

The Public Sector Wage Premium Puzzle

Yi Wang^{*}, Cardiff Metropolitan University, UK

Peng Zhou, Cardiff University, UK

Abstract

This paper investigates the public sector wage premium in the UK over the first decade of the 21st century using both econometric and economic modelling methods. A comprehensive literature review is conducted to summarise the four popular types of methods adopted by the traditional microeconomic studies. Application of these methods results in an estimated public sector wage premium equal to 6.5%. Indirect inference is then introduced as a new method of testing and estimating a microfounded economic model. All four types of econometric methods can be used as auxiliary models to summarise the data features, based on which the distance between the actual data and the model-simulated data is assessed. The selection bias can also be tested in a straightforward way under indirect inference.

Key Words

Public Sector Wage Premium, Microfoundation, Propensity Score Matching, Indirect Inference

JEL Classification

C21, C35, J31, J45

^{*} Corresponding author: ywang@cardiffmet.ac.uk O2.55a, Cardiff School of Management, Cardiff Metropolitan University, Cardiff, UK, CF5 2YB. I am grateful to Prof Gerry Makepeace, Prof Patrick Minford and Prof James Foreman-Peck for their supervision during my PhD stage, when this paper is written. We are also thankful to the useful comments received from the participants in the ICOAE 2016, Cardiff Business School research workshop, and Cardiff School of Management research seminars. All errors remain our own.

1 Introduction

Whether the public sector is paying too much compared to the private sector has a long history in applied economics literature, at least dating back to the late 1970s (e.g. Smith, 1976; Gunderson, 1979). Evidence shows that there is a significant wage premium to work in the public sector (Chatterji et al, 2010; Blackaby et al, 2012), but why this premium exists and whether this premium is fair remains a puzzle. This issue is not only related to economic efficiency, but also to political fairness, so both positive analysis and normative analysis are involved. The first step to deal with the two issues is therefore to estimate the public sector wage premium using a robust method, on which the empirical literature has never come to a consensus. Almost all the prevailing methods, no matter how complicated the techniques are, belong to the paradigm of econometric models. If individual-level microdata are involved, they are also called microeconomic models. It is ironic that very few attempts have been made to address microeconomic issues as such by microfounded economic models.

The main reason for the preference of empirical econometric models over theoretical economic models is obvious—simplicity. It is very easy and straightforward to build an econometric model such as a linear regression without much technical costs nowadays, especially after statistical software has been greatly advanced in the last several decades. The econometric models mainly follow a philosophy of “let data speak” given its weak link between these econometric models and economic theories. A common practice is that researchers start with some economic theory (and maybe even a formal economic model) and derive some relationships, which are then loosely translated into some testable hypotheses. Subsequently, instead of the economic model *per se*, the econometric model (usually a regression model) embedding these testable hypotheses is then confronted against the data. Obviously, there are two gaps between econometric models and economic models. On the one hand, such a simplified econometric model is only a subset of the original economic model, because it only test/estimate one or several hypotheses of it, not the whole. On the other hand, the linearity (or log linearity) of the regression model greatly reduces the accuracy of the predictions of a highly non-linear economic model. Therefore, it is likely that what you test/estimate by an econometric model is not what an economic model actually implies. The validity of the findings of econometric models can be questionable.

In retrospect of the history of economic thoughts, there has been a methodological separation between microeconomics and macroeconomics from 1930s (the “Keynesian revolution”) to 1970s (the “New Classical revolution”). During those four decades,

most empirical macroeconomic models were built on *ad hoc* relationships among aggregate variables—just like empirical microeconomic models nowadays are built on *ad hoc* relationships among individual variables. Nevertheless, from 1980s onwards, especially after the real business cycle (RBC) paradigm is introduced by Prescott and Kydland (1982) into macroeconomics, the modelling methods of microeconomics and macroeconomics converge more or less in the same direction—the mainstream macroeconomic models are microfounded. However, this convergence has not been synchronised in the microeconomic realm—the mainstream methods adopted by empirical microeconomic research are still regressions or its variants. This is strange, because macroeconomic models have more microfoundation than microeconomic models in the empirical research. In fact, the techniques needed to solve, test and estimate any complicated microfounded economic model with high nonlinearity are ready to use, without having to introduce an *ad hoc* gap between theoretical and empirical models.

The purpose of this paper is therefore both empirical and methodological. First, it aims to provide a robust estimate of the public sector wage premium (PSWP) in the UK during the first decade in the 21st century. Second, a critical review of existing econometric modelling methods and techniques is also conducted, while a new method (indirect inference) adapted from the frontier macroeconomic literature is introduced to estimate the wage premium and to test the selection bias. This paper attempts to combine the methods in different sub-disciplines of economics, with a hope to enhance the communication between microeconomic and macroeconomic research in terms of both methodology and techniques.

Section 2 describes the data, based on which the empirical findings are drawn. Section 3 reviews and applies the traditional econometric literature on the PSWP, while Section 4 introduces and applies the indirect inference method which is believed to be superior to the econometric methods. Section 5 concludes.

2 The Data

This section describes the dataset of the Labour Force Surveys (LFS) collected by the Office for National Statistics (ONS) in the UK for 2001-2011. The dataset used in this study only includes individuals whose economic activity is known, accounting 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.

The LFS is a unique source of information using international definitions of employment and unemployment and economic inactivity, together with a wide range of related

topics such as occupation, training, hours of work and personal characteristics of household members aged 16 years and over. It is used to inform social, economic and employment policy. The LFS was first conducted biennially from 1973-1983. Between 1984 and 1991 the survey was carried out annually and consisted of a quarterly survey conducted throughout the year and a “boost” survey in the spring quarter (data were then collected seasonally). From 1992 quarterly data were made available, with a quarterly sample size approximately equivalent to that of the previous annual data. The survey then became known as the Quarterly Labour Force Survey (QLFS). From December 1994, data gathering for Northern Ireland moved to a full quarterly cycle to match the rest of the country, so the QLFS then covered the whole of the UK (though some additional annual Northern Ireland LFS datasets are also held at the UK Data Archive). From 2006 the LFS has been run on *calendar* quarters i.e. January to March, April to June, July to September, and October to December instead of seasonal quarters i.e. March to May, June to August, September to November, and December to February.

