UNIVERSITE DE BOURGOGNE

IREDU: Institut de Recherche sur l'Éducation: Sociologie et Économie de l'Éducation, CNRS

THÈSE

Pour obtenir le grade de

Docteur de l'Université de Bourgogne

Discipline: Sciences Économiques

Spécialité: Économétrie et Statistiques Appliquée

Par

Sajjad Haider Bhatti

Estimation of the Mincerian Wage Model Addressing its Specification and Different Econometric Issues

Directeur de Thèse:

Jean Bourdon

Directeur de Recherche CNRS

IREDU, Université de Bourgogne

Co-directeur de Thèse:

Dr. Muhammad Aslam

Assistant Professor, Dept. of Statistics,

Bahauddin Zakariya University, Multan–Pakistan

Soutenue à Dijon, le 03 Décembre 2012

Jury

Christian Belzil	Directeur de Recherche CNRS, Ecole Polytechnique, Paris	**Rapporteur**
Ali Skalli	Maître de Conférences HDR, Université Paris-II, Paris	**Rapporteur**
Gérard Lassibille	Directeur de Recherche CNRS, Université de Bourgogne, Dijon	**Examinateur**
Jean Bourdon	Directeur de Recherche CNRS, Université de Bourgogne, Dijon	**Directeur de Thèse**

©

« L'Université de Bourgogne n'entend donner ni approbation, ni improbation, aux opinions émises dans les thèses. Ces opinions doivent être considérées comme propres à leurs auteurs. »

Dedication

I dedicate this dissertation

To

My *Parents, Sister,*

and To

Beloved *'Taaya Jaan'* who left us forever during my stay in France for this doctoral dissertation

but

still lives and walks beside us every day,
unseen, unheard, but always near,
still loved ……
still missed ……

Acknowledgements

Praise to Almighty ALLAH, lord of the worlds, the most gracious, merciful and beneficent Whose blessings and glory flourished my thoughts and Who gave me courage and health to finalize my study. I owe my heartiest salutations to beloved Prophet MUHAMMAD (S.A.W), who taught us to "Seek knowledge from the cradle to the grave".

It is only just to mention the fact that it would not have been possible to write this doctoral thesis without the support and help of many good people around me and that I can only pay in particular my humble gratitude to only a few of them here.

First and foremost, I pay utmost gratitude to my thesis supervisor Prof. Jean Bourdon for his guidance, kind attitude, and supportive behaviour. He was always there happily responding to my questions whenever I needed his guidance and help. He supported and encouraged me all the way during this work. He always provided me substantial guidance and valuable suggestions throughout my PhD, and never stopped inspiring me for hard work. Throughout my stay in France I found no one else of his rank. It is an honour for me to have worked under his supervision.

I would like to extend my deep gratitude to my co-supervisor Dr. Muhammad Aslam who taught me the very basics of research. I have been benefiting from his knowledge and kindness as a student for the last 8 years now. I have always dreamed to further work with him as a research student which came true when he came to France for a short period. I always found him ready for sharing his extensive experience and knowledge and I am thankful for his affectionate attitude, scholastic guidance, the best cooperation throughout this research, and for the help in the planning, execution and completion of this research work. I am grateful for his keen interest in my research work and constructive criticism which served as a stimulating factor for me. The encouragement by him is more than anything else.

Many thanks to the thesis reviewers and jury members Christian Belzil, Ali Skalli and Gérard Lassibille for their kindness to accept the responsibility of thesis review. Their comments and suggestions will help me to improve the quality of thesis and my understanding towards research.

I am also thankful to Mr. Diégo Legros who encouraged and helped me a lot not only in understanding Econometrics during masters in Economics at University of Burgundy but also at the initial stages of this doctoral work.

I am equally indebted to Prof. Bruno Suchaut, (who was director of IREDU when I joined IREDU as a PhD student) and Prof. Jean-François Giret (Director of IREDU), for hosting and providing me all the possible facilities for my research work.

Thanks to all my colleagues at IREDU especially Claire, Marie-Odile, Fabienne, Thomas, Aurélie, Lazare, Gabriela, Edang, Mitra, Léopoldine, Élodie, Aurélie, Michel, Amélie, Marielle, Fatima, Abdramane and Aliou Diop (who is no more with us but his sweet memories will remain with us forever). Special thanks to Aline and Christine for helping in French summary. All these colleagues made my stay at IREDU more pleasant, smooth and worthy.

I am grateful to Alain Gauldry, computer technician, for his assistance in maintaining the computers, providing software and related accessories. I am obliged to Bertille Theurel for providing me the documents, books, articles etc. They are true professionals in their spirit. I found them always cooperative and open.

This project would not have been possible without the enthusiastic attitude, moral support and encouragement of a galaxy of exceptional Pakistani friends and colleagues in Dijon like Sabir, Ashafqaue, Amjad, Hamid, Muhammad Ali, Farhan, Atif, Shamshir, Ahmed, Abid, Ahsan, and Farasat. They always expressed a very friendly and genial comportment. The time spent with "Dijoni-Pakistanis" will be remembered for ages.

I say without any doubt that this dissertation would have remained a rather more difficult task if I had not gotten moral support and constant encouragement from all my relatives and long time friends. It is a long list of friends from Chakwal to Islamabad to Bahauddin Zakariya University Multan, naming them all will require lots of pages. Thanks to all my relatives and friends for their generosity and kindness.

I owe special thanks to Higher Education Commission (HEC) Pakistan, for providing me with a great opportunity through this scholarship program to fulfil my dream of doctoral studies. All the concerned team at HEC provided their help when needed.

Finally, thanks to my family, for their love, support, prayers for me and encouragement over the years throughout my life. No acknowledgement could ever sufficiently express my obligations to my parents and beloved sister for their love, inspiration, and well wishing. Their prayers, encouragement and support (spiritual, emotional, intellectual, financial and otherwise) that heartened me to achieve success in every sphere of my life has no alternative.

SAJJAD HAIDER BHATTI

Dijon, France
December 2012

Table of Contents

Abstract 13

Résumé Analytique 14

Chapter 1: Introduction 18

1.1 Background and Motivation 19

1.2 Objectives 22

1.3 Outlines 23

Chapter 2: The Mincerian Wage Model and Econometric Issues: A Review 24

2.1 The Mincerian Wage Model 25

2.2 The Econometrics Issues 31

2.2.1 Endogeneity of schooling 31

2.2.1.1 Instrumental variables two- stage least squares (IV2SLS) estimation 32

2.2.1.2 Family background factors as instruments 33

2.2.1.3 Availability of educational institutions nearby as instrument 35

2.2.1.4 Different exogenous factors affecting schooling as instruments 36

2.2.2 Measurement Error 41

2.2.3 Sample Selection Bias 44

2.3 Parametric and Semi-parametric Estimation of the Mincerian Model 48

Chapter 3: Estimation of the Mincerian Wage Model for the French Data 50

3.1 A Brief Review of the Mincerian Studies for the French Data 51

3.1.1 Common data sources 51

3.1.2 Estimation and trends 52

3.2 The Data 60

3.3 Variables used in Estimation of the Mincerian Model for France 60

3.3.1 Response variable for the French wage regression 61

3.3.2 Explanatory variables for the French wage regression 61

3.4 Variables used in Estimation of Participation Equation for France _____ 68

3.4.1 Response variable for selection equation for France 68

3.4.2 Explanatory variables for selection equation for France 68

3.5 Instrumental Variables to be used in the IV2SLS Estimation Approach ____ 73

3.5.1 Definition of Instrument-1 (Z1) 73

3.5.2 Definition of Instrument-2 (Z2) 73

3.6 Estimation Results Based on the French Data _____ 74

3.6.1 Preliminary estimation 74

3.6.2 Instrumental variables 2SLS estimation using Instrument-1 77

3.6.2.1 Hausman test for endogeneity of schooling using Instrument-1 _____ 80

3.6.3 Instrumental variables 2SLS estimation using Instrument-2 80

3.6.3.1 Hausman test for endogeneity of schooling using Instrument-2 _____ 84

3.6.4 Choice between two instruments (Z1, Z2) for the French data 84

3.6.5 Sample selection model: Step-1: Estimation of participation equation for the French data 86

3.6.5.1 Marginal effects of explanatory variables on the probability of selection into the wage earners' sample in France _____ 89

3.6.6 Sample selection model: Step-2: Estimation of wage regression 92

3.6.7 Dealing with endogeneity and sample selection bias, simultaneously 94

3.6.8 Heterogeneity of the error term 97

3.6.8.1 Testing heteroscedasticity of the error term _____ 98

3.6.9 Adaptive estimation of simultaneous model 99

3.6.10 Semi-parametric estimation of the Mincerian wage regression 102

3.6.10.1 Non-parametric estimation of first-stage schooling equation _____ 103

3.6.10.2 Semi-parametric estimation of second-stage wage regression _____ 106

3.6.10.3 Testing heteroscedasticity of errors from semi-parametric model _____ 108

3.6.11 Adaptive estimation of semi-parametric model 108

3.6.12 Comparison of parametric and semi-parametric models for France 110

3.6.12.1 Choice between parametric and semi-parametric models _____ 110

3.6.12.2 Choice between simple and adaptive versions of parametric model ____ 111

3.6.13 Interval estimation of selected model for the French data 111

Chapter 4: Estimation of the Mincerian Wage Model for the Pakistani Data _____ 114

