THE PUBLIC SECTOR WAGE PREMIUM PUZZLE

by

Yi Wang

ID: C0736068

*A Thesis Submitted in Fulfilment of the Requirements for the Degree of
Doctor of Philosophy of Cardiff University*

Economics Section, Cardiff Business School, Cardiff University

September, 2016

ABSTRACT

This thesis investigates the public sector wage premium in the UK over the last decade using both econometric and economic modelling methods. A comprehensive literature review is conducted to summarise the four popular types of methods adopted by the existing microeconometric studies, which are weakly derived from some labour economic theories. A common problem of the econometric methods is the difficulty in dealing with selection bias when valid instruments are not available. All four types of econometric methods are then applied to estimating the public sector wage premium, resulting in an overall average of 6.5% with a relatively higher female's premium. In particular, propensity score matching method provides the most robust estimate against mis-specification. As a bridge between microdata and macrodata in the labour market, the wage premium is shown to be counter-cyclical.

Indirect inference is then introduced as a new method of testing and estimating a microfounded economic model in the microdata analysis context. All four types of econometric methods are 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. A calibrated model passes the test only when the propensity score matching method is used as the comparison criterion. To focus on the key properties of the model, the OLS coefficients are grouped into a smaller dimension, and the estimated model can also pass the test. The selection bias can be tested in a straightforward way under indirect inference, and we find no evidence for selection bias in the data. A Monte Carlo experiment is designed to verify the high statistical power of indirect inference test. Finally, a normative analysis is carried out and we do not find unjust factors behind the observed public sector wage premium.

Key Words

Public Sector Wage Premium, Selection Bias, Indirect Inference

JEL Classification

C21, C35, J31, J45

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GENERAL INTRODUCTION

There has been a long discussion in many countries, including the UK, whether the public sector workers are paid too much. This is both an important economic issue and a political issue, concerned with both efficiency and equity. The first step to deal with these 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 microeconometric 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 such as STATA and EVIEWS are developed. The econometric models mainly follow a philosophy of "let data speak" given its weak link between these econometric models with 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 nonlinear economic model. With these said, there is a considerable risk 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 is also under question.

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 **theoretical** 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 **empirical** realm—the mainstream methods adopted by empirical microeconomic research are still regressions or its variants. In contrast, the methods and techniques in the empirical macroeconomic research has greatly advanced in the latest decade, allowing for a tight connection between theory and empirics. A very complicated microfounded economic model with high nonlinearity can be solved, tested and estimated without having to introduce an *ad hoc* gap between theoretical models and empirical models.

The purpose of this thesis is therefore both empirical and methodological. On the one hand, it aims to provide a robust estimate of the public sector wage premium in the UK during the first decade in the 21st century (positive analysis) and to answer whether this premium is fair and justified (normative analysis). On the other hand, 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 developed to estimating the wage premium and to test the selection bias. This thesis 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.

The structure of the thesis is as follows. CHAPTER I is devoted to a comprehensive literature review, summarising the empirical findings of the public sector wage premium and comparing different microeconometric methods (based on econometric models) with the indirect inference method (based on economic models). CHAPTER II describes the data and identifies some stylised facts of the wage premium and working hours in the UK. CHAPTER III focuses on the econometric modelling and microeconometric methods, while CHAPTER IV focuses on the economic modelling and indirect inference method. Towards the end of CHAPTER III, the correlation between the public sector wage premium and the macroeconomic business cycle is discussed to bridge the microdata and microdata in the labour market. And the last section of CHAPTER IV answers the key question: is the public sector wage premium fair?

CHAPTER I: LITERATURE REVIEW

In the UK, 50% of government spending goes to wages, and the public sector employs about 20% of the total UK workforce (Chatterji et al, 2010). The public sector is the largest employer in the UK, and has an influential impact on the economy. Therefore, the efficiency and equity of the public sector pay system are of great interest to both theorists and practitioners. In particular, the recent financial crisis and the great recession revived the debate over the need to restructure the public sector, and the wage premium in public sector lies at the centre of this debate in the mass media (e.g. BBC, 09 May 2011; 10 Mar 2014; 05 Feb 2016; 28 Mar 2016).

According to the neoclassical paradigm, any pay differential in competitive markets must reflect differences in worker characteristics or job attributes. After controlling for these factors, a systematic wage premium or penalty in the public sector should not exist. However, in practice, pay differential across sectors exists almost in all economies and across all periods, even after controlling for these factors. Disney (2007) summarises that this wage premium results from different occupational composition, pay structure, unionised wage setting power and worker preferences.

The first aim of this thesis is to provide a robust estimate of the public sector wage premium (PSWP) in the UK. To answer this question is not simple, because there are various dimensions causing the wage differentials, such as time, geographic area, gender, age, education, occupation, etc. Moreover, selection bias problem may exist if the choice of working in public sector is not exogenous—there may be omitted variables that affect both a worker's sector choice and her wage. In other words, if the individuals working in the two sectors are not randomly sampled, then OLS estimates are misleading because we are not comparing like with like. Simple regression ignores the fact that the sector in which an individual is working, unlike race and gender, depends on decisions made by rational economic agents. Neglecting this selectivity effect will lead to a wrong picture of the relative earnings position of the public sector workers.

1 The Public Sector Wage Premium Puzzle

The public-private pay differential has been a central issue in labour economics and policy making since at least Smith (1976). Pay comparability between public and private sectors is important for both efficiency and equity issues in the public policy. Regarding efficiency, government needs to know whether it pays an adequate wage to attract the right workers and to motivate the right efforts. For equity, the individual workers in both private and public sectors want to know if they are equally paid, and a fair pay scheme will promote a more competitive labour market across sectors. However, a vast literature across countries and over time shows that there are systematic pay differentials between private and public sectors, especially in the

developed countries such as the US, Canada and the UK. This section provides a comprehensive review on the most influential articles in PSWP, and the key literature is summarised in Table 1 (NB: The types of method will be explained in the rest of this paper).

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

Table 1 Summary of the Key Literature on PSWP

It is noted that, during 1960s and 1970s, the federal pay systems in the US were reformed to achieve the purpose of narrowing pay differentials from comparable works in the public and private sectors. The seminal paper of Smith (1976) marks a milestone in the literature, and the empirical results indicate that the federal workers were consistently paid more than their counterparts in the private sector in the US. One of the influential contributions of this paper is that the observed pay differential is decomposed into two parts. One is due to the differences in personal characteristics between the workers in the two sectors and the other is due to the economic rent (justified or not) in the public sector. Gunderson (1979) applies the same methodology to Canadian census, with a particular focus on the wage premium in terms of gender and income distribution. Evidence shows that females enjoy a higher wage premium of working in the public sector than males. Moreover, the public sector pay advantage is found larger for low-wage workers, resulting in the basic policy dilemma that reducing the PSWP may conflict with the goal of raising low-wage workers.

In the 1980s, Carow (1981), Bellante and Long (1981) and Quinn (1982) challenged the validity of this decomposition noting that "we can never fully capture all worker-specific differences" (Venti, 1987). In other words, the unexplained component of pay differential between public

and private sectors may be more properly interpreted as unobserved individual differences and job attributes. For example, on the labour demand side, the unobservable factors may include the nonpecuniary job attributes or "fringe benefits", such as stability of employment, opportunity for internal promotion, unique nature of public service, pace of work, the bureaucratic work environment, and so on. On the labour supply side, people with different tastes may prefer to work in one sector to the other because of these nonpecuniary aspects, and the sector choice may reflect the fundamental differences in people's perception of the two types of job.

Following this argument, Venti (1987) finds that wage equality between similar workers in the public and private sectors was not achieved in the US. After adjusting for both observed and unobserved individual characteristics, the basic conclusions in previous studies are actually strengthened. Females working in the public sector earn 22% higher than their counterparts in the private sector, while males earn 4% higher than those in the private sector. It is argued that, if nonpecuniary benefits are ignored, the PSWP is likely to be underestimated. Venti (1987) also formulates and estimates a model permitting prediction of the pay differential that eliminates implicit queues for public sector jobs, noting that labour market in the two sectors does not always clear. The estimates suggest that "elimination of queues would be achieved by reducing the federal wages for males about 16% and for females by about 42%". The reduction in the public sector wages implied by the estimates is substantial, indicating a huge potential for efficiency improvement in the public sector.

Meanwhile, there is an alternative focus in the literature to investigate the impacts of union status on the wage premium in public sector. This strand of literature has brought awareness of selection bias problem and multiple-equation models into this field. The effect of unions on wage is introduced by Lewis (1963) and Freeman and Medoff (1981) for the US, and Parsley (1980) for the UK. Ashenfelter and Johnson (1972) and Schmidt (1978) show that the positive effect of unions on wage estimated in single-equation model can be eliminated in a simultaneous equations framework. Lee (1978) proposes an explicit model of endogenous choice on union status, which is then used to correct for selection bias. This idea is actually very similar to Heckman selection model (1979). Robinson and Tomes (1984) is the first study that allows for the determination of union status in estimating PSWP. It is found that the choice of union status is strongly affected by the expected wage gain from joining the unionised sector, and evidence suggests a larger union gains in the public sector than in the private sector. Therefore, a major reason for the PSWP is attributed to the stronger public sector unions.

Following a series of papers in the early 1990s, such as Chamberlain (1994) and Buchinsky (1994), quantile regression emerges as a popular technique to parsimoniously model the conditional wage distribution. This technique was immediately adopted in the studies of PSWP by Poterba and Rueben (1994), who find that the relative wages of women employed in the two

sectors changed very little during 1979-1992 in the US, while the relative wages of men rose nearly 8%. This paper also explores the distribution of wage premium in terms of education level. Evidence shows that the advantage for highly educated workers to work in the public sector disappeared in the 1980s, while those with at most a high school education still enjoy a significant wage premium. Another important conclusion from quantile regression is that the magnitude of wage premium is sensitive to the choice of quantile, but the change in the wage premium is not substantially affected.
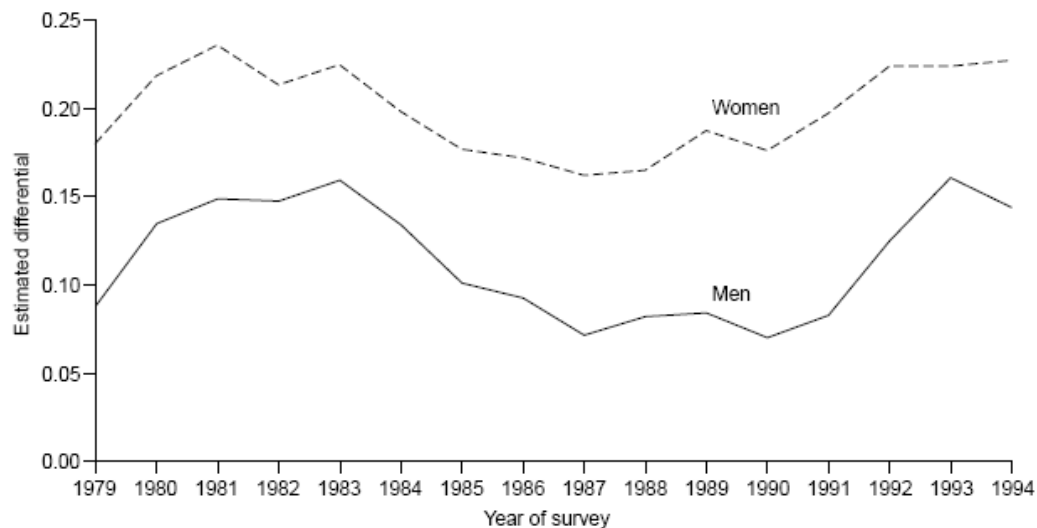
Figure 1 Public Sector Wage Premium by Gender (Disney and Gosling, 1998)

Disney and Gosling (1998) apply quantile regression technique to the UK data, and find a clear evidence of a downward trend in the PSWP over the 1980s and 1990s, as quoted in Figure 1. It is reported that wage premium for women is again persistently higher and more significant than that for men during this period. The wage premium has virtually disappeared for males by the mid-1990s. Moreover, the higher-paid workers in the public sector have a different wage premium than lower-paid workers, implying a different return to education across sectors. At the same time, Mueller (1998) uses the same technique to study the PSWP in Canada, as an extension to the previous work (Gunderson, 1979). Similar conclusions emerge—females and individuals at the lower tail of the wage distribution have higher wage premium. Following earlier literature in Canada, Mueller (1998) also takes into account the role of the public sector unions, which tend to have higher bargaining power than the private sector unions. Hence, it is a comprehensive combination of previous studies in this area. Despite the recognition of the importance of the distributional properties, these studies ignore the unobserved (residual) factors between workers and jobs. Seeing that, Blackaby et al (1999) combine quantile regression with the decomposition method proposed by Juhn et al (1993) to capture the distribution of unobserved factors. Their findings are entirely consistent with the stylised facts in the US noted by Poterba and Rueben (1994) and Mueller (1998).

Entering the 21st century, the PSWP is less significant, but the wage premium for females is still consistently higher than males. Latest literature places a special focus on the gender difference and regional difference. Chatterji et al (2010) find that the gap for male employees is less than half that for females in the UK. The major component of the earnings gap between men and women in the UK is associated with the gender effect, which suggests that the Equal Pay legislation has not been fully effective in either sector. As shown in Blackaby et al (2012), the earnings gap is significantly negative for men working in London and South East in the UK (see Figure 2), ranging from -16% (2000-2002) to -10% (2008-2010). In addition to region and gender, they also investigate the PSWP in terms of firm size, education level and specification of the earnings equation, providing a comprehensive investigation.
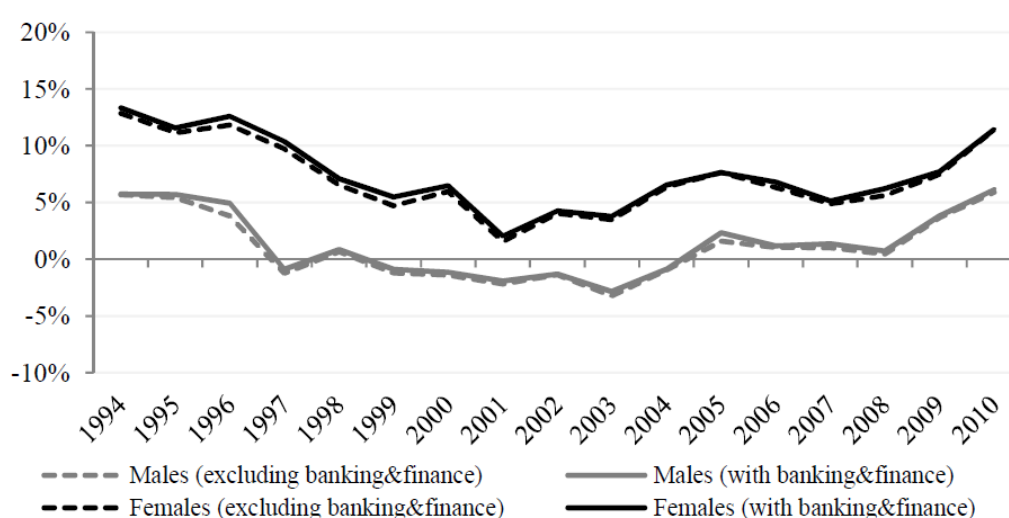
Figure 2 Public Sector Wage Premium by Gender (Blackaby et al, 2012)

Allington and Morgan (2003) conduct a literature review on the UK-based studies, not only on microeconometric studies, but also on macroeconometric studies using aggregated data on PSWP such as Trinder (1981), Elliot and Murphy (1987), Elliott and Duffus (1996), and Nickell and Quintini (2002). One of the advantages of using macrodata is that a longer sample period is available ready for time-serious analysis. However, the underlying demographic characteristics may change over time and must be taken into account. Macrodata evidence also suggests that the benefits of working in public sector were greater for females and manual workers in the 1970s and 1980s, while the premiums have gradually become penalties by the end of the 1990s, especially for males and highly skilled workers.

Different from a commonly observed public sector wage premium in the developed countries, the literature on developing countries, however, is less conclusive. For some transitional economies, the PSWP is found to be negative, such as Corbo and Stelcner (1983) for Chili, Adamchik and Bedi (2000) for Poland, Leping (2005) for Estonia, and Jovanovic and Lokshin (2004) for Russia. In contrast, countries like Tanzania (Lindauer and Sabot, 1983), Côte d'Ivoire

(Gaag and Vijverberg, 1988), Haiti (Terrell, 1993) and China (Yu and Chen, 2010, Yin and Gan, 2009) have similar features of PSWP observed in developed countries, including the distributional heterogeneity in gender and education level.

To summarise, a positive PSWP is observed from 1960s to mid-1990s on both sides of Atlantic, especially for females and low skilled/educated workers than their counterparts working in the private sectors. However, the overall wage premium has diminished since the late-1990s, suggesting a more competitive labour market across sectors. In particular, males in the public sector are more likely to have wage penalties, while females still enjoy a positive wage premium of working in the public sector.

There are four main reasons for the wage premium identified by the literature. First, private firms are profit maximisers, while public sector employers are more like vote maximisers. There is usually a wage floor for public sector wage to compete with private employers for workforce, so the disadvantaged group (e.g. females, low skilled workers) tends to benefit more from this protection. In contrast, the disadvantaged workers in private sector do not face such protection because profit is private firms' foremost objective, and the advantaged workers (e.g. males, highly skilled workers) are able to (and more motivated to) create more fortune in private sector than in public sector. Second, unions are usually more pervasive and have stronger bargaining power in the public sector, imposing a greater pressure on wage, especially for disadvantaged workers. Third, the demand for public sector services is usually regarded as inelastic, so the demand for labour in public sector is also inelastic. As such, a positive wage premium is possible and will be passed on to the customers, i.e. taxpayers. Finally, the magnitude of wage premium also depends on the phase of business cycles. Private sector wages are more sensitive to economic fluctuations, while public sector wages are less cyclical or even counter-cyclical.

2   The Selection Bias

To understand the evolution of the literature of PSWP, it is crucial to know the role of the selection bias problem in the empirical literature. It is actually a fundamental issue for applied microeconomics. Hence, this section will discuss the causes, consequences and solutions to this important issue before summarising various methods applied in the existing literature and proposing a new method in this paper.

2.1   Selection Bias VS. Endogeneity Bias

To start with, it is very easy to confuse between the bias due to selection and that due to endogeneity. These two are very distinct concepts, both of which have different solutions.

The **selection bias** refers to scenarios where the dependent variable is observed only for a non-random sample, e.g. an individual's wage within the public sector is observed only if she is working in the public sector, and vice versa. In contrast, the **endogeneity bias** means that an independent variable included in the model is an endogenous variable, correlated with unobservable factors in the error term. The dependent variable, however, is observed for all sample in the data. 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 are working in the private sector where taking risks is necessary. By nature, the selection bias is a *data problem* (non-random sample), while the endogeneity bias is a *model problem* (omitted variable).

2.2   Causes and Consequences of Selection Bias

In general, the selection bias problem may be caused by three reasons. The first is "self-selection bias", which refers to the situation where individuals have some unobserved/latent characteristics affecting the probability of being sampled. In the context of PSWP, researchers may run regression directly on all the observable individual characteristics and job attributes (including the sector choice). However, the choice of working in public sector is not like race or gender, which are determined by nature and cannot be chosen. In contrast, people have freedom to choose in which sector they work. That is to say, the observed individuals working in public sector may not be randomly sampled from population. There is an unobserved tendency or propensity leading them to choose which sector to work in, and this latent factor must take different values for different individuals. Following Heckman, this unobserved factor can be regarded as the omitted variable. As discussed later, the self-selection bias problem can be interpreted as endogeneity bias problem.

The second reason is the non-randomness due to the decisions taken by data analysts, which is also identified by Heckman (1979). Different from the first type which is caused by the observed, this type of selection bias is caused by the observer. It is common in empirical research that the researcher may drop some observations on purpose. Qualitative studies relying on small samples, collected by approaches such as questionnaires and interviews, are usually subject to this type of selection bias. This sort of problem can be avoided if the researcher is better informed about the data structure, so that stratification could be used in sampling to mitigate this risk.

The third reason for selection bias is due to the observability or tractability of some part of the population. The difference from the other two causes is that it is the nature of the data, rather than the action of the analyst, that leads to the omission of some data points. This type of selection bias often occurs when there is a threshold for individuals to be observed. For example, to estimate the return on education, the sample can only encompass those who are working,

because wages are observable only when people are employed. However, for those who are unemployed, their wages are equal to zero, but it does not mean, of course, the return to education for them is equal to zero. The wages of the unemployed individuals are not observable, and they are not likely to be a random group from the population. Therefore, omitting them will also induce selection bias.

The most important type of selection bias in analysing PSWP is the first type, i.e. self-selection bias, because the self-selection bias problem cannot be easily overcome by better sampling design, as one would do to deal with the other two types.

In conventional microeconometric literature, Hausman test and Wald exogeneity test are often used to detect the endogeneity bias, so it can also be used to test for self-selection bias. However, it is only useful for a reduced-form econometric model, which has no microeconomic theoretical foundation (microfoundation). One contribution of this thesis is to propose a new test (and estimation) procedure based on simulation of the microfounded economic model.

2.3   Solution to Selection Bias

It is common in applied microeconomics and actually in the whole social science that the data collected may not be random and are subject to selection bias. In other words, the probability of being sampled for each individual is different. Since the subjects in question are human beings with free will to make their own decisions, the individuals under observation are usually not sampled in the way carried out in controlled experiments of natural science. The distinction between the *observational data* and *experimental data* implies that the sampled individuals may not represent the whole population with equal weights.

From this point of view, selection bias can be treated as a data problem. That is to say, it causes distortion of a statistical analysis, due to the method of collecting data. If this data problem is not taken into account properly, then the estimates are biased and conclusions drawn are misleading. In other words, the estimates can only reflect the behaviour in a particular sample, and the conclusions are not generalisable to the whole population.

Alternatively, selection bias problem can also be regarded as a model specification error or endogeneity bias problem (Heckman, 1976, 1979). Given that the non-randomly selected samples are not random, then this non-randomness must be caused by some latent factor. If this latent factor is omitted, then some regressors must be correlated with the error term (containing the omitted variable). Put differently, the selection bias problem is interpreted as an endogeneity problem due to an omitted variable.

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. The following section will critically review the traditional econometric modelling methods.

3   Econometric Modelling Methods

There are four main types of method to obtain PSWP in the microeconometric 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 equation. 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 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, 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 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.