The target population of the LFS is based on the resident population in the UK. Specifically, the LFS aims to include all people resident in private households, resident in National Health Service accommodation, and young people living away from the parental home in a student hall of residence or similar institution during term time. The sample currently consists of around 41,000 responding (or imputed) households in Great Britain (GB) every quarter, representing about 0.16% the GB population. Data from approximately 1,600 households in Northern Ireland (NI) are added to this, representing about 0.23% of the NI population, allowing analysis of data relating to UK.

The LFS retains each sample household for five consecutive quarters, with a fifth of the sample replaced each quarter. The main survey was designed to produce cross-sectional data, but the data on each individual have now been linked together to provide longitudinal information. The longitudinal data comprise two types of linked datasets, created using the weighting method to adjust for non-response bias. The two-quarter datasets link data from two consecutive waves, while the five-quarter datasets link across a whole year (for example January 2010 to March 2011 inclusive) and contain data from all five waves. A full series of longitudinal data has been produced, going back to winter 1992. Linking together records to create a longitudinal dimension can, for example, provide information on gross flows over time between different labour force categories (employed, unemployed and economically inactive). This will provide detail about people who have moved between the categories. Also, longitudinal information is useful in monitoring the effects of government policies and can be used to follow the subsequent activities and circumstances of people affected by specific policy initiatives, and to compare them with other groups in the population. There are however

methodological problems which could distort the data resulting from this longitudinal linking. The ONS continues to research these issues and advises that the presentation of results should be carefully considered, and warnings should be included with outputs where necessary.

The same number of Wave 1 (new) addresses are selected each quarter. As illustrated in Figure 1, in any given quarter, about one-fifth of the addresses in the entire sample are in Wave 1, one-fifth in Wave 2, and so on. Thus, between any two consecutive quarters, about 80% of the selected addresses are in common.

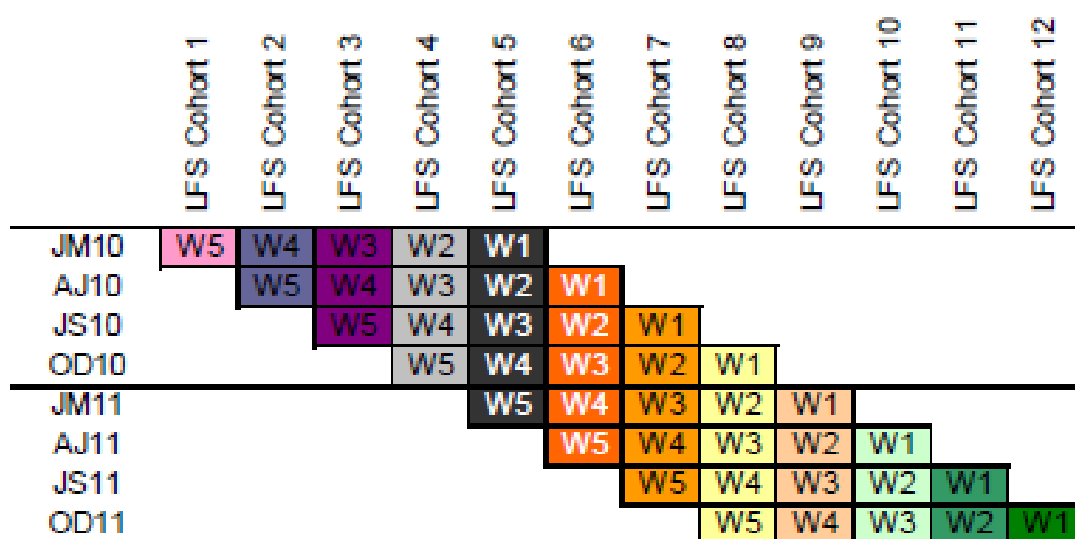


Figure 1 Illustration of Rotational Sampling Design (Source: ONS)

Since the key variable to be explained is wage, Table 2 in the Appendix summarises its important descriptive statistics for the whole sample (both weekly and hourly wage). The hourly wage is calculated based on the weekly wage and weekly working hours. It is arguable that hourly wage is preferred in later analysis because of the existence of part-time workers, who may work less and earn less on weekly basis.

One feature of wage is that the mean is greater than the median for both measures of wages and for all the years. That implies the wage distribution is not symmetric, but positively skewed. That is to say, there are more low income people than high income people in the sample.

The hourly wage data is further disaggregated by sector and by gender to show a rough picture of the pay differential between public sector and private sector (Table 2 in the Appendix). The evolution of the raw PSWP is graphed in Figure 2. There are two stylised facts consistent with other literature:

- (i) There is a persistent PSWP observed over the 2000's between public sector and private sector.
- (ii) The PSWP for females is greater than that for males.

