

Minford, Patrick; Wang, Yi; Zhou, Peng

**Working Paper**

## Resolving the public sector wage premium puzzle by indirect inference

Cardiff Economics Working Papers, No. E2017/13

**Provided in Cooperation with:**

Cardiff Business School, Cardiff University

*Suggested Citation:* Minford, Patrick; Wang, Yi; Zhou, Peng (2017) : Resolving the public sector wage premium puzzle by indirect inference, Cardiff Economics Working Papers, No. E2017/13, Cardiff University, Cardiff Business School, Cardiff

This Version is available at:

<https://hdl.handle.net/10419/174127>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# 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](http://business.cardiff.ac.uk)

This working paper is produced for discussion purpose only. These working papers are expected to be published in due course, in revised form, and should not be quoted or cited without the author's written permission.

Cardiff Economics Working Papers are available online from:

<http://econpapers.repec.org/paper/cdfwpaper/> and

[business.cardiff.ac.uk/research/academic-sections/economics/working-papers](http://business.cardiff.ac.uk/research/academic-sections/economics/working-papers)

Enquiries: [EconWP@cardiff.ac.uk](mailto:EconWP@cardiff.ac.uk)

# Resolving the Public Sector Wage Premium Puzzle by Indirect Inference

Patrick Minford<sup>1</sup>, Yi Wang<sup>2</sup> and Peng Zhou<sup>3</sup>

## Abstract

This paper investigates the public sector wage premium in the UK using a microfounded economic model and indirect inference. The neoclassical wage determination model is tested and estimated without introducing any gap between the theoretical and empirical models. To test if the model is true, four types of econometric methods are used to summarise the data features, based on which we can evaluate the distance between the observed data and the model-simulated data in the test. When the distance is minimised, we estimated a public sector wage premium between 6% and 7% using both traditional microeconometrics and indirect inference. 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. Finally, in a simulation based on the estimated model, we show that it is not the non-market factors, but the total costs and benefits of working in different sectors and the pure market force, that create the public sector wage premium. There is no inefficiency or unfairness in the labour market to justify government intervention.

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

**JEL Classification:** C21, C35, J31, J45

---

<sup>1</sup> E26, Aberconway Building, Cardiff Business School, Cardiff University, UK.

<sup>2</sup> Corresponding Author: Yi Wang, Cardiff School of Management, Cardiff Metropolitan University, UK.

Email: [ywang@cardiffmet.ac.uk](mailto:ywang@cardiffmet.ac.uk)

<sup>3</sup> D47, Aberconway Building, Cardiff Business School, Cardiff University, 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 recent 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<sup>4</sup> (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/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 models** or microeconomic 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 relationships, which are then loosely translated into testable hypotheses. Subsequently, instead of the economic model *per se*, the econometric model (usually a 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 that inspire them. 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

---

<sup>4</sup> 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).

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 been improved remarkably in the latest decade, allowing for a tighter connection between theory and empirical evidence. A complicated microfounded<sup>5</sup> 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 observed data and model-simulated data. 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. It is similar to generalised method of moments (GMM) and simulated method of moments (SMM) in the sense that they only use a set of summary information—some features of the distribution (i.e. moments) rather than the entire data distribution.

The purpose of this paper is therefore both empirical and methodological. Empirically, it aims to provide a robust estimate of the PSWP in the UK (positive analysis) and to inform what we should do about it (normative analysis). Methodologically, it critically reviews existing econometric modelling methods and techniques, and sets out the new II method for estimating the wage premium and testing for selection bias; in the process we hope to bring micro- and macroeconomic research closer together.

## 1 Econometric VS. Economic Modelling Method

In the microeconomic literature, there are four main types of method for estimating the PSWP. Though based on *ad hoc* specifications, they are useful tools for summarising the data features. To test/estimate the microfounded economic model in the coming section by II, we will employ these four types of econometric methods as the auxiliary models to measure the distance between the observed data features and the model-simulated data features.

- **Type 1:** Single-Equation-Regression Method. This directly estimates in a wage determination equation the coefficient of the dummy variable describing whether or not an

---

<sup>5</sup> “Microfoundation” is a key feature of modern macroeconomic models. It means the model is consistent with optimisation behaviour of individual consumer and firm with rational expectations. It removes the *ad hoc* gaps between microeconomics and macroeconomics in theoretical modelling methodology.

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 endogeneity and outliers.

- **Type 2: Decomposition-Based Method.** 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 Blinder-Oaxaca decomposition 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 Method.** 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 Method.** The fourth type includes the approach developed by Lee (1978) and Heckman (1979), 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.

To understand the evolution of the literature of PSWP, it is crucial to understand the role of the selection bias problem in the empirical literature. The bias due to selection and that due to endogeneity are easily confused. But the biases are very distinct and both have different solutions. **Selection bias** arises where the dependent variable is observed only for a non-random sample, e.g. an individual only earns a wage within the public sector if she is working in the public sector, and vice versa. In contrast, **endogeneity bias** means that an independent variable included in the model is an endogenous variable, correlated with unobservable factors in the error term. In the context of PSWP, those who choose to work in the public sector may have some special characteristics such as stronger risk aversion, so they would have earned less if they were working in the private sector where taking risks is more necessary. By nature, the selection bias is a *data problem* (non-random sample), while the endogeneity bias is a *model problem* (omitted variable).

If this data problem is not properly recognised and corrected the parameter estimates are biased and conclusions drawn are misleading. In other words, the estimates will only reflect the behaviour in a particular sample, and the conclusions are not generalisable to the whole population. If the selection bias problem is treated as a *data problem*, then a straightforward solution is to change the way of interpreting the estimates. It must be made clear that the conclusions drawn from the regressions are only valid within the specific range represented by the sample. For example, the estimated return on education is only for those who are working, rather than

for the whole labour force. In contrast, if the selection bias problem is treated as a *model problem*, one obvious remedy is to construct some proxy for the missing selection effect or to use instrument variables in estimation. To fundamentally resolve the problem, new modelling strategies are needed to explicitly deal with the selection bias within the model system. Heckman (1979) proposes an influential two-equation model to explicitly address selection bias. Alternative methods include treatment effects models and simultaneous equation models. One of the contributions of this paper is to develop an indirect inference test procedure capable of dealing with the selection bias problem based on a microfounded economic model.

In contrast to the typical econometric modelling strategy, the economic modelling method derives the model by strictly following representative individual agent optimisation behaviour. The resulting equation (the reduced form) explaining the endogenous variables (wage and working hours) are usually nonlinear because the components of the optimisation problems, such as the utility function and production function, are nonlinear. Microfounded models have been used to address the PSWP observed in the macroeconomic data (Finn, 1998; Ardagna, 2007; Afonso and Gomes, 2014). There is a discrepancy between the model and the data in the current macroeconomic literature on PSWP—the economic model is microfounded, but the data is aggregated. There is a great information loss due to the aggregation/averaging from the individual-level microdata to the aggregate-level macrodata, so the analysis based on the macrodata is less efficient and empirically subject to higher measurement error. In the present paper, because the main purpose is methodological, we will employ the simplest neoclassical labour economic model (i.e. ignoring union wage setting power and search frictions) to introduce indirect inference techniques to a microdata analysis.

As for the empirical strategy, there are two general ways of estimating a microfounded economic model: (i) using a data-distribution based or full information estimator, such as maximum likelihood and Bayesian, and (ii) using a data-features based estimator, such as GMM, SMM and II. The distribution based estimator is more efficient because it utilises all the distributional information of the endogenous variables in estimation, but is subject to higher possibility of mis-specification in the assumptions. The second group of estimators usually focus on the moment properties of the distribution, which are less sensitive to mis-specification.<sup>6</sup>

Within the data features estimators, GMM and SMM use the moment properties of the actual data as the criterion to estimate the structural parameters. The objective function is the weighted sum of the gap between the theoretical moments (as in GMM) or simulated moments (as in SMM) implied by the model and the observed data moments. The moments usually include means, standard deviations and correlation coefficients. They can be regarded as special cases

---

<sup>6</sup> A detailed discussion in comparing these two estimators can be found in Meenagh et al (2009), Le et al (2011) and Dai et al (2015) in the context of macroeconomic DSGE models.

of II, in which the auxiliary model can be any data properties and features, not only the moment properties. For example in Le et al (2010), the auxiliary model is a VAR(1) to summarise the joint probability of the data features of all the observables. Alternatively, impulse response functions are also used as auxiliary functions to focus on the dynamic feature of the data (Rotemberg and Woodford, 1997; Christiano et al, 2005; Uribe and Yue, 2006). II generalises the criterion to any data features one can abstract from the data, including simple moments (to capture the volatilities), impulse response functions (to capture the dynamics) and VAR (to capture both). In our case, we have four choices for the auxiliary function, i.e. these four types of econometric modelling methods. Therefore, we can systematically integrate the econometric and the economic modelling methods.

