2016). We simulate data from a structural economic model with a good claim to be appropriate and, treating it as true, see how the repeated samples from that model reject false models as they depart progressively from the true generating model. This rejection comes about via the Wald test of their match against the data behaviour as described by an auxiliary model. We can do this for the different auxiliary models we identified above, to see which of them would give us the appropriate power for our test.

The Monte Carlo experiment is designed as follows. The "true" economic model $\mathcal{M}(\widehat{\boldsymbol{\theta}})$ is constructed with $\widehat{\boldsymbol{\theta}}$ being the "true" structural parameters[8], based on which we can simulate 1000 datasets. To see the statistical power of the II test (its ability to identify false models), the parameters are manipulated up and down in an alternate fashion to create some "falsified" models $\mathcal{M}(\widetilde{\boldsymbol{\theta}})$. The degrees of falsification are chosen to be 1%, 5%, 10% and 20% higher/lower than the "true" values ($\widehat{\boldsymbol{\theta}}$). For each of the 1000 datasets, an II test is conducted based on all five types of auxiliary models. If the resulting p-value of a test is smaller than 5%, then we reject the model. Conversely, if the resulting p-value is greater than 5%, then we accept the model being true. The proportion of the 1000 tests that reject the model being true is therefore the statistical power of II test. Figure 3 summarises the following:

- **Step 1: Simulation.** Under the true model/parameters $\widehat{\boldsymbol{\theta}}$, simulate $S = 1000$ sets of data using the true model $\mathcal{M}(\widehat{\boldsymbol{\theta}})$.
- **Step 2: Falsification.** Falsify the parameters to $\widetilde{\boldsymbol{\theta}}$ by scaling the odd ones up by $+x$ and the even ones down by $-x$, where $x = 1\%, 5\%, 10\%, 20\%$.
- **Step 3: Test.** Apply the II test of the null hypothesis that "the model is true". Note that the model here refers to the ones with falsified parameters $\mathcal{M}(\widetilde{\boldsymbol{\theta}})$.
- **Step 4: Conclusion.** For all the N simulations, we can obtain N test statistics, critical values, p-values and test results (0 as true and 1 as false). The proportion of rejections is just the simulated power.

A discussion of what $\boldsymbol{\theta}$ should include is due here. We include in $\boldsymbol{\theta}$ only the structural parameters that govern the decision making of individual workers and firms, such as constant

---

[8] This is the structural parameters we will estimate following the II estimation procedures to be explained in the next section.

elasticity of substitution (s), preference weights between leisure and consumption (α), income share of labour (γ), average productivity (A). The elasticities that describe the effects of conditioning variables (individual characteristics and job attributes) on the exogenous variables are not included in $\boldsymbol{\theta}$ because they will be re-estimated when the structural parameters are falsified, in order to be consistent with the false model and the sample data.

Here the structural parameters ($\boldsymbol{\theta} \equiv s, \alpha, \gamma, A$) are falsified and the error parameters ($\boldsymbol{\eta}_S, \boldsymbol{\eta}_D$) are re-estimated for each test. The resulting power of the II test using the five auxiliary models are shown in Table 1. Type 1 and Type 4 have excessive power because both have too many auxiliary parameters to match. A slight deviation of the structural parameters will result in rejection. Type 2 and Type 3, on the other hand, only have 2 auxiliary parameters to match, so the power is too weak—when a model is 20% false, there is only about a two thirds chance of its being rejected. This leaves Type 5 (GOLS) an eclectic model lying between the two extremes. Of the five reduced forms that are available to us as auxiliary models, it comes closest to having the 'goldilocks' feature. For example, both OLS and HSM would reject virtually any model we might specify, while BOD and PSM would fail to reject all of the time models as much as 20% false. GOLS achieves that virtually total rejection rate on models 20% false, and steadily raises its rejection rate for models less false, so still giving modestly false models a chance of success. Of course, in practice the tractable models we specify will have some degree of falsity, so we would not want them to be too easily rejected. Accordingly, in our testing that follows we use GOLS as the auxiliary model.