4.1 A Brief Review of the Mincerian Studies for the Pakistani Data _____ 115

4.1.1 Common data sources 115

4.1.2 Estimation and trends 115

4.2 The Data _____ 122

4.3 Variables used in Estimation of the Mincerian Model for Pakistan _____ 122

4.3.1 Response variable for the Pakistani wage regression 122

4.3.2 Explanatory variables for the Pakistani wage regression 123

4.4 Variables used in Estimation of Participation Equation for Pakistan _____ 128

4.4.1 Response variable for selection equation for Pakistan 128

4.4.2 Explanatory variables for selection equation for Pakistan 128

4.5 Estimation Results Based on the Pakistani Data _____ 133

4.5.1 Preliminary estimation 133

4.5.2 Instrumental variables 2SLS estimation using Instrument-1 137

4.5.2.1 Hausman test for endogeneity of schooling using Instrument-1 _____ 140

4.5.3 Instrumental variables 2SLS estimation using Instrument-2 141

4.5.3.1 Hausman test for endogeneity of schooling using Instrument-2 _____ 145

4.5.4 Choice between two instruments for the Pakistani data 145

4.5.5 Sample selection model: Step-1: Estimation of participation equation for the Pakistani data 146

4.5.5.1 Marginal effects of explanatory variables on the probability of selection into the wage earners' sample in Pakistan _____ 150

4.5.6 Sample selection model: Step-2: Estimation of wage regression 151

4.5.7 Addressing endogeneity and sample selection bias, simultaneously for the Pakistani data 154

4.5.7.1 Testing heteroscedasticity of errors from simultaneous model ____ 158

4.5.8 Adaptive estimation of simultaneous model (Pakistan) 159

4.5.9 Semi-parametric estimation of the Mincerian model (Pakistan) 160

4.5.9.1 Non-parametric estimation of first stage schooling equation ____ 161

4.5.9.2 Semi-parametric estimation of the wage regression ____ 162

4.5.10 Comparison of parametric and semi-parametric models 165

4.5.10.1 Choice between models estimated for the Pakistani data ____ 165

4.5.11 Interval estimation of selected model for the Pakistan data 166

Chapter 5: The Mincerian Model as a Quantile Regression Model ____ 169

5.1 The Mincerian Wage Model in the Context of Quantile Regression ____ 170

5.2 A Short Review of the Mincerian Wage Model under Framework of Quantile Regression ____ 172

5.3 Estimation of the Mincerian Quantile Regression Model for France ____ 178

5.4 Estimation of the Mincerian Quantile Regression Model for Pakistan ____ 187

Chapter 6: Summary and Conclusions ____ 198

Appendices ____ 210

Appendix A1: Summary Statistics of the French Sample ____ 211

Appendix A2: Models used for the French Data ____ 212

Appendix B1: Summary Statistics of the Pakistani Sample ____ 216

Appendix B2: Models used for the Pakistani Data ____ 217

Bibliography ____ 220

List of Tables

Table 3.1: Brief Description of Variables used in Estimation of the Mincerian Model for the French data **67**

Table 3.2: Brief Description of Variables used in Estimation of Selection Equation for the French Data **72**

Table 3.3: OLS Estimation of the French Wage Regression **75**

Table 3.4-A: First Stage of IV2SLS Estimation for France using Z1 as Instrument **78**

Table 3.4-B: Second Stage of IV2SLS Estimation for France using Z1 as Instrument **79**

Table 3.4-C: Endogeneity Test for Schooling with Z1 as Instrument **80**

Table 3.5-A: First Stage of IV2SLS Estimation for France using Z2 as Instrument **81**

Table 3.5-B: Second Stage of IV2SLS Estimation for France using Z2 as Instrument **82**

Table 3.5-C: Endogeneity Test for Schooling with Z2 as Instrument **84**

Table 3.6: Correlation Matrix among Response, Endogenous Variable and Instruments (Z1, Z2), for the French Data **86**

Table 3.7: Estimation of Selection Equation for France **87**

Table 3.8: Proportion of Individuals According to Area of Residence **88**

Table 3.9: Marginal Effects of Different Explanatory Variables on the Probability of Selection into Wage Earner's Sample **91**

Table 3.10: Heckman (1979) Sample Selection Model for the French Data **92**

Table 3.11: Addressing Endogeneity and Sample Selection Biases Simultaneously for the French Data (Model 3.5) **96**

Table 3.12: White's Test for Heteroscedasticity on Errors from Model 3.5 **98**

Table 3.13: Model 3.5W (Adaptive Estimation of Model 3.5) **101**

Table 3.14: Principal Component Analysis on all Exogenous Variables **105**

Table 3.15: Semi-Parametric Estimation of the French Wage Regression (Model 3.6) **106**

Table 3.16: White's Test for heteroscedasticity on Errors from Model 3.6 **108**

Table 3.17: Model 3.6W (Adaptive Estimation of Model 3.6) **109**

Table 3.18 : Comparison of Parametric and Semi-Parametric Mode Estimated for the French Data **110**

Table 3.19: Interval Estimation of Model 3.5W selected for France **112**

Table 4.1: Description of Variables used in Estimation of the Mincerian Wage Model for the Pakistani Data **127**

Table 4.2: Description of Variables used in Estimation of Selection Equation for the Pakistani Data **132**

Table 4.3: OLS Estimation of the Pakistani Wage Regression **134**

Table 4.4-A: First Stage of IV2SLS Estimation for Pakistan using Z1 as Instrument **138**

Table 4.4-B: Second Stage of IV2SLS Estimation for Pakistan using Z1 as Instrument **139**

Table 4.4-C: Endogeneity Test for Schooling with Z1 as Instrument **140**

Table 4.5-A: First Stage of IV2SLS Estimation for Pakistan using Z2 as Instrument **142**

Table 4.5-B: Second Stage of IV2SLS Estimation for Pakistan using Z2 as Instrument **143**

Table 4.5-C: Endogeneity Test for Schooling with Z2 as instrument **145**

Table 4.6: Correlation Matrix among Response, Endogenous Variable and Instruments (Z1, Z2), for the Pakistani Data **146**

Table 4.7: Estimation of Selection Equation for Pakistan **147**

Table 4.8: Marginal Effects of Different Explanatory Variables on the Probability of Selection into Wage Earner's Sample **150**

Table 4.9: Heckman (1979) Sample Selection Model for the Pakistani Data **152**

Table 4.10: Addressing Endogeneity and Sample Selection Biases Simultaneously for the Pakistani data (Model 4.5) **156**

Table 4.11: Test for Heteroscedasticity on Error from Model 4.5 **158**

Table 4.12: Model 4.5W (Adaptive Estimation of Model 4.5) **159**

Table 4.13: Principal Component Analysis on all Exogenous Variables **161**

Table 4.14: Semi-Parametric Estimation of the Pakistani Wage Model (Model 4.6) **163**

Table 4.15: Comparison of Parametric and Semi-Parametric Models Estimated for the Pakistani Data **165**

Table 4.16: Interval Estimation of Model 4.5W selected for Pakistan **167**

Table 5.1: Parameter Estimates of the Mincerian Wage Model at Different Quantiles of the French Wage Distribution **180**

Table 5.2: Parameter Estimates of the Mincerian Wage Model at Different Quantiles of the Pakistani Wage Distribution **189**

Table A1.1: Summary Statistics for the French Sample **211**

Table B1.1: Summary Statistics for the Pakistani Sample **216**

Abstract

In the present doctoral thesis, we estimated Mincer's (1974) semi logarithmic wage function for the French and Pakistani labour force data. This model is considered as a standard tool in order to estimate the relationship between earnings/wages and different contributory factors. Despite of its vide and extensive use, simple estimation of the Mincerian model is biased because of different econometric problems. The main sources of bias noted in the literature are endogeneity of schooling, measurement error, and sample selectivity. We have tackled the endogeneity and measurement error biases via instrumental variables two stage least squares approach for which we have proposed two new instrumental variables. The first instrumental variable is defined as "the average years of schooling in the family of the concerned individual" and the second instrumental variable is defined as "the average years of schooling in the country, of particular age group, of particular gender, at the particular time when an individual had joined the labour force". Schooling is found to be endogenous for the both countries. Comparing two said instruments we have selected second instrument to be more appropriate. We have applied the Heckman (1979) two-step procedure to eliminate possible sample selection bias which found to be significantly positive for the both countries which means that in the both countries, people who decided not to participate in labour force as wage worker would have earned less than participants if they had decided to work as wage earner. We have estimated a specification that tackled endogeneity and sample selectivity problems together as we found in respect to present literature relative scarcity of such studies all over the globe in general and absence of such studies for France and Pakistan, in particular. Differences in coefficients proved worth of such specification. We have also estimated model semi-parametrically, but contrary to general norm in the context of the Mincerian model, our semi-parametric estimation contained non-parametric component from first-stage schooling equation instead of non-parametric component from selection equation. For both countries, we have found parametric model to be more appropriate. We found errors to be heteroscedastic for the data from both countries and then applied adaptive estimation to control adverse effects of heteroscedasticity. Comparing simple and adaptive estimations, we prefer adaptive specification of parametric model for both countries. Finally, we have applied quantile regression on the selected model from mean regression. Quantile regression exposed that different explanatory factors influence differently in different parts of the wage distribution of the two countries. For both Pakistan and France, it would be the first study that corrected both sample selectivity and endogeneity in single specification in quantile regression framework.