### 1.1 Indirect Inference Test

We now introduce the II test procedure, which forms the basis for the II estimation. Suppose the structural form of a model is a system of equations consisting of some endogenous variables ( $\mathbf{y}$ ) to be explained and exogenous variables ( $\mathbf{z}$ ) to explain  $\mathbf{y}$ , linked by parameters ( $\boldsymbol{\theta}$ ). Note that the exogenous variable vector ( $\mathbf{z}$ ) can include both conditioning variables ( $\mathbf{x}$ ) and the structural innovations ( $\boldsymbol{\varepsilon}$ ):

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

A clarification of terminology is due here. In different strands of literature terminology varies, but in economic models “**error terms**” usually refer to the exogenous variables, which are often further expressed as a *deterministic component* (a function of “conditioning variables” or “state variables”—such as other exogenous variables and predetermined variables) plus a *stochastic component* (the innovations). In many articles, innovations are also called “**shocks**” (e.g. productivity shock in RBC, markup shocks in DSGE), and error terms are sometimes treated as endogenous, because they are not mathematically different from other endogenous variables in the structural equations—depending on other variables and the shocks. In terms of this broad definition, the number of model equations equal to the number of endogenous variables. Here, the model equations include both the **structural equations** (describing the optimisation/equilibrium conditions of the endogenous control variables) and the **error structure equations** (describing how the error terms are constructed from the conditioning variables and the innovations/shocks). Equivalently, if we still treat the error terms as exogenous (as in this paper), then the number of structural equations should be equal to the number endogenous variables (narrowly and naturally defined). In contrast, in econometric terminology, “error terms” refer to the regression model’s disturbance term, which may (or preferably may not) be correlated to the regressors. Note that the structural equations (or the structural form) of an economic model are different from the structural-form econometric models—the former are derived from optimisation problems, while the latter are not.



Assume the model can be solved in a reduced form:

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

Given some calibrated parameter values  $\boldsymbol{\theta}_0$ , the observable 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  $f(\mathbf{y}^{(a)}, \mathbf{x}^{(a)}, \boldsymbol{\varepsilon}^{(a)}, \boldsymbol{\theta}_0) = 0$  if the model is identified. To achieve identification, the number of shocks must be equal to the number of 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 innovations ( $\boldsymbol{\varepsilon}^{(a)}$ ) are then bootstrapped  $S$  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:

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

Then, we can choose an appropriate auxiliary model to summarise the feature of both the actual and the simulated data of the endogenous variables. The parameter of the auxiliary model is 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:  $\text{Wald}(\boldsymbol{\theta}_0) \equiv (\boldsymbol{\vartheta}^{(a)} - \bar{\boldsymbol{\vartheta}}^{(s)})' (\text{Var}[\boldsymbol{\vartheta}^{(s)}])^{-1} (\boldsymbol{\vartheta}^{(a)} - \bar{\boldsymbol{\vartheta}}^{(s)})$ .

The Wald statistic has a  $\chi^2$  distribution with a degree of freedom equal to the dimension of the parameter 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 1 illustrates the workings of 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 the 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 will see below, some of the descriptors of the data are far too powerful by 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 could either be too great or too weak. II, as in the story of Goldilocks and the three bears, can give us a power that is “just right”.

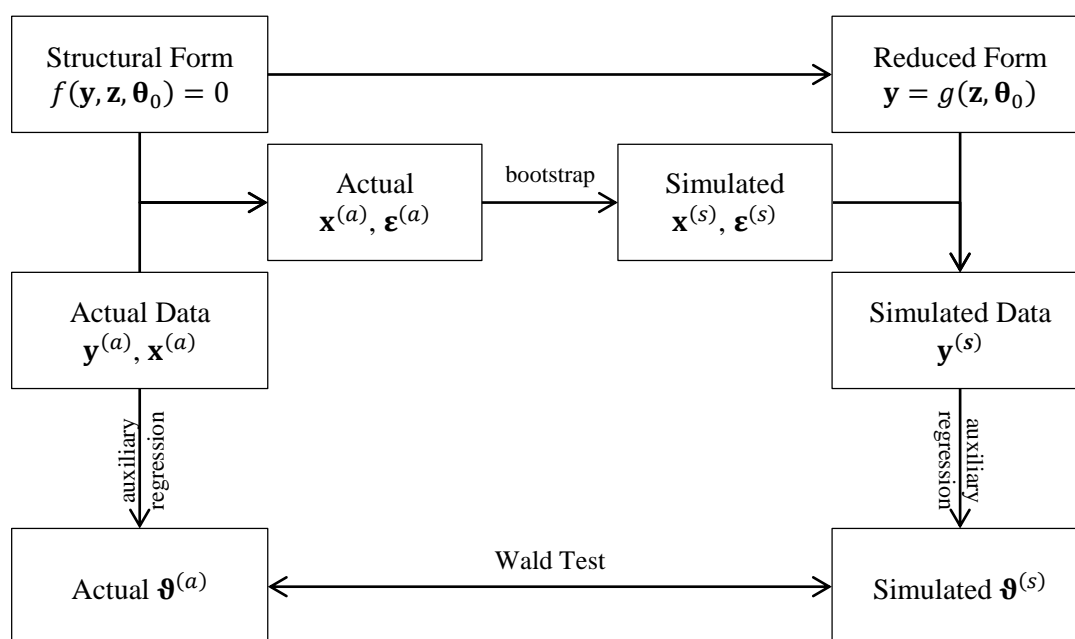


Figure 1 Flow Chart of Indirect Inference

## 1.2 Indirect Inference Estimation

We implement the II test for an initial calibration  $\theta_0$ . As a starting point, the model may be rejected because this initial calibration may not serve the model the best according to the auxiliary model criterion. An optimisation procedure can then be carried out to search for the optimal calibration  $\hat{\theta}$ , which minimises the objective function—Wald statistic. The procedure will raise the probability of accepting the model to the maximum possible. The resulting optimal calibration  $\hat{\theta}$  is therefore the II estimation of the model parameters:

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

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  $\hat{\theta}$ . 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 simulated annealing and genetic algorithm.

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.<sup>7</sup> We will use this more robust algorithm to undertake the  $\Pi$  estimation.

## 2 The Model

The model for testing the PWSP is based on a simple neoclassical labour market framework. 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).

### 2.1 The Supply Side

The representative worker faces the following standard optimisation problem:

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

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

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

For simplicity, the utility function is assumed to be 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  $\alpha$ . 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  $L$ .

---

<sup>7</sup> 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, 2014).

The first order condition is obtained by taking 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  $(\alpha_i, s_i)$  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:

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

The error term  $S_i$  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  $\varepsilon_i^S$ , supposedly to be IID:

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

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:

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

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  $\varepsilon_i^S$  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  $\varepsilon_i^S$  is uncorrelated with the terms of  $\mathbf{ind}_i$ .

## 2.2 The Demand Side

A representative firm faces the following standard optimisation problem:

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

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

$A$  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:

$$w = \gamma AL^{\gamma-1}$$

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:

$$w_i = \gamma AL_i^{\gamma-1} D_i \quad \dots(3)$$

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:

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

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, one of the variables in  $\mathbf{job}_i$  is the public sector dummy, i.e. whether the job is in public sector or private sector. Take natural logarithms to rewrite this equation into a regression-like model:

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

## 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<sup>8</sup>.

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

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).

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

The individual characteristics  $\mathbf{ind}_i$  and job attributes  $\mathbf{job}_i$  are actually the regressors in the econometric modelling method and will be termed “conditioning variables”. 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  $\boldsymbol{\beta}$  in the econometric models are the elasticities of wage. Accordingly, there are two innovations (regression error terms),  $\varepsilon_i^S$  and  $\varepsilon_i^D$ , 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 reduced-form model) should be a function of the two innovations (of structural-form model). 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<sup>8</sup> is desirable because it matches the microdata of the study.

## 3 The Indirect Inference Results

We use the dataset of 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 public sector is 9.38% higher than the private sector (23.67%

---

<sup>8</sup> 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.

for females and 6.62% for males). This crude data feature is in line with the stylised facts identified by previous econometric literature. The economic model and indirect inference will establish whether there is still a positive PSWP after taking into account individual optimisation behaviour and selection bias.

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).