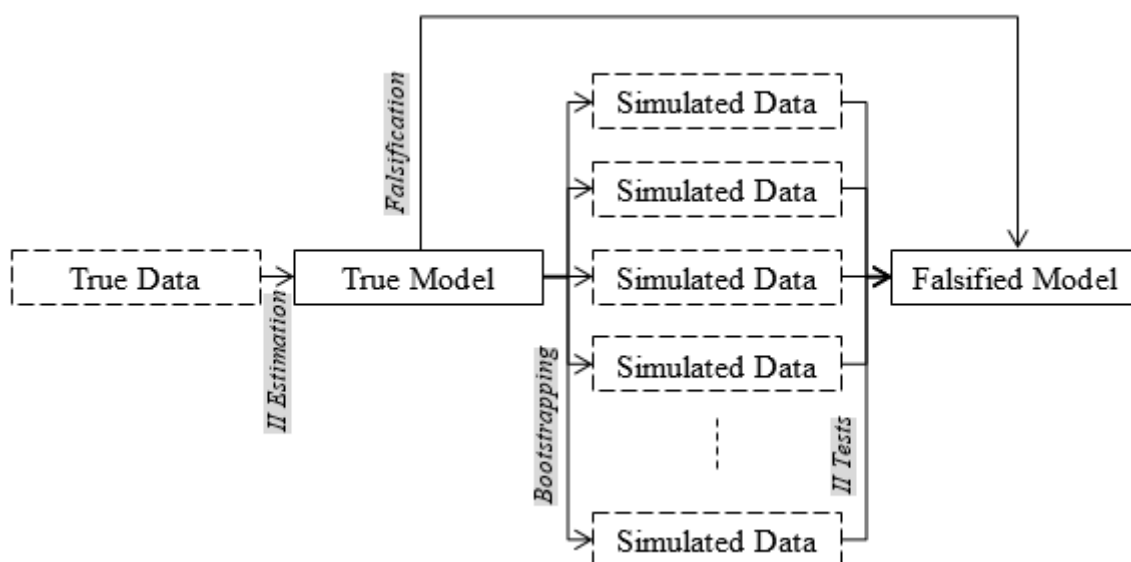*Figure 3 Illustration of the Monte Carlo Experiment*

Table 1 The Simulated Statistical Powers of Indirect Inference Tests

Falseness	Type 1: OLS	Type 2: BOD	Type 3: PSM	Type 4: HSM	Type 5: GOLS
0%	5%	5%	5%	5%	5%
1%	99.9%	7.5%	6.9%	100%	17.2%
5%	100%	36.4%	34.3%	100%	40.8%
10%	100%	61.2%	56.8%	100%	79.4%
20%	100%	77.2%	68.7%	100%	99.3%

Notes: Type 1 (OLS) is the OLS regression model with 35 auxiliary parameters. Type 2 (BOD) is the Blinder-Oaxaca decomposition with 2 auxiliary parameters, i.e. wage differentials due to (i) coefficients and (ii) endowments. Type 3 (PSM) is propensity score matching with 2 auxiliary parameters, i.e. the treatment effects of (i) the treated and (ii) the untreated. Type 4 (HSM) is the Heckman selection model with 35 coefficients from the outcome equation and 34 coefficients from the selection equation. Type 5 (GOLS) categorises the 35 OLS coefficients into 8 groups, including (i) intercept, (ii) demographic, (iii) experience, (iv) education, (v) work mode and work location, (vi) industry (SIC), (vii) occupation (SOC) and (viii) public sector dummy.

4 The Results

Any inference procedure starts with defining the null hypothesis (H0). In the II context, H0 is postulated that "the economic model (1)-(4) is the true data generating process". Under this H0, we have two further possibilities:

- H0a: The model is true and $\mathbf{x}_i$ and $\boldsymbol{\varepsilon}_i$ are uncorrelated (i.e. there is no selection bias).
- H0b: The model is true and $\mathbf{x}_i$ and $\boldsymbol{\varepsilon}_i$ are correlated (i.e. there is selection bias).

As can be seen we can check for the existence of selection bias through our choice of economic model. We can test directly a model without selection bias; this is done straightforwardly by drawing the errors on the assumption that they are random and unconnected with $\mathbf{x}_i$. The model with selection bias is created by exploiting the observed correlation found between the implied errors and the $\mathbf{x}_i$; we can argue that if there is a correlation between the two an estimate of it is provided from our sample by this observed correlation. We can then impose this correlation on our bootstrapping procedure by drawing the errors and the $\mathbf{x}_i$ jointly, so preserving their connection. We may note in passing that this ability of our approach to test directly for selection bias is a helpful feature.