Key Words: Adaptive estimation; Endogeneity; Heteroscedasticity; Instrumental variables; Mincerian model; Quantile regression; Sample selection bias; Semi-parametric estimation; Wage regression.

JEL classifications: C14, I2, J31, P52

Résumé Analytique

Dans cette thèse, notre cadre d'analyse repose sur l'estimation de la fonction de gain proposée par Mincer (1974). Le but est de reprendre la spécification de ce modèle en s'intéressant aux problèmes d'estimation liés. Le but est aussi une comparaison pour les marchés du travail français et pakistanais en utilisant une spécification plus robuste. Le modèle de Mincer est un point central, dans le cadre de la théorie du capital humain ; la relation entre les gains salariaux d'un individu, ses principales caractéristiques et les autres facteurs jouent un rôle complexe dans le processus de détermination du salaire sur le marché du travail.

Toutefois, suivant une nombreuse littérature, la simple estimation du modèle de Mincer est biaisée, ceci en raison de différents problèmes. Les sources principales des biais notés dans la littérature sont l'endogénéité de la scolarité, l'erreur de mesure, et les aléas de sélection des individus dans l'échantillon des salariés.

Généralement, dans la littérature concernée, le biais causé par l'endogénéité et l'erreur de mesure est contrôlé en utilisant méthode d'estimation en deux étapes avec variables instrumentales 'IV2SLS'.

Dans la présente thèse deux nouvelles variables instrumentales sont proposées dans une application de type IV2SLS. La première est définie comme « les années moyennes de scolarité dans la famille d'appartenance de l'individu concerné" et la seconde variable instrumentale est définie comme « les années de scolarité moyenne, pour la population en âge de travailler, dans l'économie concernée ». Cela en référence à l'année où l'individu était entré sur le marché du travail et en référence à son groupe d'âge au moment de cette entrée, mesure distincte suivant le sexe de l'individu. D'après l'analyse menée dans cette thèse, la seconde variable instrumentale apparaît être la plus appropriée, cela puisqu'elle possède un faible effet direct sur la variable de réponse par rapport à la première variable instrumentale proposée. Par ailleurs, la définition de cette variable instrumentale est plus robuste que la

première variable instrumentale. Pour les deux pays l'éducation mesurée par les années d'études, se trouve être endogène selon les conclusions du test d'Hausman (1978).

Pour éliminer une autre source potentielle de biais, dans l'estimation du modèle de Mincer, i.e. le biais de sélection, la classique méthode à deux étapes de correction proposée par Heckman (1979) a été appliquée. Par cette méthode le biais de sélection a été trouvé positif et statistiquement significatif pour les deux pays. Cela signifie que, dans les deux pays, les personnes qui se sont retirées du marché du travail auraient gagné moins que les participants effectifs, si elles avaient décidé de rejoindre la population active en tant que travailleur salarié.

Dans la littérature relative à l'estimation du modèle de Mincer, nous avons noté qu'il y a très peu d'études qui corrigent les deux sources de biais simultanément et aucune étude de cette nature n'existe pas pour la France ou le Pakistan. Par ailleurs, les changements dans les coefficients bruts, puisque non corrigés de la plupart des facteurs explicatifs, concluent dans des directions différentes suivant que sont appliquées les corrections pour l'endogénéité de scolarité ou des corrections relatives au biais de sélection d'échantillons dans les spécifications séparées. Donc, en réponse, nous estimons ici une seule spécification corrigeant de manière simultanée le biais de sélection de l'échantillon et le biais d'endogénéité de l'éducation.

Nous avons également noté, toujours d'après la littérature, que la robustesse des hypothèses du modèle linéaire utilisé pour estimer le modèle de Mincer a rarement été discutée et testée. Certaines des études se sont intéressées aux questions d'hétéroscédasticité de cohérence des erreurs types, mais les études qui ont formellement testé la présence d'hétéroscédasticité dans le terme d'erreur du modèle de Mincer sont très rares et n'existent pas, à notre connaissance, pour les deux pays que nous prenons en compte ici. Nous avons donc testé formellement la validité de l'hypothèse d'homoscédasticité, cela en appliquant le test de White (1980). Pour les deux modèles, basés sur les données françaises et les données pakistanaises respectivement, les erreurs sont jugées comme hétéroscédastiques. Donc, afin d'éviter les effets de l'hétéroscédasticité des erreurs sur le processus d'estimation, nous avons réalisé une

estimation adaptative du modèle de Mincer. L'application de l'estimation adaptative, construite pour limiter l'effet sévère de l'hétéroscédasticité, a aidé à produire une estimation plus efficace, plus fiable et plus représentative des impacts que possèdent les différentes variables explicatives du processus de détermination des salaires dans les marchés du travail de ces deux pays.

En général, dans le cadre du modèle de Mincer, une estimation semi-paramétrique contient elle-même une composante non paramétrique qui provient de l'équation de sélection. Mais notre estimation semi-paramétrique est différente, dans le sens qu'elle contient une composante non paramétrique provenant de l'équation de scolarité ; cette dernière est obtenue, en première étape, de l'estimation en doubles moindres carrés avec variables instrumentales 'IV2SLS'. À notre connaissance, c'est la première étude semi-paramétrique qui porte sur l'estimation non paramétrique de l'équation de scolarité.

Basées sur la performance globale des modèles paramétrique et semi-paramétrique, nous avons constaté que, pour la France, les deux formes d'estimation apparaissent bien spécifiées. Toujours dans l'idée de maintenir la facilité d'estimation, le modèle paramétrique a été sélectionné afin d'être le plus approprié pour les données françaises. Pour l'analyse du Pakistan, nous avons conclu que le modèle semi-paramétrique produit des résultats en désaccord avec l'agrément général au Pakistan, mais aussi en rapport à la littérature internationale pour certaines des variables. Ainsi, la performance globale du modèle semi-paramétrique n'est pas très différente du celle du modèle paramétrique. Donc, comme pour les données françaises, pour les données pakistanaises, nous avons aussi choisi le modèle paramétrique comme le plus robuste qu'afin d'estimer les impacts exercés par les différents facteurs explicatifs sur le processus de la détermination des salaires. Pour les deux pays, après avoir comparé les versions simples et adaptatives du modèle paramétrique et du modèle semi-paramétrique, nous avons trouvé que le modèle paramétrique dans la spécification adaptative est plus performant dans l'objectif d'estimer les impacts des différents facteurs contributifs au processus de détermination des salaires.

Enfin, nous avons estimé le modèle de Mincer dans une forme paramétrique choisie de ces estimations, comme le plus approprié en rapport à la forme semi-paramétrique, et à partir de l'analyse de régression en moyenne, comme pour le modèle de régression par quantile.

Pour la France et le Pakistan, à notre connaissance, ce travail serait aussi la première étude qui analyse les deux principales sources du biais simultanément (biais de la sélection et biais de l'endogénéité de l'éducation) dans le cadre de la régression par quantile.

La méthode de régression par quantile a révélé que la plupart des variables explicatives influencent les gains salariaux, ceci différemment suivant les différentes parties de la distribution des salaires, pour les deux marchés du travail considérés.

Mots clés : Biais de Sélection, Estimation adaptative, Endogénéité, Estimation semi-paramétrique, Fonction de gains, Hétéroscédasticité, Modèle de Mincer, Régression par quantile, Variables Instrumentales

JEL classifications: C14, I2, J31, P52

Chapter 1: Introduction

1.1 Background and Motivation

The term human capital is used for the knowledge, skills, or competences that individuals possess. These skills or competences come through schooling, training, social environment and network, family background, personal characteristics and many other factors. The competences or skills or more compactly one's human capital increases his/her capabilities to perform certain activities in the work market for economic gains. Therefore, we see in the job market, there are different economic rewards for different individuals depending on the stock of human capital they have.

Human capital theory (Mincer, 1958, 1974; Becker, 1964) states that education and training causes increase in the productivity of individuals by increasing their skills and knowledge. This increment in skills and knowledge results in increased earnings for individuals. So under human capital theory, education and training are the key factors for economic performance of an individual. According to this theory, like production process of any physical capital, which comes through input factors, production of human capital is also a result of investment in education and training but contrary to physical capital, human capital is not transferable without any training process. The Mincer's proposed econometric specification for relationship between wages and accumulated human capital is actually based on Ben-Porath (1967) schooling model according to which people make their decisions about schooling investments in a way that maximizes the net present value of their earnings. Putting in simple words, people's decisions about investment in acquiring more schooling are based on comparison between present possible earnings if they stop schooling and future perceived earning if they decide in favour of acquiring more human capital through schooling process.

The relationship between earnings and schooling has been studied for long. For example, Walsh (1935) computed life time earnings for people with different educational attainments. He used five different sources of data from United States. His results revealed that people with higher schooling attainments have higher expected life time earnings.