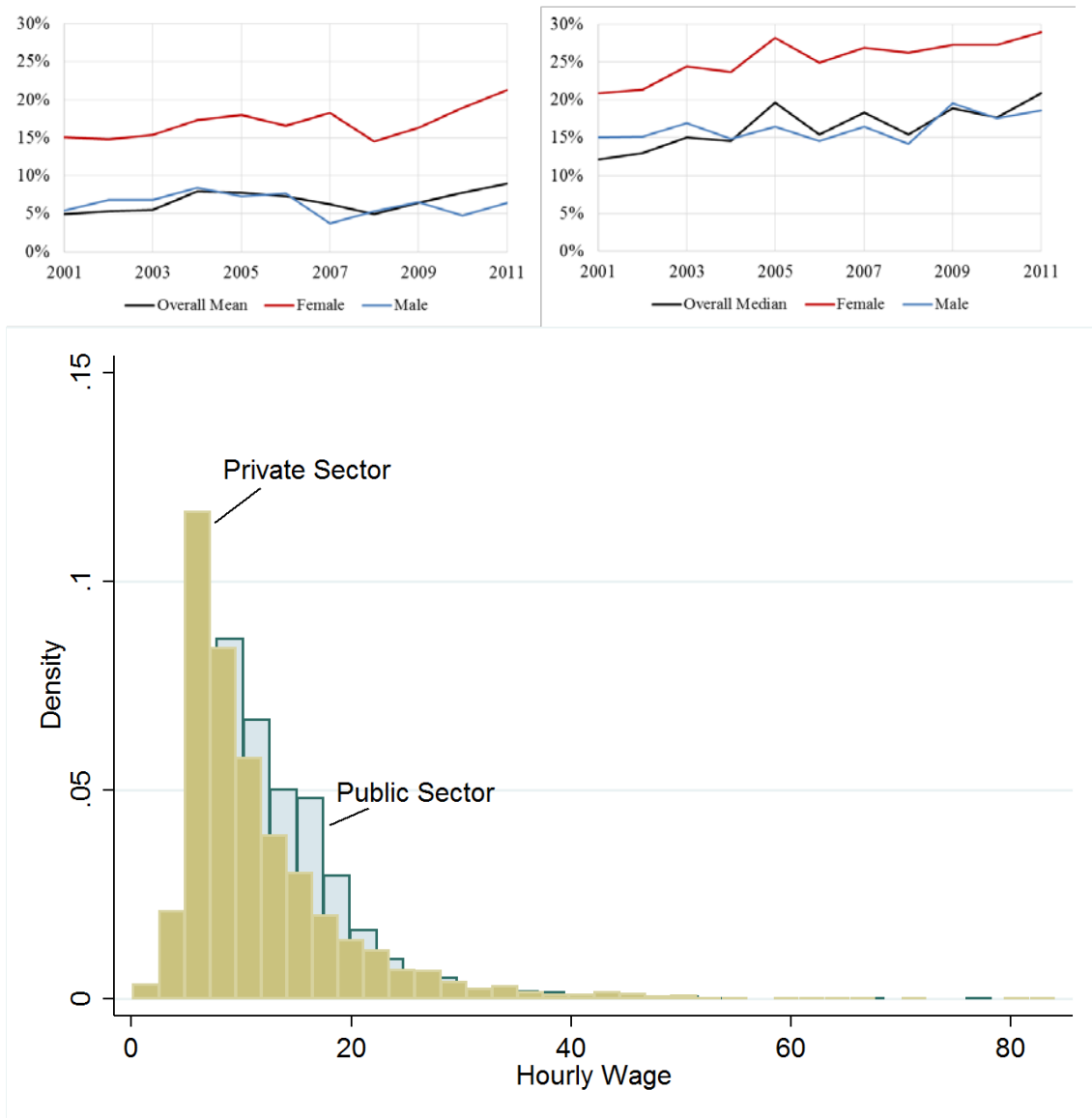


Figure 2 Raw Pay Differentials by Gender and Sector (Mean and Median)

At the first glance, it seems a bit odd that the overall raw wage premium is sometimes lower than the male's raw wage premium. But this is actually possible because the average depends on the size and the composition of the two sectors. A proof of this possibility can be found beneath Table 2 in the Appendix.

3 The Econometric Methods

To estimate the PSWP, the most naïve method is just to compare the mean wages in the two sectors, which usually lead to substantial a wage premium. This naïve method, however, ignores the fact that people working in the public sector are different in many

aspects from those working in the private sector, and the job attributes in the two sectors are quite different too. Without taking into account of these differences, the comparison is not fair.

3.1 Comparison of the Econometric Methods

To compare like-with-like, there are four main types of methods to estimate the PSWP in the microeconomic literature.

- **Type 1:** Single-Equation-Regression Method. It directly estimates the coefficient of the dummy variable describing whether or not working in public sector based on a wage determination equation². The simplest way is OLS as in Blackaby et al (2012), but instrumental variables (IV) and quantile regression are 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 literatures (Smith, 1976; Gunderson, 1979) and the later extensions by Juhn, Murphy and Pierce (1993) (JMP hereafter) 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, and simultaneous equation models. They address the problem of selection bias by adding an explicit selection equation accounting for the sector choice, so that the estimated coefficients in the wage equation are unbiased.

Though both decomposition-based methods (type 2) and matching-based methods (type 3) involve running a regression, there are some fundamental differences. First, the decomposition-based methods require running a regression of wage equation on two subsamples, while the matching-based methods require running a regression of sector choice on the entire sample. Second, the decomposition-based methods assume that the

² In our econometric model, we have 35 regressors including: a constant intercept, gender, race, marital status, sex orientation, age, age squared, migrant, work experience, work experience squared, education dummies (5 categories), work mode (full time or part time), London dummy, industry dummies (8 categories), occupation dummies (8 categories), manual job dummy, and the public sector dummy.

behaviour of individuals in the two sectors are different (so two separate regressions), while the matching-based methods believe that they follow the same behaviour (so only one pooled regression) and it is possible to find the counterfactual individual in different sectors. Third, the regressions in the first stage of decomposition-based methods are directly used to calculate the wage premium, while those in the first stage of matching-based methods are conducted just to provide a matching criterion. Thus, it does not require a strictly correct model specification, and it is more robust to mis-specification because it is based on the ranking, rather than the scale, of the estimated scores.

The multiple-equation-regression methods (type 4) integrate the features of both type 2 and type 3 methods into the benchmark methods (type 1). There are usually two equations in a multiple-equation model, with one describing how individuals make decisions on which sector to work (a feature of type 3) and the other describing how wage is determined (a feature of type 1 and 2). However, the disadvantage of multiple-equation-regression method is obvious too—if there is any mis-specification or mistakes in one of the two equations, the errors are likely to contaminate the whole system.

Both single-equation-regression methods (type 1) and multiple-equation-regression methods (type 4) specify an earnings equation or wage equation following Mincer (1974) human capital model. The difference is that the latter corrects for the selection bias using another equation, while the single-equation-regression method either ignores the problem of omitted variables (OLS and quantile regression) or deals with it using a quasi-multiple-equation method (e.g. IV). The self-selection bias problem can be interpreted as omitted variable or endogeneity problem (Heckman, 1979). To remove this bias, IV makes use of statistical relationship between “excluded instruments” and the endogenous variables, but good instruments are very difficult to find in practice. Despite being categorised as a single-equation-regression method, IV actually involves more than just one equation in estimation (that is why it is also called 2SLS). We argue that it is closer to single-equation method since the first stage of IV is usually based on statistical observation rather than economic theory. Quantile regression is another single-equation-regression method to mitigate the bias by painting a more complete picture of the distribution of wage rather than just for the mean of wage. In contrast, the multiple-equation methods, such as Heckman selection model and treatment effect models, construct and estimate a separate sector selection equation together with the wage equation. But arguably, there is a similar problem of choice of additional variables in the selection equation as with use of instruments in the IV estimation.

In addition to these techniques based on cross-sectional data, there are also various methods in estimating the wage premium using techniques for time-series data and

panel data (Disney, 2007). For example, Disney and Gosling (2003) use the privatisation programme of the 1990s in the UK as a natural experiment to avoid the problems of self-selection and measurement error. This enables them to use panel data methods to control for individual unobserved differences that do not change over time. However, the application of these methods depends on the data availability. Given this limitation, cross-sectional methods are still the most popular choice in the literature of PSWP.