4.1 II test

To initiate the II test and estimation, we need to calibrate the parameters either using the literature conventions or using the data averages consistent with the model. Since there is no microeconomic literature on these structural parameters, we use the macroeconomic literature for this. For example, the utility share of leisure α can be set at 0.5 and the constant elasticity of substitution s can be set at 0.5 to allow for greater complementarity than substitutability between consumption and leisure. The income share of labour in the production function γ is

usually estimated to be 0.6~0.8 in the macroeconomic literature (e.g. Smets and Wouters, 2007), so we set it as 0.7. Finally, the total factor productivity A can be calculated from the firm's marginal condition and the known parameters and average values of the endogenous variables:

$$w = \gamma A L^{\gamma-1} \Rightarrow 12 = 0.7 \times A \times 34^{0.7-1} \Rightarrow A = 49$$

The calibrated structural parameters give the initial values $\boldsymbol{\theta}_0 = [0.5; 0.5; 0.7; 49]$ of the vector $\boldsymbol{\theta} = [\alpha; s; \gamma; A]$. A warning over this calibration strategy is due here. The microdata may exhibit very different parameter values from those implied from the macrodata, because our microdata sample is heavily concentrated in the service sectors. Therefore, this present calibration is only used to initiate the II estimation process. The model will be reestimated by II to ensure that only the best model is tested against the auxiliary model: otherwise our test could reject the model purely on the basis of poor numerical parameter estimates.

The simulations in our II test procedure begin with obtaining the innovations $(\varepsilon_i^S, \varepsilon_i^D)$ from the implied exogenous variables (S_i, D_i) based on the structural equations (1) and (3). They are supposed to be IID across individual observations (similar to the requirement of white noise process in the time-series context), but the two structural innovations can be correlated with each other in a joint distribution.

The extracted innovations from the structural equation are apparently jointly distributed as shown in Figure 4. The estimated standard deviations of the two innovations are respectively $\sigma_S = 0.75$ and $\sigma_D = 0.42$, suggesting much more heterogeneity on the supply side (workers) than on the demand side (jobs). The correlation coefficient between the two innovations is 0.2183, which is significant at the 1% level. The implied variance-covariance matrix is:

$$\Sigma \equiv Var \begin{bmatrix} \boldsymbol{\varepsilon}_S \\ \boldsymbol{\varepsilon}_D \end{bmatrix} = \begin{bmatrix} 0.5588 & 0.0686 \\ 0.0686 & 0.1767 \end{bmatrix}$$

A non-zero correlation means that, during the bootstrapping, the innovations need to be drawn jointly rather than independently, regardless of whether there is selection bias. Similarly, there are significant correlations between conditioning variables, and the bootstrapping can deal with this too by resampling them jointly. For the selection bias model, we resample the innovations and the conditioning variables together jointly, so preserving the correlation between the two.

Figure 4 Joint Frequency Distribution of the Innovations
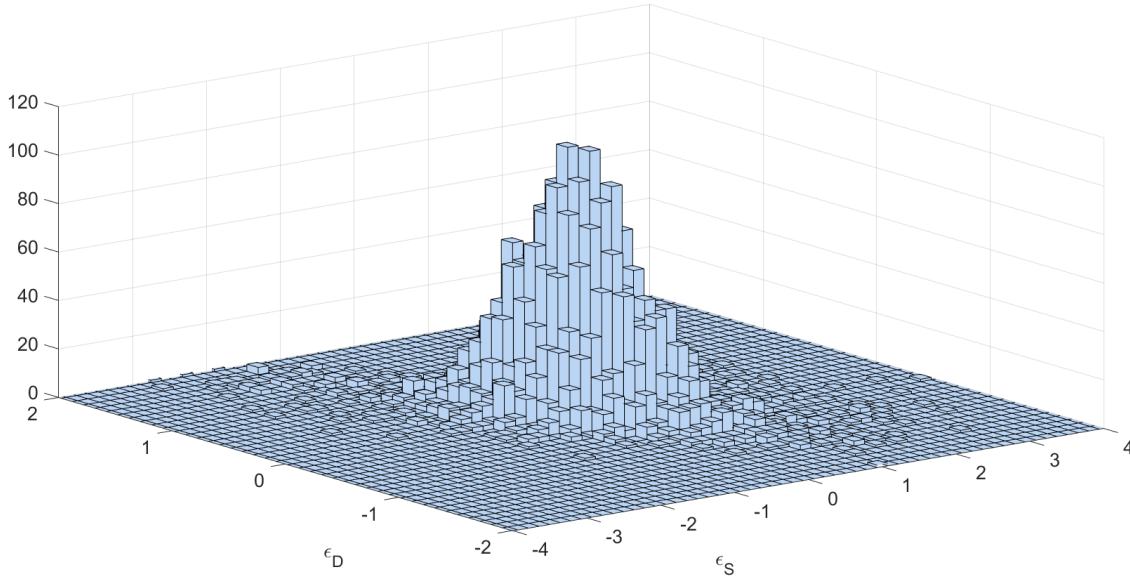
4.2 II Estimation

The economic model is estimated using the genetic algorithm to search globally for the best sets of values such that the Wald statistics under the two hypotheses are respectively minimised. Using the GOLS auxiliary model (Type 5) the estimated structural parameters under both null hypotheses are listed in Table 2.