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). As argued in the previous subsection, the self-selection bias problem can be interpreted as omitted variable or endogeneity problem. 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 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
Robust to Bias	no	no	yes	yes

Table 2 Comparison of the Econometric Modelling Methods

3.1 Type 1: Single-Equation-Regression Method

Inclusion of a dummy variable describing whether or not the individual is working in the public sector may seem to be the most straightforward way of estimating the PSWP. It ranges from the simplest OLS to more complicated procedures such as IV and quantile regression. The advantage of this type of method is that an explicit monetised "return" to working in public sector can be estimated. In contrast, the decomposition-based method only provides a relative measure of the proportion of wage differential which can be explained by just working in public sector. However, the main disadvantage of this type of method is that it is usually subject to selection bias.

3.1.1 OLS Regression

If there is no endogeneity resulting from selection bias or omitted variables, a single-equation OLS regression would answer the research question. The earnings equation or wage determination equation is usually developed based on human capital theory (Becker, 1964) and return on education (Mincer, 1974). In particular, to estimate the PSWP, intercept dummy or slope dummy or both can be added into the benchmark equation to capture the difference across sectors.

$$\ln w_i = \boldsymbol{\beta}' \mathbf{x}_i + \delta D_i + \varepsilon_i.$$

In the equation above, $\boldsymbol{\beta}$ is a vector of coefficients measuring the "returns" to the individual characteristics $\mathbf{x}_i$, such as education, experience, age, gender, marriage, and so on. D_i is the

dummy variable, which is equal to 1 if the individual works in public sector and 0 if works in private sector. Thus, δ is the PSWP to be estimated.

This method is applied by some recent studies, such as Chatterji et al (2010), Dolton and Makepeace (2011), and Blackaby et al (2012). Despite the inaccuracy due to selection bias, OLS provides a convenient tool to paint the rough picture of PSWP in terms of a variety of dimensions, such as gender, age, region and education.

3.1.2 IV Regression

It is arguable that the wage equation is often mis-specified due to omitted variables, such as ability and other unobservable propensities which affect the choice of working in public sector. This self-selection bias leads to endogeneity problem, and OLS estimator will be biased.

Since the biasedness of OLS regression can be interpreted as endogeneity or omitted variable problem, one straightforward approach to this is to use instrumental variables. Voinea and Mihaescu (2011) use the variable "whether there are any family members working in the public sector". Another example is Disney and Gosling (2003), who construct an instrument based on the difference in propensity to work in the public sector after and before the privatisation programme. However, due to data availability, it is difficult to find valid and strong instruments in practice. The inefficiency and large standard errors caused by weak instruments may bring even bigger problem than the biasedness of OLS.

As mentioned earlier, the IV estimator lies between the single-equation and multiple-equation methods. It is also called two-stage least squares (2SLS) in the sense that all the endogenous variables should be regressed on the instruments in the first stage, and apply OLS to the orthogonised variables to obtain the unbiased estimates. The first stage of IV aims to remove the correlations between the endogenous variables and the error term, in a very similar way to the multiple-equation methods. It is categorised in the single-equation-regression method, because the instruments used in the first stage may be purely statistical correlation without any economic theory behind. Therefore, orthogonisation (the first stage) in IV estimation is more a statistical procedure than an economic modelling.

3.1.3 Quantile Regression

While least squares regression techniques (including OLS and IV) estimate the partial effect of a regressor on the *mean* of dependent variable, quantile regression investigates the partial effect of a regressor on the specified *quantile* of the dependent variable. The specified quantile could be median or any percentile of the distribution of the dependent variable. Hence, quantile regression provides more information of the conditional distribution of wage.

Similar to the advantage of median over mean, quantile regression is more robust against outliers. One crucial assumption of OLS is that the error term has precisely the same distribution whatever values may be taken by the components of the regressors. This case is referred to as a pure location shift, since it assumes that regressor affects only the location of the conditional distribution of the dependent variable, not its scale or any other aspect of its distributional shape. However, in most practical cases, regressors may influence the conditional distribution of the response variable, such as expanding its dispersion, stretching one tail and compressing the other, and inducing multimodality. Explicit investigation of these effects via quantile regression can provide a more nuanced view of the stochastic relationship between variables, and therefore a more informative empirical analysis. In particular, the quantile regression coefficients in the earnings equation can be interpreted as rates of return to skills at different points of the wage distribution (Buchinsky, 1994).

This idea of quantile regression was first proposed in the 18th century by Boscovich and subsequently developed by Laplace and Edgeworth. A good introduction of quantile regression can be found in Koenker and Hallock (2001). Technically speaking, least squares regression is to minimise the symmetrically weighted sum of squares of residuals, and quantile regression is to minimise the sum of asymmetrically weighted absolute residuals. For the pth quantile:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left[\sum_{i \in \{i: y_i \geq \mathbf{x}_i \boldsymbol{\beta}\}} p \left| y_i - \boldsymbol{\beta}' \mathbf{x}_i \right| + \sum_{i \in \{i: y_i < \mathbf{x}_i \boldsymbol{\beta}\}} (1 - p) \left| y_i - \boldsymbol{\beta}' \mathbf{x}_i \right| \right].$$

Note that quantile regression is usually combined with the decomposition-based method, especially in JMP decomposition method.

3.1.4 Other Regression Methods

Many papers also use panel data models (e.g. fixed effects and random effects) when longitudinal observations are available. Depending on whether the data is individual-level or aggregate-level, different specifications are used. For example, Blanchflower and Bryson (2010) use quarterly LFS individual-level panel data in the UK and find that there is a higher union membership wage premium in the public sector. Nunzianta (2005) use country-level panel data for 18 OECD countries and find that institutional factors in labour market significantly contribute to the wage premium.

A different set of estimation and test procedures are necessary for panel data models, such as panel cointegration test and Hausman test. Since the current study only use cross-sectional data, no further details of the panel data methods are reviewed here. A recent review on these can be found in Afonso and Gomes (2014).

3.2 Type 2: Decomposition-Based Method

The regression-based method assumes that all the slope parameters are the same for individuals working in both sectors, but this is obviously not quite true. For example, the importance of education to a doctor in the NHS (public sector)[1] is much higher than that to a bartender (private sector). There are many different decomposition techniques.

3.2.1 Blinder-Oaxaca Decomposition

The decomposition-based method is originally developed by Blinder (1973) and Oaxaca (1973), and the first application to estimating the PSWP is found in Smith (1976). Its main advantage over the single-equation regression is that it does not impose the equality restriction on the other slope parameters across sectors. Though in principle slope dummies can be used in OLS to control for the slope differences, very few papers use them. The raw average pay differential can be decomposed into two portions:

$$\overline{\ln w^G} - \overline{\ln w^P} = \hat{\boldsymbol{\beta}}^{G\prime} \overline{\mathbf{x}^G} - \hat{\boldsymbol{\beta}}^{P\prime} \overline{\mathbf{x}^P}$$
$$= \hat{\boldsymbol{\beta}}^{G\prime} \overline{\mathbf{x}^G} - \hat{\boldsymbol{\beta}}^{P\prime} \overline{\mathbf{x}^G} + \hat{\boldsymbol{\beta}}^{P\prime} \overline{\mathbf{x}^G} - \hat{\boldsymbol{\beta}}^{P\prime} \overline{\mathbf{x}^P}$$
$$= \underbrace{\left(\hat{\boldsymbol{\beta}}^G - \hat{\boldsymbol{\beta}}^P \right)^\prime \overline{\mathbf{x}^G}}_{(i)} + \underbrace{\hat{\boldsymbol{\beta}}^{P\prime} \left(\overline{\mathbf{x}^G} - \overline{\mathbf{x}^P} \right)}_{(ii)}$$

The first portion (i) is due to the sectoral difference in coefficients or returns, while the second portion (ii) is due to different individual characteristics. The latter term (ii) is called "legitimate" pay differential because it reflects difference in characteristics of individuals, such as education, experience, etc. In contrast, the first term (i) is considered as wage premium or economic rent, which reflects the pure pay differential paid for the same characteristics.

A simple extension of this decomposition is proposed by Daymont and Andrisani (1984), whereby the third component (iii) is used to capture the interaction between (i) and (ii):

$$\overline{\ln w^G} - \overline{\ln w^P} = \hat{\boldsymbol{\beta}}^{G\prime} \overline{\mathbf{x}^G} - \hat{\boldsymbol{\beta}}^{P\prime} \overline{\mathbf{x}^P}$$
$$= \underbrace{\left(\hat{\boldsymbol{\beta}}^G - \hat{\boldsymbol{\beta}}^P \right)^\prime \overline{\mathbf{x}^P}}_{(i)} + \underbrace{\hat{\boldsymbol{\beta}}^{P\prime} \left(\overline{\mathbf{x}^G} - \overline{\mathbf{x}^P} \right)}_{(ii)} + \underbrace{\left(\hat{\boldsymbol{\beta}}^G - \hat{\boldsymbol{\beta}}^P \right)^\prime \left(\overline{\mathbf{x}^G} - \overline{\mathbf{x}^P} \right)}_{(iii)}$$

---

[1] ONS estimates for the NHS differ from the headline figure produced by the NHS Information Centre (IC). This reflects the wider UK coverage (IC figures are for England only) plus the exclusion by ONS of general practitioners (GPs). ONS, in accordance with National Accounts practice, classifies GPs as part of the private sector. ONS also includes 'hospital practitioners and clinical assistants' who work in hospitals on a salaried pay scale but generally work as GPs leading the IC to exclude them from their totals to avoid double counting. When these factors are allowed for, ONS and NHS data can be shown to be identical.

Whichever decomposition method is used, according to Oaxaca and Ransom (1994), the decomposed raw differences in wage can be interpreted as due to a different "productivity" or "endowment", advantage of working in the public sector, and the disadvantage of working in the private sector. There are many ways[2] of calculating the contribution of each difference to the average coefficient vector $\boldsymbol{\beta}^*$, depending on how to select the general weighting matrix $\boldsymbol{\Omega}$:

$$\boldsymbol{\beta}^* = \boldsymbol{\Omega}\hat{\boldsymbol{\beta}}^G + \left(\mathbf{I} - \boldsymbol{\Omega}\right)\hat{\boldsymbol{\beta}}^P .$$

3.2.2 Terrell Decomposition

The validity of both decomposition methods depends on the validity of the OLS regression in the first stage, which in turn depends on two key assumptions. First, given the observed individual characteristics, workers are randomly distributed across sectors. Second, pay differentials do not represent differences for non-pecuniary job attributes of each sector. Obviously, both assumptions do not hold, because the choice of sector is endogenous and sectors offer fundamentally different nonwage job attributes. Venti (1987) decomposes the pay differential between the public and private sectors into four sources: first, economic rent or overpayment by government employers; second, observed productivity or skill differences; third, unobserved productivity or skill differences; and fourth, equalising differences in pay for nonpecuniary job attributes. The original approach used in Smith (1976) only distinguishes the first two, while the latter two are ignored.

Regarding the third source ignored by Blinder-Oaxaca decomposition, one extension is to take into account the unobserved characteristics, based on Heckman selection model (Terrell, 1993). If the earnings equation omits some unobservable characteristics that are related to the choice of working in public sector, then the OLS estimates are subject to self-selection bias. Heckman selection model (detailed in the next subsection) enables one to control for this omitted variable by constructing a regressor (inverse Mills ratio) from the selection equation. Hence, there is one more component (iii) compared to Blinder-Oaxaca decomposition:

$$\overline{\ln w^G} - \overline{\ln w^P} = \hat{\boldsymbol{\beta}}^{G\prime}\overline{\mathbf{x}^G} + \hat{\delta}^G\overline{\hat{\lambda}}^G - \hat{\boldsymbol{\beta}}^{P\prime}\overline{\mathbf{x}^P} - \hat{\delta}^P\overline{\hat{\lambda}}^P$$

$$= \underbrace{\left(\hat{\boldsymbol{\beta}}^G - \hat{\boldsymbol{\beta}}^P\right)'\overline{\mathbf{x}^G}}_{(i)} + \underbrace{\hat{\boldsymbol{\beta}}^{P\prime}\left(\overline{\mathbf{x}^G} - \overline{\mathbf{x}^P}\right)}_{(ii)} + \underbrace{\left(\hat{\delta}^G\overline{\hat{\lambda}}^G - \hat{\delta}^P\overline{\hat{\lambda}}^P\right)}_{(iii)}$$

3.2.3 JMP Decomposition

Another extension of Blinder-Oaxaca decomposition is proposed by Juhn, Murphy and Pierce (1993) (JMP decomposition hereafter). It takes into account the distribution of residuals, based

---

[2] For example, Cotton (1988), Neumark (1988) and Reimers (1983). We will use Cotton (1988) method, setting the weight equal to the proportion of the "treated", i.e. public sector workers.

on quantile regression. The overall observed wage differential across sectors can be decomposed in three components, using the quantile rather than the mean of the dependent variable:

$$\ln w^G - \ln w^P = \hat{\boldsymbol{\beta}}^{G\prime}\mathbf{x}^G + \hat{\varepsilon}^G - \hat{\boldsymbol{\beta}}^{P\prime}\mathbf{x}^P - \hat{\varepsilon}^P$$

$$= \underbrace{\left(\hat{\boldsymbol{\beta}}^G - \hat{\boldsymbol{\beta}}^P\right)'\mathbf{x}^G}_{(i)} + \underbrace{\hat{\boldsymbol{\beta}}^{P\prime}\left(\mathbf{x}^G - \mathbf{x}^P\right)}_{(ii)} + \underbrace{\left(\hat{\varepsilon}^G - \hat{\varepsilon}^P\right)}_{(iii)}$$

This method is usually based on separate quantile regression for each sector. The first term (i) and the second term (ii) are similar to those in Smith (1976), except for the interpretation is in terms of some specified quantile rather than mean. In addition to the identified contributions from difference across sectoral returns and individual characteristics, there is an extra term (iii) corresponding to an unmeasured component of the difference and consists of unidentified sectoral and individual effects.

A problem of JMP decomposition is that it does not account for heteroscedasticity of the error term (Melly, 2005). If the error term is independently, identically and normally distributed, then JMP decomposition is efficient. However, if this assumption does not hold, this procedure may produce misleading results. Based on DiNardo et al (1996) and Lemieux (2002), Melly (2005) develops an extension to JMP decomposition to deal with heteroscedasticity. First, quantile regression is used to estimate the conditional wage distribution. Then, the conditional distribution is integrated over the range of regressors to obtain an estimate of the unconditional distribution.

3.3   Type 3: Matching-Based Method

The ultimate problem is to estimate the wage premium between public and private sectors, and a key principle is to compare like with like. It is not meaningful to compare a highly educated individual in public sector with a very poorly educated individual in private sector, since these two persons are different in various dimensions, i.e. "comparing oranges with apples". Unfortunately, naive methods of using simple descriptive statistics will result in the problem of selection bias. Of course, the most ideal way is to use the wage of an individual if he works in public sector minus the wage of the same individual if he works in private sector. However, it is impossible to observe the counterfactual wage if he works in the other sector. Nevertheless, it is possible to find the closest match for him in the other sector to mitigate the selection bias.

An individual, whether he works in public sector or private sector, has several dimensions of properties, such as sex, age, education, region, industry, occupation, etc. It is almost impossible to find an exact match for an individual with so many dimensions of characteristics. As a result, it is attractive to use a single-valued "propensity score" to summarise these characteristics, and match individuals in the two sectors in terms of this propensity score. Once matches are found,

it is then easy to estimate the average treatment effect for the treated (ATT), average treatment effect for the untreated (ATU), and the average treatment effect (ATE).

In the context of PSWP, "working in public sector" can be regarded as "treatment", so those working in private sector are the "control group". The ATT here can be interpreted as "the wage differential if one changed her current work from public sector to private sector, given that she is actually now working in the public sector". The ATU can be interpreted as "the wage differential if one changed his current work from private sector to public sector, given that she is actually now working in the private sector". The ATE is just an average between ATT and ATU. In practice, there are two popular matching techniques to obtain the treatment effects (ATT/ATU/ATE).

3.3.1 Propensity Score Matching (PSM)

First introduced by Paul Rosenbaum and Donald Rubin (1983), PSM estimates the ATE from observational data by imputing the missing potential outcome for each subject using an average of the outcomes of similar subjects that receive the other treatment level. Similarity between subjects is based on estimated treatment probabilities, known as propensity scores. There are two steps in implementation:

First, the propensity score for each observation in both public and private sectors is to be obtained. The logit or probit models can be used for this purpose, with "working in public sector" as the dependent variable and a variety of exogenous variables (such as sex, age, education, industry, etc.) as the independent variables. The predicted probability of working in public sector is used as the propensity score.

The second step is to choose the algorithm of matching individuals in the two groups. The most popular approach is to use calliper matching. That is to find the two individuals with closest propensity scores in the two sectors, but only use them when the difference of propensity scores is within the pre-set calliper range. In most studies, a convention of 0.01 is used as the range within which matches are valid.

The range of propensity scores with both "control group" (those working in private sector) and "treated group" (those working in public sector) is referred to as "common support". While there are cases where the range of propensity scores for the control group is different from that of the treated group, it is called "off common support". Only matches within the common support are used to estimate the treatment effects.

3.3.2 *Nearest-Neighbour Matching (NNM)*

Similarly, NNM also estimates the ATE by imputing the missing potential outcome for each subject using an average of the outcomes of similar subjects that receive the other treatment level. The difference is measured by a weighted function of the covariates for each observation. The NNM method determines the "nearest" by the Mahalanobis distance, in which the weights are based on the inverse of the covariates' variance-covariance matrix **V**:

$$Distance_{Mahalanobis} = \sqrt{(\mathbf{x}-\boldsymbol{\mu})' \mathbf{V}^{-1} (\mathbf{x}-\boldsymbol{\mu})} \text{ , where } \boldsymbol{\mu} \text{ is the mean vector of } \mathbf{x}.$$

It includes the Euclidean distance as a special case where **V** is an identity matrix. After matching criterion is estimated, the rest of the NNM method is the same as PSM. In some statistical software, NNM is incorporated into PSM, but NNM does not require estimating a choice model before generating the matching criterion.

3.4 Type 4: Multiple-Equation-Regression Model

The original purpose of this method is to solve the problem of selection bias problem, which can be interpreted as an omitted variable problem (Heckman, 1979). Heckman selection model is usually applied to construct a proxy for the omitted latent variable in order to solve the selection bias problem. Another popular method within multiple-equation framework is treatment effects model, which considers the effect of an endogenously chosen binary treatment, conditional on two sets of regressors. Other methods with simultaneous equations system also exist in literature, but Heckman selection model and treatment effects model are the most popular techniques.

3.4.1 *Heckman Selection Model*

The Heckman selection model is developed originally to address the selection bias problem due to unobserved factors affecting the probability of being "treated". If the individuals in the treated group share some specific characteristics, then the OLS estimates of the treatment effect as well as all the other slopes are biased. Heckman (1979) proposes a two-equation model to account for this selection bias.

To see this, the first stage of Heckman selection model is to estimate a "selection equation", with the selectivity variable (usually the "treatment" dummy) being determined by a set of exogenous regressors. Usually, probit is used because the error term is assumed to have a standard normal distribution. Following that, for each observation in the selected sample, compute the inverse Mills ratio λ_i (conditional probability), which is used to account for the selection bias in the outcome equation. Note that $\Phi(\cdot)$ is the cumulative distribution function of the

standard normal distribution and $\phi(\cdot)$ is the corresponding probability density function. The second stage is to run OLS on the original equation, including λ_i as an additional regressor.

$$\text{Selection Equation: } \Pr\left(D_i = 1 \middle| \mathbf{z}_i\right) = \Phi\left(\boldsymbol{\alpha}'\mathbf{z}_i + \varepsilon_{1i}\right) \Rightarrow \hat{\lambda}_i \equiv \frac{\phi\left(\hat{\boldsymbol{\alpha}}'\mathbf{z}_i\right)}{\Phi\left(\hat{\boldsymbol{\alpha}}'\mathbf{z}_i\right)}$$

$$\text{Outcome Equation: } y_i = \boldsymbol{\beta}'\mathbf{x}_i + \hat{\delta}\hat{\lambda}_i + \varepsilon_{2i}$$

The coefficient of λ_i is defined as $\delta = \rho\sigma$, where ρ is the correlation coefficient between the two error terms, and σ is the adjusted standard error of the outcome equation. That is to say, If the error term in the selection equation (ϵ_{1i}) is correlated with that in the outcome equation (ϵ_{2i}), then single equation OLS estimator is biased due to omitting the inverse Mill ratio. However, if the coefficient γ of selection bias (λ_i) is not significant, which implies a zero correlation coefficient (ρ), then there is no selection bias.

However, it may be useful for estimating the $\boldsymbol{\beta}$ of the model, but it is not very useful in estimating the PSWP, which is the ultimate purpose. This is because the key regressor D_i does not enter the wage determination equation at all. Note that, strictly speaking, the coefficient in front of $\hat{\lambda}_i$ cannot be interpreted as the wage premium, because $\hat{\lambda}_i$ means the propensity to work in the public sector, not the status of working in the public sector. Loosely speaking, nevertheless, if this propensity is equal to 1, then the individual works in the public sector with 100% probability, so we could interpret the coefficient δ as the counterpart of the "treatment effect". The difference here is that the treatment effect is the effect of a binary variable (D_i), while the coefficient δ here is the effect of a continuous variable (λ_i).

As shown later, the first stage of Heckman selection model might be exactly the same as that of PSM, i.e. a probit regression to obtain the fitted value of propensity for working in the public sector. However, the purposes of this step are different. In Heckman selection model, the first stage is used in the second regression to in order to remove selection bias problem. In contrast, in PSM, the fitted propensity is used as the matching criterion to link individuals in different sectors. The first stage of PSM only serves as a matching criterion, rather than a serious estimation of any causal relationship.

3.4.2 Endogenous Treatment Effects Model

Another multiple-equation-regression method is endogenous treatment effects model, because working in the public sector (the "treatment") is not randomly allocated. In other words, the "*conditional independence*" is violated[3]. That is, after conditioning on covariates, when no

---

[3] If the conditional independence assumption is satisfied, then the usual treatment effects model can be used, including the two matching-based methods (type 3) as special cases.

unobservable variable affects both treatment assignment and the potential outcomes, the potential outcomes are conditionally independent of the treatment. The intuition of the conditional independence is that only the covariates **x** affect both the treatment and the potential outcomes. Any other factors that affect the treatment must be independent of the treatment. This assumption is also known as "unconfoundedness" or "selection-on-observables" in the literature (Rosenbaum and Rubin, 1983; Heckman, 1997; Heckman and Navarro-Lozano, 2004; Cameron and Trivedi, 2005; Tsiatis, 2006; Angrist and Pischke, 2009; Imbens and Wooldridge, 2009; Wooldridge, 2010).