The following subsections describe the literature history, the mechanisms, advantages and disadvantages of the four types of econometric methods. A comparison among the four types of methods are put upfront in Table 1, and a list of key literatures on the PSWP is compared in terms of time, country, data source and method in Table 2.

	Type 1	Type 2	Type 3	Type 4
Sectoral heterogeneity	no	yes	no	no
System-Based	no	no	no	yes
Wage Determination	yes	yes	no	yes
Sector Choice	no	no	yes	yes
Robustness to Bias	no	no	yes	yes

Table 1 Comparison of the Econometric Modelling Methods

3.2 The Findings of Econometric Methods

We have conducted all the four types of econometric methods to estimate the PSWP. To smooth the estimates over time, we take average of the estimated PSWP over the 11 sample years (2001-2011). They are compared in Figure 4. Different methods give different magnitudes ranging from 3.43% (JMP) to 10% (treatment effects). Every method has its advantages and disadvantages. Taking into account all these estimates, the average estimate for the PSWP is around 6.5%.

In addition, we have seen from the time path of the estimated PSWP and one remarkable feature is its relationship over the business cycles. It is reasonable to conjecture that the PSWP is negatively correlated to the business cycle, because the wage in the public sector is less cyclical. To verify this hypothesis, the real GDP of the UK during this period is decomposed into the cyclical component and trend component using Hodrick-Prescott filter, and the percentage deviation of the cycles from the trend are used to measure the business cycles (sometimes called the “output gap” in the macroeconomic literature). **Figure 3** contrasts this measure of business cycles with the estimated PSWP.

A simple eyeballing suggests that, during the booms (2001-2007), the PSWP tends to drop, while during the recession (2008-2011), the PSWP trends up. To formally test the hypothesis of negative relationship, the correlation coefficients are estimated between

the output gap and each estimated PSWP. The results confirm the conjecture: all estimated PSWP have negative correlation coefficients (ranging from -0.05 to -0.42) with the output gap.

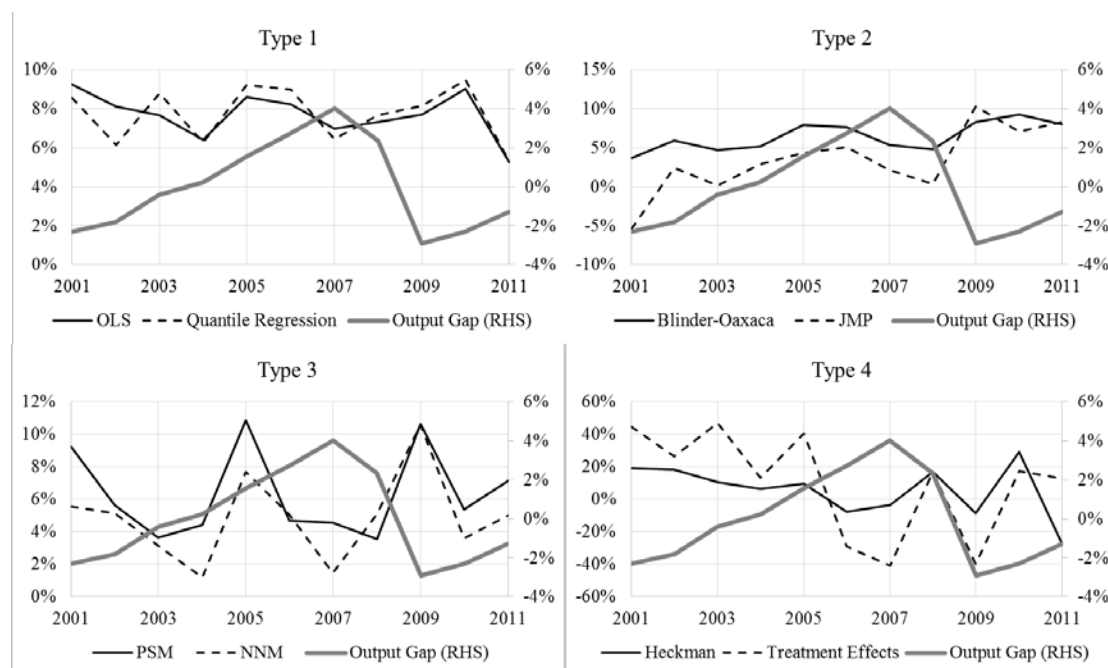


Figure 3 The Relationship between the PSWP and the Business Cycles

This implication is very informative for macroeconomic modelling, because it links the microdata evidence with the macrodata evidence. It can be used to incorporate both theoretical and empirical microfoundations into the macroeconomic models with two-sector labour market. A model allowing for heterogeneous agents is usually more powerful in explaining the persistence of the business cycles.

To summarise, the four types of econometric modelling methods give qualitatively similar results of the PSWP in the UK over the first decade of 21st century. There are four stylised facts:

- (i) There is a positive and significant PSWP (an average estimate is 6.5%).
- (ii) The females enjoy a higher PSWP than males.
- (iii) The observed raw PSWP is mainly due to the coefficient differences across sectors, especially after the financial crisis.
- (iv) The fluctuations of PSWP over time is negatively correlated with the macroeconomic business cycles (around -0.24).

The first two stylised facts are in line with the existing literature in PSWP based on the earlier datasets and the other countries. The latter two are newly identified by this paper. Methodologically speaking, though OLS is subject to selection bias or endogeneity bias,

it actually gives very stable and reliable estimates. The Blinder-Oaxaca decomposition (1973) and propensity score matching seem to give the estimates closest to the average, while the multiple-equation-regression results are much more volatile. As argued earlier, the bias due to mis-specification may be more serious than that due to selection bias or endogeneity bias. Therefore, the matching based methods (especially propensity score matching) are favourable due to its robustness to mis-specification bias, and it is indeed very close to the grand mean PSWP. The countercyclical relationship comes from the job security and pay security in the public sector. A recession is more likely to depress the private sector's wage but has little effect on the public sector's wage, resulting in a countercyclical PSWP.