Table 2 II Estimation Results

Structural Parameters θ		Lower Bound	Upper Bound	Calibration	H0a no bias	H0b bias
α	Leisure weight	0.2	1.2	0.50	**0.4650**	0.4710
s	Elasticity of Substitution	0.1	10	0.50	**6.5479**	1.1712
γ	Labour Share	0.6	0.95	0.70	**0.9366**	0.6036
A	Productivity	24.75	74.24	49.50	**56.76**	27.21
				Wald	**4.69**	84.25
				C-Value	**15.07**	15.69
				P-Value	**79.02%**	0.00%

Notes: H0a (the model is true and there is no selection bias), H0b (the model is true and there is selection bias). The auxiliary model is GOLS (Type 5). The parameters are estimated under the two hypotheses to minimise the Wald statistic or to maximise the potential of passing the test. C-value is the critical value for the Wald statistic to pass the test, i.e. the simulated 95th percentile. P-value gives the exact probability of each hypothesis being true.

It is clear that the selection bias model is totally rejected, with a zero p-value, so in what follows we only discuss the model with no selection bias. The constant elasticity of substitution, s, is very high, indicating that individuals treat consumption and leisure as substitutes more than complements. In a CES utility function, as $s \to 0$ the complementarity is greater while as $s \to$

∞ the substitutability is greater, with $s = 1$ being the Cobb-Douglas specification with equal degrees of complementarity and substitutability. The II estimate of s (6.5 under H0a) is actually at odds with the macroeconomics literature, where s is usually set close to 1. The reason is the inability of this neoclassical model to capture the fluctuations in working hours at the individual level if $s \to 1$. In the macroeconomics literature, there are many other complicated mechanisms besides a low s that help to match the lower aggregate fluctuations (e.g. habit persistence, price rigidity and adjustment costs as introduced by Christiano et al, 2005, into DSGE models). However, the micro-data context is very different, and our model interprets the wide cross-section variations in hours across people as their rather sharp response to differences in the rewards to work. As we see, this passes the test of matching a wide range of data features as summarised by the 8 groups.

The estimated utility weight of leisure (α) is higher than the calibrated value (0.5 under H0a), so it is about half of the utility weight on consumption (fixed at 1). The estimated share of labour in the production function (γ) is very close to 1 (0.94 under H0a), but this is not surprising as our sample is highly concentrated in labour-intensive industries. Finally, the productivity (A) is calculated to match the other parameters in the production function.

4.3   Determining the PSWP

There is as noted above no economic theory of optimising agents implying that there is a PSWP. Instead what we have is an auxiliary model which obtains a positive coefficient on the PSWP dummy variable. That is to say we find that the facts are best represented by a regression in which this coefficient is included. What we therefore need to discover is whether a model *without* a PSWP dummy can account for this auxiliary model. Of course, it is obvious that inserting a PSWP dummy into the structural model will assist it to match the auxiliary model complete with its dummy. However, this of course would cease to be a model: hence the interesting question is whether an optimising model, in which therefore there is no such ad hoc addition, can match the auxiliary model. If it can and so we cannot reject it, we may consider it an adequate structural model and use this model for our analysis. The method of economics is to use theory-based models provided they pass our empirical tests. If we cannot find such a model for our purposes in an investigation, then strictly speaking we have not found an adequate explanation.

So now we ask: can a structural neoclassical model with no public-sector dummy explain the observed data features? Throughout we ignore the model with selection bias since it is heavily rejected. The column headed 'H0 No Bias' of Table 2 above showed the answer. The model matches the auxiliary model with a p-value of 79%.

We now ask a further question: could our structural model do better in matching the auxiliary model if we were to supply it with an ad hoc PSWP dummy and re-estimate the resulting model? We can think of this as asking whether the theory alone is adequate to provide a *full* matching of the auxiliary model description of the data behaviour. If we found that the 'PSWP-augmented' model provided a substantially closer match, we could possibly regard this as evidence of 'missing theory'; as we noted above, it must produce some improvement in the match since we are supplying ad hoc the very dummy we are trying to explain. Notice though that if it did make a material improvement in the match, we would still have trouble justifying this theory, since we have no basis for it in the theory we are using.