The outcome equation is augmented with a selection equation (which is usually a probit or logit model):

$$\text{Selection Equation: } \Pr\left(D_i = 1\right) = F\left(D_i^* \equiv \boldsymbol{\alpha}'\mathbf{z}_i + \varepsilon_{1i}\right)$$

$$\text{Outcome Equation: } y_i = \boldsymbol{\beta}'\mathbf{x}_i + \delta D_i + \varepsilon_{2i}$$

In the selection equation, D_i is the sector choice variable which is assumed to stem from an unobservable latent variable:

$$D_i = \begin{cases} 1, \text{ if } D_i^* > 0 \\ 0, \text{ otherwise} \end{cases}$$

The variance-covariance matrix for the two error terms ϵ_{1i} and ϵ_{2i} is:

$$\text{var}\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}$$

Different from Heckman selection model, the sector choice dummy is explicitly included in the outcome equation. The ATE and the other parameters of the model can be estimated by two ways: the consistent two-step estimator and full information maximum likelihood estimator. The likelihood function is given in Maddala (1983) and Greene (2012).

3.4.3   Simultaneous Equations Model

Another early attempt to correct for selection bias using multiple-equation method is Venti (1987). He develops a simultaneous equation model, with a job acceptance decision equation (demand), job offer decision equation (supply), as well as equations describing the probabilities of working in the two sectors. Maximum likelihood is applied to estimate this simultaneous equation model.

An innovative aspect of his method is that the labour supply does not need to equal the labour demand, i.e. there may be a queue waiting to work in public sector. This enables to measure

the degree of distortion in public sector wage, because there should be little queue if there is no PSWP.

Lamo et al (2005) and Afonso and Gomes (2014) use system-based estimation such as 3SLS to explain the interactions between wages in the two sectors, which are modelled as two separate equations. The drawback of simultaneous equation models is that, if there is a problem with one of the equations, the biasedness will be spilled over to the whole system.

4   Economic Modelling Methods

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.

It is noticed that there has already been some attempt to combine microeconomic model with the microeconometric techniques in the industrial organisation literature (e.g. Berry et al, 1995). However, they apply IV (or GMM) technique to estimate the structural microeconomic model. It can of course estimate the model by minimising the gap between model/population moments and the sample moments, but it CANNOT address a more fundamental question—is the model true? Any model can be estimated by any method (including GMM, ML or Bayesian), but how do we know the estimated model is a good model? GMM, ML or Bayesian methods can only test model against model, or specification against specification (e.g. t test, F test, likelihood ratio test, etc.). They cannot test the model against the data. This is the methodological advantage of indirect inference here. By using a widely accepted auxiliary model (linear regression here), we can directly measure the distance between the data and the model. If the distance between the model-simulated data behaviour is too far from the actual data behaviour, and this distance is quantified by a strictly defined Wald statistic.

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. Furthermore, one advantage of II is its flexibility. The auxiliary model needs not be correctly specified; when it is, II 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 ($\boldsymbol{\varepsilon}$):

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

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:

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

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 $f\left(\mathbf{y}^{(a)}, \mathbf{x}^{(a)}, \boldsymbol{\varepsilon}^{(a)}, \boldsymbol{\theta}_0\right) = 0$. 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:

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

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: $\text{Wald}(\boldsymbol{\theta}_0) \equiv \left(\boldsymbol{\vartheta}^{(a)} - \overline{\boldsymbol{\vartheta}}^{(s)}\right)'\left(Var[\boldsymbol{\vartheta}^{(s)}]\right)^{-1}\left(\boldsymbol{\vartheta}^{(a)} - \overline{\boldsymbol{\vartheta}}^{(s)}\right).$

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 $f\left(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}_0\right) = 0$ 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 3 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.
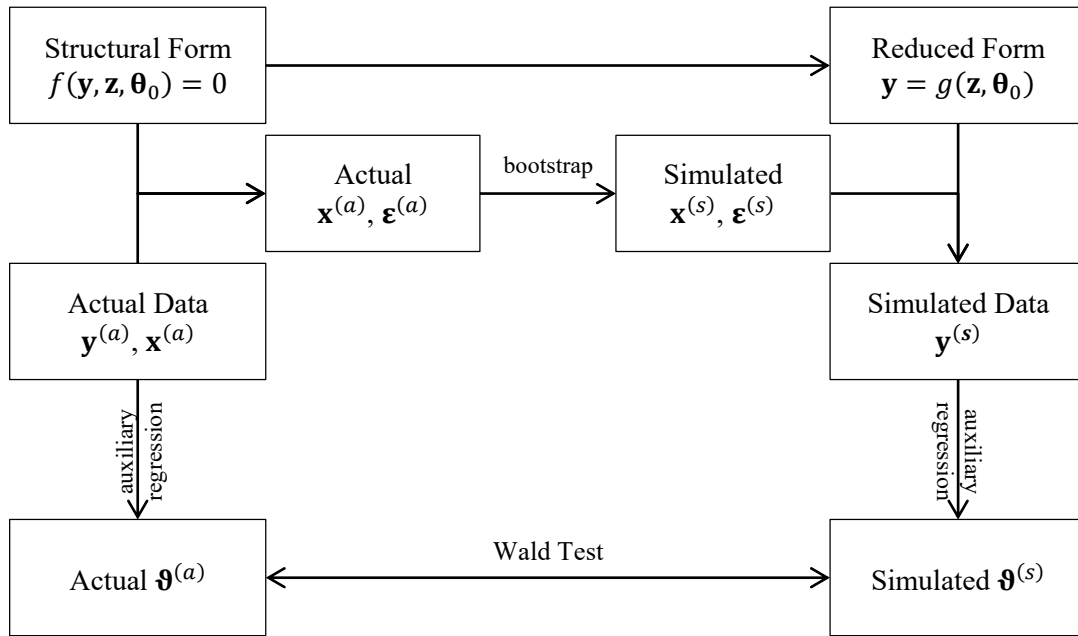
Figure 3 Flow Chart of Indirect Inference

4.2 Indirect Inference Estimation

Furthermore, as noted above, we implement the II test for a given calibration $\boldsymbol{\theta}_0$. 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 $\hat{\boldsymbol{\theta}}$, 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 $\hat{\boldsymbol{\theta}}$ is therefore the II estimation of the model parameter:

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

Note that the estimation here is a multivariate global optimisation problem, which has a stochastic and non-smooth objective function. It is usually impossible to derive the analytical solution for $\hat{\theta}$. 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.

CHAPTER II: THE DATA

This chapter 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.

1   The Sampling Design

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 4, 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.
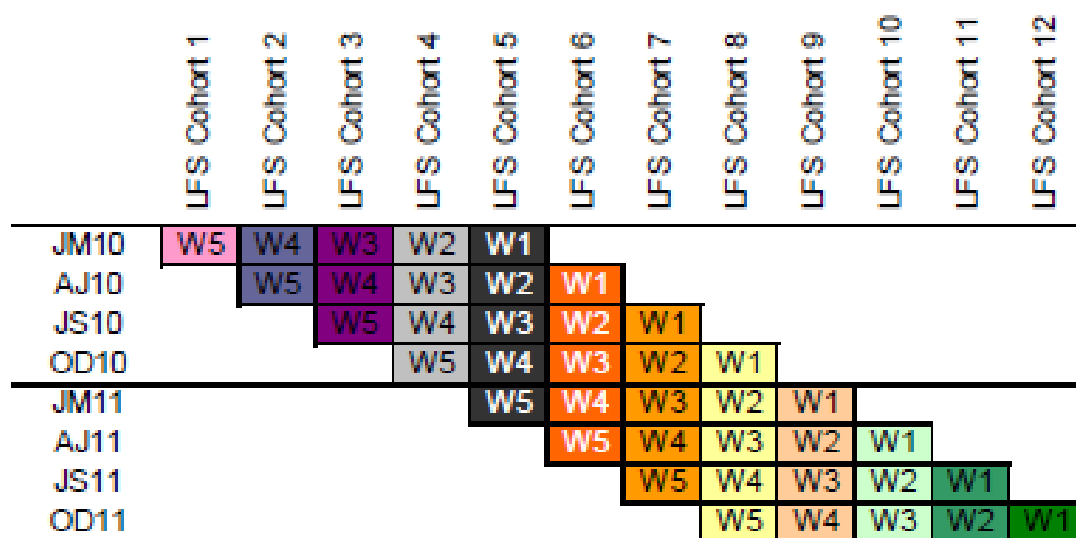
Figure 4 Illustration of Rotational Sampling Design (Source: ONS)

2 The Raw Wage Premium

Since the key variable to be explained is wage, Table 12 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 12 in the Appendix). The evolution of the raw PSWP is graphed in Figure 5. 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.
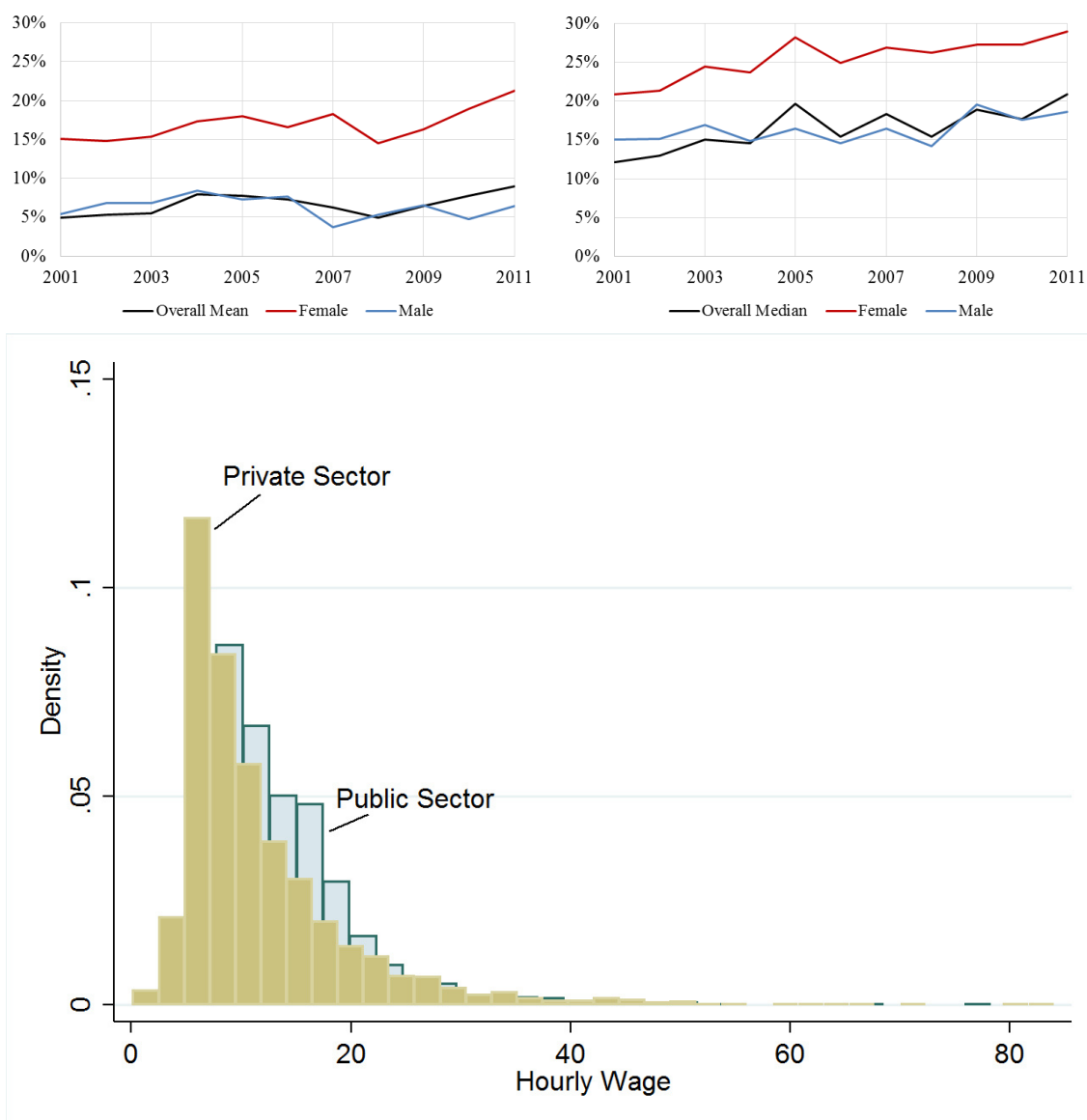
Figure 5 Raw Pay Differentials by Gender and Sector (Mean and Median)

3 The Working Hours

Another endogenous variable to be explained within the general equilibrium framework is the "quantity" of labour (working hour), which is determined with the "price" of labour (wage). According to Hansen (1985), there are two types of decision to be made: one is whether or not to work (the "extensive margin") and the other is how much to work (the "intensive margin"). In this study, we only focus on the latter, because all the data available are about the employed individuals.

As shown in Figure 6, there are four stylised facts of working hours consistent with the evidence identified by the existing literature.

(i)   The private sector workers tend to work longer hours than the public sector workers.
(ii)   Male workers are working longer hours than the female workers.
(iii)   There is a slight downward trend of working hours over time.
(iv)   The intensive margin is more stable than the extensive margin (measured by the unemployment rate, UR, on the right axis).
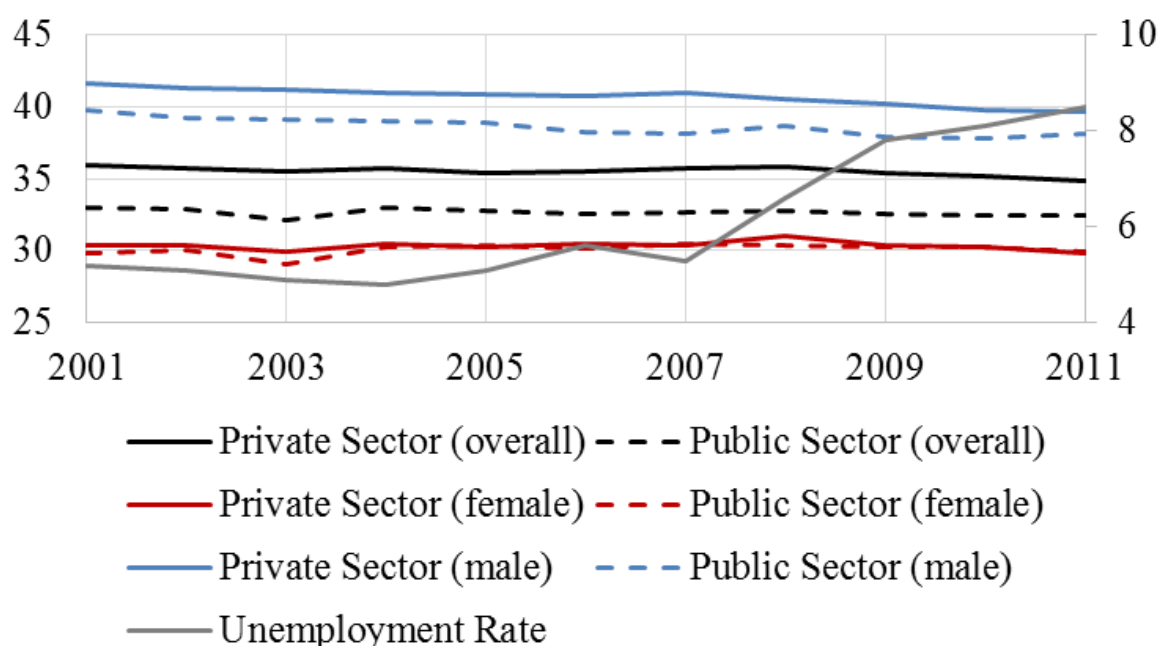
Figure 6 Working Hours by Gender and Sector (Mean)

Notes: The left axis is the working hours, and the right axis is the unemployment rate (percentage point).

The descriptive statistics are informative, but the evidence does not control for the differences between individuals, such as education, age, work experience, etc. That is why we need the econometric models and economic models to identify the effect of working in the public sector

to explain whether and why PSWP exists. The covariates to be used in econometric modelling and economic modelling are defined in Table 15 in the Appendix.

CHAPTER III: ECONOMETRIC MODELLING METHOD

This chapter applies all four types of econometric method reviewed in CHAPTER I to estimating the PSWP in the UK based on the LFS data described in CHAPTER II.

1 The Econometric Model of Wage Differential

To empirically estimate the PSWP, some (loosely linked) theoretical grounds are needed for the econometric model. There are two key variables to be explained in the current study, i.e. how wage is determined and how sector is chosen. For simplicity, we will use the Classical (as well as neoclassical) equilibrium theory to develop the econometric model of wage determination, and the consumer theory to develop the econometric model of sector choice.

Indeed, more sophisticated theories, such as wage bargaining (for wage determination) and search friction (for sector choice), can be used to provide a more realistic theoretical basis, but the main objective of this paper is methodological, and the neoclassical model still serves as the benchmark and contributes the most in explanatory power. Moreover, data is not available for incorporating the variables from those complicated theories into the econometric models. Thus, we will only focus on the two simple theories.

1.1 The Econometric Model of Wage Determination

According to the Classical (as well as neoclassical) labour economic theory, the equilibrium wage is determined by the interactions between the supply side (households or workers, who rationally maximise their utility) and the demand side (firms or employers, who rationally maximise their profit). A competitive labour market means all the individuals in the labour market have no wage setting power. All individuals in both parties are wage takers, not makers. The competitive equilibrium occurs when labour supply and labour demand are equal, i.e. market clears, generating a competitive wage and employment level.

As assumed in the Classical theory, if workers are mobile and entry and exit of workers to the labour market is free, then there would be a single wage paid to all workers. The allocation of workers to firms equating the wage to the value of marginal product is also the allocation that maximises national income (this is also known as allocative efficiency). The "single wage" property of a competitive equilibrium has important implications for economic efficiency. In a competitive equilibrium the wage equals the value of marginal product of labour as a result of optimisation behaviour. As firms and workers move to the sector that provides the best opportunities, they eliminate wage differentials between private and public sectors. Therefore, workers of given skills have the same value of marginal product of labour in all markets. The allocation of workers to firms that equates the value of marginal product across markets is also the sorting that leads to an efficient allocation of labour resources.

However, the single equilibrium wage in the labour market relies on the assumption that all jobs are alike and all workers are alike across sectors. This is not true in the reality. Usually, the existence of the public sector is to provide public goods and services, which would not be provided efficiently in the private sector markets due to market failure (e.g. imperfect competition, externalities and asymmetric information). This is why the "invisible hand" (the market force) should be complemented by the "visible hand" (the government).

As Adam Smith pointed out, job attributes (the demand side) and the individual characteristics (the supply side) will lead to different wages. It is not the monetary wage that is equated across jobs in a competitive market, but the "whole of the advantages and disadvantages" of the job (*The Wealth of Nations*, 1776). These two sets of variables act as shifting factors of the labour supply curve and labour demand curve, resulting in different wages.

As illustrated in Figure 7, assume that the overall market equilibrium is $\bar{w}$ where the aggregate labour supply curve (S) and the aggregate labour demand curve (D) meets. For different individuals and different jobs, there are some specific factors shifting the two curves to S_i and D_i, resulting in a specific equilibrium wage w_i for each individual/job.
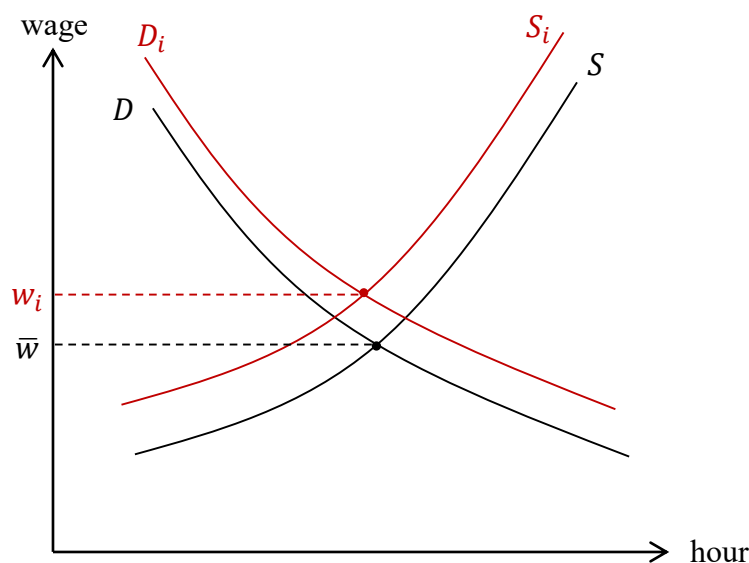
Figure 7 Wage Differentials due to Shifting Factors in the Labour Market

To see if there is any wage differentials between the public and private sectors, we need to take into account the two class of factors that contribute to the wage differentials apart from working in the public sector. Some of them are justified such as education level (individual characteristics on the supply side) and riskiness (job attributes on the demand side), but others may not be justified discriminations such as race and gender—if we believe that race and gender are not correlated with productivity.

The econometric model of the wage determination is implied from the neoclassical labour economic theory above. The dependent variable $\ln w_i$ (log of hourly wage) depends on the following two sets of regressors:

- **Individual Characteristics** (supply side): gender, race, marital status, sex orientation, age, age squared, migrant, work experience, work experience squared and education.
- **Job Attributes** (demand side): work mode (full-time or part-time), London dummy, industry dummies (SIC), occupation dummies (OCC), job type (manual work dummy) and *sector dummy*.

Note that the sector dummy ("whether the job is in the public sector") is the key regressor in the model, and the ultimate purpose is to estimate the effect of this dummy variable—the PSWP.

1.2 The Econometric Model of Sector Choice

In the competitive labour market, which sector an individual chooses to work for depends on which sector gives a higher utility. To be consistent with the neoclassical model in the previous subsection, it is straightforward to conclude that if all the individual characteristics and job attributes are exactly the same, then the individual should be indifferent between working in the public sector and the private sector, because the competitive equilibrium wage and working hours must be equalised, so is the utility which is a function of wage and working hours.

Therefore, any preference of one sector over the other must be due to the differences either in the individual characteristics or in the job attributes. Since the dependent variable to be explained is a dummy variable ("whether the job is in the public sector"), we can use either probit or logit function form. Given the similarity between the results, we will stick to logit in all the analysis below due to its lighter computational burden.

2 Results

In this section, we are going to apply all the aforementioned four types of methods in the existing literature to provide a robust estimate for the PSWP in the UK. Only two representative techniques within each type are used. Others are omitted either because they give similar results (e.g. Daymont-Andrisani decomposition, Terrell decomposition) or because some required data are not available (e.g. IV).

