Cardiff Economics Working Papers

Working Paper No. E2017/13

Resolving the Public Sector Wage Premium Puzzle by Indirect Inference

Patrick Minford, Yi Wang and Peng Zhou

October 2017

ISSN 1749-6010

Cardiff Business School
Cardiff University
Colum Drive
Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
business.cardiff.ac.uk

Resolving the Public-Sector Wage Premium Puzzle by Indirect Inference*

Patrick Minford[†], Yi Wang[‡] and Peng Zhou[§]

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

---

* We are grateful for Prof Casper G. de Vries's critical comments and constructive suggestions. All remaining mistakes are our own responsibility.

† E26, Aberconway Building, Cardiff Business School, Cardiff University, UK. Email: minfordp@cardiff.ac.uk

‡ Cardiff School of Management, Cardiff Metropolitan University, UK. Email: ywang@cardiffmet.ac.uk

§ Corresponding Author: D47, Aberconway Building, Cardiff Business School, Cardiff University, UK. Email: zhoup1@cardiff.ac.uk

Resolving the Public-Sector Wage Premium Puzzle by Indirect Inference

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; Chatterji et al, 2010). 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 (Afonso and Gomes, 2014; Morikawa, 2016). Though most studies agree that the public-sector wage premium[1] (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 has never come to a consensus on how to estimate the wage premium, nor on 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 belong to the paradigm of econometric or microeconometric models. Very few attempts have been made to confront the microdata with the economic models *per se*.

The main reason for this preference for empirical econometric models over theoretical economic models is convenience. It is very easy and straightforward to build an econometric model such as a linear regression without much technical cost nowadays. Econometric models mainly follow a philosophy of "let the data speak", given the weak links between these econometric models and economic theories. A common practice is for researchers to start with some economic theory (and sometimes a formal economic model involving optimisation behaviour) and derive some qualitative relationships, which are then loosely translated into testable hypotheses. Subsequently, instead of the economic model *per se*, the econometric model (usually a reduced-form regression model) embedding these testable hypotheses is then estimated and tested with the data. There are three gaps between these econometric models and the economic models. First, a typical econometric model only uses a subset of the original economic model, because it only tests or estimates one or several implications of it, not all of them. Second, the linearity (or log linearity) of the regression model greatly reduces the accuracy of the predictions of a highly nonlinear economic model. With these deficiencies, there is a considerable risk that what is tested or estimated by an econometric model is not what the corresponding economic model actually implies. Third, and perhaps most worrying, is the problem of identification: the

[1] 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.).

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). Thus, please "mind the gap" between the econometric and the economic models.

In this paper, to answer the research question "whether there is PSWP" without creating such a gap, we propose an alternative approach to econometric modelling. We ask this question in an equivalent but inverse way, "whether a model with perfect competition and no market distortions can explain the data". If it can, then we can conclude that the labour market is by and large efficient, and no government intervention is needed. Therefore, the PSWP estimated in econometric models is only a statistical phenomenon. Note that there may be other models in the labour economics literature (e.g. imperfect competition or search) that could possibly match the data better and produce a different conclusion on the role of the PSWP. But this remains to be demonstrated and we leave it as an issue for the future; here we assume the standard model of labour market supply and demand that is usually thought to underlie the reduced forms that are estimated in the econometric literature.

The theoretical modelling methods of microeconomics and macroeconomics have converged in recent years, but this convergence has not been synchronised in the empirical realm; the mainstream methods adopted by empirical microeconomic research have been regressions or its variants. In contrast, the methods and techniques of empirical macroeconomic research have improved remarkably in the last decade, allowing for a tighter connection between theory and empirical evidence. A complicated microfounded[2] economic model with high nonlinearity can be solved, tested and estimated without introducing any discrepancy between theoretical and empirical models. Indirect Inference (II) is one of these powerful techniques. Instead of the distribution of the data (as in maximum likelihood and Bayesian inference), it uses the features of the data summarised by an "auxiliary model" (e.g. moments, regression coefficients, impulse responses, etc.) to measure the distance between the data and the model. The model can be regarded as the "true" data generating process if the distance is not too big, and the parameters can be estimated by minimising this distance.

The purpose of this paper is to use these techniques of microfounded modelling and testing to examine the PSWP puzzle: specifically, we will use Indirect Inference. We hope thereby to contribute both to the solution of this empirical puzzle and to the methodology of microeconomic modelling.

[2] "Microfoundation" is a key feature of modern macroeconomic models. It means the model is consistent with the optimisation behaviour of individual consumers and firms with rational expectations. It removes the ad hoc gaps between microeconomics and macroeconomics in theoretical modelling methodology.