Therefore, the alternative hypothesis (H1) is that “the economic model is false”, and there is no point discussing if there is selection bias. The economic model can be true or false according to the chosen II criterion, and selection bias (or interpreted as endogeneity bias) can exist, so there are four possible combinations:

	<b>The model is true</b>	<b>The model is false</b>
<b>No selection bias</b>	$Wald_a = \min(Wald_a, Wald_b) \leq c$	$c < \min(Wald_a, Wald_b)$
<b>Selection bias</b>	$Wald_b = \min(Wald_a, Wald_b) \leq c$	$c < \min(Wald_a, Wald_b)$

Notes:  $c$  is the critical value corresponding to the 95% p-value of the  $\chi^2$  distribution.

In the application of the II procedure, there is a wide choice of auxiliary model. All four types of econometric modelling methods are used to extract the information on PSWP from both actual data and simulated data. To keep the argument succinct, only one representative technique in each type is used: type 1 (linear regression model, OLS), type 2 (Blinder-Oaxaca decomposition, BOD), type 3 (propensity score matching, PSM) and type 4 (Heckman selection model, HSM). The corresponding auxiliary parameter vector ( $\boldsymbol{\theta}$ ) is:

- Type 1 (OLS): the 35 coefficients of the linear regression model.
- Type 2 (BOD): wage differentials due to different (i) coefficients and (ii) endowments.
- Type 3 (PSM): the treatment effects of (i) the treated and (ii) the untreated.
- Type 4 (HSM): the 35 coefficients of the outcome equation of Heckman model.

As noted above, the power of the II test rises with the dimension of the auxiliary parameter vector. In our case, type 1 and type 4 have very large numbers of auxiliary parameters to be matched between the actual data and the simulated data. II tests based on type 1 and type 4 auxiliary models will therefore be far more powerful than type 2 and type 3, each of which only has two auxiliary parameters; with these two the power will be rather low.

We therefore propose a compromise approach, type 5, to obtain sufficient but not excessive power of the II test: “Grouped OLS” (GOLS), which 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 is shown in Table 7 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.<sup>9</sup>

### 3.1 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  $\alpha$  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  $\gamma$  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:

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

The calibrated structural parameters give the initial values  $\theta_0 = [0.5; 0.5; 0.7; 49]$  of the vector  $\theta = [\alpha; s; \gamma; A]$ . 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. A more formal II estimation procedure will be done in the next section to provide a more robust conclusion.

The simulation in II test begins 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 2. The estimated standard deviations of the two innovations are respectively

---

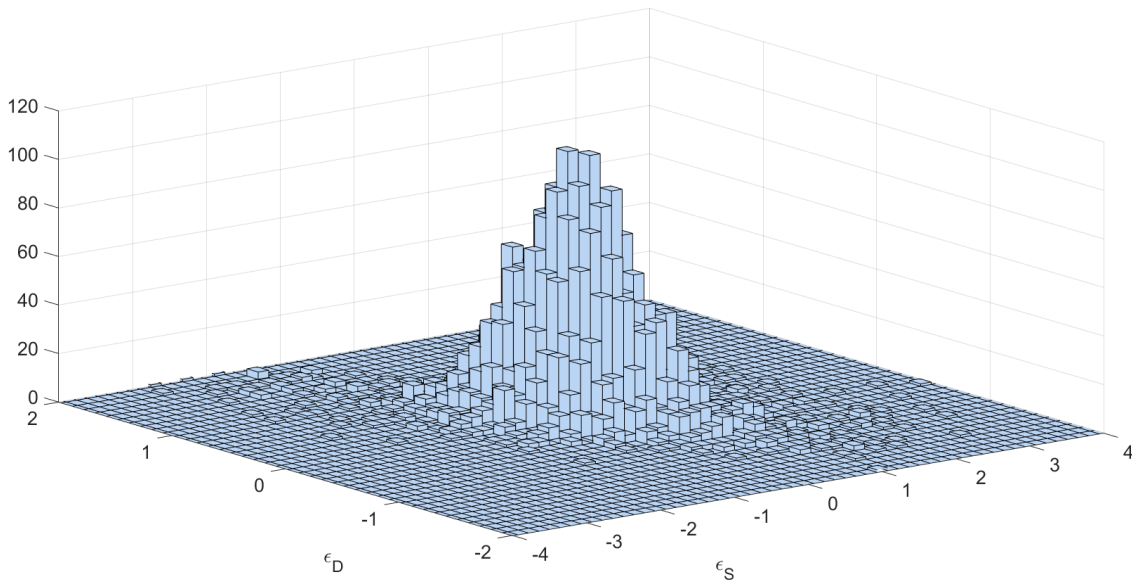
<sup>9</sup> We have done a Monte Carlo simulation to quantify the power of GOLS; it is shown below.



$\sigma_S = 0.75$  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:

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

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 cannot ignore that either. A simple solution to the dependent resampling is to bundle all the dependent variables for each observation. Bundling can maintain the observed correlations between the dependent variables in bootstrapping, but the sample variation will be greatly reduced.



*Figure 2 Joint Frequency Distribution of the Innovations*

Based on the bootstrapped innovations, we can simulate  $S$  datasets under both H0a and H0b. All four types of econometric modelling methods are used as the auxiliary regression in the II test. The simulated Wald statistics are supposed to follow a  $\chi^2$  distribution with  $K$  degrees of freedom, where  $K$  is the dimension of the auxiliary regression parameter vector ( $\boldsymbol{\theta}$ ). If the economic model is true, then the corresponding Wald statistic based on the actual data should be quite close to 0, indicating that the difference between simulated data features and the actual data features are very small. Otherwise, if the actual Wald statistic lies at the far right end, e.g. to the right of the critical value of 95% percentile, then we will have to reject the model being the true data generating process.

As an illustration, Figure 3 shows the distribution of the Wald statistics based on the  $S$  sets of simulated data using OLS as auxiliary regression. The actual Wald statistics under both H0a

and  $H0b$  are far beyond the 95% percentile, so the model is false according to this particular auxiliary regression criterion. Note that the distribution of the simulated Wald is supposed to be  $\chi^2$ , but as the degree of freedom gets larger ( $K = 35$  as in Figure 3), it converges in distribution to a normal distribution.

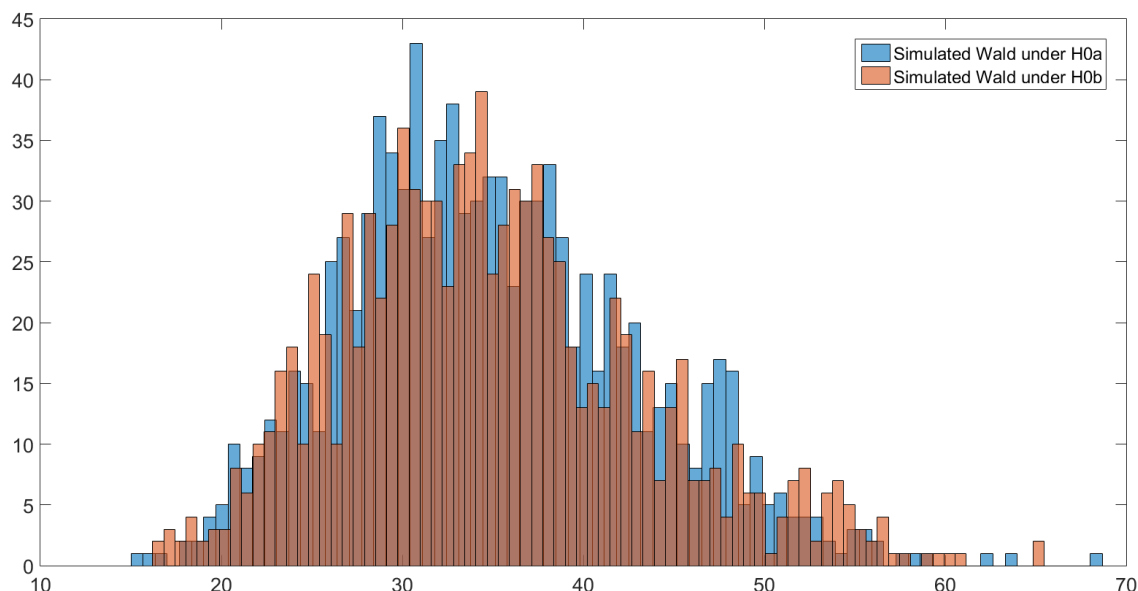


Figure 3 The Distribution of Simulated Wald Statistics (Type 1)

In the case of BOD (type 2), the mean wage differential is decomposed into a component due to different coefficients (0.0806) and a component due to different endowments (0.0862). As shown in Figure 4, these two auxiliary parameters turn out to be negatively correlated in the simulated distribution, and the actual auxiliary parameters lie outside the concentrated area of the distribution mainly due to the failure in matching the component of the endowment differences. As for the case of PSM (type 3), the estimated PSWP is different for those who work in the public sector, i.e. the average treatment effect for the treated (ATT), and for those who work in the private sector, i.e. the average treatment effect for the untreated (ATU). The two seem to be positively correlated in the joint distribution (Figure 5), with the actual auxiliary parameters ( $ATT = 0.2421$ ,  $ATU = -0.0120$ ) lying right in the most concentrated area of the histogram, indicating that the model is very likely to be true. This also makes economic sense—people are better off staying in the sector they are currently working in, so ATT is positive and ATU is negative.