2.1 Type 1: Single-Equation-Regression Method

This type of method essentially uses the public sector dummy (D_i) to control for the sector wage differential in the intercept. Figure 8 shows the OLS estimates of the PSWP in contrast with the raw wage differential, and Figure 9 is based on the subsamples by gender.
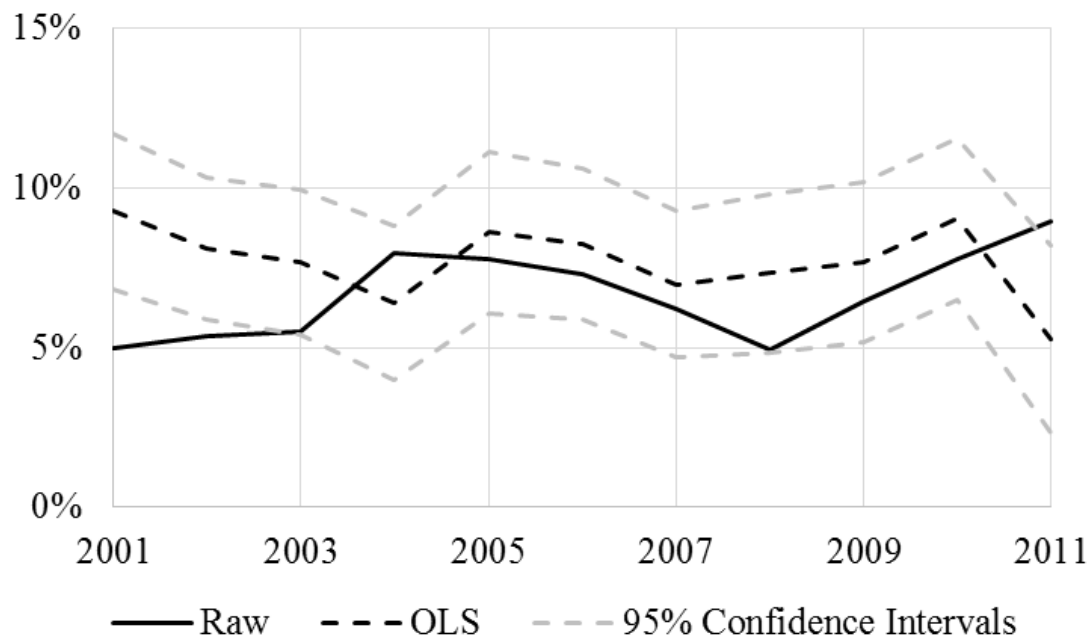
Figure 8 OLS Estimates of the PSWP (Pooled Sample)

The PSWP exists even after controlling for all the differences in individual characteristics and job attributes. This is true for both male and female workers, but female workers tend to enjoy a higher PSWP. This finding is consistent with the literature (Dolton and Makepeace, 2011; Blackaby et al, 2012). Over the first decade of the 21st century, the wage premium is quite stable. One interesting implication is that, although the raw wage differential suggests that working in the public sector is more and more privileged after the financial crisis (an increasing PSWP after 2008), the OLS estimates suggest a much weaker trend and even a drop in 2011.
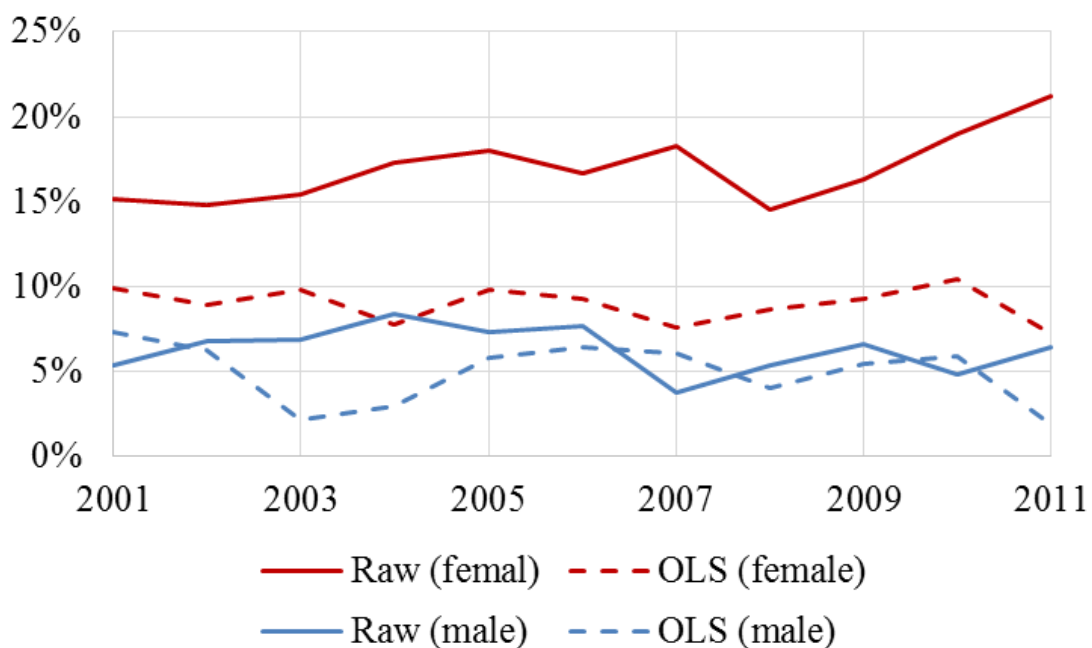
Figure 9 OLS Estimates of the PSWP by Gender
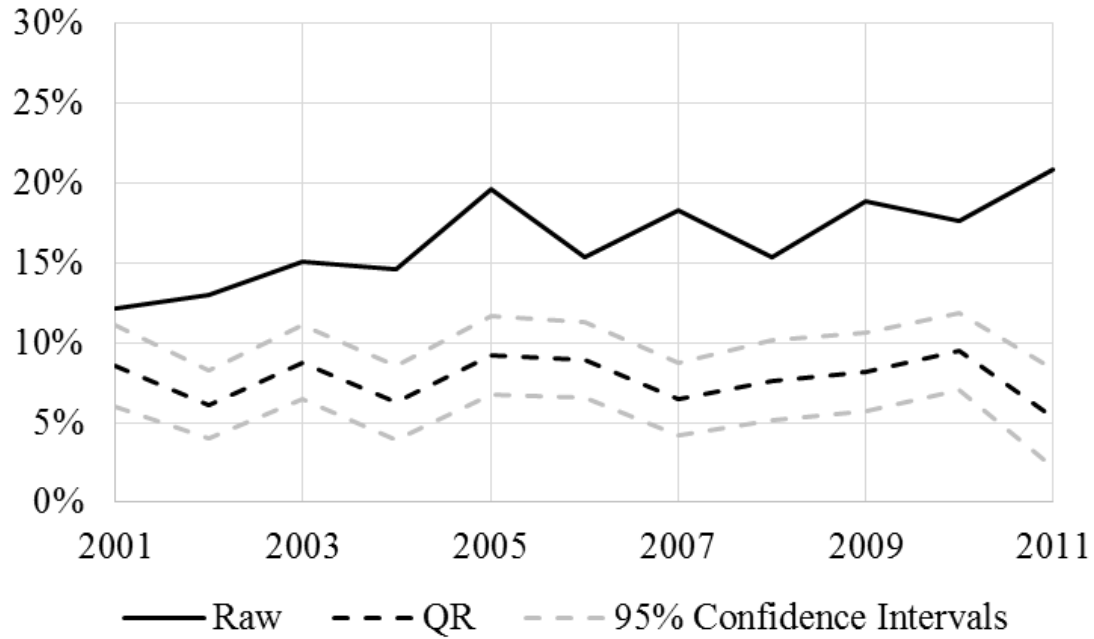
Figure 10 Quantile Regression Estimates of the PSWP (Pooled Sample)

The OLS tells the difference in mean, but it is usually criticised to be sensitive to outliers. As shown in the descriptive statistics in CHAPTER II, the distribution of wage is indeed quite skewed, and there are many very high and very low wages in the data. To deal with this problem, we also conduct the quantile regressions and contrast it with the median wage premium.
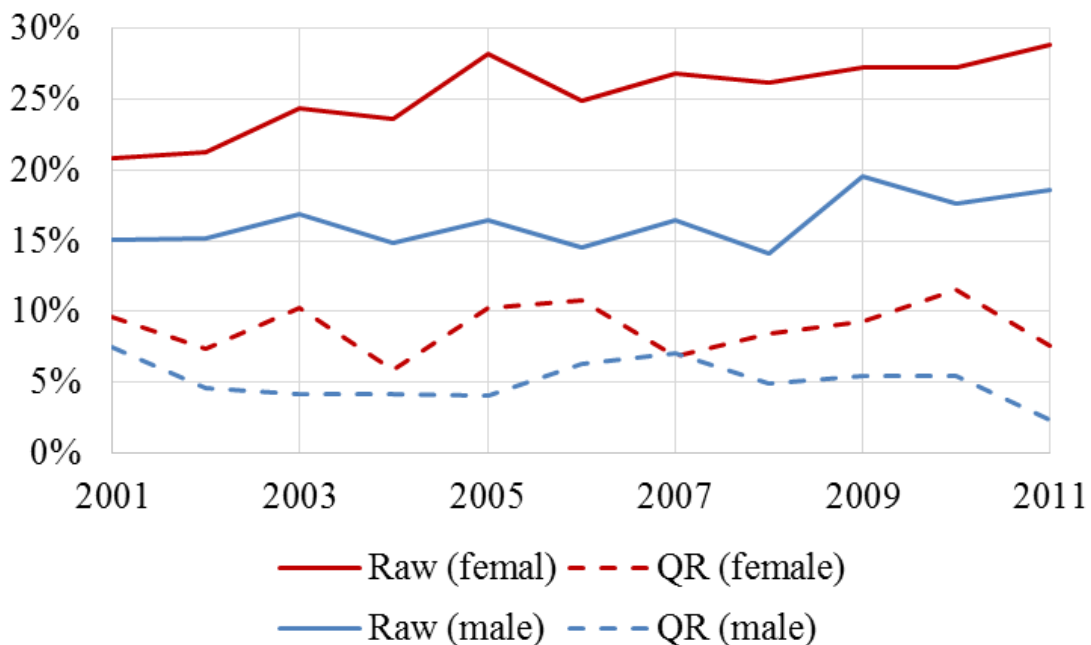
Figure 11 Quantile Regression Estimates of the PSWP by Gender

It is obvious that the median PSWP (around 16%) is remarkably higher than the mean PSWP (around 7%), but the quantile regression estimates at median are quite similar to those in OLS

(around 8%). This finding suggests that the median of sector wage differential is not a good measure of the average PSWP, but the quantile regression confirms the robustness of the OLS results. Also, the females benefit more from working in the public sector than males, but the estimated median PSWP is narrower than the estimated mean PSWP.

2.2 Type 2: Decomposition-Based Method

We use the same model specification to apply the Blinder-Oaxaca decomposition. The predicted PSWP estimated based on the two subsamples is decomposed into two components: (i) the component due to the differences in coefficients, and (ii) the component due to the differences in endowments.

As shown in Figure 12, the predicted PSWP is kept above 10% since 2001, and keeps rising after the financial crisis. The main reason for a rising PSWP is the rising contribution of the component due to differences in coefficients. The slight drop in the coefficient component in 2011 is in line with the findings in type 1 methods.
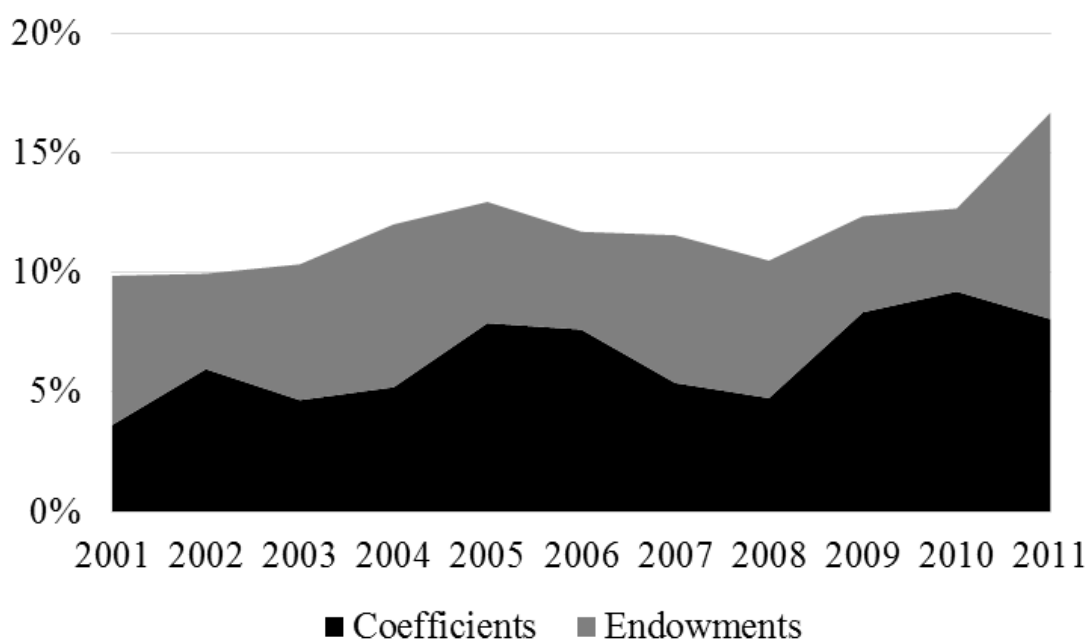
Figure 12 Blinder-Oaxaca Decomposition of PSWP

To be fully comparable to the type 1 methods in the last subsection, the JMP decomposition is also conducted based on the predicted median wage premium. The raw PSWP is therefore decomposed into three components: (i) the component due to the differences in observed "prices" (coefficients), (ii) the component due to the observed "quantities" (endowments), and (iii) the component due to the differences in unobserved prices or quantities.

Figure 13 shows the JMP decomposition on the median of the raw PSWP, which is again higher than the mean as before. However, the contribution of the differences in coefficients is significantly less than the evidence found based on the Blinder-Oaxaca decomposition, although the rising trend of this component is still maintained. Note that the unexplained component is not accounting for much of the wage premium.
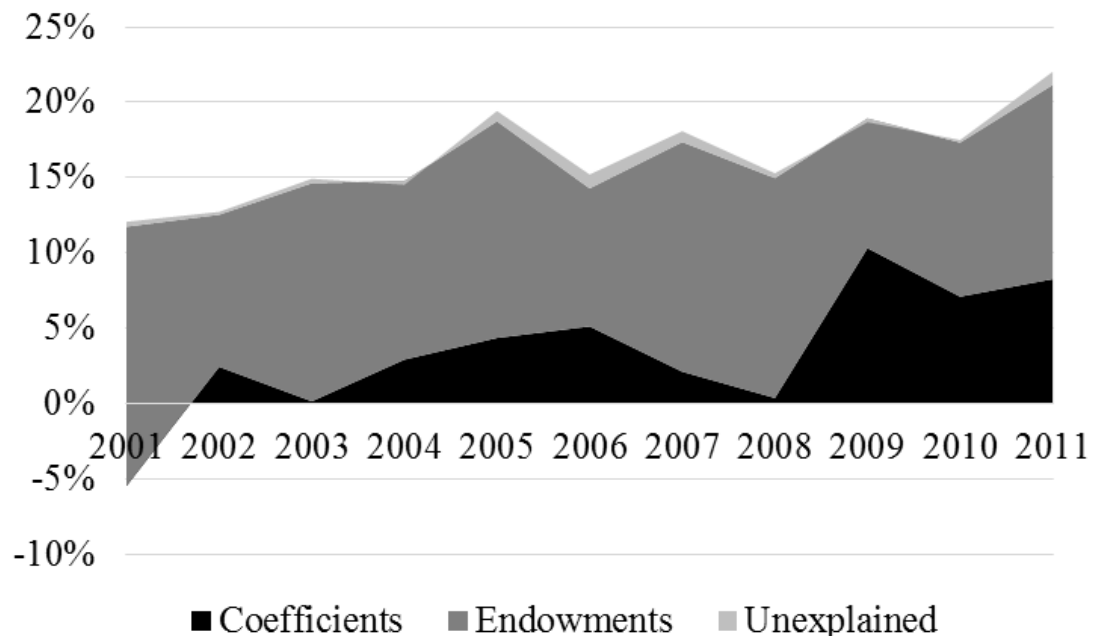
Figure 13 JMP Decomposition of PSWP

2.3 Type 3: Matching-Based Method

Different from the decomposition-based method whereby the wage determination equation is explicitly specified, the matching-based method instead spells out the sector choice equation. The sector choice equation is derived from the rational worker's utility maximisation problem, but the implied econometric model is not necessarily correct in specification. It is because the "treatment" equation (i.e. the sector choice equation) only serves as a matching criterion to link individuals with their counterparts in the other sector. The PSM method applied in this paper is based on Edwin and Babara (2003).

Figure 14 shows the results of the estimated average treatment effects (ATE) and those for the untreated (ATU) and the treated (ATT). The ATT tends to be higher than the ATU, which makes sense because whoever chooses to work in the public sector tends to be those who benefit the most. Those who could have earned more if they switch from private sector to public sector may suffer from the excess labour supply in the public sector labour market, similar to the "queuing model" in Venti (1987). This is actually contrary to the assumed Neoclassical model, where market always clear.
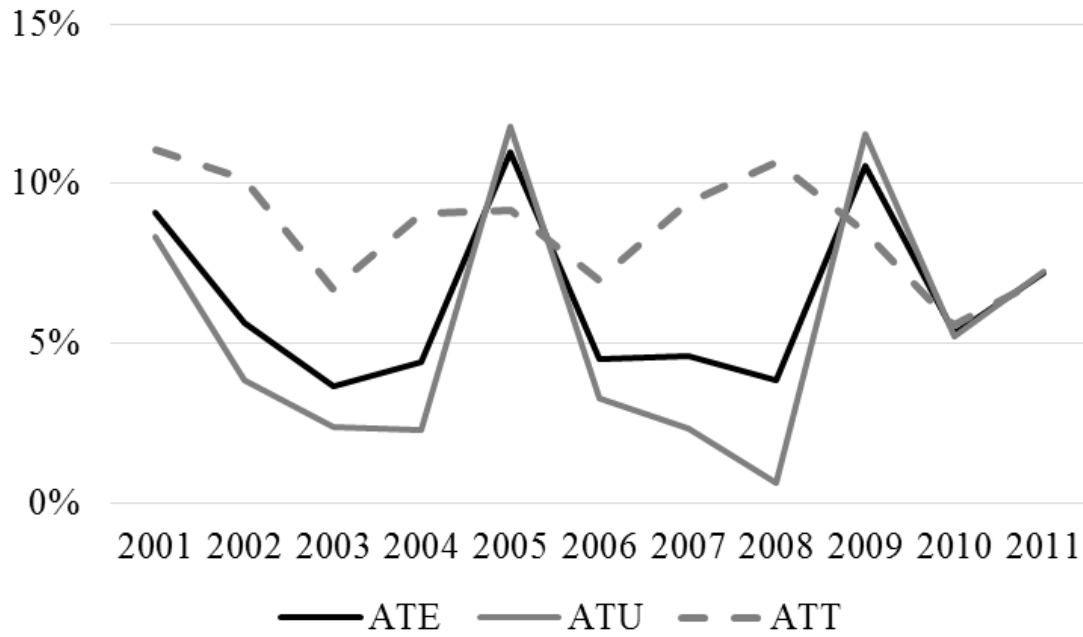
Figure 14 Estimated Treatment Effects using PSM
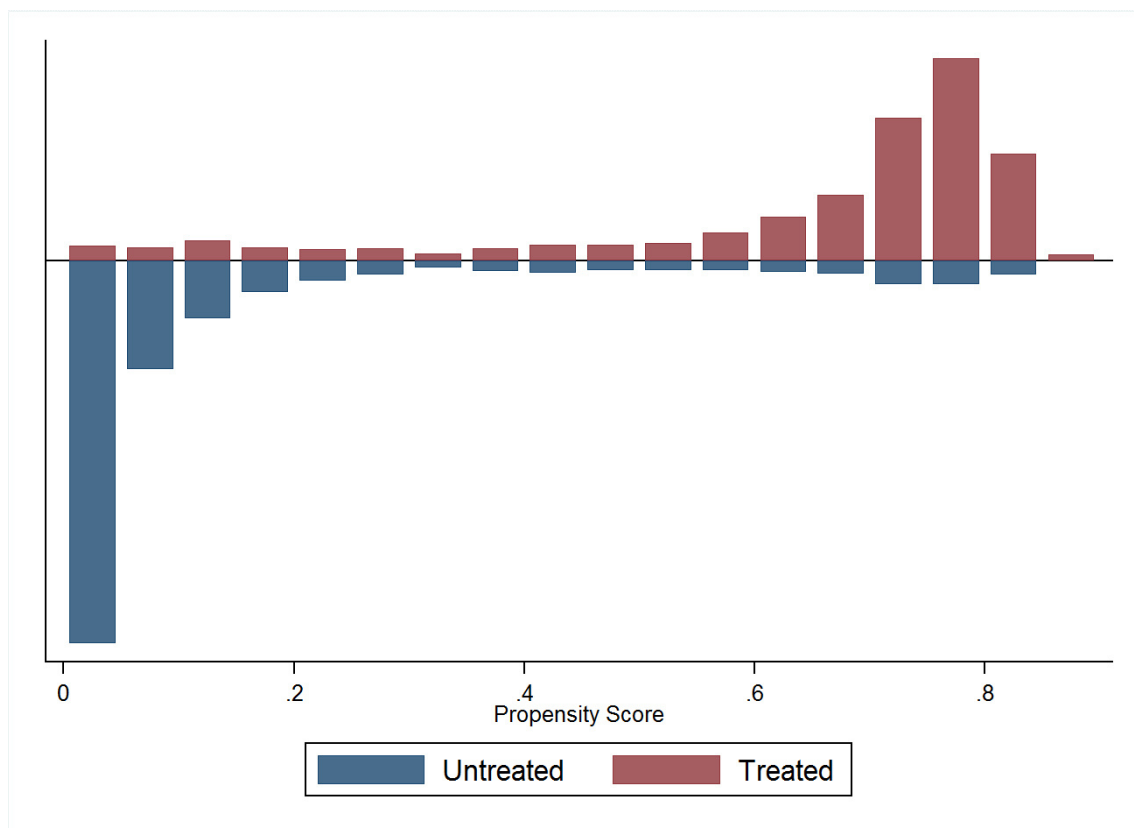
Figure 15 Distribution of Propensity Score (2011)

One problem with PSM is that the distributions of propensity score between public sector and private sector are substantially different. That causes the common support problem in matching. For example, Figure 15 graphs the propensity score histogram by treatment status for the year

2011. The distribution of propensity score for those who are working in public sector (the treated group) are more skewed towards 1, while the distribution for those who are working in private sector (the untreated group) are more skewed towards 0. Ramoni-Perazzi and Bellante (2006) also find that the data is too heterogeneous to be used to compare wages across sectors based on PSM.