The pay-off to education is studied in different ways, before the emergence of human capital theory; most studies estimated the rates of returns to education via cost-benefit approach. In this approach, rates of returns has been computed by equating the costs of educational

investment (which includes direct costs of schooling process and foregone earning during the investment period for more schooling) and discounted benefits in the form of expected future earnings after the investment for extra schooling. This cost-benefit approach has been used in many studies around the world (see Psacharopoulos, 1973 for a detailed review). But after the concept of human capital theory by Mincer (1958) and Becker (1964) and proposal of econometric model by Mincer (1974) in his famous book "Schooling, Experience, and Earnings", this cost-benefit approach has rarely been used. Mincer's (1974) proposed earnings function is so widely used that now it is also referred as the Mincerian function or Mincerian wage regression model. Covering different theoretical aspects Chiswick (2003) has documented the development of Mincer's human capital earnings function and its long time impacts on the empirical literature

In this model, natural logarithm of earnings or wage is taken as a function of the key determinants of the accumulated human capital. Mincer's (1974) proposed specification considers log of observed wages as a function of linear term in schooling and both linear and quadratic terms in labour market experience. Algebraically,

$$\ln W_i = \beta_0 + \beta_1 (\text{Schooling})_i + \beta_2 (\text{Experience})_i + \beta_3 (\text{Experience})_i^2 + \varepsilon_i \qquad (1.1)$$

The schooling is measured in the number of years of schooling. Contrary to the said cost-benefit approach, Mincer's model assumes that only costs due to investment for more schooling are the foregone earnings which an individual would have earned if he had stopped schooling earlier.

Rates of returns to different human capital factors are helpful for the students to make their decisions regarding investing in education by viewing their future possible earnings. Also important to policy makers to decide about the resource allocation, decision to discourage or eliminate the discrimination in economic rewards against different geographic, ethnic belongings of the people as well as sex and different age cohorts. Moreover , to invest more resources to particular sectors of economy with respect to economic, regional, social, political or demographic factors. Particularly, the returns to education measure can be used to see the productive performance of education which is believed to increase productivity and monetary gains not only at individual level but to enhance the economic growth at macro level as well. Development of human capital is a must for any economy to progress. In fact in this era of

globalized economy, any economy cannot progress without having a labour force possessing strong base in human capital factors like education, skills, competences and professional training. The importance of human capital for the growth and development of countries and nations is stressed by many economists, as in the words of Robert E. Lucas (1993) "The main engine of growth is the accumulation of human capital—of knowledge—and the main source of differences in living standards among nations is a difference in human capital".

After addition of the Mincerian wage regression to the literature devoted to relationship between education and earnings, a number of studies have used this model to explore the wage gaps between groups that differentiate with each other due to different factors. For example Chiswick (1983b) compared rates of returns to education for American born Jews with those of white people belonging to other ethnic groups. Tomes (1983) and Kuch and Haessel (1979) are other examples that used the Mincerian earnings function to see the differences in returns to education based on religious affiliation of the people.

The Mincerian model is also used to estimate the difference in the effects of urban and rural area labour market conditions on wages of workers (Snipp and Sandefur, 1988; Johnson and Chow, 1997; McLaughlin and Perman, 1991). Studies that used the Mincerian wage regression to capture the difference in wages or returns due to ethnic belongings include Gwartney and Long (1978), Kimmel (1997) and some others. Some authors have also used the Mincerian wage regression in order to see the changes in economic rewards to education using the data from before and post communist transitions in a number of countries (see Svejnar, 1999 for a review).

Similarly, Papanicolaou and Psacharopoulos, (1979) investigated the impact of father's occupation on earning of their children via estimating the Mincerian function. They estimated wage regression for 9 different occupational groups (of father's) separately.

The Mincerian wage model has been used in many studies focused to see the differences between earnings of different racial groups, differentials due to working in public or private sector , wage gaps between males and females (Chiswick, 1983a; Quinn, 1979; Shapiro & Stelcner, 1989; Assaad, 1997 among many others). Mincer's human capital wage model has also been used in some studies (like Kahn, 1998; Robinson & Tomes, 1984; Simpson, 1985 and some others) to investigate the impact of union status of workers on their wages.

Much growing interest in the Mincerian wage regression in the empirical economics literature, lead us to estimate the above said model for the French and Pakistani Labour Markets' data. This will help us to produce the impacts that different cofactors put on the process of wage determination in the labour markets of the two countries under consideration in the present research work.

We have chosen to estimate the Mincerian model for the French and Pakistani labour force data because we found labour force surveys of these two countries to be very similar in nature and in data collection process. Moreover, as we are interested in specification and econometric problems, so we believe it as worthy to apply the proposed methodology on data coming from a developed and a developing country.

1.2 Objectives

The present thesis is about the estimation of the Mincerian wage model for data from the French as well as Pakistani labour markets by addressing its econometric issues. Precisely, objectives of the thesis are as follows:

1. Proposal of two new instrumental variables in order to address the possible bias due to the endogeneity of schooling.

2. Inclusion of some new variables like sector of employment, type of contract etc.

3. Addressing major sources of bias (endogeneity of schooling, measurement error and sample selectivity) in a single specification.

4. Comparison between the parametric and semi-parametric estimation techniques. Semi-parametric approach will be based on non-parametric component from the first stage schooling equation instead of that from selection equation.

5. Testing homoscedasticity of error term and applying the adaptive estimation approach on the Mincerian wage regression if there is violation of the assumption of homoscedasticity.

6. Application of interval estimation to give a range of coefficients associated with different independent variables.

7. A quantile regression analysis for the Mincerian model for both countries; France and Pakistan, in order to get a picture that how different explanatory factors are related to wages of individuals in different parts of conditional wage distributions of the two labour markets.

1.3 Outlines

The present thesis is structured as follows, in Chapter 2; we review the existing literature concerning estimation of the Mincerian wage regression, its problems and techniques used to overcome these problems. In Chapter 3, we review some of the studies based on the Mincerian framework with special reference to France. This chapter also deals with description of data and estimation of the Mincerian wage regression as well as estimation of the selection equation for the French labour force data. Chapter 3 also gives description about adaptive estimation of the model under consideration and about the semi-parametric estimation of the Mincerian wage model which includes non-parametric estimation of first stage schooling equation through LOESS regression. In Chapter 4, some studies regarding estimation of the Mincerian model for Pakistan are reviewed. Then we describe the data used for estimation of selection equation and wage regression for the Pakistani labour force data. In both of the chapters, 3 and 4, we try to propose the most suitable specification for estimation of the Mincerian wage regression for the French and Pakistani labour markets, respectively. In Chapter 5, we apply a quantile regression approach to the preferred (parametric or semi-parametric) models for both countries. Finally, Chapter 6 briefly summarizes and concludes the findings of the present thesis.

Chapter 2: The Mincerian Wage Model and Econometric Issues: A Review

2.1 The Mincerian Wage Model

The relationship between wages of individuals and the accumulated human capital is an important topic of research in the present era. Mincer (1958, 1974) developed a relationship in which he used the main components of accumulated human capital as determinants of the wages earned by individuals in the labour market. As we discussed in Chapter 1 that Mincer's (1974) proposed model takes natural log of wages as a function of linear term for schooling and linear and quadratic terms for experience (Eq. 1.1) but now it is very common to include other social, demographic, and regional wage affecting factors in the set of explanatory variables of the Mincerian semi-log specification. The extended version of the Mincerian wage model (Eq. 1.1) can be algebraically represented as,

$$\ln W_i = \beta_0 + \beta_1 (SCH1)_i + \sum \beta_K (X_K)_i + \varepsilon_i \qquad (2.1)$$

where W_i represents monthly wage of *i*th individual, *SCH*1 represents educational attainment measured in number of years of schooling, X_{Ki} is the value of *i*th individual for *k*th explanatory variable, β_K are the coefficients associated to K explanatory variables respectively and ε_i is error term of the model assumed to follow normal distribution with zero mean and a constant variance. The *Xs* are the variables that are included in design matrix other than schooling, like linear and quadratic terms for labour market experience and other variables believed to affect the wage determination process in relevant labour market at individual level. As we are estimating the Mincerian model for labour markets of France and Pakistan, so there may be some variables that differ in both analyses. The data source and the description of the variables are given in the relevant chapters dedicated for the estimation of the Mincerian model for France and Pakistan.

Some assumptions of the basic Mincerian model are given as follows:

1) All individuals are identical other than their difference in education and training (which Mincer captured through work experience). In other words the Mincerian

model assumes that all individual are with equal ability, have equal opportunities and had their brought up in similar environment.

2) Only cost of schooling is the forgone earnings due to additional schooling. In other words there are no direct costs of schooling.

3) In addition to 1) and 2), all assumption of a standard linear model must also be fulfilled in order to have the true impacts of different explanatory factors on wages estimated and to draw correct and reliable inferences and conclusions.

Under the above assumptions, the Mincerian model given in (2.1) is generally estimated using the OLS approach to get the impacts that different variables have on the wages of workers in labour market.