1 Econometric Model as Auxiliary Models

In the traditional empirical literature, econometric models are usually adopted to estimate the PSWP. Though based on *ad hoc* linear or log-linear specifications, they are useful tools for summarising the data features. We summarise four main types of econometric models in the PSWP literature. To test/estimate the economic model in the coming section by II procedures, we will consider employing these econometric methods as the "auxiliary models" to summarise the features of both observed and model-simulated data.

- **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 subsamples, 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.

Based on the Type 1 model, we develop the Type 5 model for II testing and estimation.

- **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 microfounded economic model to pass the test of matching this behaviour. At one extreme, only the "real world" 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 justify the use of Type 5 auxiliary model in terms of its merit in statistical power.

2 Indirect Inference

2.1 The Power of the II Test: A Monte Carlo Experiment

The validity and reliability of the estimation/test results using II depend on its statistical power (the probability of rejecting a false model). we will adopt Le et al's (2016) approach to investigating the power of the II test in this context of microeconomic models. We can simulate data from a microfounded 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. 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.

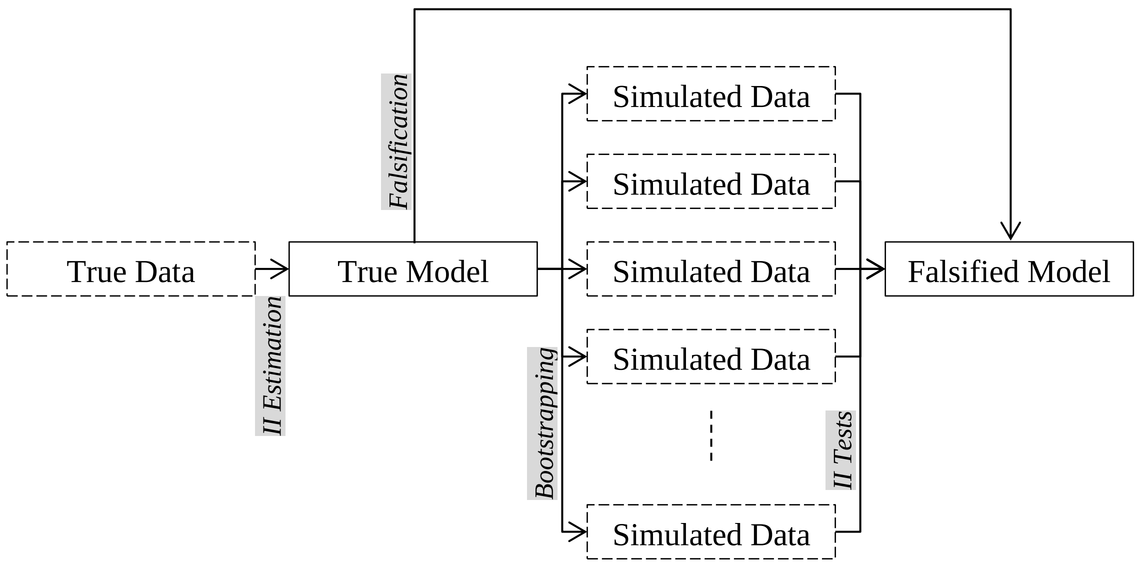
The Monte Carlo experiment is designed as follows. The "true" microfounded economic model $\mathcal{M}(\widehat{\boldsymbol{\theta}})$ is constructed with $\widehat{\boldsymbol{\theta}}$ being the "true" structural parameters[3], based on which we can simulate 1000 datasets. To see the statistical power 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—the II test correctly distinguishes the false model and contributes to a higher power. Conversely, if the resulting p-value is greater than 5%, then we accept the model being true—the II test fails to spot the false model and lowers the power. The proportion of the 1000 tests that reject the model being true is therefore the statistical power of II test. Figure 1 summarises the following:

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

- **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 S simulations, we can obtain S 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 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.

Figure 1 Illustration of the Monte Carlo Experiment



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. It has the 'goldilocks' feature that its power is neither too strong to allow for analysing the key features of the microfounded model, nor too weak to reduce type-II error.

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.