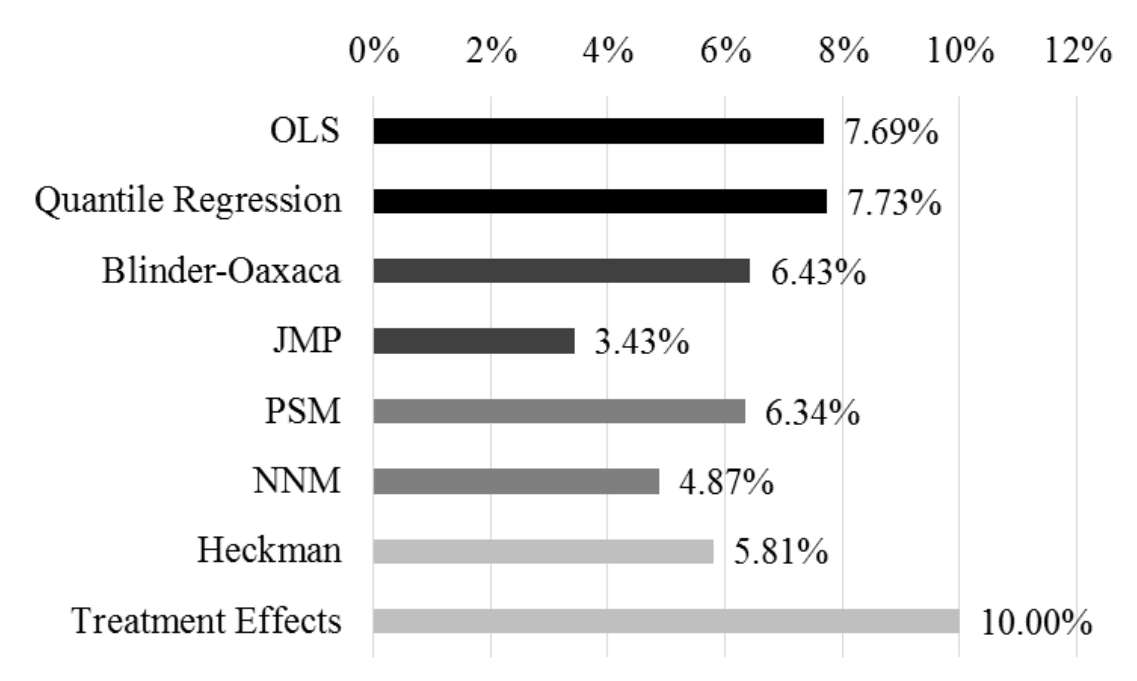


Figure 4 Summary of the Estimated PSWP of Hourly Wage (Average over Time)

Notes: Type 1, single-equation-regression method (including OLS and quantile regression); Type 2, decomposition-based method (including Blinder-Oaxaca decomposition and Juhn-Murphy-Pierce (JMP) decomposition), Type 3, matching-based method (including Propensity Score Matching (PSM) and Nearest Neighbour Matching (NNM)); Type 4, multiple-equation-regression method (including Heckman Selection Model and Treatment Effect Model). The detailed estimation results (including the estimates and standard errors) and Stata codes are available on request.

4 The Economic Method: Indirect Inference

The economic modelling method is a leap forward to directly confront the economic theories against the empirical evidence. The econometric modelling methods are usually criticised as being *ad hoc*, because the empirical models are only loosely related to the economic theory, either in a reduced-form model (type 1 to type 3) or in a structural-form model (type 4). This is a philosophy of “let data speak” because the modelling process is mainly guided by the information contained in the data. At the other end of

the methodological spectrum is the economic modelling method, which derives the model strictly following the microfounded optimisation behaviour. The resulting equation explaining the endogenous variables (wage and working hour) are nonlinear because the setup of the optimisation problem, such as utility function, is nonlinear.

This modelling strategy is often criticised to be too restrictive, because a parametric model is very difficult to capture the complicated reality. However, a loosely implied econometric model (usually linearised) is not able to capture the complicated reality either, while losing the strict theoretical foundation. Therefore, it is not convincing to adopt econometric modelling methods on this ground. Moreover, it is also logically coherent to keep the economic model as it is set up if we are to empirically verify or falsify the associated economic theory. This trend of “let theory speak” has been pushed forward in the latest two decades in the macroeconomic DSGE literature with the advances in computing power. One motivation of this paper is to bring this new trend of microfoundation “back” to microeconomics.

In the macroeconomic literature, there has already been an increasing interest in addressing the PSWP at the aggregate level. These models usually have two sectors to obtain differentiated wages, though an early RBC model proposed by Finn (1998) imposes unified wage across sectors due to its assumption of competitive market. In contrast, Ardagna (2007) develops a dynamic general equilibrium model with a unionised labour market allowing for separate wages in private and public sectors. Afonso and Gomes (2014) establish a dynamic two-sector labour market equilibrium model with search and matching frictions. In this paper, because the main purpose is methodological, we will employ the simplest neoclassical labour economic model (i.e. ignoring the union wage setting power and the searching frictions) to introduce Indirect Inference (II) techniques to microdata analysis. A more complicated non-competitive labour market model can be utilised to the II test and estimation with macrodata analysis following the DSGE literature.

Moreover, note that 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 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 logically less efficient and empirically subject to higher measurement error.

In summary, the traditional econometric modelling method links *ad hoc* econometric model with microdata (with a weak theoretical basis), while the traditional economic modelling method links microfounded economic model with macrodata (with a weak

empirical basis). As a middle way, I will use a very simple neoclassical labour economic model to introduce the idea of estimating the microfounded model using the microdata. This pushes the theoretical and empirical studies in microeconomics closer, a trend in macroeconomics since the 1980s led by the New Classical school of thought (e.g. Kydland and Prescott, 1982).

Regarding the inference techniques, there are two general ways of estimating a microfounded economic model: (i) data distribution based estimator, such as maximum likelihood and Bayesian, and (ii) data behaviour based estimator, such as Generalised Method of Moments (GMM), Simulated Method of Moments (SMM) and Indirect Inference (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 only focuses on the moment properties of the distribution, so have much more flexibility and robustness. One limitation is that it is not capable of capturing the nonlinearities in the distribution, such as asymmetry and kurtosis. Luckily, the labour market data do not show such abnormalities in the distribution of wage, so a data behaviour based estimator is preferred because it is robust to mis-specifications. 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.

Within the data behaviour 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 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 VAR(1) to summarise the joint probability of the data behaviour 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 behaviour 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. the aforementioned four types of econometric modelling methods. Therefore, we can systematically integrate the econometric and the economic modelling methods. Fur-

thermore, one advantage of Π is its flexibility. The auxiliary model needs not be correctly specified; when it is, Π is equivalent to maximum likelihood (Gourieroux et al, 1993; Smith, n.d.).