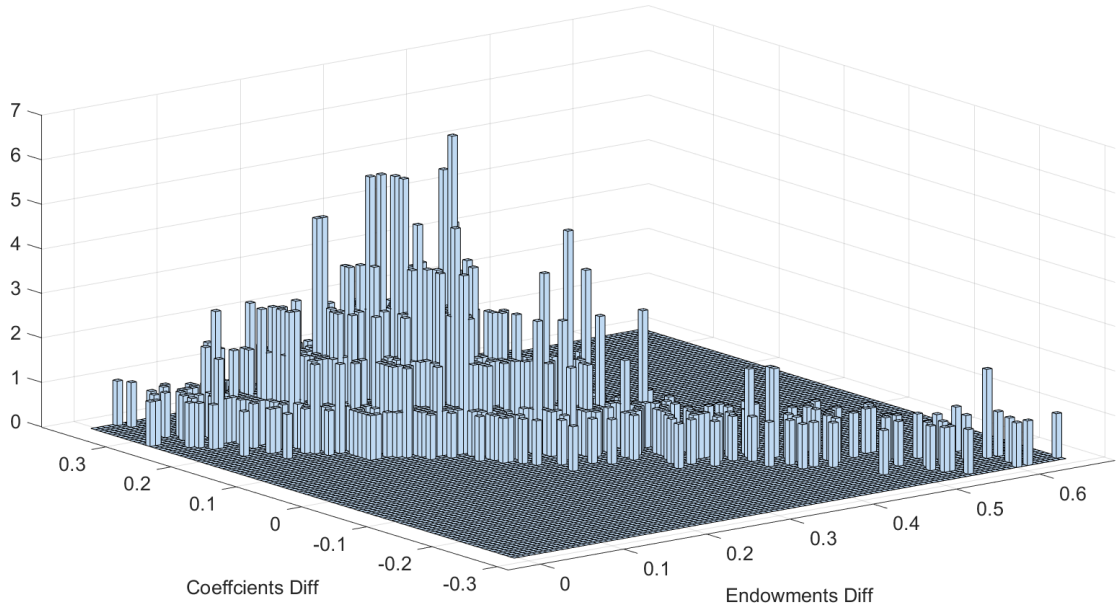


Figure 4 The Distribution of Wage Differentials (Blinder-Oaxaca Decomposition)

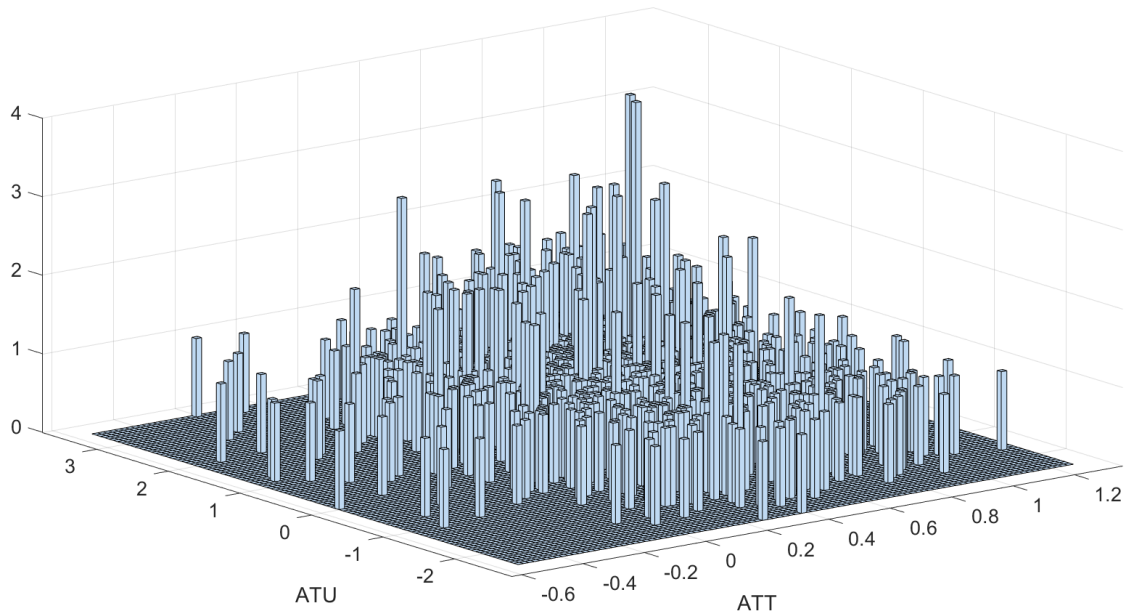


Figure 5 The Distribution of Wage Differentials (Propensity Score Matching)

Under the initial calibration  $\theta_0$ , the actual Wald statistics, the associated P-values and the critical values at 5% significance level (C-values) are reported in Table 1.

Auxiliary Regression	H0a			H0b		
	Wald	C-value	P-value	Wald	C-value	P-value
Type 1: OLS	3453.97	49.53	0.00%	4144.39	49.29	0.00%
Type 2: BOD	22.04	6.74	0.00%	156.75	6.26	0.00%
Type 3: PSM	0.91	5.83	63.40%	17.33	6.31	0.02%
Type 4: HSM	3416.85	49.70	0.00%	3988.66	49.21	0.00%
Type 5: GOLS	382.84	15.18	0.00%	102.06	15.82	0.00%

Table 1 II Test under the Initial Calibration

It is not surprising that a heavily parameterised auxiliary models (such as OLS and HSM, each with 35 auxiliary parameters) are imposing a higher bar to pass the model (implying a higher power of the tests), because there are more requirements for the model to achieve to be a “true” model. In contrast, type 2 and type 3 only have two auxiliary parameters to be compared, so the dimension of auxiliary parameters is much smaller and it is much easier for the model to pass the test. However, type 2 and type 3 auxiliary models have the advantage of providing more detailed information on the PSWP per se, unlike type 1 and type 4 in which only places a very small weight on the PSWP feature of the data. Therefore, type 5 is a middle way between the two extremes.

There are two basic conclusions that can be drawn from the II test. First, although both hypotheses are rejected to be the true data generating process in most cases (so there is no point discussing which one is less false), but we can see that most Wald statistics under  $H_{0a}$  are smaller than  $H_{0b}$  (with an exception for type 5), so there is a higher chance to accept the hypothesis of no selection bias. Second, the only type of auxiliary regression under which the model passes is PSM (type 3) under  $H_{0a}$ . It implies that the model can offer a very good explanation for the PSWP issue we are originally interested in, but may not do a good job in matching the other features of the data.

Is the PSM passing the test by chance, or is there some deeper reason behind this outperformance? As argued earlier, the propensity score matching only requires a sector choice model, and it does not have to be correctly specified, because the probit or logit equation “modelling” the probability of choosing public sector is nothing but a way of generating a matching criterion between individuals in the two sectors. This is exactly the same logic behind indirect inference test—the auxiliary model does not have to be correctly specified and only serves as a comparison “ruler”. This “ruler” may have imprecise measurements, say, stretched somehow, but we are using the same ruler to compare both data and the model generated data, so it is a fair comparison. This relative robustness exists in both propensity score matching and indirect inference, so it is not mere good luck to find this result. As shown later, the matching-based method always outperforms the other types in indirect inference, even after the model is estimated following a different auxiliary regression.

### **3.2 II Estimation**

The economic model is estimated using the genetic algorithm to search globally the best sets of values such that the Wald statistics under the two hypotheses are respectively minimised. We only adopt the GOLS auxiliary model (type 5) for the estimation purpose because of its eclectic advantages of high test power, reasonable weight on PSWP and light computational burden. The estimated structural parameters under both null hypotheses are listed in Table 2.