Table 3 shows that making this addition to the theory results in virtually no improvement in the model's ability to match the data behaviour. The second column shows the match when the Original parameters of column 1 are kept constant and a PSWP factor added to the demand equation; here the model falls short of the Original model with no PSWP. The third column shows the match when this model is re-estimated by II; here it obtains essentially the same match with a virtually identical p-value. Our theory-based model therefore cannot be improved upon in its matching of the facts even by adding some non-theory designed to help it fit better. This is a rather remarkable result since we would expect that ad hoc non-theory additions which mimic the data behaviour closely would enhance the data behaviour match. In effect we have discovered here that proper theory is both sufficient to explain the data behaviour and its match cannot significantly be improved upon by any further additions.

It follows that the PSWP dummy found in the reduced form regressions is reflecting the effect of the factors driving demand and supply, not a special factor separate from these. When such a special factor is included in the model, the model generates the same reduced form, indicating that it is a redundant factor mimicking the same joint effects of the demand/supply factors. But the identifying restrictions of economic theory rule out adding such a special factor. The fact that a PSWP dummy shows up in the reduced form comes about simply because this picks up an effect of demand/supply factors beyond that directly found in the reduced form; it cannot reflect a structural special factor, as we have seen. The fact that this dummy is needed in the reduced form could arise from approximating the nonlinearity of the structural model.

Table 3 The Indirect Inference Results of a Different Specification

Specification	*Original*		Ad Hoc		Ad Hoc	
Parameters	Estimated $\widehat{\theta}$		Estimated $\widehat{\theta}$		Re-Estimated $\widehat{\widehat{\theta}}$	
Hypothesis	H0a	H0b	H0a	H0b	H0a	H0b
P-Values	79.02%	0.00%	59.50%	0.00%	80.50%	0.10%

To summarise, we find a neoclassical economic model without any ad hoc public-sector dummy can very well explain the wage data summarised by auxiliary models (including the PSWP). That is to say, the estimated 6%-7% wage premium in the public sector is not a mystery. It comes about only because the people and jobs in the public sector require higher wages. The pure economics of the public sector and the workers creates this premium in a competitive labour market. There is no "bias" or "non-economic inequality" or "injustice due to political pressure" going on.

5 Conclusion

A simple neoclassical labour economic model derived from optimisation behaviour is shown to be able to match most data features in UK wage setting, especially in the wage premium summarised by popular types of methods in the microeconometric literature. The model passes our chosen test where the test is set so that a high level of power is delivered. The power function of our test implies that our structural model would have been rejected effectively all the time had it been 20% false in numerical terms. Hence, the statistically true structural model lies within a range of 20% of our estimated structural parameters. As we have seen across the whole of this range there is no role for a PSWP dummy in the structural model, though there would be variation in the PSWP implied in the data behaviour.

What we have discovered therefore is that the empirically estimated PSWP is due to the operation of individual preferences and productivity within the marketplace. Policy intervention to change it would push the market away from this non-distorted equilibrium where welfare is maximised. Of course our paper has limitations. First, our study is limited by the data we have, which is cross-sectional UK data for 2011. Second, we cannot totally rule out the possibility that some other structural model with numerically different parameters could also pass our test; however the virtually total power of our test procedure against mis-specification makes it highly unlikely that any such model would overturn our policy conclusions.

REFERENCES

AFONSO, A. & GOMES, P. 2014. Interactions between private and public sector wages. *Journal of Macroeconomics,* 39**,** 97-112.

ARDAGNA, S. 2007. Fiscal policy in unionized labor markets. *Journal of Economic Dynamics and Control,* 31**,** 1498-1534.

BLACKABY, D. H., MURPHY, P. D., O'LEARY, N. C. & STANEVA, A. V. 2012. An Investigation of the IFS Public-Private Sector Pay Differential: A Robustness Check.

Swansea University Working Paper.

CHATTERJI, M., MUMFORD, K. & SMITH, P. N. 2011. The Public-Private Sector Gender Wage Differential in Britain: Evidence from Matched Employee-Workplace Data. *Applied Economics,* 43**,** 3819-3833.

CHRISTIANO, L., EICHENBAUM, M. & EVANS, C. 2005. Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy,* 113**,** 1-45.

DAI, L., MINFORD, P. & ZHOU, P. 2015. A DSGE Model of China. *Applied Economics,* 47**,** 6438-6460.

DISNEY, R. & GOSLING, A. 1998. Does It Pay to Work in the Public Sector? *Fiscal Studies,* 19**,** 347-374.

FINN, M. G. 1998. Cyclical Effects of Government's Employment and Goods Purchases. *International Economic Review,* 39**,** 635-657.