2.2 Indirect Inference Test

We now detail the II test procedure, which is the basis of the Monte Carlo experiment above and the subsequent II estimation. The structural form of a microfounded 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 $f\left(\mathbf{y}^{(a)},\mathbf{x}^{(a)},\boldsymbol{\varepsilon}^{(a)},\boldsymbol{\theta}_0\right)=0$ if the model is identified. To achieve identification, 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 $\mathbf{z}^{(s)}$. Using these S sets of exogenous variables, we simulate S sets of endogenous variables $\mathbf{y}^{(s)}$ by substituting the bootstrapped exogenous variables and calibrated parameters into the reduced form:

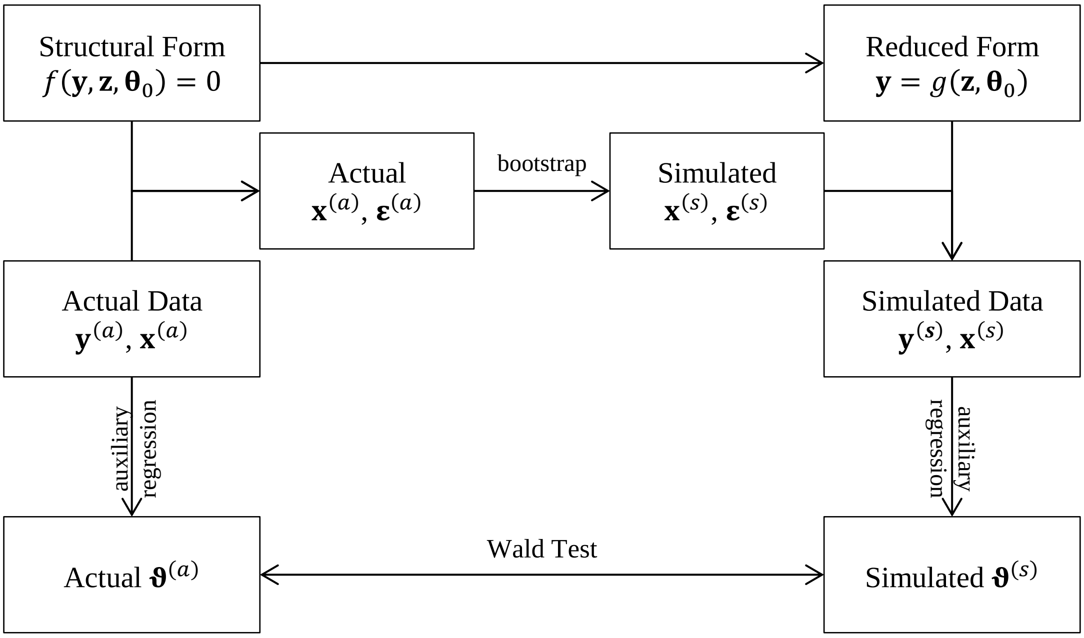
$$\mathbf{y}^{(s)} = g\left(\mathbf{z}^{(s)}\left(\mathbf{x}^{(s)}, \boldsymbol{\varepsilon}^{(s)}\right), \boldsymbol{\theta}_0\right).$$

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 S sets of $\boldsymbol{\vartheta}^{(s)}$ based on the simulated data $\mathbf{x}^{(s)}$. A standard Wald test can be implemented by computing the Wald statistic: $\text{Wald}(\boldsymbol{\theta}_0) \equiv \left(\boldsymbol{\vartheta}^{(a)} - \overline{\boldsymbol{\vartheta}}^{(s)}\right)' \left(Var\left[\boldsymbol{\vartheta}^{(s)}\right]\right)^{-1} \left(\boldsymbol{\vartheta}^{(a)} - \overline{\boldsymbol{\vartheta}}^{(s)}\right)$.

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.

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, for a recent survey) 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 have seen above in our 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. IIcan give us a Goldilocks power that is "just right". Hence here we chose the GOLS auxiliary model as discussed above.

Figure 2 Flow Chart of Indirect Inference



2.3 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 $\widehat{\boldsymbol{\theta}}$, 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 $\widehat{\boldsymbol{\theta}}$ is therefore the II estimation of the model parameters:

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

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 $\widehat{\boldsymbol{\theta}}$. 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.[4] We will use this more robust algorithm to undertake the II estimation.

3 The Microfounded 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).

3.1 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

[4] 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).

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:

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

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

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

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

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

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 $\boldsymbol{\eta}_\mathbf{S}$ 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 $\mathbf{ind}_i$.

3.2 The Demand Side

A representative firm faces the following standard optimisation problem:

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

$$\text{Technology Constraint: } Y = AL^\gamma .$$

A is to capture the average total factor productivity level in the production function. 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:

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

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:

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

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:

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

Here, $\mathbf{job}_i$ is a vector of job attributes, such as industry, sector, occupation, work mode and location, and $\boldsymbol{\eta}_D$ 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:

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

3.3 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[5].

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

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

[5] 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.

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.

4 The Results

Any inference procedure starts with defining the null hypothesis (H0). In the II test/estimation 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.

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, the macroeconomic literature is used for the calibration purpose. 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 practice is only to initiate and illustrate the II test. 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.