Structural Parameters $\theta$		Lower Bound	Upper Bound	Calibration	H0a	H0b
$\alpha$	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
$\gamma$	Labour Share	0.6	0.95	0.70	0.9366	0.6036
$A$	Productivity	24.75	74.24	49.50	56.76	27.21
<b>Wald Statistic</b>					4.69	84.25
<b>C-Value</b>					15.07	15.69
<b>P-Value</b>					79.02%	0.00%

*Table 2 II Estimation of the Structural Parameters under H0 (Type 5 GOLS)*

It is clear that H0a is favoured against H0b and H1, i.e. the model is very likely to be true and there is no evidence for selection bias, with a probability of 79.02%. Therefore, we will just focus on the estimates under H0a hereinafter.

The constant elasticity of substitution,  $s$ , is very high, indicating that the individuals treat consumption and leisure as substitutes more than complements. In a CES utility function, as  $s \rightarrow 0$  the complementarity is greater while as  $s \rightarrow \infty$  the substitutability is greater, with  $s = 1$  being the Cobb-Douglas specification with equal degrees of complementarity and substitutability. The II estimate of  $s$  (6.55 under H0a) is actually at odds with the macroeconomic 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  $s \rightarrow 1$ . In the macroeconomic 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 to match a wide range of data features as summarised by the 8 groups.

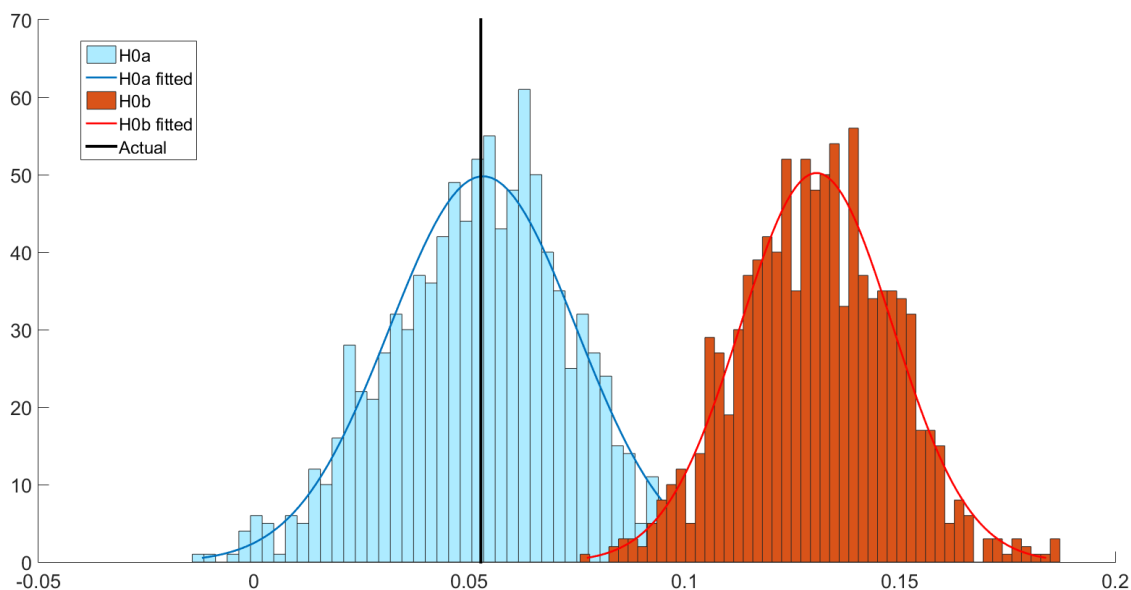
The estimated utility weight of leisure ( $\alpha$ ) is lower than the calibrated value (0.47 under H0a), so the weight on consumption is about twice of that on leisure. The estimated share of labour in the production function ( $\gamma$ ) is very close to 1 (0.94 under H0a), but this is not surprising as our sample is highly concentrated in the labour intensive industries. Finally, the productivity ( $A$ ) is calculated to match the other parameters in the production function.

### **The Auxiliary PSWP**

Based on the estimated structural parameters, many implications can be drawn with the help of the structural model. For example, the unobserved endogenous variables, such as consumption

and leisure can also be calculated, but we will focus on the comparison between the observed and simulated wage premium in the public sector, which is the main theme of this study.

The simulated wages under both hypotheses H0a and H0b are used to run the auxiliary regression (OLS or GOLS) in addition to running the same regression on the actual data (Figure 6). It is shown that the estimated auxiliary PSWP (i.e. the coefficient of the public sector dummy) based on the actual data lies right in the centre of the distribution of the PSWP estimates based on the simulated data of H0a.



*Figure 6 The PSWP of the Actual and Simulated Data (Single-Equation-Regression)*

The estimated structural parameters are also used to conduct the II test for other types of auxiliary regressions. For example, Figure 7 shows the estimated PSWP (due to differences in coefficients) using the Blinder-Oaxaca decomposition method. Again, the hypothesis H0a is preferred over both H0b and H1. A similar conclusion is found for the type 3 (Figure 8) and type 4 (results omitted, because the coefficient of the inverse Mill's ratio is not exactly a measure of the PSWP). If we only care about the model's ability of matching the PSWP feature, then the neoclassical model can do a very good job no matter what auxiliary model is chosen, and there is no evidence for selection bias. This conclusion very robust in both II tests and II estimation.

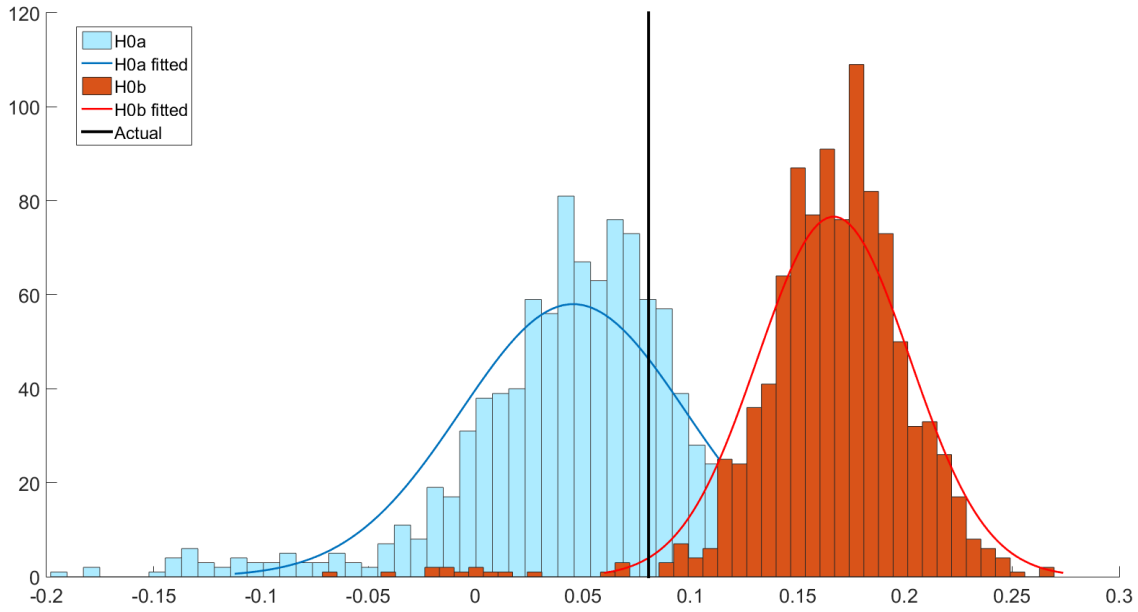


Figure 7 The PSWP of the Actual and Simulated Data (Decomposition-Based)