GOMES, P. 2014. Optimal Public Sector Wages. *Economic Journal*, 125(587), 1425-1451.

GIBSON, J. 2009. Public Sector Pay Premium and Compensating Differentials in the New Zealand Labour Market. *New Zealand Journal of Employment Relations,* 32**,** 56-68.

GOURIEROUX, C., MONFORT, A. & RENAULT, E. 1993. Indirect Inference. *Journal of Applied Econometrics,* 8**,** S85-S118.

GUNDERSON, M. 1979. Earnings Differentials between the Public and Private Sectors. *The Canadian Journal of Economics / Revue canadienne d'Economique,* 12**,** 228-242.

HANSEN, G. D. 1985. Indivisible Labor and the Business Cycle. *Journal of Monetary Economics,* 16**,** 309-327.

HECKMAN, J. J. 1979. Sample Selection Bias as a Specification Error. *Econometrica,* 47**,** 153-161.

JONES, M., MAKEPEACE, G. & WASS, V. 2018. The UK gender pay gap 1997-2015: What is the role of the public sector? *Industrial Relations*, 57(2), 296-319.

JUHN, C., MURPHY, K. M. & PIERCE, B. 1993. Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy,* 101**,** 410-442.

KYDLAND, F. & PRESCOTT, E. 1982. Time to Build and Aggregate Fluctuations.

Econometrica, 50, 1345-1370.

LE, V. P. M., MEENAGH, D., MINFORD, P. & WICKENS, M. 2011. How much nominal rigidity is there in the US economy? Testing a new Keynesian DSGE model using indirect inference. *Journal of Economic Dynamics and Control,* 35, 2078-2104.

LE, V. P. M., MEENAGH, D., MINFORD, P., WICKENS, M. & XU, Y. 2016. Testing Macro Models by Indirect Inference: A Survey for Users. *Open Economies Review,* 27, 1-38.

LE, V. P. M., MINFORD, P. & WICKENS, M., 2010. The 'Puzzles' methodology: En route to Indirect Inference? *Economic Modelling,* 27, 1417-1428.

MEENAGH, D., MINFORD, P., WICKENS, M., and XU, Y., 2019 'Testing DSGE models by Indirect Inference: a survey of recent findings', Open Economies Review, Open Economies Review, July 2019, Volume 30, Issue 3, pp 593–620

LEE, L.-F. 1978. Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables. *International Economic Review,* 19, 415-433.

MEENAGH, D., MINFORD, P. & WICKENS, M. 2009. Testing a DSGE Model of the EU Using Indirect Inference. *Open Economies Review,* 20, 435-471.

MELLY, B. 2005. Decomposition of Differences in Distribution Using Quantile Regression. *Labour Economics,* 12, 577-590.

MINFORD, P., XU, Y. & ZHOU, P. 2015. How Good are Out of Sample Forecasting Tests on DSGE Models? *Italian Economic Journal,* 1, 333-351.

MORIKAWA, M. 2016. A comparison of the wage structure between the public and private sectors in Japan. *Journal of the Japanese and International Economies,* 39, 73-90.

RAMONI-PERAZZI, J. & BELLANTE, D. 2006. Wage Differentials between the Public and the Private Sector: How Comparable are the Workers? *Journal of Business and Economics Research,* 4, 43-57.

ROBINSON, C. & TOMES, N. 1984. Union Wage Differentials in the Public and Private Sectors: A Simultaneous Equations Specification. *Journal of Labor Economics,* 2, 106-127.

ROTEMBERG, J. & WOODFORD, M. 1997. An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy. National Bureau of Economic Research,

Inc.

SMETS, F. & WOUTERS, R. 2007. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *The American Economic Review,* 97**,** 586-606.

SMITH, A. Indirect Inference. *The New Palgrave Dictionary of Economics.* Palgrave.

SMITH, S. P. 1976. Pay Differentials between Federal Government and Private Sector Workers. *Industrial and Labor Relations Review,* 29**,** 179-197.

URIBE, M. & YUE, V. 2006. Country Spreads and Emerging Markets: Who Drives Whom? *Journal of International Economics,* 69**,** 6-36.

WANG, Y. & ZHOU, P. 2019. The Public Wage Premium Puzzle. *International Journal of Computational Economics and Econometrics*, 9(4): 287-307.

WANG, Y. & ZHOU, P. 2017. Are We Better Off Working in the Public Sector? In: Tsounis, N. & Vlachvei, A. (eds.) *Advances in Applied Economic Research.* Berlin: Springer International Publishing, 379-409. ISBN: 978-3-319-48453-2.