The proposed regression has extensively been used in the literature devoted to returns to education and wage differences (Belzil, 2006). In fact, recently almost all studies focusing on the relationship between education and wage determination process have been conducted via estimation of the Mincerian Wage regression (Card, 1999). In the words of Robert Willis (1986), "Mincer earning function has been one of the great success stories of the modern labour economics. It has been used in hundreds of studies using data from virtually every historical period and country for which suitable data exists". Mincerian equation's popularity lies in its simplicity and availability of relevant data (Guille & Skalli, 1999).

Due to its wide range of applicability in the present day labour economics and economics of education as a research tool, there exist a large number of studies that used the Mincerian earnings function in order to estimate the rates of returns to schooling, hence it is not possible that every study using the Mincerian model to get covered but here, we give a brief review of some of these studies.

Psacharopoulos (1977) estimated the Mincerian wage model for a sample of the Moroccan male workers. They used potential experience calculated by (Age - schooling - 6). After entering schooling and experience in different forms, they confirmed the Mincerian specification by comparing the proportion of explained variation from different specifications they used. They reported rate of return for each additional year of schooling as 15.8%. A

relatively small sample of 1600 was used in this study and also this study was restricted to males only.

Using data for British males, a study is carried out by Psacharopoulos and Layard (1979) who estimated the Mincerian wage regression and reported marginal returns to schooling of about 9%. The explanatory power of the model remained near one-third for the different specifications. Although they had a sample of reasonable size but it was limited to men and also they did not included any other factors which might have worked as determinants of wages. Their results from weekly and annually earnings were not largely different. Behrman et al. (1985) also applied the Mincerian approach using a sample of pre-revolutionary Nicaragua in order to estimate the rate of return to schooling. The Mincerian model was separately estimated for men and women and also for three categories of urban, rural or metropolitan areas. In addition to the standard human capital variables, they have used region, migration, health and nutrition as controls. They reported higher schooling returns for females (12-13% in urban and 5% in rural areas) as compared to those for men (7-10 % in urban and 3.7% in rural areas). To see the changes in returns to education in Israel, Weisberg (1995) estimated Mincer's model using data collected in 1974 and 1983 respectively. They found that returns to education increase over the period considered. They reported higher returns to education for higher educational levels.

Chiswick (1983b) compared the rates of returns to schooling for the American born Jews with those of White people from other ethnic belongings. Beyond the standard human capital variables of schooling and experience they estimated the Mincerian wage regression including some other control variables for region and type of residential area. They reported higher returns for Jews as compared to other White ethnic groups. Their results provide support for the inclusion of hours worked and regional controls. Papanicolaou & Psacharopoulos (1979) using data from UK males, compared educational effects on wages for different groups depending on father's occupation via estimation of the Mincerian function. Estimates from wage regressions of 9 different groups gave different rates of returns that varied between 7 to 14% depending on the occupational group of father.

There are many researches that used Mincer's model to capture the ethnic differences. For example, Chiswick (1983a) used Mincer's framework to find the wage differentials between Whites and Asian Americans as well as the differentials among three major groups of Asian

Americans. The sample covered only 25-64 years old men. They found no support for any discrimination against Asian Americans compared to Whites as they reported similar returns to schooling for Whites and Asian Americans except for Filipino group but this group was also lower in schooling attainments and other factors affecting earnings. Their results showed significant effects of regional and urban-rural differences but these effects were different for different racial groups. This creates a support for the inclusion of controls for rural-urban as well as for different regions. For Pakistani analysis, our data permit us to have both these variables included in the human capital earnings function but due to data availability constraint, we will only be able to control for rural, urban and Parisian region in case of French analysis.

There are many studies that have used Mincer's human capital wage regression to explain the differences between different groups due their ethnicity. For example, Gwartney and Long (1978) used the Mincerian model for investigating the differences in earnings due to racial differences in the US labour market. By comparing 9 different regressions for different racial groups, they found substantial differences in returns to schooling in different racial groups for both males and females. They used hours worked and it was found to have significant effects on earnings in all regressions. Marital status found significant for males except for one minority group while insignificant for females. Location of individuals was significant with expected signs except two minority groups. Age is used as an indicator for the labour market experience. An interesting result supporting human capital theory was that, they found high relative earnings for Japanese minority consistent with their higher investment in acquiring human capital via schooling, and lower returns for Mexican Americans consistent with their lesser schooling. Kimmel (1997) also compared the wages in rural US workers. They found that in rural areas, males belonging to American Indians community and females from Black ethnicity are economically less rewarded in comparison to males and females from other groups of males and females of White ethnicity. Using data from the Canadian labour market, Patrinos & Sakellariou (1992) decomposed the wage differential based on ethnic affiliation of the workers. They used the Mincerian model for this purpose and compared results for the people belonging to Indian ethnicity with non-Indians. They reported a considerable difference in returns to human capital indicators of schooling and labour market experience in favour of non-Indians. Contrary to this Canadian evidence , Sandefur and Scott (1983), using the Mincerian wage model to explain the wage differences among minorities in US, showed

higher returns to schooling in favour of Indians in comparison to other ethnic groups and in favour of Whites in comparison to Black people. Mwabu & Schultz (1998) also worked on the racial differences in rates of returns to education in the South African labour market.

The Mincerian wage regression has been used by many researchers in order to estimate the wage gaps between public and private sector employees. Like, Smith (1976) while examining the impacts of the reforms policy for federal US employees estimated wage differentials between federal and private employees, using human capital earnings model. They estimated separate regressions for the federal and private sector workers using data for 1960 and for 1970. They found that returns to schooling were similar or higher for private workers in 1960 but higher for federal employees in 1970. As far as the returns to experience are concerned, it was same or higher in favour of federal workers for data sets from both years. It provides support for inclusion of public and private sector dummies as controls in our analysis for the French and Pakistani labour force data. Lassibille (1998) used Spanish data to investigate the differences between returns to human capital paid by public and private sectors of employment. Researcher revealed that private sector pays higher returns to human capital measures as compared to public sector. His results were similar for both genders. The similar direction of differences between public and private sectors rewards towards human capital factors are also found for Paraguay (Psacharopoulos et al., 1994) and Kuwait (Al-Qudsi, 1989).

The Mincerian function has also been used in literature devoted to effect of religion on the wages of individuals. Korsun (2010) is a recent example that applied the Mincerian earnings regression to see the impact of being religious on the wages of individuals in Ukraine. They reported that being religious lowers the wages in Ukraine. The negative relationship between wages and religiosity has also reported for Canada using human capital earnings function (Dilmaghani, 2011). She also found that Muslims earn lesser compared to people with other religious affiliations. As far as the returns to schooling are concerned, she did not found any significant differences among different religious communities. However, Muslims face lower returns to experience while Jews enjoy higher returns to experience than workers having other religious beliefs. Ewing (2000) is another example that used the Mincerian model to capture the effects of religion on wages of workers.

Mincer's wage equation has also been used for the explanation of wage gaps and differences in effects of different factors affecting wages due to urban or rural locality of the labour market. Snipp & Sandefur (1988) applied human capital earnings function to see the effect of metropolitan or non-metropolitan location of house hold on returns to education using a sample of Indian house holders in Alaska. They found that wages in metropolitan areas were higher while returns to schooling were also higher in metropolitan areas but statistically non-significantly different from non-metropolitan areas. Another study that applied the Mincerian model to see the urban-rural differences between wages and returns to human capital measures is conducted by Johnson & Chow (1997) by using 1988 Chinese House Hold Income Project data. They concluded that being a resident of rural area makes no significant differences in wages. But separate urban-rural estimations uncovered the fact that returns to schooling were higher for rural areas (about 4.02%) than those in urban areas (about 3.29%). However, they found higher returns to experience in urban China as compared to rural areas. Being a female had a negative effect on wages for both urban as well as in rural labour markets.

The Mincerian earnings regression has also been used to see the impact of changes in education earning relationship due to different political or economic transitions, like end of apartheid regime in South Africa (Burger, 2011) and impact due to fall of communist era in many countries on economic benefits of education at individual level (Jolliffe & Campos, 2005; Filer et al., 1999; Pastore & Verashchagina, 2006 and many others). One such study is carried out by Zhang et al. (2005) using data sampled from 6 representative provinces of China, they estimated the Mincer equation for different years' data and concluded that rates of returns to schooling were on the rise after the economic reforms in China. In addition to human capital variables they have also used some controls such as dummies for provincial effects on wages. Chase (1998) investigated the changes in the economic benefits of education for individuals due to changes in the political and economic systems from communist to non-communist systems in Czech Republic and Slovakia. They used Mincer's earning function for the said purpose. According to their estimation, returns to schooling increased in both countries for both genders after ending of communist era. However, the rewards for experience have dropped for both economies. The similar investigations were also done for Bulgaria (Jones & Simon, 2005) and Slovenia (Orazem & Vodopivec, 1994).

2.2 The Econometrics Issues

Despite of the wide ranged applications of the Mincerian wage regression, its simple estimation may be biased due to some econometric problems like endogeneity of schooling, measurement error and sample selection bias (see Griliches, 1977; Griliches, 19779; Ashenfelter and Krueger, 1994; Card, 1999; Bhalotra & Sanhueza, 2004 among many others). In this section we briefly review the concerning literature focusing on the nature and the way that how these issues have been addressed.