As a robustness check, the NNM is also conducted and shown in Figure 16. The NNM method is very similar to PSM, so are the estimation results. Note that the ATU is even lower under NNM algorithm, and sometimes it is negative. That means people would be better off if they stay in the private sector, and this supports the rational optimisation behaviour in the sector choice model.

There is a common feature in both PSM and NNM—the ATU is especially high right after the financial crisis (i.e. 2009). This implies a nonpecuniary benefit of working in the public sector, the job security and wage stability. During the recession, the private sector tends to have higher chance to dismiss employees and those who stay usually face pay cuts. In the contrary, the PSWP tends to be lower during the booms, because private sector is more driven by profit maximisation and wage is more flexible due to incentive pay scheme. The PSWP is counter-cyclical and can be used as an indicator of the business cycles. The hypothesis of a negative correlation between the business cycles and the PSWP is formally tested in the next section.
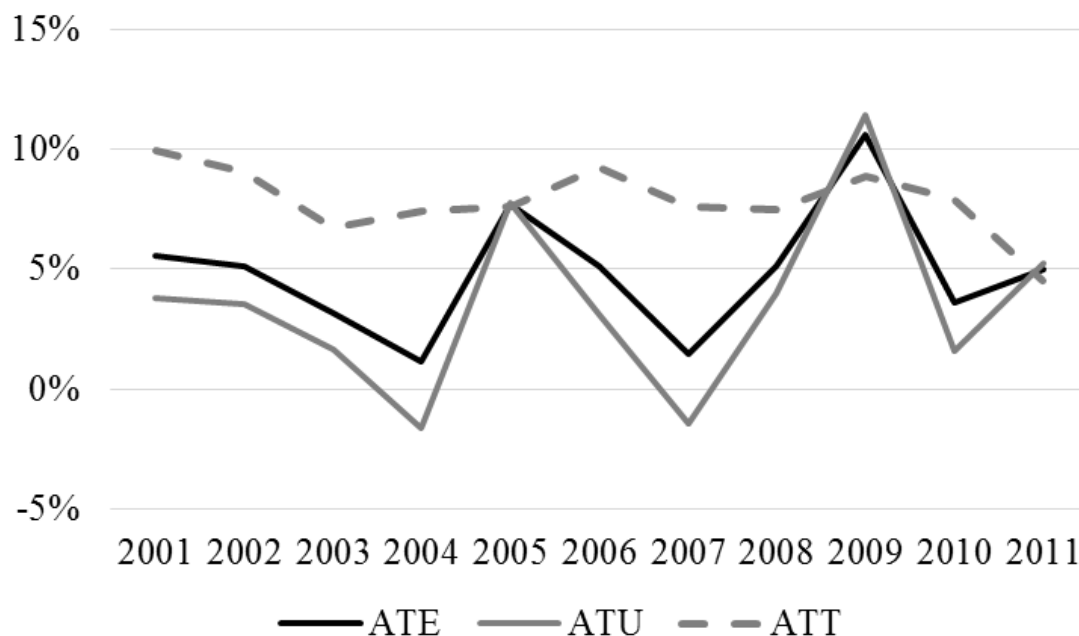
Figure 16 Estimated Treatment Effects using NNM

2.4   Type 4: Multiple-Equation-Regression Method

Following Gronau (1974), Lewis (1974) and Heckman (1976, 1979), the Heckman selection model becomes the most influential method to resolve the selection bias problem in microeconometric literature. This is essentially a multiple-equation-regression method, because it involves explicitly specifying both the wage determination equation (the "outcome equation") and the sector choice equation (the "selection equation"), so it combines the features of single-equation-regression method, decomposition-based method and the matching-based method into one system.

The selection equation is specified as a probit model using the econometric model developed in 1.2. After that, we calculate the inverse Mill's ratio (λ_i) according to the formula below, and then include it as an extra regressor in the wage determination equation.

$$\text{Selection Equation: } \Pr\left(D_i = 1 | \mathbf{z}_i\right) = \Phi\left(\boldsymbol{\alpha}'\mathbf{z}_i + \varepsilon_{1i}\right) \Rightarrow \hat{\lambda}_i \equiv \frac{\phi\left(\hat{\boldsymbol{\alpha}}'\mathbf{z}_i\right)}{\Phi\left(\hat{\boldsymbol{\alpha}}'\mathbf{z}_i\right)}$$

$$\text{Outcome Equation: } y_i = \boldsymbol{\beta}'\mathbf{x}_i + \hat{\delta}\hat{\lambda}_i + \varepsilon_{2i}$$

This is done for both pooled sample and the separate samples by gender, and the estimated coefficient $\hat{\delta}$ over time is shown in Figure 17. As mentioned previously, the inverse Mill's ratio is not exactly the same as the sector dummy D_i, and the interpretation of the coefficient $\hat{\delta}$ is not exactly the same as the public sector wage premium. The inclusion of $\hat{\lambda}_i$ is mainly to correct for the selection bias (or endogeneity bias due to the omitted variable), rather than estimating the PSWP. Nevertheless, the trend of $\hat{\delta}$ is similar to the estimated PSWP in other methods.
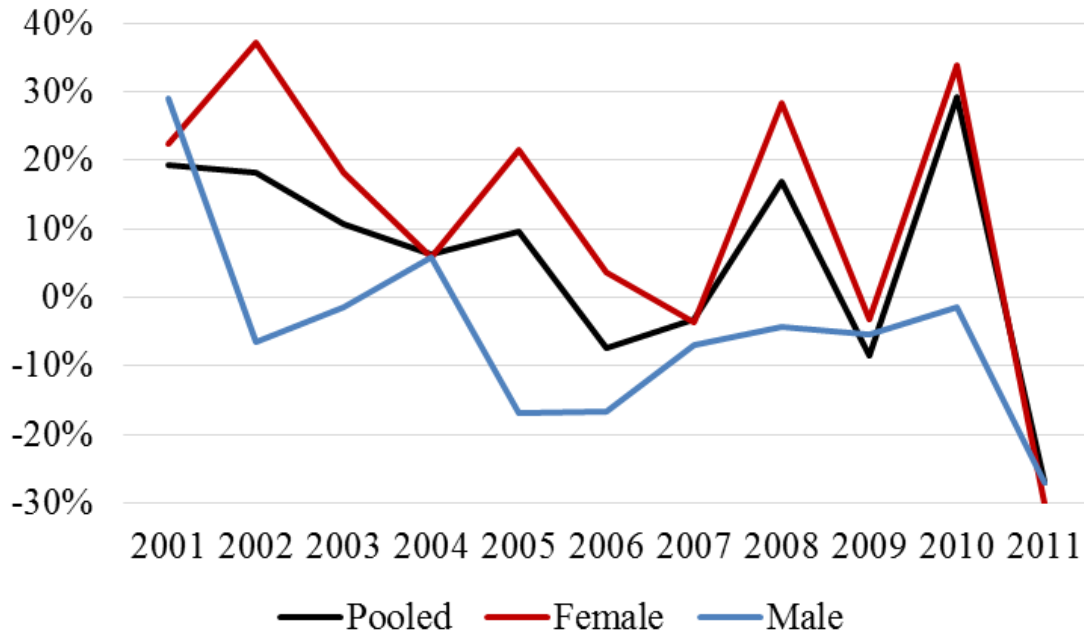
Figure 17 Estimated Coefficient of Inverse Mill's Ratio using Heckman Selection Model
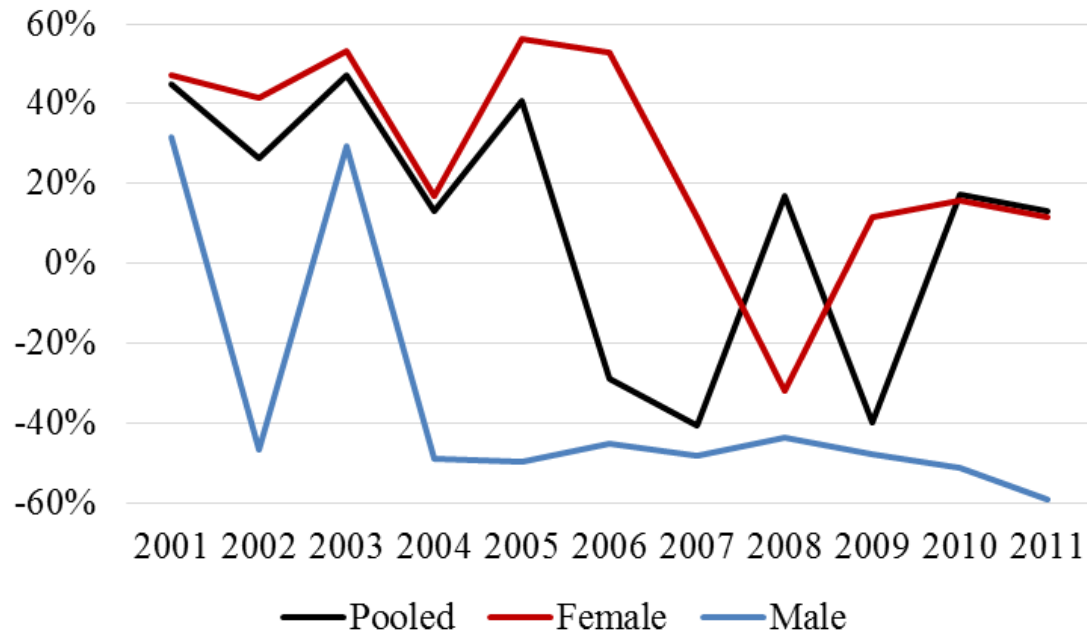
Figure 18 Estimated Treatment Effects using Endogenous Treatment Effects Model

Similarly, the endogenous treatment effects model is estimated using both pooled sample and separate sample by gender (Figure 18). The estimated ATE is more volatile than the other methods, but stylised fact that females enjoy a higher "treatment effect" (i.e. PSWP) is well maintained.

As argued in CHAPTER I, there is a vital drawback of the multiple-equation-regression method, i.e. any mis-specification in one of the equations will contaminate the whole system. The mis-specification bias may be more serious than the consequence due to selection bias. As a result, we only use type 4 method to provide a robustness check, without placing too much attention to the magnitudes of the estimates.

To smooth the estimates over time, we take average of the estimated PSWP over the 11 sample years (2001-2011). They are compared in the bar chart below. 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%.
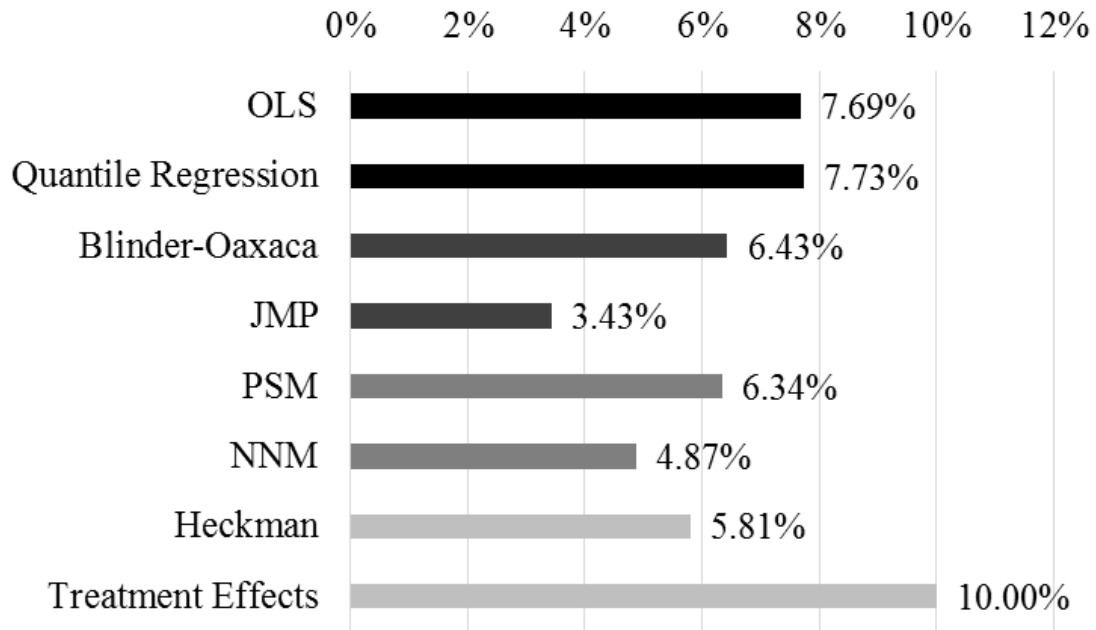
Figure 19 Summary of the Estimated PSWP (Average over Time)

3   From Microdata to Macrodata

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 21 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 conclusion confirms the conjecture: all estimated PSWP have negative correlation coefficients with the output gap.

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.
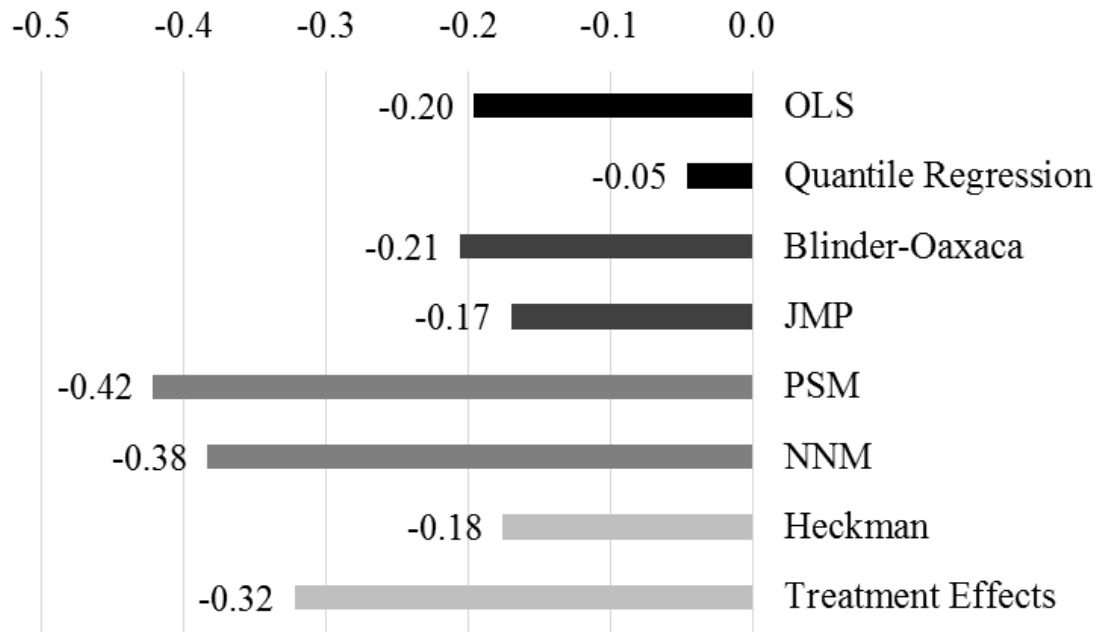
Figure 20 Correlation Coefficients between the PSWP and the Business Cycles
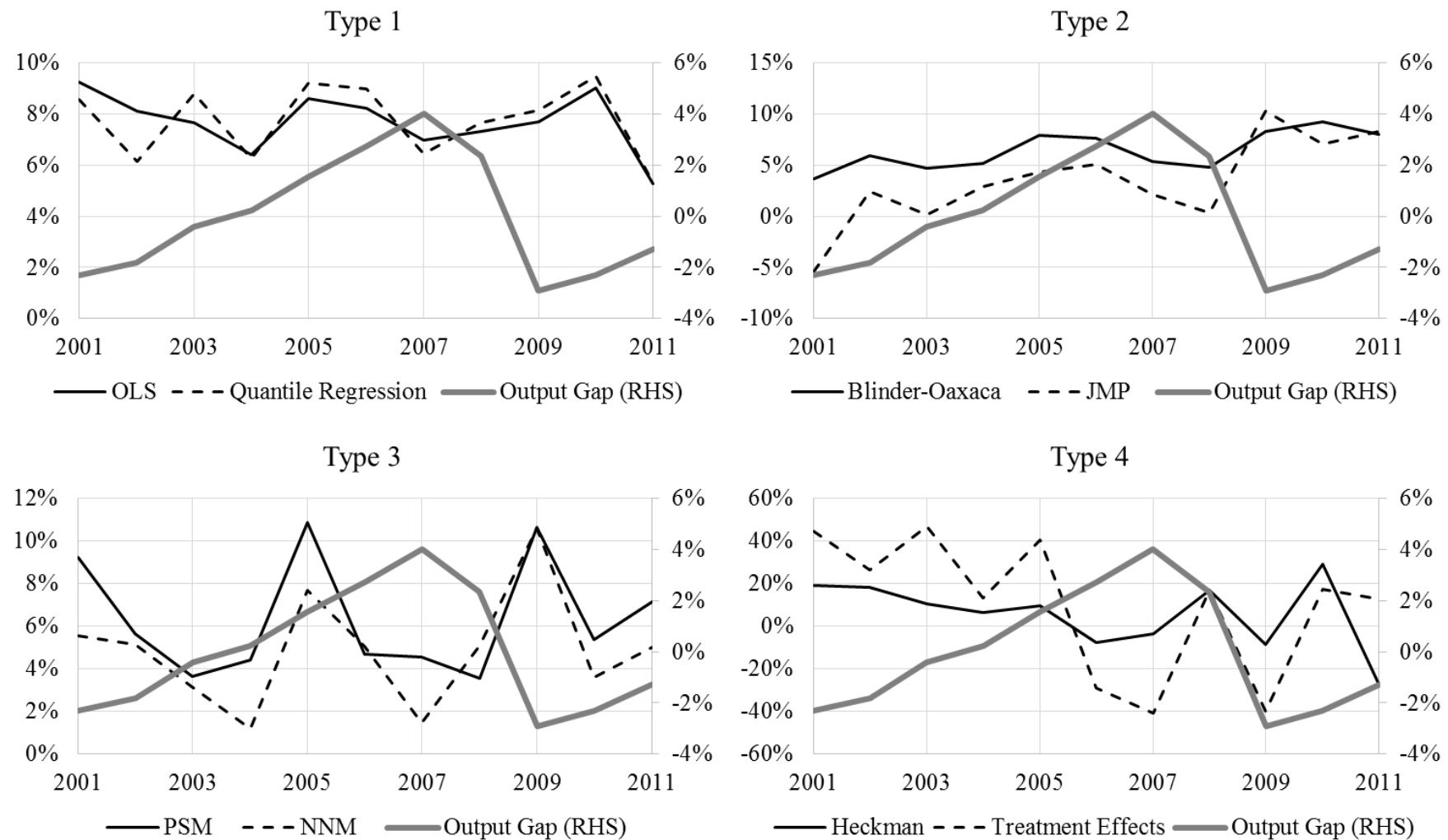
Figure 21 The Relationship between the PSWP and the Business Cycles

4   Conclusion

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 PSWP (an average estimate is around 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 the most stable and therefore more reliable estimates. The Blinder-Oaxaca decomposition 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.

In CHAPTER IV that follows, we are going to develop a new approach to testing and estimating the microfounded theoretical economic model. It is also capable of testing if there is selection bias or endogeneity bias problem within the same methodology.

CHAPTER IV: ECONOMIC MODELLING METHODS

This chapter applies the II tests and estimation described in CHAPTER I, built on the econometric modelling methods conducted in CHAPTER III.

1 The Model

The model is microfounded 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).

1.1 The Supply Side

The representative worker faces the following standard optimisation problem:

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

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

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

For simplicity, the utility function is assumed to 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 α. The budget constraint is expressed in real terms, so wL is real wage income. The time endowment T is allocated between leisure X and labour L.

If the two constraints are substituted into the utility function, then the optimisation problem with constraints become one without, and only one control variable is left:

$$\max_{L} U\left(L\right)=\left[\left(wL\right)^{\frac{s-1}{s}} +\alpha\left(T-L\right)^{\frac{s-1}{s}} \right]^{\frac{s}{s-1}}$$

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 homogenous. 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. When labour market clears, the marginal condition above only holds for the very marginal ones, and the other individual workers enjoy some positive surplus over the market-agreed wages. Thus, the wages and hours we observed in the data do not exactly satisfy the marginal condition for the representative worker. Instead, the derived marginal condition of wage above is like a lower bound, and the actual condition for each individual should be an inequality:

$$w_i \geq \alpha \left(\frac{w_i L_i}{T - L_i} \right)^{\frac{1}{s}}$$

To make the condition an equation, a supply-side surplus term S_i is needed to include the effects of individual characteristics:

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

The surplus term S_i can be interpreted as an exogenous shock or error term similar to those in a macroeconomic DSGE model. Usually, these shocks are not identically and independently distributed (IID) and are assumed to have an ARMA structure to generate a white noise innovation (Smets and Wouters, 2003, 2007). In a similar logic, we can break this exogenous supply shock S_i into a deterministic component capturing the differences in individual characteristics and a stochastic component ε_i^S, supposedly to be IID:

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

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

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

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 $\mathbf{\eta}_S$ is the coefficient vector of each term inside $\mathbf{ind}_i$. The innovation term ε_i^S is supposed to be an IID random variable under the null hypothesis (there is no selection bias, or equivalently there is no endogeneity bias), so ε_i^S is uncorrelated with the terms of $\mathbf{ind}_i$.

1.2 The Demand Side

A representative firm faces the following standard optimisation problem:

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

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

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.[4] Substitute the constraint into the objective function, and the optimisation problem becomes:

$$\max_{L} \pi = AL^{\gamma} - wL$$

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 we have homogenous jobs. In reality, job attributes, such as industry, sector, occupation, work mode and location, are all different across jobs. 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. When labour market clears, the marginal condition above only holds for the very marginal ones, and the other firms enjoy some positive surplus over the market-agreed wages. Therefore, the derived marginal condition of wage on the demand side is like an upper bound, and the actual condition for each job should be an inequality:

$$w_i \leq \gamma AL_i^{\gamma-1}$$

To make the condition an equation, a demand-side surplus term (D_i) is needed to include the effects of job attributes:

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

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

---

[4] To see how capital is implicitly included in the model, note that the production function $Y = AL^{\gamma}$ can be treated as a short-run case with fixed capital (K) of $Y = \tilde{A}L^{\gamma}K^{1-\gamma}$. If $K = \bar{K}$, the production function becomes $Y = (\tilde{A}\bar{K}^{1-\gamma})L^{\gamma}$ with $A \equiv \tilde{A}\bar{K}^{1-\gamma}$.