APPENDIX 1

Note that in general there is no analytical solution to this nonlinear equation system, but there are two methods to deal with this problem.

First, note that in a special case $s = 1$ which actually implies a Cobb-Douglas utility function, the reduced form of this equation system can be solved analytically:

$$
\begin{cases} w_i = \alpha \left(\dfrac{w_i L_i}{T - L_i} \right) S_i \\ w_i = \gamma A L_i^{\gamma-1} D_i \end{cases}
\Rightarrow
\begin{cases} L_i = \dfrac{1}{1 + \alpha S_i} T \\ w_i = \gamma A \left(\dfrac{1}{1 + \alpha S_i} T \right)^{\gamma-1} D_i \end{cases}
$$

One remarkable feature of the reduced form is that the equilibrium working hour L_i does not depend on the total factor productivity A (but varies due to the different individual characteristics S_i), which is a typical feature in neoclassical models. It is because a change in productivity will lead to both substitution effect and income effect, which offset each other perfectly. The original production function (blue dash) shifts out to the higher level (bold blue dash) due to a higher productivity, and we can construct a hypothetical production function (black dotted) with the new productivity level but tangent to the original utility level.
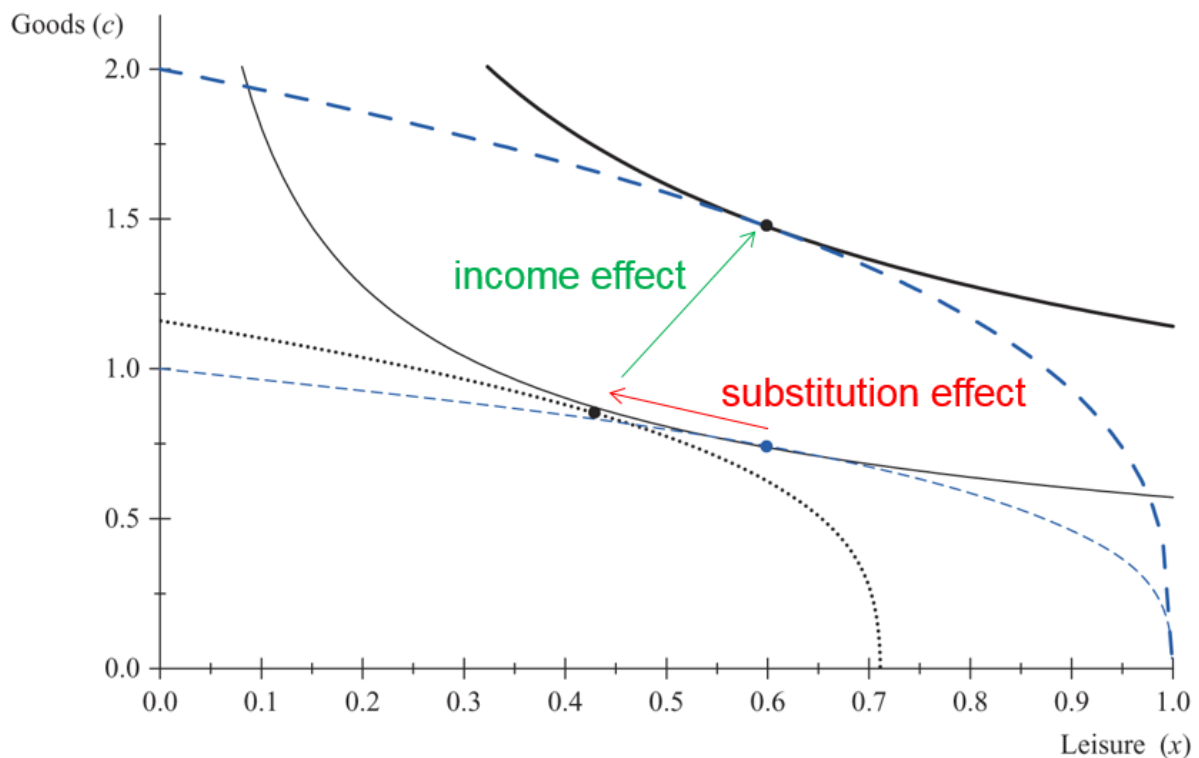

Figure 5 The Perfect Offset between Income Effect and Substitution Effect (s = 1)

In general when $s \neq 1$, however, the nonlinear equation system (1) and (3), or equivalently the consolidated equation (5), does not have analytical solution.