2.2.1 Endogeneity of schooling

Endogeneity problem arises when some determinants of response variable are also correlated with an explanatory variable and they are not observable. The endogeneity in the Mincerian model arises due to violation of assumption (1) given in Section 2.1. The violation of assumption (1) is evident as different people cannot be identical with respect to some of their unobservable characteristics coming through different sources like social environment, location, and family background etc. For example, ability can be seen as a determinant of wages in the labour market and on the other hand it may also be correlated with schooling, i.e. more able people tend to get more schooling and also more able people will be more productive at their jobs and hence they will be better paid. If the unobserved ability affects schooling and wages both, then OLS estimation will yield biased results (Griliches, 1977; also supported by Card, 2001 and proven in many other studies). Now in such situation if the ability is not observable (which is very common in real situations) then schooling variable will be correlated with the error term in the wage equation and as a result, the coefficient associated with schooling will be biased. If the unobservable factors or characteristics can be made available, then their inclusion in the wage equation will be sufficient to eliminate the problem of endogenous schooling but most of the time it is not possible to have these unobservables, particularly ability, to get measured.

This kind of bias is known as endogeneity bias and is tackled by using instrumental variables two-stage least squares (IV2SLS) estimation approach. For this, we need instrumental variables which affect schooling but otherwise uncorrelated with wages. An extensive increase in the literature using instrumental variables technique has been seen after Griliches

(1977) who has pointed out endogeneity of schooling as a serious issue that has kept researchers away from estimating the true causal effect of schooling on wages.

2.2.1.1 Instrumental variables two- stage least squares (IV2SLS) estimation

The IV2SLS estimation approach used to avoid endogeneity bias was originally developed by Thiel (1953) and Basmann (1957). The basic idea behind IV2SLS is that endogenous explanatory variable is replaced by its estimate obtained by regressing it on its instruments and all exogenous variables in the model. So the IV2SLS estimation technique is based on two step estimation. In first stage, we estimate the endogenous explanatory variable (schooling in our case) on the instruments and all other exogenous variables in the model. Final wage model is estimated in second stage replacing schooling variable by its values fitted from first stage schooling equation.

We can summaries these two steps as below,

Step-1: First stage schooling regression

$$SCH1_i = \beta_0 + \sum \beta_K (X_K)_i + B_{INS.}(Z)_i + \varepsilon_i \qquad (2.2)$$

where Z represents the instrumental variable X_K is vector of all explanatory variables other than schooling ($SCH1$).

Get $\hat{SCH}1_Z_i$ (predicted value of $SCH1_i$ based on equation 2.2i.e. First-stage Regression)

Step-2: Second stage wage regression

Replace $SCH1_i$ by $\hat{SCH}1_Z_i$ in the wage equation as,

$$\ln W_i = \beta_0 + \beta_1 (\hat{SCH}1_Z)_i + \sum \beta_K (X_K)_i + \varepsilon_i \qquad (2.3)$$

In order to correct for bias due to endogeneity of schooling, the above described IV2SLS estimation approach has been used by many researchers for different countries of the world.

Different kind of instrumental variables have been used in the literature devoted to estimation of the Mincerian wage model via IV2SLS technique. In the following lines we give a review of the studies that tackled endogeneity bias by applying IV2SLS method using different instruments.

2.2.1.2 Family background factors as instruments

Measures on individual's family background has been used as instruments in many studies like, Blackburn and Neumark (1993) estimated wage regression while exploring the role of ability-schooling relationship in the US labour market during the period 1979-1987. They tackled the issues of ability bias and other factors causing endogeneity bias, separately. They used scores in academic, technical and computational tests as proxies for unobserved ability. The magnitude of the schooling coefficient decreased once proxies for ability have been used. For the endogeneity bias coming from sources other than ability, they applied IV2SLS technique using education of parents, variables representing education related material (like magazines, newspapers, library membership etc.) at home and other family background variables as instruments for education. Their results showed that returns to schooling drop from 3.2% to 2.4% when education is instrumented in the presence of ability proxy as a regressor in the wage function. They have used age, marital status, union membership, and area (rural or urban) as control variables in addition to the human capital variables in wage regression. Conneely and Uusitalo (1997) worked on the Finish data set using family background variables and military test scores as instruments to tackle endogeneity of education. They reported higher IV estimates for returns to schooling as compared to those from OLS estimation.

Family background variables have also been used as instruments for individual's schooling by Trostel et al. (2002). They estimated the Mincerian equation for micro data coming from 28 different countries of the world. They used 3 different instruments coming from education of family members namely, spouse's education, father's education and mother's education. They concluded like, Card (1994) and Ashenfelter et al. (1999) that IV estimates are roughly about 2 percentage points higher than the corresponding OLS estimates. Moreover, this finding was found robust to the use of all 3 instruments used for different countries. Zhang (2011) also used spouse's educational level for instrumenting endogenous schooling. They used data from urban Chinese labour markets. They found about 8% returns to education from the OLS

which were lower from IV estimates. Butcher and Case (1994) used sex composition of sibling as an instrument for educational attainment while estimating wage regression for the White women in US. Their idea as they cited comes from the literature saying that women who had a brought up with brothers show masculine behaviours, compared to women who had their brought up with more sisters (Koch, 1955; Brim, 1958) and Becker (1991) suggested "education as more masculine trait". Therefore, it can be thought that women with more brothers may have more education, so use of presence of sisters and number of siblings as instruments may produce an exogenous source of variation for education. Using the above said instruments, their results are quite similar to those of Card (1993), as they reported IV estimates for returns to schooling, as double than those from the OLS specification (18% compared to 9%). Their set of explanatory variables also included different indicators of parental education and other family background variables, like dummies for oldest child and poor household which we believe should have their effect only on participation decision concerning labour market waged work activity.

Despite their use in many empirical studies, family background characteristics have been criticized for being used as instruments because of their direct influence on wages (Flabbi, 1999). Family background factors can directly influence returns to schooling. This may be due to their correlation with other factors for example, area of residence, as different areas may have different schooling as well as labour market conditions. The direct correlation between family background factors and returns to schooling leads to their correlation with wage and that may lead to non validity of such factors as instruments for schooling in the wage equation.

It is also found that family background factors contribute significantly in determination of wages even when used as explanatory variables in the log wage model with schooling. Like Liu et al. (2000) included family background factors as additional regressors and found that these factors have significant effects on wages, particularly in private sector for Taiwan. This may be due to more chances of entrance in better jobs in private sector due to social networking of their families, as families who are well educated and well placed in society, have social connections with people already in better and prestigious jobs. So children belonging to such families may get better jobs due to these social connections (Montgomery, 1991; Schultz, 1988). Similarly, Armitage and Sabot (1987) used the Mincer equation and

showed similar evidence from Kenya and Tanzania by reporting a positive correlation of the schooling effect on wages with rise in the parental educational status. However, evidence of this direct effect of family background factors such as parental education is mixed. For example, a contrasting evidence using the Turkish data is due to Ozdural (1993) who concluded that the parental education only affects one's schooling and do not exert any effects on wages other than through their influence on schooling attainments. A similar result is found for the Brazilian data (Lam and Schoeni, 1993).

Now in the presence of these contrasting evidences, if more educated families affect wages by providing education friendly environment or by providing more financial help for their children, then family background or parental education may serve as reliable instruments for schooling in IV2SLS approach. But on the other hand if such people affect wages of their children by securing better and well paid jobs for them using their social relationships, then family factors may not be good or valid candidates to be used as instrumental variables as they are directly (other than through their effect on schooling) affecting wages in such case.

2.2.1.3 Availability of educational institutions nearby as instrument

There are many studies which have used the presence of educational institution in community or nearby as instrument for endogenous schooling. The idea comes from the notion that existence of educational institution in a locality increases level of schooling in general, in community nearby. The distance to an educational institution may affect schooling levels because living far from school, college or university increases the cost of schooling in different ways like, transportation cost, fatigue, homesickness etc. On the other hand, it can be regarded as a source of exogenous variation keeping in view its independence with wages offered in the labour market. These studies include, Card (1993) who estimated earnings regression for US labour market, by instrumenting schooling on presence of nearby college. They reported results from IV2SLS approach based on instruments of presence 4 year as well as 2 year college, as about 13% while about 7% by using the simple OLS technique. Their results are quite robust with and without different control variables like region, race, parental education and other family background variables. They stated exogeneity of proposed instruments by proving absence of any direct significant effect of instruments on earnings. Kane and Rouse (1993) also used IV2SLS technique in order to identify the returns to education. In their study, distance to nearest 2-year and 4-year College and the state specific

tuition fees were taken as instruments. Their IV estimates for schooling returns were above their OLS counterparts. Maluccio (1998) tried to evaluate the effect of education on wages using the data from rural Philippines. He tackled with endogeneity of education by using different instruments, like distance from secondary school, parents' educational level, household wealth etc. They found that the conventional OLS estimates are downward biased, as the estimates from IV2SLS method are significantly higher than the OLS estimates. Another work using such type of instruments is done by Bhalotra and Sanhueza (2004) who estimated the Mincerian wage regression for South African women. In order to control the bias in the schooling coefficient due to endogeneity, they used scores in mathematics and reading tests to proxy for ability as additional regressors and by applying IV2SLS technique by using dummies for presence, and log of distance for primary and secondary schools in the community as instruments for schooling. The idea is same as that of Kane and Rouse (1993) and Card (1993) that the presence of educational institution nearer plays a role in individual's decision to stay for some more period of time in the schooling process. They reported IV estimates as considerably higher compared to those from the OLS for both specifications when ability proxies used as well as when ability proxies were not used. In addition to the well known human capital variables of schooling and labour market experience, they have used number of children interacted with experience and measures on casual worker and unemployment rates by community as independent variables. But in our opinion these two community variables affect wages through their effect on schooling decisions of the people. So these could be potential instruments for education. These variables can also affect decisions to work in the labour market. The indicator for presence of university/teacher training institution is also used by Warunsiri and Mcnown (2010) as instrument for schooling attainments in Thailand. The idea is that presence of these institutions lowers the cost of and increases motivation for education.