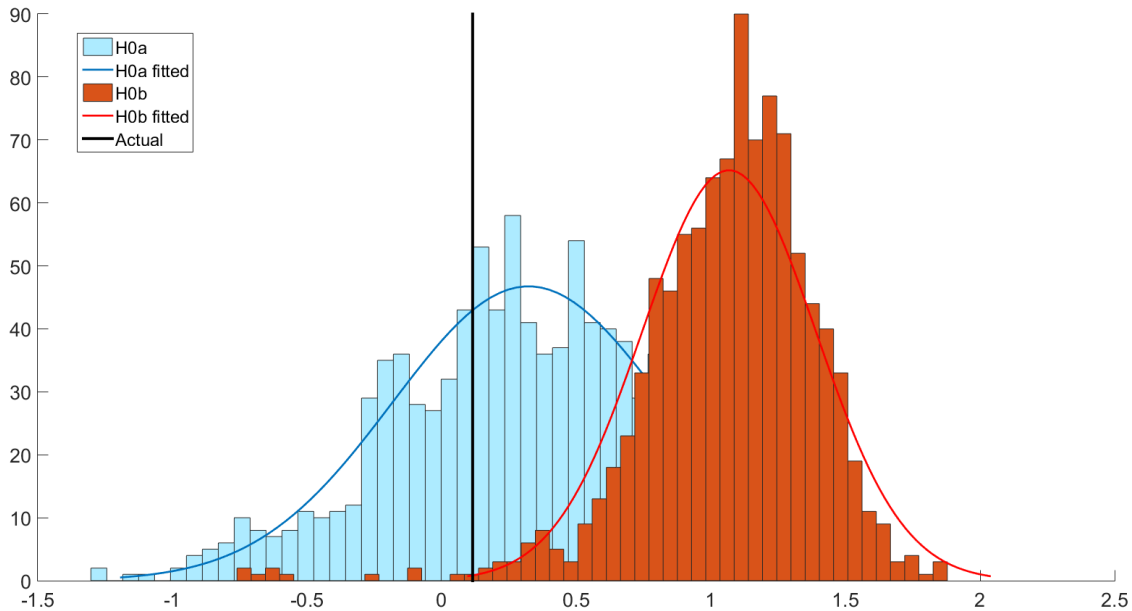


Figure 8 The PSWP of the Actual and Simulated Data (Matching-Based)

### The Structural PSWP

We should distinguish between the auxiliary PSWP (or the PSWP estimated by the auxiliary models) and the structural PSWP (or the PSWP estimated by the structural model). The former is the coefficient of the auxiliary regression based on the reduced form, i.e.  $\beta$  (an element of the auxiliary parameter vector  $\Theta$ ), while the latter is the coefficient of the regression based on the structural form, i.e.  $\eta$  (an element of the structural elasticity vector  $\eta_D$ ).

As analysed earlier, the coefficient  $\eta$  of the public sector dummy in the structural model can be interpreted as the elasticity of demand-side extra surplus with respect to working in the

public sector. Strictly speaking, it is not equivalent to  $\beta$ , which is the wage premium paid to the worker, because  $\eta$  is a surplus measure received by the firm. By these two concepts make no difference in a world without distortionary income tax. The structural PSWP is estimated to be 0.0672 (or 6.72%) with a standard deviation of 0.0158, so it is highly significant.

	<b>Model</b>	<b>Technique</b>	<b>PSWP</b>
Econometric Model	<b>Type 1</b>	OLS	5.27%
	Single-Equation-Regression	Quantile Regression	5.31%
	<b>Type 2</b>	Blinder-Oaxaca Decomposition	8.06%
	Decomposition-Based	JMP Decomposition	8.27%
	<b>Type 3</b>	Propensity Score Matching	7.17%
	Matching-Based	Nearest Neighbour Matching	5.01%
	<b>Type 4</b>	Heckman Selection Model	5.81%
	Multiple-Equation-Regression	Treatment Effects Model	12.96%
Economic Model	Neoclassical Model	Indirect Inference	6.72%

*Table 3 Summary of the Estimated PSWP*

Table 3 summarises and contrasts the different measures of PSWP from both reduced-form econometric models (auxiliary models) and the microfounded economic model for the same data. The economic model provides a quite robust estimate, lying in the middle of the estimates from various econometric models.

### 3.3 The Post-Estimation Test

The estimated parameters  $\hat{\theta}$  are used to conduct the II test for all the five types of auxiliary models, and the test results are summarised in Table 4.

For type 1, the resulting Wald statistics are halved, compared to the calibrated parameters (Table 1), but both hypotheses are still strongly rejected. The same holds for another regression-based method (type 4), because both type 1 and type 4 have 35 auxiliary parameters to be matched so it is very difficult to pass the powerful test. In contrast, under the estimated parameters, the model under the no selection bias hypotheses can pass the II test with auxiliary models of types 2, 3 and 5. In particular, the GOLS (type 5) shows that the model is capable of matching other (grouped) data features apart from PSWP.



Auxiliary Regression	H0a			H0b		
	Wald	C-value	P-value	Wald	C-value	P-value
Type 1: OLS	1909.45	49.60	0.00%	3113.23	49.82	0.00%
Type 2: BOD	0.23	6.56	88.98%	57.29	6.59	0.00%
Type 3: PSM	0.51	5.44	77.51%	17.59	6.74	0.02%
Type 4: HSM	1881.93	50.19	0.00%	3082.42	49.58	0.00%
Type 5: GOLDS	4.69	15.07	79.02%	84.25	15.69	0.00%

*Table 4 II Tests of the Model under the Estimated Parameters*

It is also informative to see how the model performs in more details by spelling out which auxiliary parameters the model can and cannot match the data counterparts. The in-out tables for the five types of auxiliary regressions are reported in Table 8 and Table 9 in Appendix 2.

It is obvious that the simulated data under the estimated parameters can match most aspects of the auxiliary regression, including all the education dummies, industry dummies, most occupational dummies and especially the public sector wage premium measures in both single-equation-regression (type 1) and multiple-equation-regression (type 4). The model is rejected overall mainly because of the discrepancies in matching some job attributes (demand side), such as industry and occupation dummies. If we only care about the model's ability to match the estimated PSWP, then the model can actually pass the II test. This is confirmed by the type 5 auxiliary model, with grouped regression coefficients. The model under no selection bias can match all the grouped data features, but the model under the pre-assumption of selection bias fail to match the return to work experience, occupation differences as well as the PSWP.

Looking at the decomposition-based (type 2) and matching-based (type 3) auxiliary regressions in Table 9, which only focus on the PSWP, both Wald statistics calculated from the actual data lie within the critical values. Even if the parameters are estimated to minimise the gap between the simulated data and the actual data in terms of type 1 auxiliary regression, it also improves the capability of the estimated model to match the other types of auxiliary regressions.

#### **4 The Power of Indirect Inference: A Monte Carlo Experiment**

The validity and reliability of the estimation/test results using II depend on its statistical power (the probability of correctly rejecting a false model). This section will adopt Le et al's (2016) approach to investigating the power of II test in the context of microeconomic models.

The Monte Carlo experiment is designed as follows. The estimated model with  $\hat{\theta}$  is assumed to be the "true" model (the true data generating process), based on which we can simulate 1000 datasets. To see the power of II test (its ability to identify false models), the parameters are manipulated up and down in an alternate fashion to create some "falsified" models. The degrees of falsification are chosen to be 1%, 5%, 10% and 20% higher/lower than the estimated values. For each of the 1000 dataset, an II test is conducted based on the type 5 auxiliary model

(GOLS)<sup>10</sup>. 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 9 summarises the following:

- **Step 1: Simulation.** Under the true model/parameters ( $\hat{\theta}$ ), i.e. the “true” DGP, simulate  $S = 1000$  sets of data.
- **Step 2: Falsification.** Adjust the parameter ( $\hat{\theta}$ ) 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.
- **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.

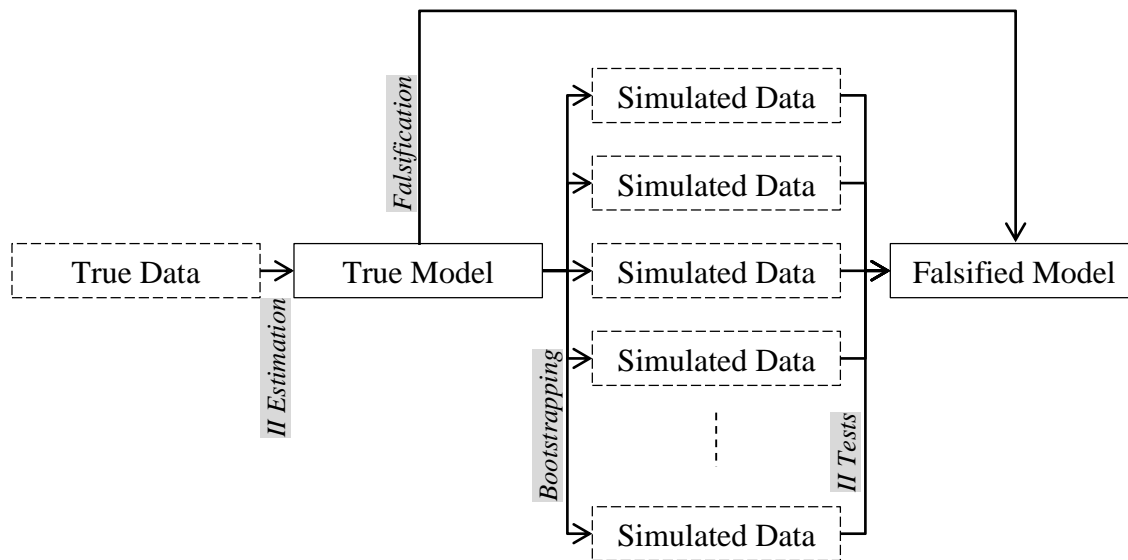


Figure 9 Illustration of the Monte Carlo Experiment

A discussion of what  $\hat{\theta}$  should include is due here. In Le et al (2016), their main results are based on different procedures applied to LR test and II test. To evaluate the power of LR test, they falsify the structural parameters while re-estimating the error parameters; but they falsify both structural and error parameters in evaluating indirect inference test. There are many arguable reasons for this difference in their study and they do check the robustness of their conclusions (Table 8, p21), but the inclusion of error parameters does raise the powers of indirect

<sup>10</sup> As argued earlier, other auxiliary models are too heavily parameterised and it is very difficult to pass a model. GOLS provides the highest chance of accepting a model, so it is basically the lower bound of the power of II.

inference test. Moreover, as shown in the results, for a much simpler microeconomic model, the error parameters have a much more weight than the structural parameters, so falsifying all parameters is neither fair to the structural parameters nor consistent with the II test procedure.