$$\alpha S_i \left[w_i \left(\frac{w_i}{\gamma AD_i} \right)^{\frac{1}{\gamma-1}} \right]^{\frac{1}{s}} - w_i \left[T - \left(\frac{w_i}{\gamma AD_i} \right)^{\frac{1}{\gamma-1}} \right]^{\frac{1}{s}} = 0 \qquad \ldots(5)$$

One possibility is to use a numerical method (e.g. Newton-Raphson algorithm) to solve for w_i and L_i. Nevertheless, despite that the numerical method is not very difficult to solve the nonlinear equation system once, it will induce an extremely heavy computation burden due to the simulation of the II procedures. To see this, consider a particular simulation in the II test procedure, there will be about 7,000 observations to be solved (each observation i implies a nonlinear equation system). For a typical II test, we usually run 1,000 simulations, so there will be 7,000,000 nonlinear equation systems to be solved for one test. Even if it only takes 1 second for each solution, it will take about 81 days to finish one test. Let alone the II estimation, which involves at least several thousands of II tests.

Alternatively, we can linearise the equation system around some point and then solve the linear equation analytically. A straightforward choice for the expansion point is the average wage of the whole sample, on the basis that the individual equilibrium should not be too far away from the population equilibrium.
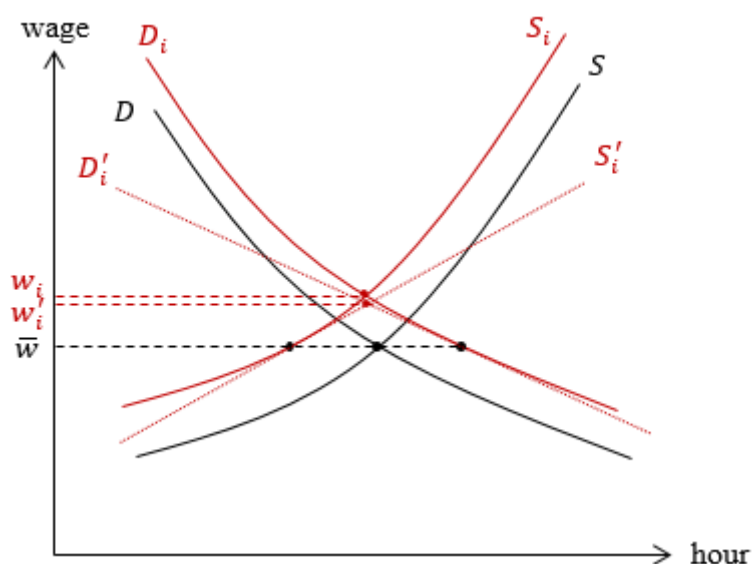

Figure 6 Linear Approximation of the Equilibrium

Figure 6 illustrates the linear approximation of the solution of the nonlinear equation system. The aggregate/average labour demand curve (D) and labour supply (S) intersect at the market equilibrium wage ($\overline{w}$), which is observable in the data. For each specific individual/job, due to

25

shifting factors captured by $\mathbf{ind}_i$ and $\mathbf{job}_i$, the specific equilibrium wage (w_i) will be different. To solve this specific wage, we expand the supply curve and demand curve at $\bar{w}$, ending up with the linearised supply "curve" (S_i') and demand "curve" (D_i'). The approximate solution w_i' is very easy to obtain because the nonlinear equation system is now a linear equation system. The closer are w_i and $\bar{w}$, the closer are the approximate solution w_i' and the true solution w_i. This linearisation method is a special case of local approximation, which is widely used in the macroeconomic DSGE literature. Its counterpart in the dynamic stochastic model setting is called perturbation method, see for example Uhlig (1998) for more details.

Arguably, the local approximation method is more general because not all economic models have unique analytical solutions, and the restriction of parameter values may not be reasonable. In contrast, for any model, the average wage (or any other endogenous variables) always exists, so linear approximation always works. Its disadvantage is also clear, because the approximate solution may lie very far away from the true solution due to the high degree of nonlinearity.

APPENDIX 2

	Grouped ϑ	OLS Regressors
	ϑ_1: Intercept	intercept
Individual Characteristics	ϑ_2: Demographic	male white married homosexual age age^2 migrant
	ϑ_3: Experience	work experience work experience^2
	ϑ_4: Education	low education GCSE A-level higher education degree
Job Attributes	ϑ_5: Temporospatical	full time London
	ϑ_6: Industry	energy & water manufacturing construction distribution transport banking public admin other services
	ϑ_7: Occupation	professional technical administrative skilled trades personal service customer service processing elementary
	ϑ_8: PSWP	public sector dummy

Table 4 The Grouped Auxiliary Parameters of Type 5 Auxiliary Model