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

Here, $\mathbf{job}_i$ is a vector of job attributes, such as industry, sector, occupation, work mode and location, and $\boldsymbol{\eta}_D$ is the coefficient vector of each term of $\mathbf{job}_i$. In particular, 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:

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

1.3 Market Equilibrium

If the labour market clears, the supply of a particular sort of labour L_i is equal to the demand for it. To summarise, equation (1) and equation (3) describe the equilibrium[5].

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

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 (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. Let's term them as "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 with respect to individual characteristics", and that of $\boldsymbol{\eta}^D$ is the "elasticities of demand-side surplus with respect to job attributes". In contrast, the $\boldsymbol{\beta}$ in the econometric models in the last CHAPTER are the elasticities of wage. The surplus is only a part of wage. Accordingly, there are two innovations (regression error terms), ε_i^S and ε_i^D, respectively describing the idiosyncratic disturbances on the supply-side surplus and demand-side surplus. Again, they are different from the error terms in the regressions. In fact, the error term should be a function of the two innovations.

Another point to be stressed is that the "intensive margin" must not to be confused with the "extensive margin" (Hansen, 1985). In our model, the focus is on the former, i.e. the working

---

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

hours, rather than the participation decision—whether or not to work at all. This feature of our model on "intensive margin" is desirable because it matches the microdata at hand.

1.4 Solution Method

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

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

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

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 (see the illustration of Figure 22). 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.
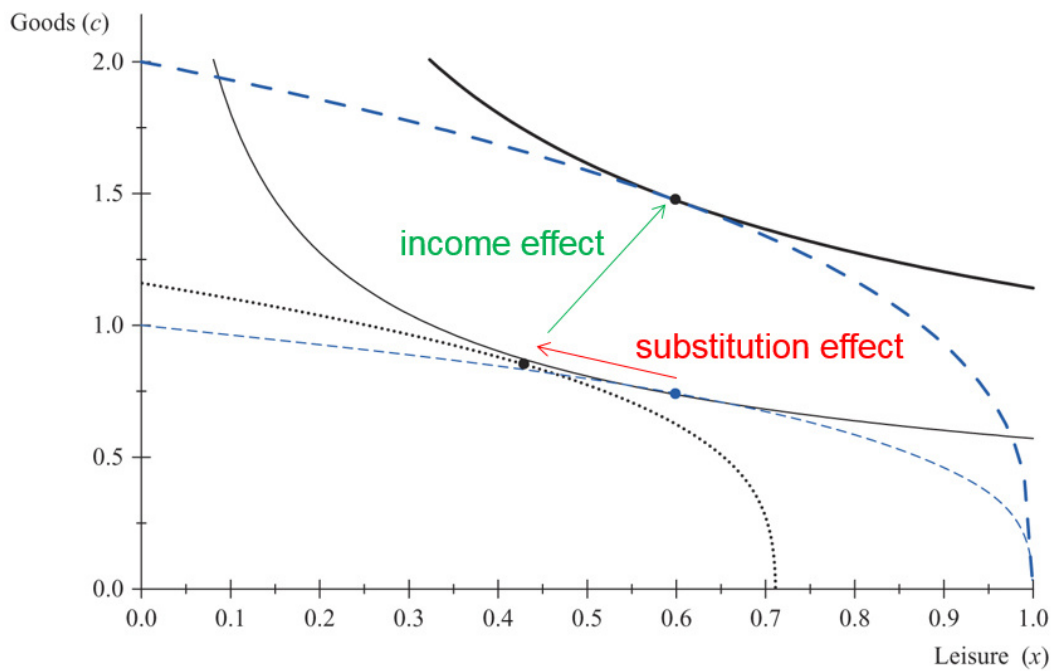
Figure 22 The Perfect Offset between Income Effect and Substitution Effect ($s = 1$)

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.

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

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

Alternatively, we can linearise the equation system around some point and then solve the linear equation analytically. A straightforward choice for the expansion point is the average wage of the whole sample, on the basis that the individual equilibrium should not be too far away from the population equilibrium.
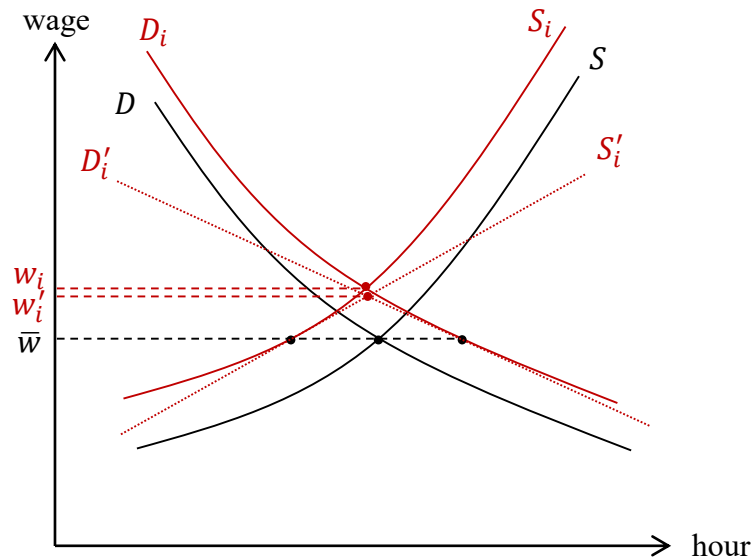
Figure 23 Linear Approximation of the Equilibrium

Figure 23 illustrates the linear approximation of the solution of the nonlinear equation system, built on Figure 7. 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 $\overline{w}$, 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 $\overline{w}$, 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.

2   Application of Indirect Inference

The structural form of the model $f(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) = 0$ is just the two equations derived from the marginal conditions (1) and (3), or equivalently the condensed equation (5) or the linearised form of it. Using the notation introduced in CHAPTER I, the endogenous variables $\mathbf{y}_i$ are w_i and L_i, while the exogenous variables $\mathbf{z}_i$ include both S_i and D_i, which in turn depend on the conditioning variables $\mathbf{x}_i$ ($\mathbf{ind}_i$ and $\mathbf{job}_i$) and the two IID innovations $\boldsymbol{\varepsilon}$ (ε_i^S and ε_i^D). The structural parameter $\boldsymbol{\theta}$ comprises α, s, γ and A. In principle, the shock structure parameters, i.e. $\boldsymbol{\eta}_S$, $\boldsymbol{\eta}_D$, σ_S and σ_S, are also part of $\boldsymbol{\theta}$. To reduce the dimensionality of the parameters to be estimated, we set them to the OLS estimates.

The data is 11 years' cross-section. For simplicity, we only focus on the latest year (2011), but the same procedure can be used for all years. The common sample size includes 6,216 observations with a mean wage of £12 per hour and a mean working hour of 34 hours a week.

2.1 Calibration

To initiate the II test and estimation, we need to calibrate the parameters either using the literature convention 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[6] α can be set at 0.5 and the constant elasticity of substitution s can be set at 0.5 to allow for greater complementarity than substitutability between consumption and leisure. The income share of labour in the production function γ is usually estimated to be 0.6~0.8 in the macroeconomic literature (e.g. Smets and Wouters, 2007), so we set it as 0.7. Finally, the total factor productivity A can be calculated from the firm's marginal condition and the known parameters and average values of the endogenous variables:

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

The calibrated structural parameters give the initial values $\boldsymbol{\theta}_0 = [0.5; 0.5; 0.7; 49]$ of the vector $\boldsymbol{\theta} = [\alpha; s; \gamma; A]$.

A warning over this calibration strategy is due here. The microdata may exhibit very different parameter values from those implied from the macrodata, especially when the aggregate macroeconomic data cover the whole economy, while our microdata are heavily concentrated in the service sectors. Therefore, this calibration practice is only done 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.

2.2 The Hypotheses

The null hypothesis (H0) is that the economic model is the true data generating process. Under this null, we have two further possibilities[7]:

- 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:

---

[6] This is based on a textbook written by Gillman (2011).
[7] See appendix for the details of the indirect inference test on selection bias.

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

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

It would be straightforward if we have concluded a false model or a true model with no selection bias. The tricky one is the case of true model with selection bias. In this case, some omitted conditioning variables are correlated with the error terms, either the specification of the model needs to be revised (e.g. Heckman) or the estimation method needs to be updated (e.g. IV).

2.3   The Auxiliary Model

In the application of II procedure, the choice of auxiliary model is vital for II. As mentioned in CHAPTER I, the auxiliary model does not have to be correctly specified, because it is merely a way of summarising the data feature. However, if it is correctly specified, then the estimates are consistent and asymptotically converging the maximum likelihood estimates (Gourieroux et al, 1993). To build connections with the previous CHAPTER, 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 ($\boldsymbol{\vartheta}$) is respectively:

- 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.

It is implied by a series of studies that the power of the II test is lower when the dimension of the auxiliary parameter vector is smaller (Meenagh et al, 2009; Minford et al 2015). In our case, type 1 and type 4 have very big number 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 are arguably more powerful than type 2 and type 3, each of which only has two auxiliary parameters. However, type 2 and type 3 provide a more detailed and informative way of analysing the PSWP, but type 1 and type 4 models place too much weights on irrelevant information of the data other than PSWP (only one of the 35 coefficients). Thus, there seems to be a trade-off between the power of the test and the usefulness of the test.

To accommodate this trade-off, we propose an eclectic approach in order to maintain a relatively high power of the II test while raising the importance of PSWP in the auxiliary parameter vector. We name it Grouped OLS (GOLS) or type 5, but it is actually a variant of type 1. This

is done by grouping the 35 coefficients of OLS regression into 8 categories, one of which is the PSWP. The details of the grouping is shown in Table 3.

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

Table 3 The Grouped Auxiliary Parameters of Type 5 Auxiliary Model

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 greatly reduced to emphasise the focus of the model—the PSWP, with a mild sacrifice of the power. A Monte Carlo simulation can be done in the future to quantify the decrease in power of the test due to this reduction in the dimensionality from 35 to 8, but this is beyond the scope of this thesis.

2.4   Procedure

Following CHAPTER I and the discussion above, the technical procedures of II test in this context are summarised as follows:

Step 1: Given the initial parameter values $\boldsymbol{\theta}_0$ and the actual observed endogenous variables $\mathbf{y}_i \equiv [w_i; L_i]$, calculate the exogenous variables/shocks $\mathbf{z}_i \equiv [S_i; D_i]$ for each observation $i = 1, \dots, N$ based on the structural equation (1) and (3), or just (5).

Step 2: Regress the calculated $\mathbf{z}_i$ on the conditioning variables $\mathbf{x}_i$ and obtain the elasticities vectors $\boldsymbol{\eta}_S$ and $\boldsymbol{\eta}_D$ as well as the innovations $\boldsymbol{\varepsilon}_i \equiv \left[\varepsilon_i^S; \varepsilon_i^D\right]$ for each observation i. Then we have the actual conditioning variable vector $(\mathbf{x}^{(a)})$ and innovation vector $(\boldsymbol{\varepsilon}^{(a)})$ based on equation (2) and (4).

Step 3: Bootstrap $S = 1000$ sets of both $\mathbf{x}^{(s)}$ and $\boldsymbol{\varepsilon}^{(s)}$ from the actual counterparts $(\mathbf{x}^{(a)}$ and $\boldsymbol{\varepsilon}^{(a)})$ under both H0a and H0b, where $s = 1, \dots, S$. Under H0a, the bootstrap is done independently on $\mathbf{x}^{(a)}$ and $\boldsymbol{\varepsilon}^{(a)}$, while under H0b, the bootstrap is done by bundling[8] the public sector dummy with the innovation for each observation.

Step 4: Run auxiliary regression (type 1-type 5) for the $1 + S$ sets of data, including 1 actual dataset and S simulated datasets.

Step 5: Calculate the Wald statistics and p-values for both H0a and H0b, based on which the conclusion is drawn.

The whole II test procedure is dependent of the structural parameter vector $\boldsymbol{\theta}_0$. The auxiliary parameters are $\boldsymbol{\vartheta}^{(a)}$ and $\boldsymbol{\vartheta}^{(s)}$, with the latter also depending on $\boldsymbol{\theta}_0$. Therefore the Wald statistics basically depend on $\boldsymbol{\theta}_0$. For a different set of values of $\boldsymbol{\theta}$, there might be different conclusions. Thus, to push the explanation power of the model to the limit, we use global optimisation algorithms to search for the best parameter values $\widehat{\boldsymbol{\theta}}$ which can minimise the Wald statistics under both H0a and H0b.

Note that the computational costs become substantial if any step during the S simulation involves numerical solution. This is indeed the case for both propensity score matching and

---

[8] By bundling, we are effectively bootstrapping the vector of $[\mathbf{x}; \boldsymbol{\varepsilon}]$, rather than $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ separately. In this way, we actually impose correlations between them, indicating the existence of selection bias (endogeneity).

Heckman selection model, where a generalised linear regression model has to be estimated using maximum likelihood via numerical optimisation. For each simulation (out of 1000), there is one Newton-Raphson conducted to maximise the nonlinear likelihood function.

This is marginally acceptable[9] for II testing, but it would be impossibly costly for II estimation. Therefore, we will omit Type 3 and Type 4 in the next section when we try to estimate the four structural parameters using the global optimisation algorithm, which usually involves thousands of tests (if we are lucky to achieve convergence by then).

3   II Test Results

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). However, there are two structural innovations, and they might be correlated with each other in a joint distribution.

3.1   The Innovations

The extracted innovations from the structural equation are apparently jointly distributed as shown in Figure 24 (Figure 25 is the implied bivariate normal distribution). The estimated standard deviations of the two innovations are respectively $\sigma_S = 0.7475$ and $\sigma_D = 0.4204$, suggesting a much bigger heterogeneity on the supply side (workers) than that on the demand side (jobs). The correlation coefficient between the two innovations is 0.2183, which is significant at the 1% level. The implied variance-covariance matrix is:

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

A non-zero correlation means that, during the bootstrapping, the innovations need to be drawn jointly rather than independently, regardless of whether there is selection bias. Similarly, there are significant correlations between conditioning variables, and the bootstrapping 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.

An alternative resampling strategy is to use Monte Carlo simulation rather than bootstrapping. The advantage of Monte Carlo is that it can generate new possible observations while maintaining the dependence between variables. However, the problem is that it involves another

---

[9] On an average desktop, it takes about 20 minutes for each test of type 3 and type 4.

estimation procedure before the simulation, so there might be more errors accumulated. For example, if we compare Figure 24 and Figure 25, we find that the estimated multivariate normal distribution is "over-smoothed" compared to the observed distribution. Following the convention in the II literature, we adopt the bootstrapping resampling in this paper.
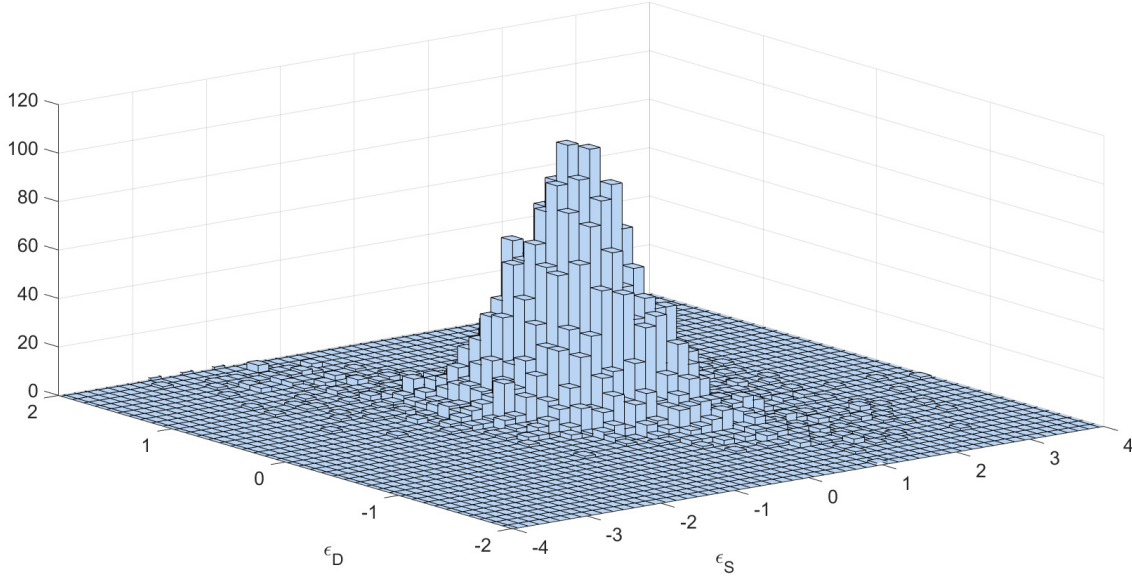
Figure 24 Joint Frequency Distribution of the Innovations
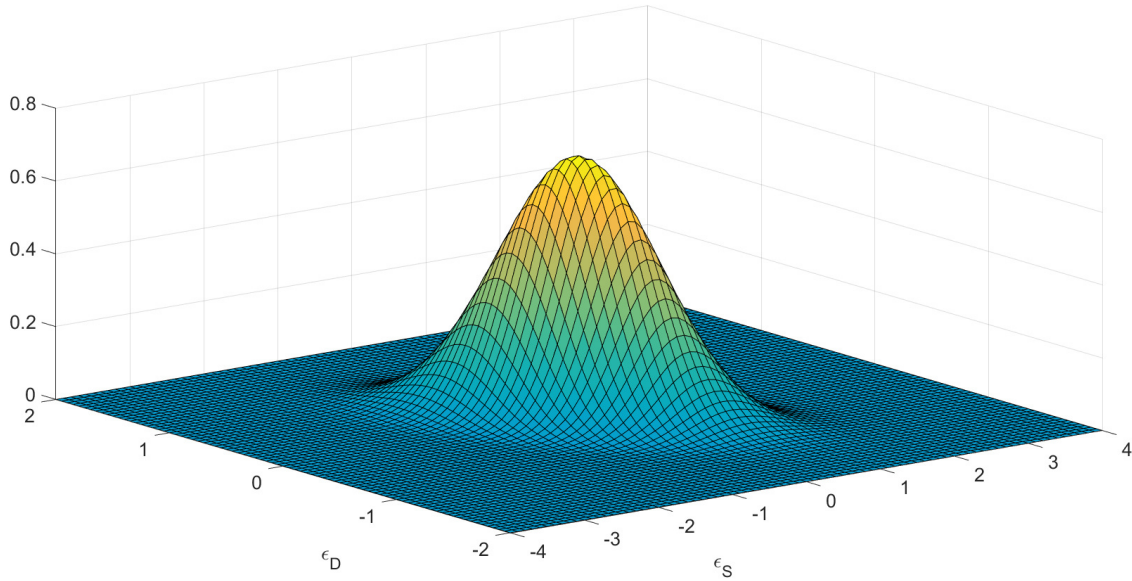
Figure 25 Implied Joint Normal Distribution of the Innovations

3.2 The Wald Statistics

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 χ^2 distribution with K degrees of

freedom, where K is the dimension of the auxiliary regression parameter vector ($\boldsymbol{\vartheta}$). 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 feature and the actual data feature 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 26 shows the distribution of the Wald statistics based on the S sets of simulated data using OLS as auxiliary regression. Unfortunately, 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 χ^2, but as the degree of freedom gets larger ($K = 35$ as in Figure 26), it converges in distribution to a normal distribution.
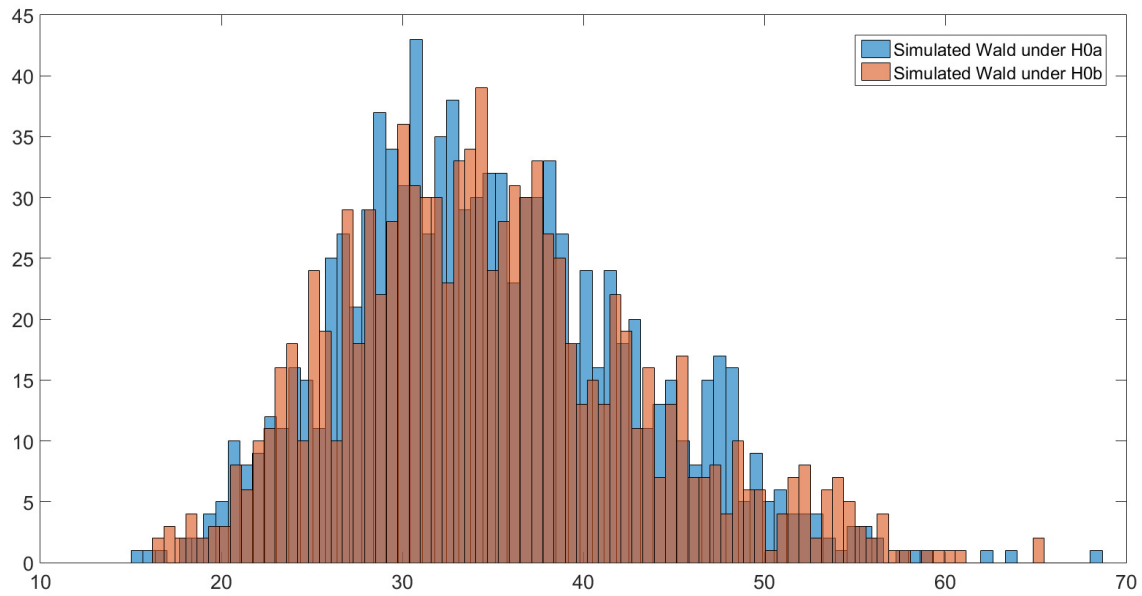
Figure 26 The Distribution of Simulated Wald Statistics (Type 1)