In this paper, both methods are implemented to falsify the model, but the main conclusions are drawn based on the first: (i) only structural parameters ( $s, \alpha, \gamma, A$ ) are falsified but the error parameters ( $\eta_S, \eta_D$ ) are re-estimated for each test; (ii) all parameters are falsified without re-estimating the error parameters. The resulting average p-values and simulated powers of the two methods are summarised in Table 5.

Falsification	(i) Structural Parameters Only		(ii) Structural & Error Parameters	
	Mean P-Value	Power	Mean P-Value	Power
1%	59.03%	7.0%	0.1%	100%
5%	20.87%	35.8%	0.0%	100%
10%	5.92%	74.4%	0.0%	100%
20%	1.35%	94.3%	0.0%	100%

*Table 5 The Simulated Powers of Indirect Inference Tests*

If we falsify all the parameters in the model, the indirect inference test has a too high power—any small deviation will lead to a rejection. This finding is not a surprise, because by fiddling with both structural and error parameters, the falsified model can behave in a very different way from the true data simulated from the true model—remember that there are 4 structural parameters but 35 error parameters (8 groups).

It is therefore more plausible to adopt the first method (only falsify the structural parameters and re-estimate the error parameters) in evaluating the power of indirect inference. The main argument is that this choice is consistent with the procedures used in indirect inference. In both test and estimation, we allow for re-estimation of the error parameters, such that the resulting innovations are IID. Without re-estimation of the error parameters, a small degree of deviation from the true parameters actually implies a huge degree of deviation from the true model, because of the high dimensionality of the error parameters<sup>11</sup>.

According to Table 5, we have seen that the power of indirect inference test is higher as the degree of falsification rises, and the p-values of accepting a false model is lower as the structural parameters are more false. Compared to the findings of Le et al (2016), the powers at the same degrees of falsification are relatively lower in a microeconomic model. This is because a macroeconomic model is typically heavily parameterised.

<sup>11</sup> This is a smaller issue for a complicated DSGE model, where the error parameters have a relatively smaller dimension than the structural parameters. Therefore, the conclusion in Le et al (2016) is robust despite the unfairness in implementing the likelihood ratio and II tests.

## 5 A Level Playing Field?

Various approaches arrive at a similar positive analysis conclusion: there is indeed a positive PSWP. However, what should we do about it? This normative enquiry is effectively asking whether this observed wage premium is fairly determined in a competitive labour market or it is unfairly manipulated by political power. Here is how we are going to approach this question.

Previously, to maximise the model's capability of matching the data features, both the econometric and economic models include the *ad hoc* public sector dummy to explain the wage data features—the sector dummy is both part of the regressors in the econometric models and part of the error structure in the economic model. Therefore, though we managed to obtain a robust estimate of the PSWP by including the public sector dummy in both models, we do not know what causes it. To answer the normative question, we asked in a slightly different way via model simulation: can a pure neoclassical model with no public sector dummy still explain the observed data features? In other words, if only meaningful economic parameters enter such an economic model and there is no “catch-all” public sector dummy in the error structure, can the model still pass the test? If it can, then the “pure” economic model is able to explain what is happening in the data features in which we found a “dummy” effect for public sector in all those econometric methods. Based on this argument, we re-test and re-estimate this “pure” economic model with no public sector dummy in the error structure, while keeping the auxiliary models unchanged to summarise the data features including the PSWP.

Firstly, we re-test this “pure” economic model under the previously estimated structural parameters (Table 2) to see the impact on the II test conclusions. Then, we re-estimate the “pure” model to see if the II estimates are significantly different from those based on the original model with the public sector dummy (the *ad hoc* specification). Finally, we re-test the re-estimated “pure” model to obtain the maximum probabilities of passing the model.

In Table 6, the two sets of tests based on the “pure” economic specification are contrasted with the ones under the original *ad hoc* specification. Under the previous estimates, the P-values are inevitably smaller because those estimates are chosen to maximise the probabilities of passing the *ad hoc* specification. But the test conclusions are well maintained—under both specification, the null hypothesis H0a is still decisively accepted against the other alternatives. After re-estimation under the “pure” economic specification, the P-values increase marginally, and the re-estimated structural parameters are very close to those under the original specification. It also suggests that the II procedures are fairly robust to different specifications.

To summarise, we find a pure neoclassical economic model without a 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.

<b>Specification</b>	<i>ad hoc</i>		pure		pure	
<b>Parameters</b>	estimated		estimated		re-estimated	
	H0a	H0b	H0a	H0b	H0a	H0b
$\alpha$	0.2926	0.6788	0.2926	0.6788	0.3414	1.1308
$s$	5.4729	1.1673	5.4729	1.1673	6.4743	1.1712
$\gamma$	0.9204	0.6029	0.9204	0.6029	0.9376	0.6042
$A$	29.50	27.22	29.50	27.22	31.12	27.11
<b>P-Values</b>	H0a	H0b	H0a	H0b	H0a	H0b
Type 1: OLS	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Type 2: BOD	88.93%	0.00%	46.01%	0.00%	48.76%	0.00%
Type 3: PSM	80.82%	0.10%	38.19%	0.02%	36.97%	0.02%
Type 4: HSM	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Type 5: GOLS	<b>79.04%</b>	0.00%	<b>59.50%</b>	0.00%	<b>60.50%</b>	0.00%

Table 6 The Indirect Inference Results of a Different Specification

Notes: The estimation results and test results under the original specification (cells in shade) are respectively extracted from Table 2 and Table 4.

## 6 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 the four popular types of methods in the microeconomic literature. Although the model cannot pass the test in the strictest sense if the complete linear regression is used as the auxiliary model, it can successfully mimic the data features if PSWP oriented auxiliary models or grouped regression coefficients are used. In particular, propensity score matching leads to a high probability of passing the model, because of its robustness against mis-specification. This is logically coherent with the indirect inference test, which uses the auxiliary regression only to provide a comparison basis between observed and model-simulated data rather than to seriously model of the data.

With the help of a global optimisation algorithm (the genetic algorithm), the structural parameters are formally estimated using the grouped OLS as the auxiliary regression. The estimates are then used to simulate data and run other types of auxiliary regressions, resulting in a very robust conclusion that there is no selection bias in this particular dataset. If decomposition-based or matching-based methods are used as auxiliary regressions, then the model can be verified as the true data generating process. Moreover, with a mild sacrifice of test power by grouping the regression coefficients, the model can also pass the test with a strong ability to

match both the PSWP feature and other data features such as return to education. The model-consistent estimate of PSWP is 6.72%, in line with the evidence drawn from econometric models. A Monte Carlo experiment is also conducted to evaluate the statistical power of the II test, and it confirms that the II test and estimation procedures can provide a “formidable weapon in the armoury” of the users of micro models, as well as of “macro models” (Le et al, 2016).

A normative analysis is conducted with the help of II applied to a pure economic model specification without the *ad hoc* public sector dummy in the error structure. It is argued that if such a pure neoclassical model can explain the data features including the PSWP summarised by the auxiliary models, then the estimated PSWP is not caused by unfair political arrangements. The estimation and test results suggest that the observed wage premium in the public sector is economically justified and the workers in different sectors are on a level playing field competing for wages.

Methodologically, this paper attempts to bridge the microeconomic and macroeconomic research at the technical level. As reviewed in the introduction, the methodological convergence between the two sub-disciplines has begun in the 1980s, but most efforts are invested in building a microfoundation for macrodata analysis. This paper, however, is trying to provide a microfoundation for microdata analysis, which is long ignored in the empirical literature. It would provide a closer link between the 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:

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

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.