4.1 Indirect Inference Test

Suppose the structural form of a model is a system of equations composing 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, the use of jargons 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, structural innovations are also called “**shocks**” (e.g. productivity shock in RBC, markup shocks in DSGE), but 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 thesis), then the number of structural equations should be equal to the number endogenous variables (narrowly and naturally defined). Besides, in econometric terminology, “error terms” refer to the disturbance, which may (or preferably may not) be correlated to the regressors or covariates in the reduced-form or structural-form econometric models. 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, while the latter are *ad hoc*.

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 $f(\mathbf{y}^{(a)}, \mathbf{x}^{(a)}, \boldsymbol{\varepsilon}^{(a)}, \boldsymbol{\theta}_0) = 0$. To achieve identification, the number of shocks must be equal to the number of endogenous variables; otherwise, we will have “stochastic singularity”.

Following that, 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 will be able to simulate S sets of endogenous variables $\mathbf{y}^{(s)}$. This is done by simply substituting in 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 χ^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 5 illustrates the workings of II test procedures.

Note that the conclusion of the test does not depend on the likelihood of the data, but 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 on the data. It implies that an accepted model may only do a good job in matching some specific data features, so II is a weaker test of the model, compared to the likelihood ratio test. As a result, there are two advantages of II test. On the one hand, it provides a formal test of a model against the data, while the conventional likelihood ratio test can only relatively test one model against another model. On the other hand, II test is more flexible and customisable for different modelling purposes. Instead of trying to match the whole data distribution, one can choose any feature of the data to be matched.

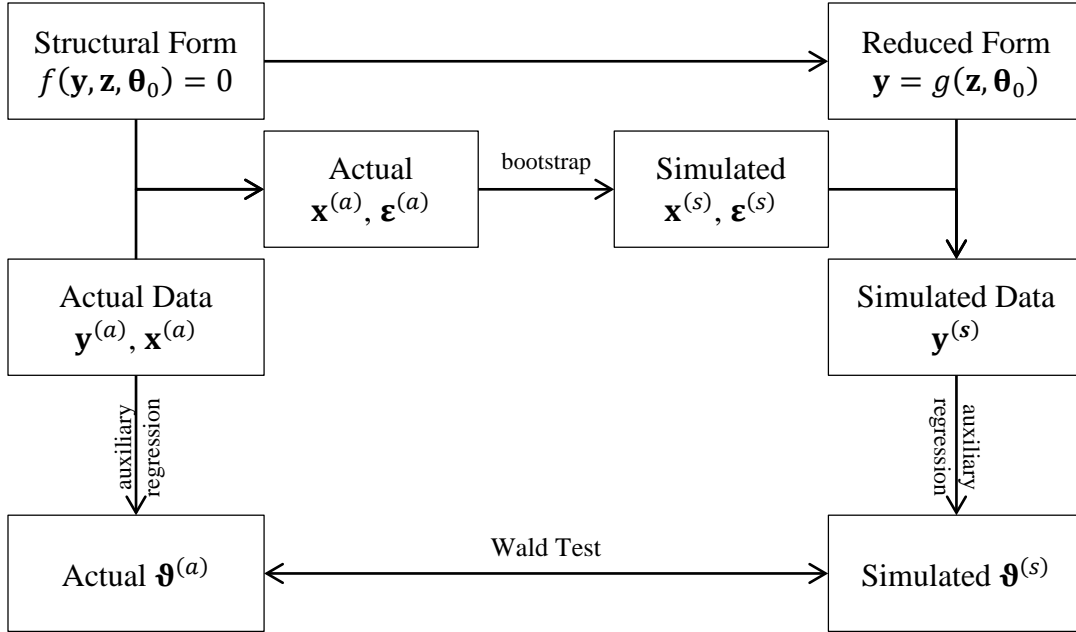


Figure 5 Flow Chart of Indirect Inference

4.2 Indirect Inference Estimation

Furthermore, as noted above, we implement the II test for a given calibration θ_0 . As a starting point, the model may be rejected because the 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 parameter:

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

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 usually 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 is widely used by macroeconomic research, such as Le et al (2010, 2011). One disadvantage of simulated annealing is 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. 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). We will use this more robust algorithm to apply the II estimation.

4.3 The Findings of Economic Models

Based on the neoclassical theory, we establish a microfounded economic model with a representative consumer who maximises utility defined over consumption and leisure, and a representative firm who maximises profit subject to a labour-only production function. The consumer's surplus and firm's surplus are further modelled as functions of individual characteristics and job attributes respectively. This structural model is specified in the Appendix.

With the help of indirect inference procedures introduced above, many implications can be drawn. 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 ("the model is true and there is no selection bias") and H0b ("the model is true and there is selection bias") are used to run the auxiliary regression (OLS) in addition to running the same regression on the actual data (Figure 6). To operationalise the assumption of selection bias (or omitted variable bias), we use bundle bootstrapping, so that the conditioning variables and innovations are sampled together. In other words, there is a dependence between the "regressors" and the "error terms". 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. Therefore, the neoclassical model is accepted as the true data generating process, and there is no evidence for selection bias.

Furthermore, 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. β (an element of the auxiliary parameter vector Θ), while the latter is the coefficient of the regression based on the structural form, i.e. η (an element of the structural elasticity vector η_D).

As analysed earlier, the coefficient η of the public sector dummy in the structural model can be interpreted as the elasticity of demand-side surplus with respect to working in the public sector. Strictly speaking, it is not equivalent to β , which is the wage premium paid to the worker, in two senses. Firstly, η is a welfare measure rather than a monetary

measure. Secondly, η is a measure from the firm's (demand-side) perspective, so the sign should be reversed if we are to measure the net welfare gain from the worker's (supply side) perspective. Nonetheless, in the neoclassical model, welfare is monetised and surplus is denominated by the same unit of wage, so the first difference is resolved. After reversing the sign of the estimated elasticity, the structural PSWP is estimated to be 0.0672 (or 6.72%) with a standard deviation of 0.0158, so it is highly significant.