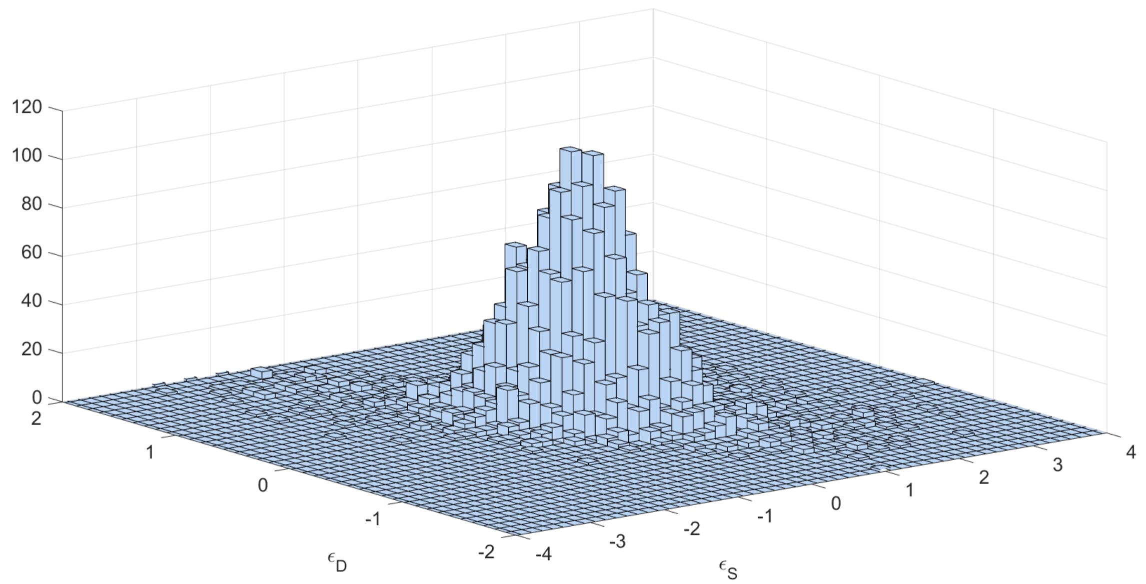
The simulations in our II test procedure begin with obtaining the innovations $(\varepsilon_i^S, \varepsilon_i^D)$ from the implied exogenous variables (S_i, D_i) based on the structural equations (1) and (3). They are supposed to be IID across individual observations (similar to the requirement of white 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 3. 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 3 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 $s \to \infty$ 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. One reason is the inability of the neoclassical model to capture the fluctuations in working hours if $s \to 1$. In the macroeconomics literature, there are many other complicated mechanisms (e.g. habit persistence, price rigidity and adjustment costs as introduced by Christiano et al (2005) into DSGE models) to make up for this drawback, but in our simple microeconomic model, the only way to improve the model's ability to generate fluctuations in working hours is to drive s away from 1. The higher the substitutability, the more widely spread the working hours will be. This is indeed one of the limitations of the neoclassical model due to its simplicity. However, even under this simple model, it can still pass 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 micro-founded theory 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 micro-founded 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 micro-founded model: hence the interesting question is whether a truly micro-founded 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 micro-founded 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 closer match, we could possibly regard this as evidence of 'missing theory'; notice though that we would still have trouble justifying this theory, since we have no basis for it in the theory we are using.

What Table 3 shows is 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.

Table 3 The Indirect Inference Results of a Different Specification

Specification	*Original*		Ad Hoc		Ad Hoc	
Parameters	Estimated $\hat{\boldsymbol{\theta}}$		Estimated $\hat{\boldsymbol{\theta}}$		Re-Estimated $\hat{\hat{\boldsymbol{\theta}}}$	
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 we can be sure that the 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, 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.

Methodologically, this paper attempts to bridge the gap between microeconomic and macroeconomic research approaches. As reviewed in the introduction, while a methodological convergence between the two sub-disciplines began in the 1980s, most efforts were then invested in building a microfoundation for macrodata analysis. Here however we are trying to provide a microfoundation for microdata analysis- a matter largely ignored in the empirical micro literature. We believe this approach can provide us with a closer link between microeconomic theory and microdata.

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. 2010. 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.

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.

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.

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.

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.