2.2.1.4 Different exogenous factors affecting schooling as instruments

In addition to above mentioned types of instrumental variables, we find in the literature, studies which have used different exogenous sources of variation as instruments for the schooling attainments. For example, Harmon and Walker (1995) using British data reported estimates based on IV2SLS estimation as higher from the traditional OLS estimates. They have used minimum legal school leaving age as an instrument for education. Minimum school leaving age has been changed twice in UK in 1947 and in 1971 from 14 to 15 and 16 years

respectively. The idea is that cohort of people who faced different legislation of minimum school leaving age will have different levels of education which is free of ability and any other effects directly related to wages in the labour market. Their results show a considerable difference between the OLS and IV estimates as IV2SLS approach gave a return to schooling of about 15% compared to about 6% from the OLS method, which suggest the existence of endogeneity bias in the estimates from OLS approach. The differences in the schooling attainments of cohorts facing different laws concerning minimum schooling leaving age found apparent in different studies (Oreopoulos, 2003; Aakvik et al., 2003; Meghir and Palme, 2003). Brunello and Miniaci (1999) also exploited the exogenous variation due to reforms/changes in the educational system as an instrument for schooling attainment of individuals. While investigating the contributory factors in determination of individual's wages in the Italian labour market, they instrumented schooling by an educational reform in 1969 that allowed secondary school diploma holders to get enrolled in college irrespective of the curriculum they studied at secondary school level. The idea comes from the belief that the said reform enhanced the college enrolment and hence schooling level of individuals born in 1951 or after, because they reached the age of college enrolment in 1969 or onwards. They also used family background factors like, education and occupation of father and mother as instruments. As they had more instruments than endogenous variable, they established the exogeneity of instruments by Sargan (1964) test. For schooling coefficient, they reported higher IV estimates (5.7%) than the corresponding OLS (4.8%) estimates. Age used as proxy for experience found to contribute significantly while results about significance of other controls for region and town size gave a mixed picture. The sample was limited to male heads of house hold. They excluded individuals less than 30 years of age because their data contained very less number of such individuals. Individuals more than 53 years of age were also excluded in order to avoid those who started their schooling during world war-II. Another study is due to Flabbi (1999) who estimated the Mincerian wage regression and tried to solve the problem of endogenous schooling by IV2SLS approach as well, for the Italian data. Using similar idea like Card (1993) and Brunello and Miniaci (1999), they used easy access to university and educational system reforms as instruments for schooling respectively. Precisely, they used dummy variables that weather individual at age of 19 lived in a province that had university at that time. For educational reforms instrument, they used dummy variable indicating weather individual had the chance to enter in a schooling level due to reforms. Their results produced higher IV estimates as compared to those from the OLS

approach, but amount of downward bias found different depending on gender and instrument used. The gender differences in the amount of bias were further confirmed by Hausman (1978) test which established instrument validity only for female's sample. Indicator of reforms in educational system had also been used as instrument of schooling for estimating the Mincerian wage function for Malaysia (Ismail, 2007)). This was based on educational system reforms of 1970 in which language of instruction changed from English to Malaysian. The idea was that due to increased understanding of national language, people may feel easy or motivated to remain in schooling process for some more time resulting higher schooling attainment. They documented a downward bias in the OLS estimates for the both data sets used.

Angrist and Krueger (1991) used the individual's quarter of birth in connection to the compulsory schooling laws. As quarter of birth is independent of ability, their assumption was that a child who born earlier in the year will reach the compulsory schooling age earlier with a lower grade as compared to the people born later in the year. Using controls of age, age squared, race, marital status, and work location dummies, they reported IV estimates for returns to schooling to be little higher but statistically non-significantly different from the corresponding OLS estimates. But their reliability was questioned due to low correlation between instrument and schooling by Bound and Jaeger (1996). The relationship between higher and lower schooling attainments and months of birth are also empirically justified (Bono and Galindo-Rueda, 2004), which corroborates the use of such a measure as instrument for schooling attainments of individuals. Month or season of birth had also been used as instruments of schooling in Plug (2001) and Webbink & Wassenberg (2004). Angrist and Krueger (1992) estimated the Mincerian wage regression using draft lottery numbers for inclusion in the military service as an instrument for schooling. The idea is that people are likely to remain in schooling process for some more duration in order to avoid their induction in the military service during Vietnam era. Their reported IV estimates were higher than the OLS estimates, 6.6% as compared to 5.9%.

Following the positive association between low educational levels and bad health habits reported by Fuchs (1982), the smoking status has been used as instrument for schooling in some studies like, Evans and Montgomery (1994) used whether a person smoked or not at age 18 as instrument for schooling attainments. Using five different data sets they have shown

that those who smoked at the said age had relatively fewer years of schooling than non smokers. They reported IV estimates of the returns to schooling ranging between 8% and 10%. The studies by Chevalier & Walker (1999) and Fersterer & Winter-Ebmer (2003) are other examples that used IV2SLS technique by instrumenting schooling on smoking status at some given age. The smoking habit as an instrument is used due its health effects and possible negative impact on schooling levels.

Arabsheibani and Mussurov (2007) used spouse' education and smoking habit as instrument for worker's schooling for estimating the Mincerian returns in Kazakhstan. The idea for spouse's schooling as instrument comes from assortative nature of marriage (Weiss, 1999; Mancuso, 1997; Mare, 1991; Pencavel, 1998; Sviatova et al., 1988, and Liu et al., 2000) that leads to more similarity or positive association between education levels of couples. Another study on the same lines is conducted by Chen and Hamori (2009) for the urban Chinese labour market. They tackled the endogeneity by taking spouse's education as instrument for schooling. However, it differed with Arabsheibani and Mussurov (2007) in the choice of some control variables. They have used belonging to minority group, coastal region, metropolitan area, and sub-urban area as control variables and age instead of experience in log earnings function. They reported IV estimates of schooling coefficient as about 4 percentage points higher than the corresponding OLS one for males while for females the IV estimates were found to be double than the OLS estimates. They reported male's returns to schooling as higher when schooling is treated as exogenous but returns were higher for females after applying the IV2SLS techniques.

As there exist always a potential problem with the use of IV estimation that instruments may affect different sub groups of the population. Taking this issue, Harmon and Walker (1999) used the data from UK and applied IV2SLS estimation using different sets of instruments some affecting educational participation at lower levels and some affecting at higher levels. They confirmed the well know notion that the OLS estimates are downward biased.

The search for the valid instruments for the endogenous explanatory variable is hard. When we have some instruments which are weakly correlated with the endogenous variable, then IV estimates may not be good. Because in such case even a small correlation between instruments and error can lead to biased IV estimates even if large samples were used (Bound et al., 1995). The problem of ability bias in OLS setting still persists in the IV estimates as

they are not very precise in many studies using this technique. Because of the absence of the true randomization, even good instruments may reveal the returns to education of a sub-population affected by instruments (Card, 2001). Despite many efforts using different instruments in IV2SLS technique, the problem of endogeneity bias is still not settled (Angrist and Krueger, 1992).

Different instruments used in the IV2SLS approach are compared in Pons & Gonzalo (2003). They used parental education, father's nature of job, father's type of contract, season of birth, and availability of college nearby as instruments separately for different IV2SLS estimations. Robust to different instruments used, they concluded that the IV2SLS approach produce higher returns to schooling compared to those from the OLS estimation.

From above discussion it comes out that for majority of the cases returns to schooling found higher from the IV2SLS estimation compared to schooling coefficient found from the OLS estimation. Despite of huge literature stating that IV estimates are higher compared to the OLS estimates for returns to schooling, Leigh and Ryan (2008) found that the IV estimates as lower to their OLS counterparts. They compared OLS and IV2SLS estimation results using two different specifications with month of birth and compulsory schooling law as instruments for schooling for Australian data. Such non-significant differences in estimates from OLS and those obtained through IV2SLS estimations are also reported by Duflo (2001, for Indonesia) and by Callan & Harmon (1999, for Ireland) using availability of schooling opportunities (measured by intensity of school construction programme) and family background factors (parents' education and social class) as instruments, respectively.

The above discussion reveals that although, many instruments have been used in the literature to overcome the problem endogenous schooling but this area of research still remains open. This leads us try to compare the results from the simple OLS setting and applying IV2SLS technique with two new instruments for suspected endogenous schooling.