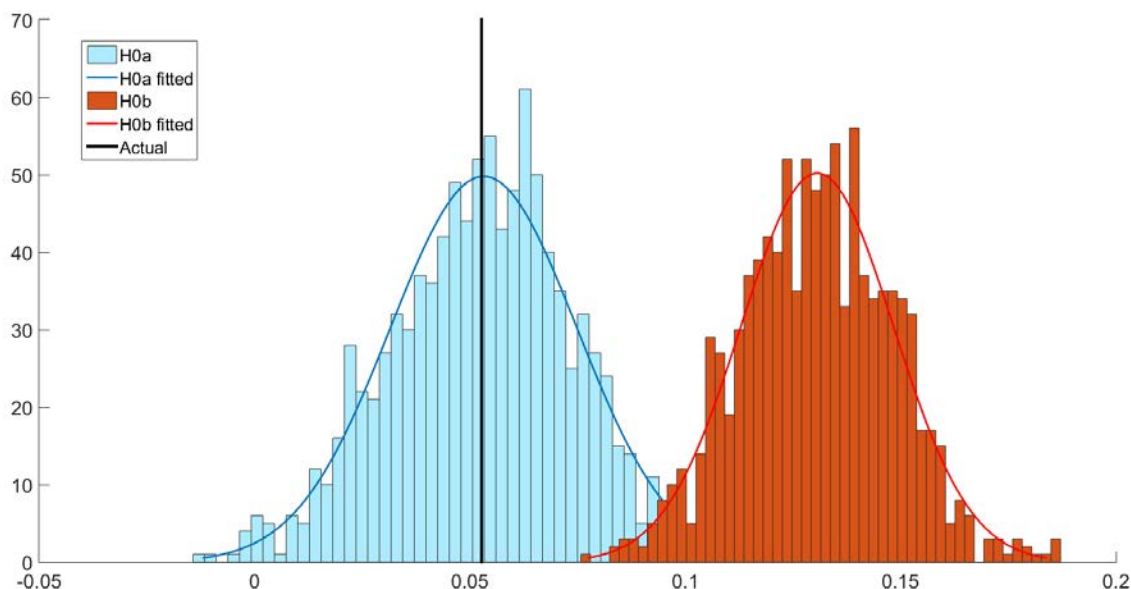


Figure 6 The PSWP of the Actual and Simulated Data

5 Conclusion

This paper reviews and compares the econometric and economic methodologies and methods used in estimating the public sector wage premium literature. We provide a robust estimate (6%-7%) using the four types of microeconomic methods and the indirect inference method. It is found that propensity score matching gives the most reasonable estimate thanks to its robustness to mis-specification. The estimate provided by decomposition-based method shows that the observed wage differentials across sectors is mainly accounted for by the economic rent of working in the public sector, rather than the individual differences.

Furthermore, built on the neoclassical labour economic theory, we derived a micro-founded structural economic model, which is confronted with the individual-level wage data using indirect inference. All the four types of microeconomic methods are used as auxiliary models to summarise the data behaviour, in terms of which the economic model is tested. Under a reasonable calibration, the model can only pass the test if propensity score matching is used as the comparison criterion between the observed and simulated data features. A common feature of indirect inference and propensity score matching is that both involve a procedure robust to mis-specification—the auxiliary

model for indirect inference and the selection equation for propensity score matching. To maximise the probability of accepting the model, estimation is carried out using OLS as the auxiliary model. The estimated model implies a 6.72% public sector wage premium, very close to the grand average of the microeconomic estimate (which is 6.5%). It is also straightforward to test selection bias using bundling bootstrap. For a specific dataset of 2011, the hypothesis of the neoclassical labour market model with no selection bias can be accepted with a very high probability. Finally, with the help of indirect inference, it is shown that the estimated public sector wage premium is not likely to be a result of unfair political arrangement, and it can be consistent with the normal economic interactions in a competitive labour market.

Methodologically, this paper is an innovative attempt to bridge the microeconomic and macroeconomic research. As reviewed in the general 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 thesis, however, is trying to provide a microfoundation for microdata analysis, which is long ignored in the empirical literature.

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*.
- BLINDER, A. S. 1973. Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8, 436-455.
- 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. 2007. Public-Private Sector Wage Differentials around the World: Methods and Evidence. *Working Paper*.
- DISNEY, R. & GOSLING, A. 2003. A New Method for Estimating Public Sector Pay Premia: Evidence from Britain in the 1990s. *CEPR Discussion Paper*.
- FINN, M. G. 1998. Cyclical Effects of Government's Employment and Goods Purchases. *International Economic Review*, 39, 635-657.
- FOREMAN-PECK, J. & ZHOU, P. 2014. The Rise of the English Economy 1300-1900: A Lasting Response to Demographic Shocks, Cardiff Economics Working Papers E2014/3, Cardiff University, Cardiff Business School, Economics Section.
- 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.
- 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., 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.
- MINCER, J. 1974. *Schooling, Experience, and Earnings*, National Bureau of Economic Research, Inc.
- OAXACA, R. 1973. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14, 693-709.
- PRESCOTT, E. & KYDLAND, F. 1982. Time to Build and Aggregate Fluctuations. *Econometrica*, 50, 1345-1370.
- 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.
- ROTEMBERG, J. & WOODFORD, M. 1997. An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy. National Bureau of Economic Research, Inc.
- 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

Table 2 Descriptive Statistics of Wage (Aggregate)

	Year	N	Mean	SD	P25	P50	P75
Weekly Wage	2001	9191	423.63	312.03	220.80	364.16	554.45
	2002	11082	425.71	315.06	222.54	363.82	554.58
	2003	10655	428.54	322.73	218.04	362.57	563.74
	2004	9066	433.91	311.10	228.74	371.33	558.72
	2005	8129	446.75	327.08	234.36	381.21	578.82
	2006	9399	438.42	318.85	227.60	372.47	573.66
	2007	9633	442.44	318.42	237.31	376.01	577.24
	2008	9218	443.24	325.74	229.92	373.97	570.81
	2009	8640	449.90	330.57	233.61	380.87	577.77
	2010	8046	433.51	319.02	223.66	366.86	562.48
	2011	7732	419.33	309.07	212.30	349.78	543.32
Hourly Wage	2001	9191	11.76	7.72	6.86	9.69	14.28
	2002	11082	11.84	7.70	6.94	9.81	14.42
	2003	10655	12.05	7.93	6.96	9.85	14.71
	2004	9066	12.11	7.45	7.18	10.07	14.80
	2005	8129	12.55	7.96	7.35	10.39	15.35
	2006	9399	12.30	7.63	7.25	10.24	15.28
	2007	9633	12.40	7.74	7.29	10.26	15.33
	2008	9218	12.44	8.16	7.19	10.24	15.29
	2009	8640	12.73	8.12	7.42	10.55	15.56
	2010	8046	12.33	7.94	7.10	10.20	15.19
	2011	7732	11.96	7.58	6.90	9.78	14.97