Among all the five auxiliary regressions, only the Blinder-Oaxaca decomposition (type 2) and the propensity score matching (type 3) can be visualised in a 3D graph, because there are only two auxiliary parameters estimated. In the case of BOD (type 2), the PSWP due to different coefficients and the pay differential due to different endowments, the joint distribution of which is shown in Figure 27. These two parameters are negatively correlated, and the actual auxiliary parameters ($Endowments\ Diff = 0.0862, Coefficients\ Diff = 0.0806$) 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 28), 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.
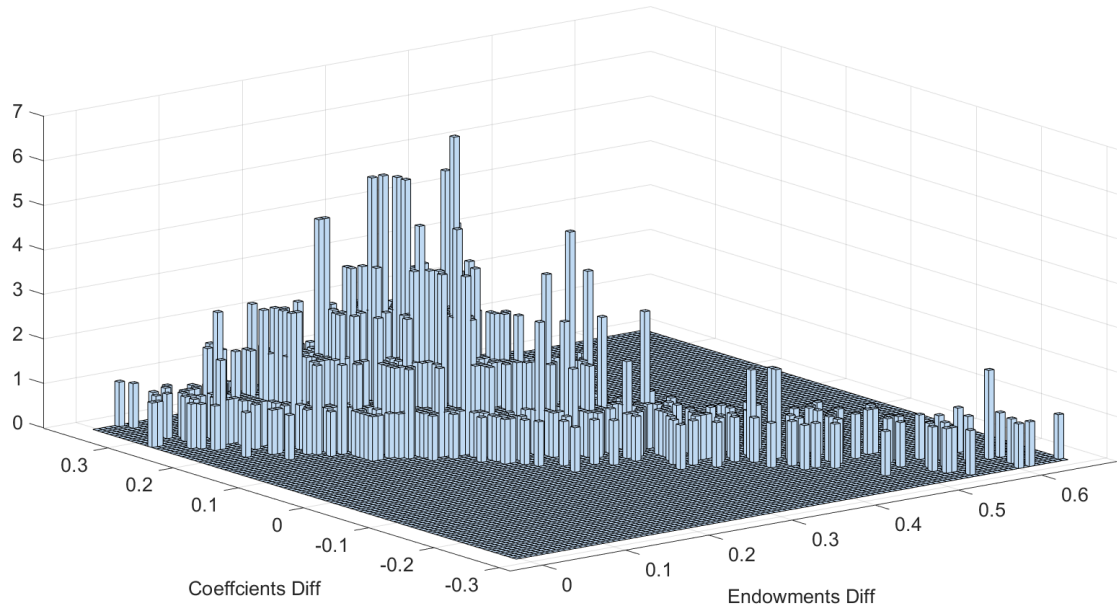
Figure 27 The Distribution of Wage Differentials following Blinder-Oaxaca Decomposition
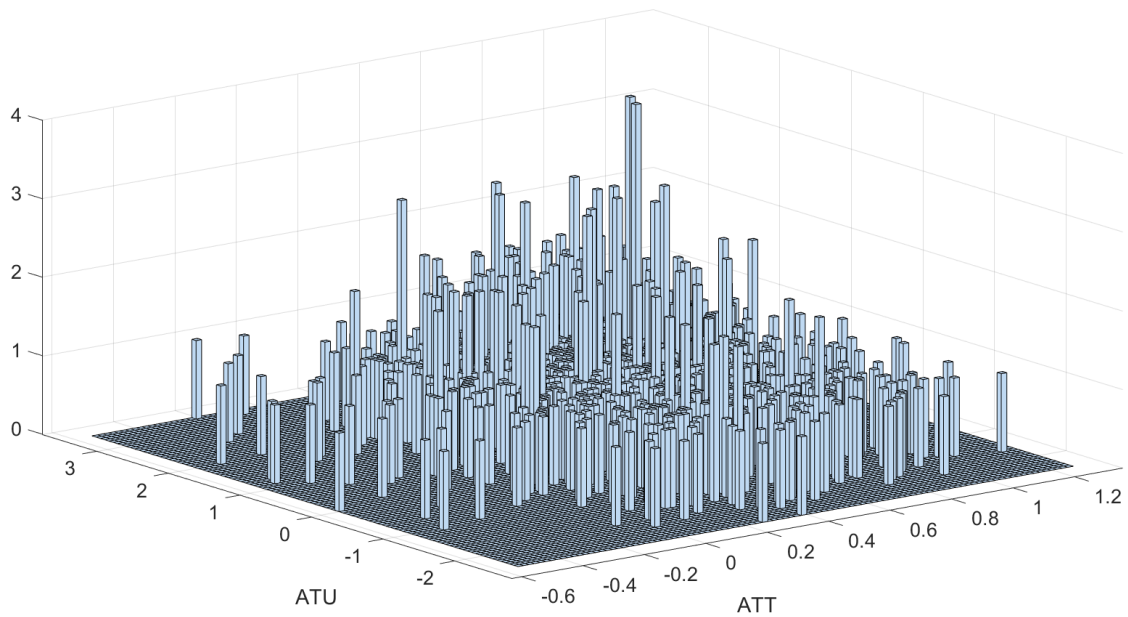
Figure 28 The Distribution of Wage Differentials following Propensity Score Matching

3.3 The Test Conclusions

Under the initial calibration $\boldsymbol{\theta}_0$, the actual Wald statistics, the associated P-values and the critical values at 5% significance level (C-values) are reported in Table 4.

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 4 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 H0a are smaller than H0b (with an exception for type 5). Second, the only type of auxiliary regression under which the model passes is PSM (type 3) under H0a. 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 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.

4 II Estimation Results

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 5.

Structural Parameters θ		Lower Bound	Upper Bound	Calibration	H0a	H0b
α	Leisure weight	0.1	10	0.50	0.4650	0.4710
s	Elasticity of Substitution	0.2	1.2	0.70	6.5479	1.1712
γ	Labour Share	0.6	0.95	0.70	0.9366	0.6036
A	Productivity	24.75	74.24	49.50	56.76	27.21
				Wald Statistic	4.69	84.25
				C-Value	15.07	15.69
				P-Value	79.02%	0.00%

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

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 $s \to 0$ the complementarity is greater while as $s \to \infty$ the substitutability is greater, with $s = 1$ being the Cobb-Douglas specification with equal degrees of complementarity and substitutability. The II estimate of s (6.55 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 $s \to 1$ (see Figure 22 for the details). 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 (α) 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 (γ) 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.

4.1 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 29). 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.
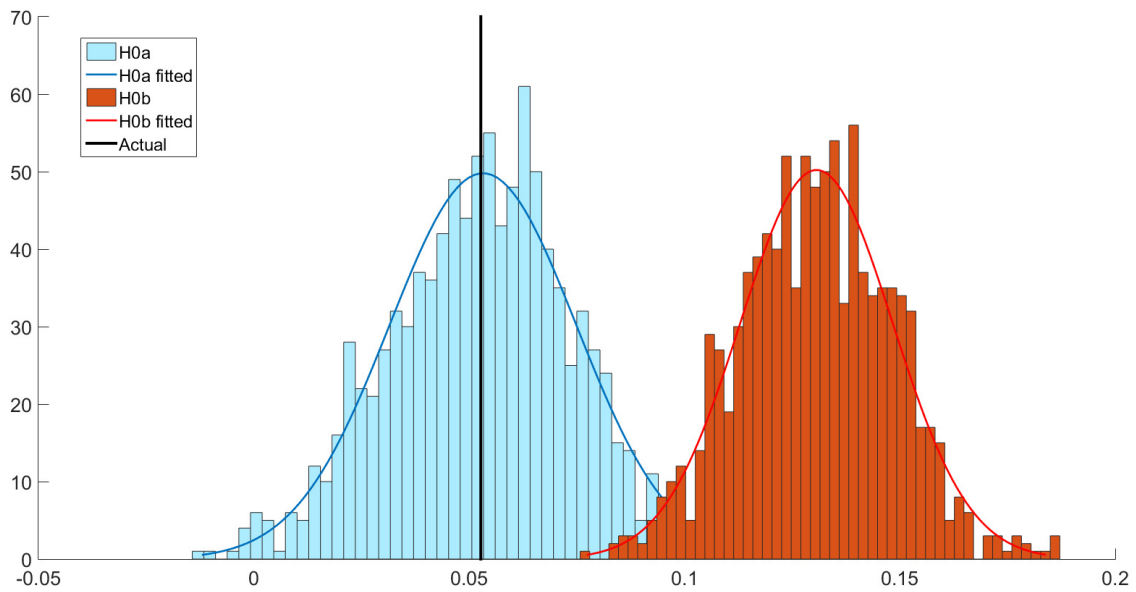
Figure 29 The PSWP of the Actual and Simulated Data (Single-Equation-Regression)

The estimated structural parameters are also used to conduct the II test for other types of auxiliary regressions. For example, Figure 30 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 31) 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.
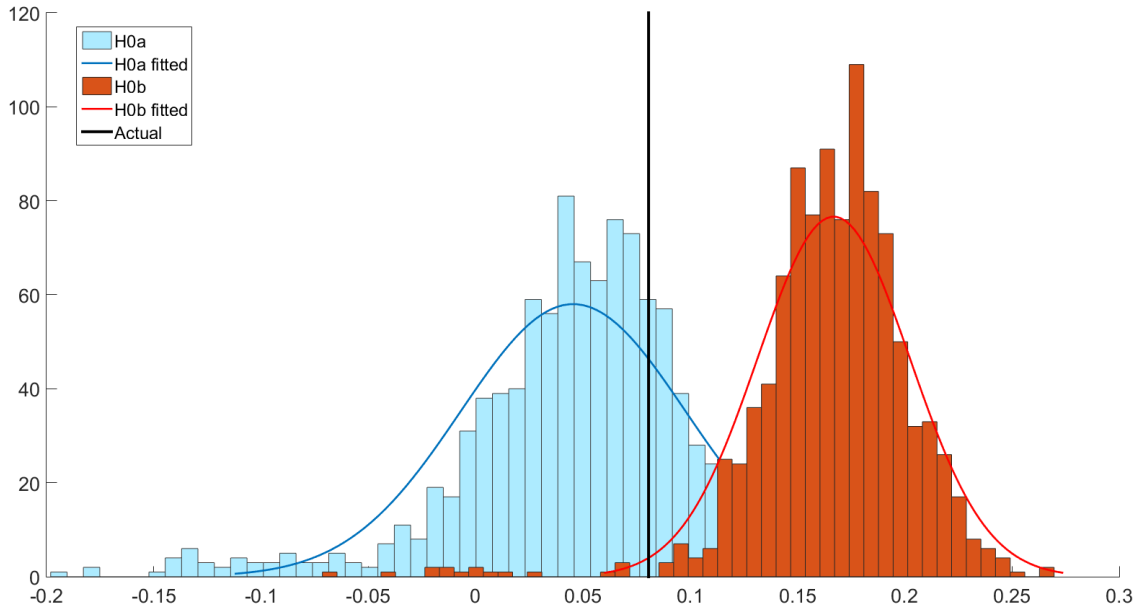
Figure 30 The PSWP of the Actual and Simulated Data (Decomposition-Based)
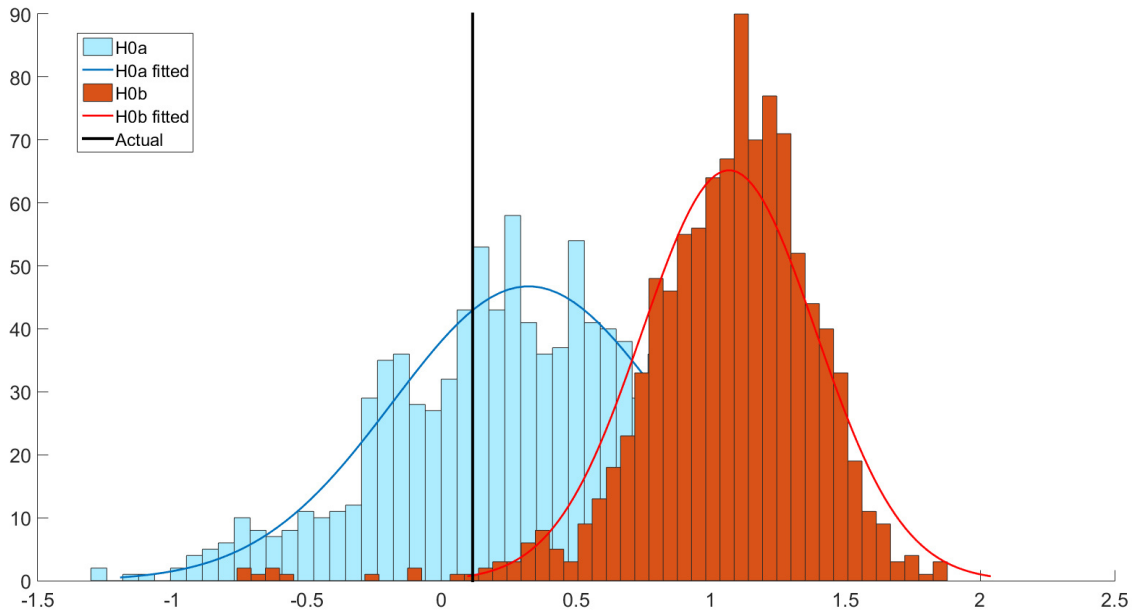
Figure 31 The PSWP of the Actual and Simulated Data (Matching-Based)

4.2 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. 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.

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

Table 6 Summary of the Estimated PSWP

Table 6 summarises and contrasts the different measures of PSWP from both reduced-form econometric models (auxiliary models) and the microfounded economic model for the year 2011. The economic model provides a quite robust estimate, lying in the middle of the estimates from various econometric models.

4.3 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 7.

For type 1, the resulting Wald statistics are halved, compared to the calibrated parameters (Table 4), 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: GOLS	4.69	15.07	79.02%	84.25	15.69	0.00%

Table 7 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.

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.

Auxiliary Parameters ϑ	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	IN	IN
white	IN	IN	IN	IN		
married	IN	IN	IN	IN		
homosexual	IN	IN	IN	IN		
age	IN	OUT	IN	OUT		
age^2	IN	OUT	IN	OUT		
migrant	IN	IN	IN	IN		
work experience	IN	OUT	IN	OUT	IN	OUT
work experience^2	IN	OUT	IN	OUT		
low education	IN	IN	IN	IN	IN	IN
GCSE	IN	IN	IN	IN		
A-level	IN	OUT	IN	IN		
higher education	IN	IN	IN	IN		
degree	IN	IN	IN	IN		
full time	OUT	OUT	OUT	OUT	IN	IN
London	IN	IN	IN	IN		
energy & water	IN	IN	IN	IN	IN	IN
manufacturing	IN	IN	IN	IN		
construction	IN	IN	IN	IN		
distribution	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	IN	OUT
technical	IN	OUT	IN	OUT		
administrative	IN	OUT	IN	OUT		
skilled trades	OUT	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		
OVERALL	**OUT**	**OUT**	**OUT**	**OUT**	**IN**	**OUT**

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

Auxiliary Parameters	Type 2: Decomposition			Type 3: Matching		
	ϑ	H0a	H0b	ϑ	H0a	H0b
Endow. Diff.	0.0862	IN	IN			
Coeff. Diff.	0.0806	IN	OUT			
ATT				0.2421	IN	OUT
ATU				-0.0120	IN	OUT
OVERALL		**IN**	**OUT**		**IN**	**OUT**

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

5 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). Le et al (2016) use Monte Carlo experiments to evaluate the power of II in the context of macroeconomic DSGE models, concluding that II test has a greater power than the classical likelihood ratio test. This section will adopt a similar 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 spot out 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)[10]. If the resulting p-value of a test is smaller than 5%, then we reject the model—the II test correctly spots out the false model and contributes to a higher power. Reversely, if the resulting p-value is greater than 5%, then we accept the model being true—the II test fails to spot out 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 32 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.

---

[10] 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.
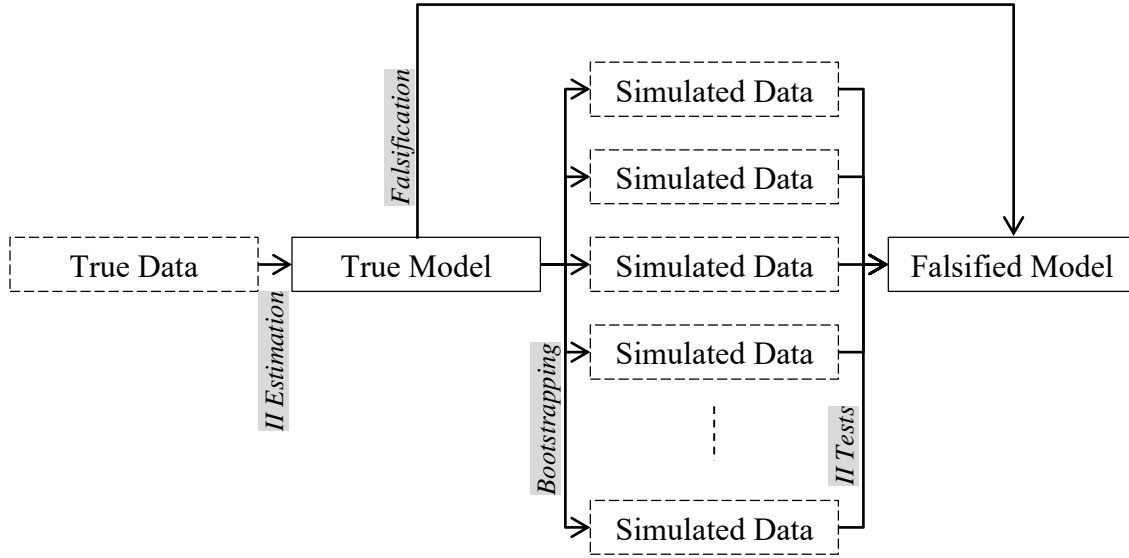
Figure 32 Illustration of the Monte Carlo Experiment

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 this difference does raise the powers of indirect inference test by more falsified and more restricted parameters. 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, α, γ, A) are falsified but the error parameters (η_S, η_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 10 and the distributions of p-values under the former method is shown in Figure 33 (the distributions of the latter method are simply clustered at 0).

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 10 The Simulated Powers of Indirect Inference Tests
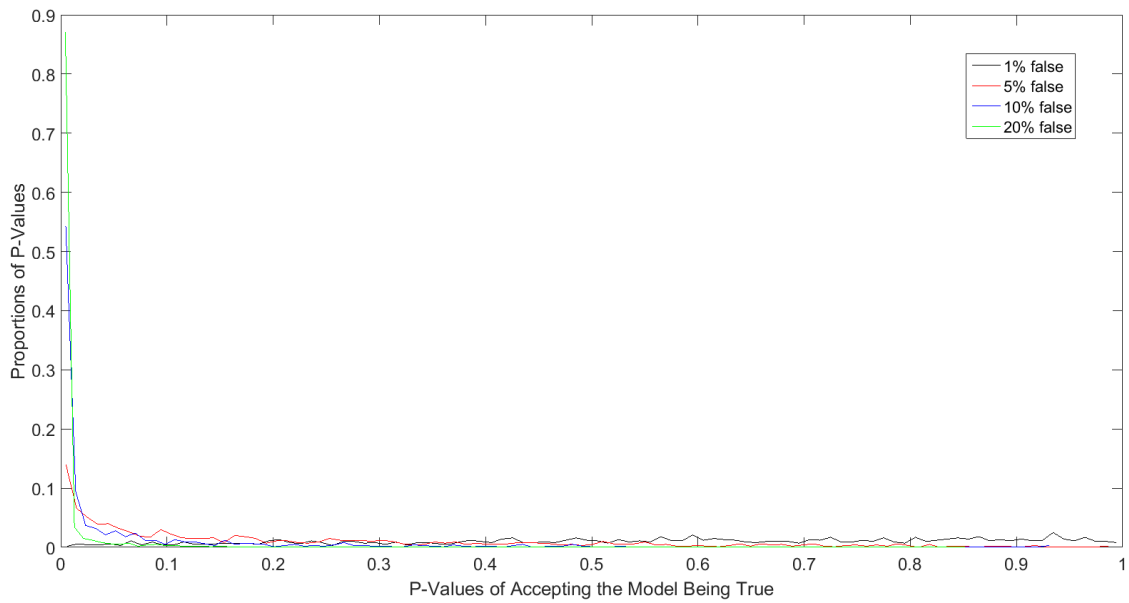
Figure 33 The Distributions of P-Values of the Monte Carlo Experiments

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[11].

According to Table 10 and Figure 33, 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 Le et al (2016), the powers at the same degrees of falsification are relatively lower in a microeconomic model. It is because a macroeconomic model is typically more heavily parameterised and more difficult to pass the test.

---

[11] 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.

6 A Level Playing Field?

Both the econometric and economic models include the *ad hoc* public sector dummy to explain the wage data feature/behaviour—the sector dummy is explicitly part of the regressors in the econometric models, while it is implicitly specified in the error structure of the economic model. To some extent, it is not a huge surprise that the economic model can well explain the data behaviour summarised by auxiliary models with a similar specification. Moreover, though we managed to obtain a robust estimate of the PSWP by including the public sector dummy in both econometric and economic models, we don't actually know what causes it. Is it fair to pay such a PSWP? Are the public sector workers on a level playing field? Different from the positive analysis (which is the focus of the thesis up to now), this section deals with the normative inquiry into the PSWP.

This question can be asked in a slightly different but essentially equivalent way: can a pure neoclassical model with no *ad hoc* public sector dummy explain the data behaviour? In other words, if only meaningful economic parameters enter such an economic model and there is no "hold-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 behaviour where we found a "dummy" effect for public sector by all those econometric methods. Based on this argument, we re-test and re-estimate this "pure" economic model with no public sector dummy, while keeping the auxiliary models unchanged to summarise the data behaviour including the PSWP.

Firstly, we re-test this "pure" economic model under the previously estimated structural parameters (Table 5) 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 11, 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 triumphantly find a pure neoclassical economic model without *ad hoc* public sector dummy and it can very well explain the wage data behaviour 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. There is no "bias" or "non-economic inequality" or "injustice due to political pressure" going on.

Specification	*ad hoc*		pure		pure	
Parameters	estimated		estimated		re-estimated	
	H0a	H0b	H0a	H0b	H0a	H0b
α	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
γ	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
P-Values	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	**79.04%**	0.00%	**59.50%**	0.00%	**60.50%**	0.00%

Table 11 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 5 and Table 7.

7 Conclusion

The neoclassical labour economic model with microfoundation is shown to be able to match most data features in the wage setting, especially in the wage premium summarised by the four popular types of methods in the microeconometric literature. Although the model cannot pass the test in the strict sense if the complete linear regression is used as the auxiliary model, it can successfully mimic the data feature if PSWP oriented auxiliary models or grouped regression coefficients are used. In particular, the propensity score matching leads to a high probability to pass the model, because of its robustness against mis-specification. This is logically coherent with the indirect inference test, which uses the auxiliary regression to provide a comparison basis rather than a serious model.

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 behaviour 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.