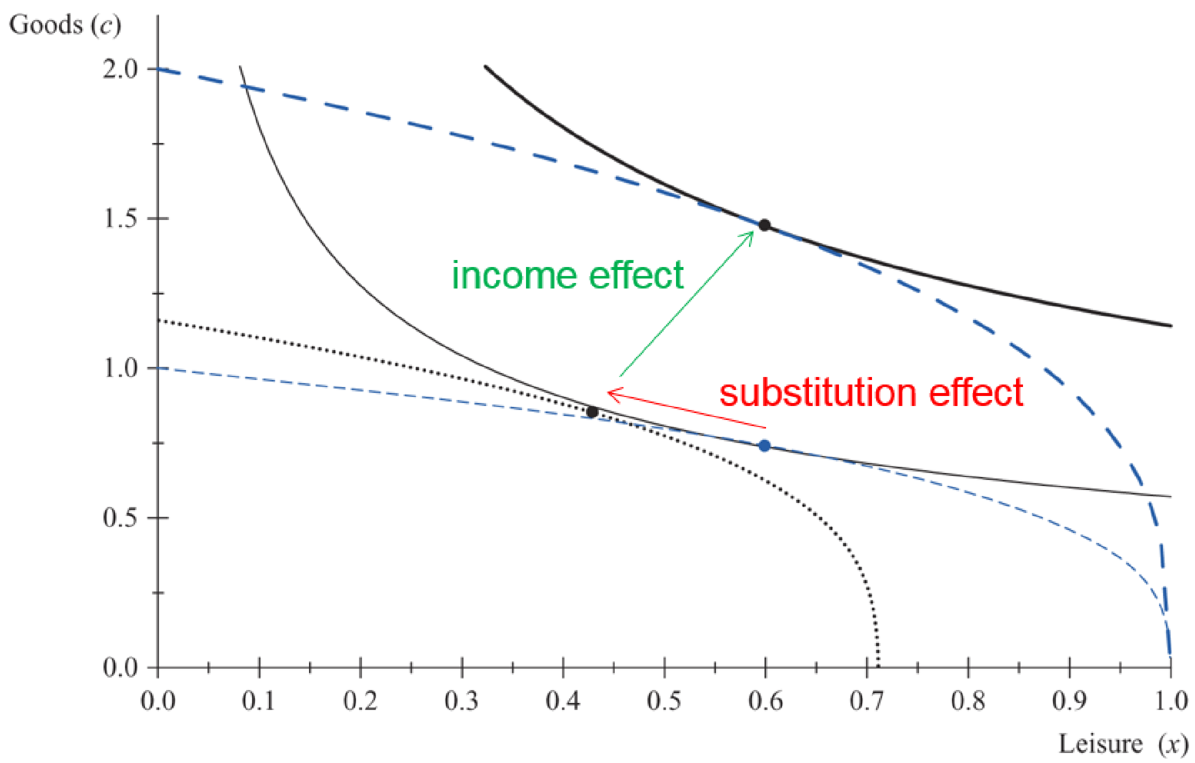


Figure 4 The Perfect Offset between Income Effect and Substitution Effect ($s = 1$)

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.

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

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 non-linear 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 non-linear 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.

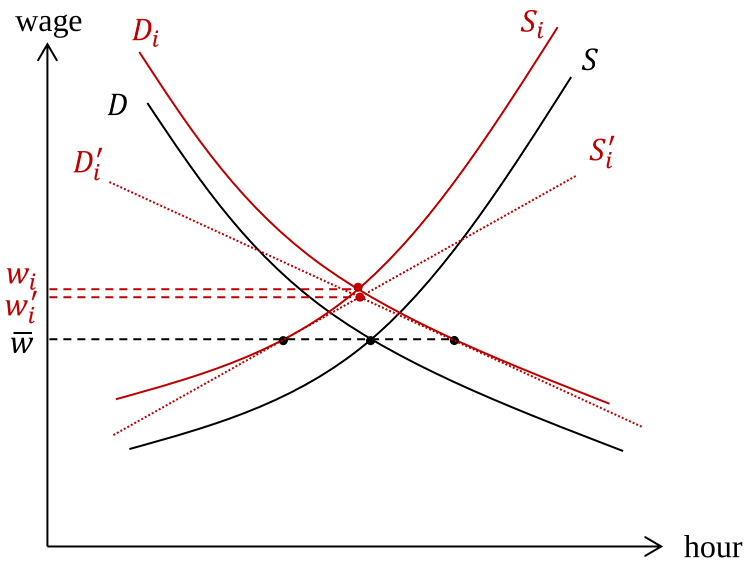


Figure 5 Linear Approximation of the Equilibrium

Figure 5 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 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.

To summarise, there are two methods to solve the nonlinear equation system:

A. parameter restriction to make it analytically solvable;
B. local approximation of nonlinear equation system to linear equation system.

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. Therefore, we will focus on method (B) in this paper, while method (A) is equivalent to method (B) if the estimated s is equal to 1.

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