The proof of the possibility of $(\bar{w}_M^G - \bar{w}_M^P) - (\bar{w}^G - \bar{w}^P) < 0$:

The following proof explains why it is possible to have lower overall raw wage premium than the male raw wage premium. Assume the numbers of workers are as follows:

	Public Sector	Private Sector
Male	N_M^G	N_M^P
Female	N_F^G	N_F^P

We want to see if it is possible that male's PSWP $\bar{w}_M^G - \bar{w}_M^P$ is lower than the overall PSWP $\bar{w}^G - \bar{w}^P$, i.e. the condition for $(\bar{w}_M^G - \bar{w}_M^P) - (\bar{w}^G - \bar{w}^P) < 0$.

$$\begin{aligned}
& (\bar{w}_M^G - \bar{w}_M^P) - (\bar{w}^G - \bar{w}^P) \\
&= \left(\frac{\sum_{i=1}^{N_M^G} w_i}{N_M^G} - \frac{\sum_{i=1}^{N_M^P} w_i}{N_M^P} \right) - \left(\frac{\sum_{i=1}^{N_M^G + N_F^G} w_i}{N_M^G + N_F^G} - \frac{\sum_{i=1}^{N_M^P + N_F^P} w_i}{N_M^P + N_F^P} \right) \\
&= \frac{\sum_{i=1}^{N_M^G} w_i}{N_M^G} - \frac{\sum_{i=1}^{N_M^G + N_F^G} w_i}{N_M^G + N_F^G} + \frac{\sum_{i=1}^{N_M^P + N_F^P} w_i}{N_M^P + N_F^P} - \frac{\sum_{i=1}^{N_M^P} w_i}{N_M^P} \\
&= \frac{\sum_{i=1}^{N_M^G} w_i}{N_M^G} - \frac{\sum_{i=1}^{N_M^G} w_i + \sum_{i=1}^{N_F^G} w_i}{N_M^G + N_F^G} + \frac{\sum_{i=1}^{N_M^P} w_i + \sum_{i=1}^{N_F^P} w_i}{N_M^P + N_F^P} - \frac{\sum_{i=1}^{N_M^P} w_i}{N_M^P} \\
&= \frac{N_M^G \sum_{i=1}^{N_M^G} w_i + N_F^G \sum_{i=1}^{N_M^G} w_i - N_M^G \sum_{i=1}^{N_M^G} w_i - N_M^G \sum_{i=1}^{N_F^G} w_i}{N_M^G (N_M^G + N_F^G)} \\
&\quad - \frac{N_M^P \sum_{i=1}^{N_M^P} w_i + N_F^P \sum_{i=1}^{N_M^P} w_i - N_M^P \sum_{i=1}^{N_M^P} w_i - N_M^P \sum_{i=1}^{N_F^P} w_i}{N_M^P (N_M^P + N_F^P)} \\
&= \frac{\sum_{i=1}^{N_M^G} w_i - \sum_{i=1}^{N_F^G} w_i}{(N_M^G + N_F^G)} - \frac{\sum_{i=1}^{N_M^P} w_i - \sum_{i=1}^{N_F^P} w_i}{(N_M^P + N_F^P)} < 0
\end{aligned}$$

That is to say, as long as the last line is possible, then $(\bar{w}_M^G - \bar{w}_M^P) - (\bar{w}^G - \bar{w}^P) < 0$ is possible. The numerator of the first term is the difference between the total wage of males and females in the public sector, and the numerator of the second term is the difference between the total wage of males and females in the private sector. There is nothing stopping the first term being smaller than the second term, so it is possible that $(\bar{w}_M^G - \bar{w}_M^P) - (\bar{w}^G - \bar{w}^P) < 0$. The bottom line of the above can be further expressed in a more meaningful way:

$$\begin{aligned}
& \left(\sum_{i=1}^{N_M^G} w_i - \sum_{i=1}^{N_F^G} w_i \right) - \left(\frac{N_M^G + N_F^G}{N_M^P + N_F^P} \right) \left(\sum_{i=1}^{N_M^P} w_i - \sum_{i=1}^{N_F^P} w_i \right) < 0 \\
& \frac{\sum_{i=1}^{N_M^G} w_i - \sum_{i=1}^{N_F^G} w_i}{\sum_{i=1}^{N_M^P} w_i - \sum_{i=1}^{N_F^P} w_i} < \frac{N_M^G + N_F^G}{N_M^P + N_F^P}
\end{aligned}$$

Table 3 Key Literature

Literature	Data	Country	Period	Method
Smith (1976)	US Census	US	1960s-1970s	Type 2
Gunderson (1979)	Canadian Census	Canada	1971	Type 2
Robinson and Tomes (1984)	Social Change in Canada Survey	Canada	1979	Type 4
Venti (1987)	Current Population Survey	US	1982	Type 2, 4
Poterba and Rueben (1994)	Employer Cost Index; Current Population Survey	US	1979-1992	Type 1, 2
Disney and Gosling (1998)	General Household Survey & British Household Panel Survey	UK	1980s-1990s	Type 1, 2
Blackaby et al (1999)	Labour Force Survey	UK	1993-1995	Type 1, 2
Melly (2005)	Current Population Survey	US	1973-1989	Type 1, 2
Ramoni-Perazzi and Bellante (2006)	Current Population Survey	US	1992-2000	Type 3, 4
Gibson (2009)	International Social Survey Program Work Orientations Survey	New Zealand	2005	Type 3
Chatterji et al (2010)	British Workplace; Employee Relations Survey	UK	2004	Type 1, 2
Voinea and Mihaescu (2011)	Household Budget Survey	Romania	2004-2009	Type 1, 2
Blackaby et al (2012)	Labour Force Survey	UK	1994-2011	Type 2
Maczulskij (2013)	Labour Force Survey	Finland	1977-2008	Type 1, 4
Afonso and Gomes (2014)	Macrodata	OECD	1973-2000	Type 1, 4
Anton et al (2015)	Wage Structure Survey	Spain	2010	Type 2
Morikawa (2016)	Employment Status Survey	Japan	2007	Type 1, 2

Notes: Type 1, single-equation-regression method; Type 2, decomposition-based method; Type 3, matching-based method; Type 4, multiple-equation-regression method.

The neoclassical microfounded model:

The representative worker faces the following standard optimisation problem:

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

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

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

A representative firm faces the following standard optimisation problem:

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

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

If the labour market clears, the supply of a particular sort of labour L_i is equal to the demand for it. The equilibrium is described by the two marginal conditions:

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

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:

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