GENERAL CONCLUSION

This thesis contributes to the literature on public sector wage premium by providing a robust estimate (6.5%) using the four types of microeconometric methods and a newly introduced 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. The wage premium is greater for females and during economic downturns—evidence shows that there is a negative correlation between the estimated public sector wage premium and the macroeconomic business cycle (around -0.24).

Built on the neoclassical labour economic theory, we derived a microfounded economic model, which is confronted with the individual-level wage data using indirect inference. All the four types of microeconometric 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 grouped OLS as the auxiliary model. The estimated model implies a 6.72% public sector wage premium, very close to the average of the microeconometric estimate. 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. A Monte Carlo experiment based on the estimated model is conducted, which verifies the high statistical power of indirect inference method. 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 thesis 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

ADAMCHIK, V. A. & BEDI, A. S. 2000. Wage Differentials between the Public and the Private Sectors: Evidence from an Economy in Transition. *Labour Economics,* 7**,** 203-224.

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

ALLINGTON, N. & MORGAN, P. 2003. Does It Pay to Work in the Public Sector? Evidence from Three Decades of Econometric Analyses. *Public Money and Management***,** 253-262.

ANGRIST, J. & PISCHKE, J.-S. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton, NJ, Princeton University Press.

ANTÓN, J.-I. & BUSTILLO, R. M. D. 2015. Public-private sector wage differentials in Spain. An updated picture in the midst of the Great Recession. *Investigación Económica,* 74**,** 115-157.

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

ASHENFELTER, O. & JOHNSON, G. E. 1972. Unionism, Relative Wages, and Labor Quality in U.S. Manufacturing Industries. *International Economic Review,* 13**,** 488-508.

BBC 2011. Public sector pay gap 'increases'. *Business.*

BBC 2014. Public sector hourly pay outstrips private sector pay. *Business.*

BBC 2016. Ministers to look at further curbs on public sector pay-offs. *UK Politics.*

BBC 2016. Public sector workers to face pay cut. *Business.*

BECKER, G. S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education,* Chicargo, University of Chicargo Press.

BELLANTE, D. & LONG, J. 1981. The Political Economy of the Rent-Seeking Society: The Case of Public Employees and Their Unions. *Journal of Labor Research,* 2**,** 1-14.

BLACKABY, D. H., MURPHY, P. D. & O'LEARY, N. C. 1999. The Payment of Public Sector Workers in the UK: Reconciliation with North American Findings. *Economics Letters,* 65**,** 239-243.

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*.

BLANCHFLOWER, D. G. & BRYSON, A. 2010. The Wage Impact of Trade Unions in the UK Public and Private Sectors. *Economica,* 77**,** 92-109.

BLINDER, A. S. 1973. Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources,* 8**,** 436-455.

BRADLEY, J., POSTEL-VINAY, F. & TURON, H. 2011. Public Sector Wage Policy and Labor Market Equilibrium: A Structural Model. *Working Paper*.

BUCHINSKY, M. 1994. Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression. *Econometrica,* 62**,** 405-458.

CAMERON, A. & TRIVEDI, P. 2005. *Microeconometrics: Methods and Applications,* New York, Cambridge University Press.

CAROW, R. 1981. Total Compensation Comparability in the Evolution of Federal Compensation Policy. *In:* MIESZKOWSKI, P. & PETERSON, G. (eds.) *Public Sector Labor Markets.* Washington, D. C.: Urban Institute Press.

CHAMBERLAIN, G. 1994. Quantile Regression, Censoring, and the Structure of Wage. *In:* SIMS, C. A. (ed.) *Advances in Econometrics: Volume 1: Sixth World Congress.* Cambridge University Press.

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.

CORBO, V. & STELCNER, M. 1983. Earning Determination and Labor Markets: Gran Santiago. *Journal of Development Economics,* 12**,** 251-266.

COTTON, J. 1988. On the decomposition of wage differentials. *Review of Economics and Statistics,* 70**,** 236-243.

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

47**,** 6438-6460.

DAYMONT, T. N. & ANDRISANI, P. J. 1984. Job Preferences, College Major, and the Gender Gap in Earnings. *The Journal of Human Resources,* 19**,** 408-428.

DINARDO, J., FORTIN, N. M. & LEMIEUX, T. 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica,* 64**,** 1001-1044.

DISNEY, R. 2007. Public-Private Sector Wage Differentials around the World: Methods and Evidence. *Working Paper*.

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

DISNEY, R. & GOSLING, A. 2003. A New Method for Estimating Public Sector Pay Premia: Evidence from Britain in the 1990s. *CEPR Discussion Paper*.

DOLTON, P. & MAKEPEACE, G. 2011. Public and Private Sector Labour Markets. *In:* GREGG, P. & WADSWORTH, J. (eds.) *The State of Working Britain.* Oxford University Press.

ELLIOT, R. & MURPHY, P. D. 1987. The Relative Pay of Public and Private Sector Employees, 1970-1984. *Cambridge Journal of Economics,* 11**,** 107-132.

ELLIOTT, R. F. & DUFFUS, K. 1996. What Has Been Happening to Pay in the Public-Service Sector of the British Economy? Developments over the Period 1970–1992. *British Journal of Industrial Relations,* 34**,** 51-85.

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

FREEMAN, R. B. & MEDOFF, J. L. 1981. The Impact of the Percentage Organized on Union and Nonunion Wages. *The Review of Economics and Statistics,* 63**,** 561-572.

GAAG, J. V. D. & VIJVERBERG, W. 1988. A Switching Regression Model for Wage Determinants in the Public and Private Sectors of a Developing Country. *The Review of Economics and Statistics,* 70**,** 244-252.

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.

GILLMAN, M. 2011. *Advanced Modern Macroeconomics: Analysis and Application*, Pearson.

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

GRONAU, R. 1974. Wage comparisons: A selectivity bias. *Journal of Political Economy,* 82**,** 1119-1143.

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. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement,* 5**,** 475-492.

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

HECKMAN, J. J. 1997. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources,* 32**,** 441-462.

HECKMAN, J. J. & NAVARRO-LOZANO, S. 2004. Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics,* 86**,** 30-57.

IMBENS, G. & WOOLDRIDGE, J. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature,* 47**,** 5-86.

JOVANOVIC, B. & LOKSHIN, M. M. 2004. Wage Differentials between the State and Private Sectors in Moscow. *Review of Income and Wealth,* 50**,** 107-123.

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

KEANE, M. P. & WOLPIN, K. I. 1997. The Career Decisions of Young Men. *Journal of Political Economy,* 105**,** 473-522.

KEANE, M. P. & WOLPIN, K. I. 2009. Empirical applications of discrete choice dynamic programming models. *Review of Economic Dynamics,* 12**,** 1-22.

KYDLAND, F. & PRESCOTT, E. 1982. Time to Build and Aggregate Fluctuations. *Econometrica,* 50**,** 1345-1370.

LAMO, A., PÉREZ, J. J. & SCHUKNECHT, L. 2012. Public or Private Sector Wage Leadership? An International Perspective. *The Scandinavian Journal of Economics,* 114**,** 228-244.

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.

LEMIEUX, T. 2002. Decomposing Changes in Wage Distributions: A Unified Approach. *The Canadian Journal of Economics / Revue canadienne d'Economique,* 35**,** 646-688.

LEPING, K. O. 2005. *Public-Private Sector Wage Differential in Estonia: Evidence from Quantile Regression*, Tartu University Press.

LEUVEN, E. & SIANESI, B. 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.

LEWIS, H. G. 1963. *Unionism and Relative Wages in the United States: An Empirical Inquiry,* Chicargo, University of Chicago Press.

LEWIS, H. G. 1974. Comments on selectivity biases in wage comparisons. *Journal of Political Economy,* 82**,** 1145-1155.

LINDAUER, D. L. & SABOT, R. H. 1983. The Public/Private Wage Differential in a Poor Urban Economy. *Journal of Development Economics,* 12**,** 137-152.

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.

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.

MUELLER, R. E. 1998. Public-Private Sector Wage Differentials in Canada: Evidence from Quantile Regressions. *Economics Letters,* 60**,** 229-235.

NEUMARK, D. 1988. Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources,* 23**,** 279-709.

NICKELL, S. & QUINTINI, G. 2002. The Consequences of the Decline in Public Sector Pay in Britain: A Little Bit of Evidence. *The Economic Journal,* 112**,** F107-F118.

NUNZIATA, L. 2005. Institutions and Wage Determination: a Multi-country Approach. *Oxford Bulletin of Economics and Statistics,* 67**,** 435-466.

OAXACA, R. 1973. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review,* 14**,** 693-709.

OAXACA, R. & RANSOM, M. R. 1994. On discrimination and the decomposition of wage differentials. *Journal of Econometrics,* 61**,** 5-21.

PARSLEY, C. J. 1980. Labor Union Effects on Wage Gains: A Survey of Recent Literature. *Journal of Economic Literature,* 18**,** 1-31.

POTERBA, J. M. & RUEBEN, K. S. 1994. The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods. *NBER Working Paper Series*.

PRESCOTT, E. & KYDLAND, F. 1982. Time to Build and Aggregate Fluctuations. *Econometrica,* 50**,** 1345-1370.

QUINN, J. F. 1982. Pension Wealth of Government and Private Sector Workers. *American Economic Review,* 72**,** 283-87.

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.

REIMERS, C. W. 1983. Labor market discrimination against Hispanic and black men. *Review of Economics and Statistics* 65**,** 570-579.

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.

ROSENBAUM, P. & RUBIN, D. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70**,** 41-55.

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

SCHMIDT, P. 1978. Estimation of a Simultaneous Equations Model with Jointly Dependent Continuous and Qualitative Variables: The Union-Earnings Question Revisited. *International Economic Review,* 19**,** 453-465.

SMETS, F. & WOUTERS, R. 2003. An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association,* 1**,** 1123-1175.

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.

TERRELL, K. 1993. Public-Private Wage Differentials in Haiti: Do Public Servants Earn a Rent? *Journal of Development Economics,* 42**,** 293-314.

TRINDER, C. 1981. Pay of Employees in the Public and Private Sector. *National Institute Economic Review,* 97**,** 48-56.

TSIATIS, A. 2006. *Semiparametric Theory and Missing Data,* New York, Springer.

UHLIG, H. 1998. A Toolkit for Analysing Nonlinear Dynamic Stochastic Models Easily. *In:*

MARIMON, R. & SCOTT, A. (eds.) *Computational Methods for the Study of Dynamic Economies.* Oxford: Oxford University Press.

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

VENTI, S. 1987. Wages in the Federal and Private Sectors. *In:* WISE, D. A. (ed.) *Public sector payrolls.* Cambridge, MA: National Bureau of Economic Research.

VOINEA, L. & MIHAESCU, F. 2011. A Contribution to the Public-Private Wage Inequality Debate: the Iconic Case of Romania. *The wiiw Balkan Observatory Working Papers*.

WOOLDRIDGE, J. 2010. *Econometric Analysis of Cross Section and Panel Data,* Cambridge, MA, MIT Press.

YIN, Z. & GAN, L. 2009. Wage Differentials between Public and Nonpublic Sector in China. *Economic Research Journal***,** 129-140.

YU, X.-H. & CHEN, X.-J. 2010. An Empirical Research on China Sector Wage Gap: 1993-2006. *2010 International Conference on Management Science & Engineering (17th).* Melbourne, Australia.

APPENDIX

Table 12 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

Table 13 Descriptive Statistics of Pay Differentials by Gender

	Overall		Female		Male	
	Private	Public	Private	Public	Private	Public
2001	11.59	12.18	9.62	11.19	13.58	14.33
	9.40	10.61	7.76	9.56	11.14	12.95
2002	11.65	12.29	9.79	11.35	13.55	14.50
	9.45	10.76	7.96	9.85	11.18	13.01
2003	11.85	12.52	9.91	11.56	13.81	14.79
	9.43	10.96	7.85	10.02	11.34	13.43
2004	11.80	12.78	10.00	11.89	13.59	14.78
	9.57	11.07	8.16	10.34	11.25	13.05
2005	12.24	13.23	10.38	12.43	14.21	15.28
	9.75	11.86	8.25	10.94	11.69	13.78
2006	12.00	12.91	10.18	12.02	13.89	14.99
	9.69	11.30	8.24	10.57	11.42	13.21
2007	12.15	12.93	10.20	12.24	14.08	14.62
	9.66	11.60	8.25	10.79	11.44	13.48
2008	12.24	12.86	10.43	12.06	13.98	14.75
	9.69	11.30	8.12	10.55	11.38	13.11
2009	12.46	13.29	10.60	12.48	14.27	15.24
	9.90	11.96	8.44	11.08	11.51	14.00
2010	12.02	12.99	10.26	12.40	13.69	14.36
	9.59	11.44	8.28	10.87	10.97	13.08
2011	11.62	12.71	9.72	12.02	13.44	14.33
	9.10	11.21	7.79	10.40	10.81	13.02

NB: The means are at the top of each cell, and medians are at the bottom.

Table 14 Descriptive Statistics of Working Hours by Gender

	Overall		**Female**		**Male**	
	Private	Public	Private	Public	Private	Public
2001	35.99	32.95	30.35	29.78	41.68	39.83
	38	37	35	33	40	38
2002	35.77	32.85	30.32	30.1	41.33	39.26
	38	37	35	33	40	38
2003	35.52	32.08	29.92	29.1	41.19	39.13
	38	36	35	32	40	38
2004	35.75	32.96	30.48	30.25	40.95	39.06
	38	36	35	35	40	38
2005	35.43	32.74	30.21	30.37	40.93	38.86
	38	37	35	35	40	38
2006	35.53	32.57	30.49	30.11	40.73	38.25
	38	36	35	34	40	38
2007	35.69	32.7	30.41	30.46	40.95	38.17
	38	37	35	35	40	38
2008	35.85	32.82	30.98	30.32	40.55	38.7
	38	37	35	35	40	38
2009	35.38	32.54	30.36	30.31	40.26	37.98
	38	37	35	35	40	38
2010	35.15	32.48	30.21	30.22	39.84	37.8
	38	37	35	35	40	38
2011	34.87	32.4	29.86	29.94	39.64	38.17
	38	36	35	32	40	38

NB: The means are at the top of each cell, and medians are at the bottom.

Table 15 Description of the Variables in the Data

Variable name	Description
gender	Female = 0 and male = 1
married	Married (and living together) or cohabiting, heterosexual
white_ghm	White ethnic group using ethcen6
ww	Real weekly gross pay.
ln_ww	Natural log of real weekly gross pay.
wh	Real hourly gross pay.
ln_wh	Natural log of real hourly gross pay.
child	One or more dependent children in family aged 0-4.
qualdgroup	Highest qualification includes vocational qualifications 0 No qualifications 1 Other qualifications 2 GCSE A*-C or equivalent 3 GCE A Level or equivalent 4 Higher Education 5 Degree or equivalent
age	Age in years
regiongroup	Categorical variable for region of work (Standard Regions) 1 North 2 Yorks and Humberside 3 North West 4 East Midlands 5 West Midlands 6 East Anglia 7 London 8 South East 9 South West 10 Scotland 11 Wales 12 N. Ireland 13 Overseas
ftime_ptime	Works full time or part time (in main job) or not at all 0 Not working 1 Part Time Worker 2 Full Time Worker
public_sector	Works in public sector
sic92_sector	Industry sector using SIC92 (same as SIC 2003)- 9 categories 1 Agriculture, farming & fishing 2 Energy & water 3 Manufacturing 4 Construction 5 Distribution, hotels & restaurants 6 Transport & communication 7 Banking, finance & insurance 8 Public admin, education & health 9 Other services 10 Workplace outside UK

soc2000_1_digit	Occupation groups, SOC2000, 9 groups
	1 Managers and senior officials
	2 Professional occupations
	3 Associate professional and technical
	4 Administrative and secretarial
	5 Skilled trades occupations
	6 Personal service occupations
	7 Sales and customer service occupations
	8 Process, plant and machine operatives
	9 Elementary occupations
year	Year using calendar data. (Jan to Dec). Equals refwky
manual	Manual worker dummy
ftime	Works full time
married_same	Same sex couple, civil partners or cohabiting
edage	Age when completed continuous full time education
workexp	Work experience = age - edage

Selection Bias and Indirect Inference

Selection bias occurs when there is an unobservable correlation between individual idiosyncratic and unobservable errors (e) and observable characteristics (X) in a cross section. In direct inference it causes a bias in the estimator. Various suggestions exist for reducing or eliminating the bias, such as Hausman's two-stage procedure. But of course there is no way of knowing whether in the end it has been effectively eliminated.

In Indirect Inference (II) the auxiliary model equations, say as here a GOLS, are separate from the structural model. The auxiliary equations represent the data behaviour and therefore include the effect of any such unobservable correlation between e and X. The question then is what the true structural relationships are: do they include such a correlation or do they not? If they do we would like to estimate the 'underlying' effect of X on the agents after allowing for this correlation. If they do not, we would like to estimate it without any such correlation. We would also like to know whether there is indeed any such correlation in case we wish to do direct inference, because then we will know how much we can regard those results as bias-free.

Under II we can carry out a test of whether the model with the correlation or without the correlation is the true one. The way we do this is to set the model up in two different ways:

1. without any such correlation where e is random;
2. with such a correlation where e is related to X.

In the first case we simulate the model with random e: viz. by bootstrapping e randomly.

In the second we simulate the model allowing for e's correlation for X by bootstrapping e and X together as a vector, so as to preserve any correlation observable in the sample. This sample correlation gives us our best estimate of the population correlation.

In each case we can estimate the resulting model by II (i.e. by a minimum distance estimator where the distance is that between the auxiliary coefficients produced by the data and produced by the simulated data). Thus we get the best estimate of each model on the two assumptions about the correlation: zero or existent.

Notice that the 'underlying' structural coefficients will in general be different estimates in the two cases, so that we can achieve our objective of estimating them whichever model is the true one. We may then test the two models against the data behaviour and choose the one that is not rejected or the one with the highest p-value (the model yielding the highest probability of generating the data).

This is a small sample method. We know the asymptotic properties of the estimator are consistent and normal. The test needs to be evaluated in small samples and its asymptotic properties can be obtained numerically- Meenagh et al (2016) explain that the asymptotic power of the Wald test we use cannot be gauged analytically because it uses the distribution of the auxiliary parameters generated by a false structural model.

We can assess the power of the test by Monte Carlo methods. Since this is a test where the true model can be either the one with the correlation or the one without we need to check its power under both possibilities. It is plain enough that when the true model is one of selection bias the power against the false model of no selection bias is increasing steadily as the selection bias gets larger.

However, it is not so clear what the power is when the true model is of no selection bias: it might well seem that the test would have little power owing to the fact that any correlation found in the sample would occur by chance.

One may raise a basic question about how one can be rejecting the selection-bias model when the no-selection-bias model is true. We must re-emphasise that the II process involves no OLS estimation; rather the structural model is bootstrapped under two different assumptions about its bias, in each case this assumption being treated as true. Effectively this basic question can be interpreted as that the two bootstrapping procedures should not be distinguishable when the true model (generating the data) has no bias. As a result the 'selection-bias' model should not be rejected more frequently than the true no-selection-bias model, i.e. only by 5% on a 5%-size test. In other words the II test has no power when the true model has no selection bias.

It might well be thought that because there is in truth no such correlation that is what the sample bootstrapping must also reveal. However, in any given sample randomness will produce correlations by chance; there are many characteristics in the X vector and so many possibilities for the error to be by chance related to one or more of these. We know that particularly if one has very large samples such patterns are more likely to be found: thus, if one constantly repeats experiments with the intent of finding a particular 'result' one is likely in the end to find one. This is similar to the idea of setting a bunch of monkeys to hit typewriters for weeks on end and to find 'some Shakespeare'.

Thus one should note that there is a considerable difference between simulating a model under the restriction of a zero correlation (no selection bias bootstrap) and doing so under no restriction on the correlation (selection bias assumed); in the latter case any chance correlation found in the sample will be generalised in the simulations even though it is not the true correlation. Thus if we test the model on the assumption it has selection bias, then we treat this

sample correlation as the best estimate of the true correlation: we then reestimate the model on this basis producing different underlying structural coefficient estimates.

The only way to resolve this issue in the context of small sample procedures is via Monte Carlo experiment. We can also consider via such an experiment how the power changes as the sample size rises, to investigate the asymptotic properties of the test.

In the following Monte Carlo experiment we test the power holding these underlying coefficient estimates constant at their assumed true values. This power calculation therefore underestimates the true power because changes in these coefficient estimates are excluded. Therefore, the Monte Carlo experiment gives the lower bound on the power. It is possible and desirable to go further and examine by Monte Carlo just how much power is enhanced by reestimation of the model: but this is highly computer-intensive and not possible in the time available now and so rather is left for future work.

The Monte Carlo Experiment

I have investigated the power issue raised by doing a Monte Carlo experiment, in which I have taken the situation as above: the 'true' model with no bias and the bootstrap test of the selection bias model using the bundled (or 'vector') bootstrap. I have taken the true model to be one with no selection bias and I have generated 1000 data samples from this model. I have then simulated both the no-selection-bias model and the selection-bias model as explained, by using the two different bootstrapping methods. I then test each model on the 1000 data samples.

It turns out that there is power though not in a very small sample of only 500. Once the sample reaches the size of my sample here (6216), the rejection rate reaches 15% on a 5% test. Were I to have a sample nearly three times as large the rejection rate would reach nearly 100%. See the Table that follows. It is also plain that as the sample size rises even further the power reaches its maximum of 100%, suggesting that its power rises asymptotically.

As noted above, the power here is at the lower bound because the structural model coefficients are held at their true value in the selection bias model. Were they to be re-estimated they would be moved to values away from the true ones which would raise the rejection rate. This is the reason that I find the selection bias model to be strongly rejected in my work.

TRUE: no selection bias	Proportion of Rejection of the Null			
Sample Size	Small (N = 500)	Original (N = 6,216)	Big (N = 15,000)	Asymptotic (N = 30,000)
Null: H0a (no selection bias)	5%	5%	5%	5%
Null: H0b (selection bias)	5%	14.5%	99.5%	100%

The II test distinguishes two structural models, one with a random error (no selection bias therefore), one where the error is correlated with people's characteristics, X (creating selection bias). It checks the match of each model's implied auxiliary model when simulated to the auxiliary model found in the data sample. The true model may either have or not have such an error correlation and this will be reflected in the auxiliary model coefficients. To simulate the model with such a correlation we bootstrap the error together with X since the sample correlation is the best estimate of the population correlation. What we have shown here by Monte Carlo experiment is that even when the true model has a random error (no selection bias), the data will reject with greater than the size frequency the model where the error is assumed to be correlated with X. The reason is that since data samples typically generate such correlations by chance, only by ignoring them will the simulated model behave like the data.