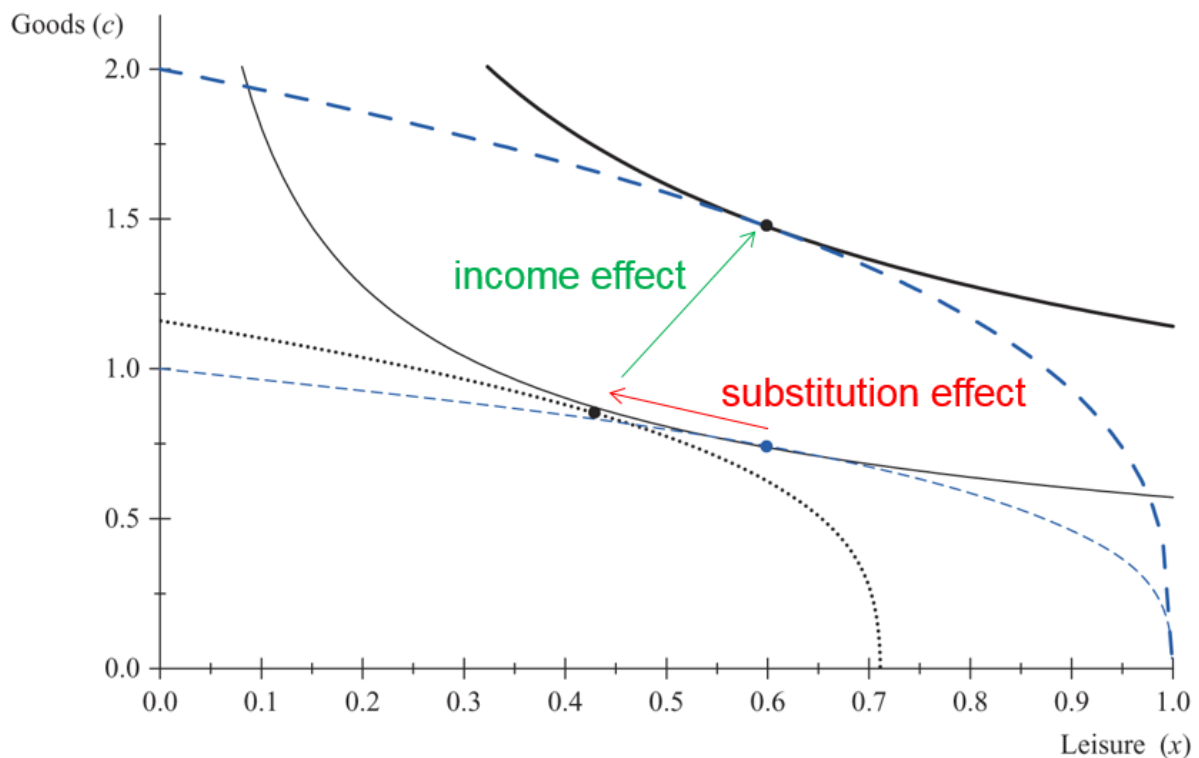


Figure 10 The Perfect Offset between Income Effect and Substitution Effect ( $s = 1$ )

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.

$$\alpha S_i \left( w_i \left( \frac{w_i}{\gamma A D_i} \right)^{\frac{1}{\gamma-1}} \right)^{\frac{1}{s}} - w_i \left( T - \left( \frac{w_i}{\gamma A D_i} \right)^{\frac{1}{\gamma-1}} \right)^{\frac{1}{s}} = 0 \quad \dots(5)$$

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.

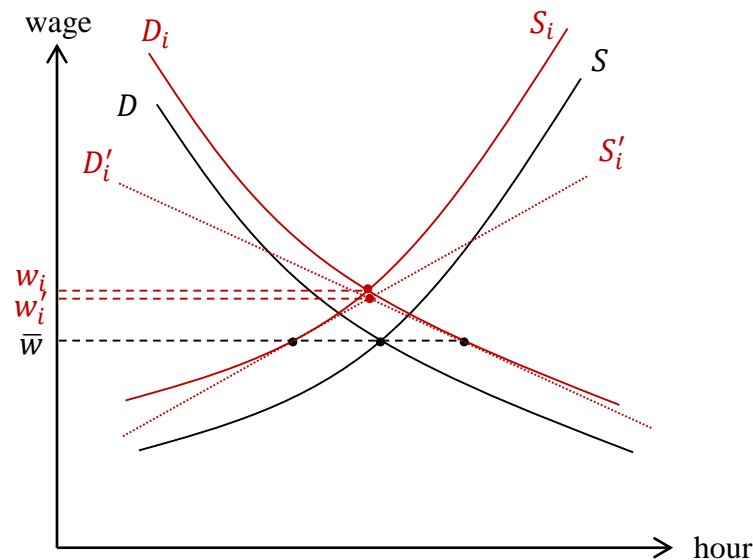


Figure 11 Linear Approximation of the Equilibrium

Figure 11 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 ( $\bar{w}$ ), 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

	<b>Grouped <math>\vartheta</math></b>	<b>OLS Regressors</b>
	$\vartheta_1$ : Intercept	intercept
<b>Individual Characteristics</b>	$\vartheta_2$ : Demographic	male white married homosexual age age <sup>2</sup> migrant
	$\vartheta_3$ : Experience	work experience work experience <sup>2</sup>
	$\vartheta_4$ : Education	low education GCSE A-level higher education degree
	$\vartheta_5$ : Temporospatical	full time London
<b>Job Attributes</b>	$\vartheta_6$ : Industry	energy & water manufacturing construction distribution transport banking public admin other services
	$\vartheta_7$ : Occupation	professional technical administrative skilled trades personal service customer service processing elementary manual job
	$\vartheta_8$ : PSWP	public sector dummy

*Table 7 The Grouped Auxiliary Parameters of Type 5 Auxiliary Model*

Auxiliary Parameters $\theta$	Type 1 OLS		Type 4 HSM		Type 5 GOLS	
	H0a	H0b	H0a	H0b	H0a	H0b
intercept	IN	OUT	IN	IN	IN	OUT
male	OUT	IN	OUT	IN		
white	IN	IN	IN	IN		
married	IN	IN	IN	IN		
homosexual	IN	IN	IN	IN	IN	IN
age	IN	OUT	IN	OUT		
age <sup>2</sup>	IN	OUT	IN	OUT		
migrant	IN	IN	IN	IN		
work experience	IN	OUT	IN	OUT	IN	OUT
work experience <sup>2</sup>	IN	OUT	IN	OUT		
low education	IN	IN	IN	IN		
GCSE	IN	IN	IN	IN		
A-level	IN	OUT	IN	IN	IN	IN
higher education degree	IN	IN	IN	IN		
full time	OUT	OUT	OUT	OUT	IN	IN
London	IN	IN	IN	IN		
energy & water	IN	IN	IN	IN		
manufacturing	IN	IN	IN	IN		
construction	IN	IN	IN	IN		
distribution	IN	IN	IN	IN	IN	IN
transport	IN	IN	IN	IN		
banking	IN	IN	IN	IN		
public admin	IN	IN	IN	IN		
other services	IN	IN	IN	IN		
professional	IN	OUT	OUT	OUT		
technical	IN	OUT	IN	OUT		
administrative	IN	OUT	IN	OUT		
skilled trades	OUT	OUT	IN	OUT	IN	OUT
personal service	IN	OUT	IN	OUT		
customer service	IN	OUT	IN	OUT		
processing	IN	OUT	IN	OUT		
elementary	IN	OUT	IN	OUT		
manual job	IN	IN	IN	IN		
public sector dummy	IN	OUT			IN	OUT
inverse Mill's ratio			IN	IN		
<b>OVERALL</b>	<b>OUT</b>	<b>OUT</b>	<b>OUT</b>	<b>OUT</b>	<b>IN</b>	<b>OUT</b>

Table 8 The Details of Post-Estimation II Tests (Type 1, Type 4 and Type 5)

Auxiliary Parameters	Type 2: Decomposition			Type 3: Matching		
	$\theta$	H0a	H0b	$\theta$	H0a	H0b
Endow. Diff.	0.0862	IN	IN			
Coeff. Diff.	0.0806	IN	OUT			
ATT				0.2421	IN	OUT
ATU				-0.0120	IN	OUT
<b>OVERALL</b>		<b>IN</b>	<b>OUT</b>		<b>IN</b>	<b>OUT</b>

Table 9 The Details of Post-Estimation II Tests (Type 2 and Type 3)