Other than IV2SLS approach some studies also used proxy variables for ability to coup with problem of inherent ability bias like Griliches (1977) used IQ test scores in the wage regression. He argued that the widely used term ability bias in the coefficient of schooling need not to be necessarily positive but it can also be negative. The addition of proxy variables for ability in the wage function (as Griliches did) to overcome ability bias and endogeneity of

education have been used commonly in the literature but it is difficult to search a proxy for the ability such that it is not correlated with education. Kenny et al. (1979) also used student's scores in composite math's tests as proxy for the ability in order to control for ability bias.

The ability bias has also encountered by many studies through fixed effects model using data on sibling and twins (Flores-Lagunes and Light, 2006; Ashenfelter and Krueger, 1994; Behrman ad Rosenzweig, 1999; Miller et al., 1995; and Rouse, 1999). The study conducted using fixed effects model by Ashenfelter and Krueger (1994) used data on identical twins from US to avoid the effects due to ability bias. They used self-reported schooling level and that reported by his/her twin to tackle the measurement error bias. They reported that measurement error makes conventional estimates as downward biased. However, they also used IV technique to account for measurement error. For this they used individual's schooling as reported by other twin sibling as instrument. Their results are according to well known behaviour of returns to schooling literature, that is, the IV estimates are about 2 percentage points higher than the respective OLS estimates. But the method of proxy variables for eliminating ability bias is criticized because it is difficult to have such ability proxies which are not influenced by schooling attainment (Sanroman, 2006). Similarly, the family fixed effects model has been criticized because the bias due to measurement error will be increased than the bias would have been due to measurement error if conventional estimation method is being used (Sanroman, 2006; Behrman and Rosenzweig, 1999). Moreover, ability bias can be tackled with the use of IV2SLS approach as Uusitalo (1999) used Finish dataset to estimate the returns to schooling. They controlled for ability bias by proxy variables and with IV2SLS method. Comparing with the OLS, returns to education and experience were found a little lower when different test scores used as proxy for unobserved ability but both schooling and experience returns were reported higher when the IV2SLS approach used with father's earnings, education and job related measures as instruments for schooling. An interesting result in their study is that test score measures which were significantly affecting schooling when used as proxy for ability, have lost their significance in the IV2SLS estimation. So it hints at the fact that IV2SLS approach corrected the ability bias as well.

2.2.2 Measurement Error

Another potential problem in estimation of the Mincerian wage model is possible measurement error in schooling variable that leads to biased and inconsistent OLS estimates.

The measurement error comes in when the schooling variable is believed to be reported with error. It this case, the schooling variable may have correlation with the error term in the wage model. Bias due to measurement error in the schooling variable is generally known to produce an attenuation bias in the coefficient of schooling (Hertz, 2003; Sanroman, 2006; Miller et al., 1995; Rouse, 1999). Many studies found evidence of measurement error in schooling (Griliches, 1977; Griliches, 1979; Ashenfelter & Krueger, 1994). However, some researchers like Trostel et al. (2002) view this bias as to have very little effect due to high reliability of the schooling data. Moreover, some researchers analyzed sibling's data by assuming schooling data as free of error (for example, see Taubman, 1976; Olneck, 1977).

From the literature review, we see that the measurement error issue is normally tackled in the studies that use fixed effects model or proxy variables in order to eliminate ability bias in returns to schooling. This is because measurement error may be more sensitive in such studies as measurement error is believed to be heightened in the fixed effect models (Griliches, 1977). In fact, these studies tackle the endogeneity that comes only from omitted ability and cope with measurement error separately.

Like in these studies (studies using fixed effects model), to address the bias due to measurement error, we must have two measures/information/values on schooling attainments of each individual. These two measures may come from different points in time, or reported by different people (self-reported and reported by any other person like sibling, parents, husband etc). These two measures are necessarily required in order to calculate the reliability ratio of schooling variable which is believed to be helpful in uncovering the bias in schooling coefficient due to measurement error problem. The conventional OLS estimates are believed to be biased due to measurement error by an amount approximately equal to one minus reliability ratio (Ashenfelter and Krueger, 1994; Isacsson, 1999). The studies which used fixed effects model had always two measures on schooling. For example Flores-Lagunes and Light (2006) used one measure on schooling by individual himself and one from his sibling. Ashenfelter and Krueger (1994), Ashenfelter and Rouse (1998) and Rouse (1999) used one reported by individual himself and other by his/her identical twin. Isacsson (1999) used two schooling reports one by getting information by sending questionnaire to twins and other by register of educational attainments.

Focusing on present thesis, as for both French and Pakistani data sets, we don't have any complementary information on individuals' schooling attainments, so it is not possible to address biases due to measurement error and other unobserved factors including ability separately.

As it is well known that the endogeneity problem arises due to the violation of an assumption relevant to the standard linear model. According to this assumption, the error term and explanatory variables must be uncorrelated. Any violation of this assumption, that is, any correlation between any explanatory variable and error term will result in biased and inconsistent estimates if estimated via OLS method. No matter, what is the reason for this correlation, it may be due to omission of any variables that are correlated with both schooling and error term or it may be due to measurement error in an explanatory variable or due to simultaneous equation bias. In the literature concerning the Mincerian wage regression, endogeneity may be present due to the omission of unobserved ability and other factors correlated with schooling that also influence wages and hence correlated with error term, and it may be present due to the measurement error in schooling variable (Maluccio, 1998; Brunello and Miniaci, 1999).

The bias due to error of measurement can be coped with in two ways. One is to find another measure on the suspected mis-measured variable. The other and mostly used solution in econometrics is to use instrumental variables estimation method, particularly when instruments are highly correlated with the variable possibly affected from measurement error (Hausman, 2001). Bias in the OLS estimates due to unobservable factors and measurement error can be corrected for by the use of IV2SLS technique with suitable instruments (Pons and Gonzalo, 2003; Uusitalo, 1999). In general the standard solution to the endogeneity problem is application of instrumental variables approach (Koutsoyiannis, 1977; Gujarati and Sangeetha, 2007; Greene, 2002).

Moreover, Sanroman (2006) estimated the Mincerian returns to schooling by using Uruguayan house-hold survey data. He tackled the issue of measurement error by applying IV2SLS estimation. They used dummy variable for availability of internet connection at home as an instrument for schooling. They make use of the independence between dummy instrumental variable and wage and correlation between instrument and true schooling. Their sample was limited only to private sector workers working in the capital city of Montevideo.

Brunello and Miniaci (1999) also dealt measurement error and bias due to unobserved factors omitted, by the IV2SLS approach.

The IV2SLS approach for correcting measurement error problem has also been used even in framework of fixed effects model instrumenting first report on schooling by second schooling measure or vice-versa (for example, see Ashenfelter and Krueger, 1994; Ashenfelter and Rouse, 1998; Miller et al., 1995).

Bhalotra and Sanhueza (2004) stated that the IV2SLS approach can also be used for correction of measurement error bias in schooling coefficient in the Mincerian wage regression. Similarly, Flores-Lagunes and Light (2006) concluded that researchers may trust on the IV2SLS approach for eliminating bias in the schooling coefficient due the possibility of measurement error in schooling variable.

So keeping above discussion in view, it seems reasonable to rely on the IV2SLS estimation for eliminating measurement error problem assuming it as a part of overall endogeneity bias. Therefore, we will address the issue of measurement error by using the IV2SLS estimation, that is, as a part of overall endogeneity ("statistical endogeneity" in words of Maluccio, 1998) instead of separately addressing ability bias and measurement error.

2.2.3 Sample Selection Bias

Another very common problem in the literature concerning returns to schooling is bias due to sample selection. The problem of sample selection bias arises due to non random selection of the sample used for the estimation process i.e. when the sample used for estimation is based on only a sub-population. For example in wage regressions we only get the estimates based on those individuals for whom the wages are observed, i.e. who did choose to participate in the labour market as a wage earner. So the differences between characteristics of actives and non-actives may cause the sample selection bias.

If the decision to participate or not, was a random decision then the OLS would be appropriate estimating procedure but it is not a random decision instead it is driven by some other factors.

In the literature dedicated to the Mincerian wage model, the issue of sample selection bias was first prompted by Gronau (1974) by showing that the US females participating in labour force were different in characteristics from those who decided not to be included in the labour force. So Gronau (1974) concluded that due to these differences simple OLS estimation may produce biased estimates for different factors influencing wages in the labour market. The regression applied only on wage workers can produce incorrect inferences about the factors affecting wages due to their possible suffering from the problem of non random selection of the workers in to labour market (Bagheri and Kara, 2005). By using the corrective measures in order to avoid the possibility of sample selection bias, the results may be considered for the whole of the target population. While ignoring this correction, means that results are valid only for the sub-population of people who decided to work in the labour market. This issue may gain further severity when we exclude self-employed, agricultural workers and non-paid workers from analysis because in many economies there may be a sufficient number of people who decide to be self-employed based on the different contributing factors.

In order to overcome this problem, Heckman (1976, 1979) first proposed maximum likelihood (ML) estimator (1976) and then a two step estimation procedure (1979). In empirical literature the two step procedure by Heckman (1979) is more commonly employed than the ML estimator (Vella, 1998; Broekhuizen, 2011) and the model eliminating sample selection bias by this method is sometimes referred as Heckman sample selection model (HSSM).

Heckman's (1979) sample selection model that eliminates the possible sample selection bias is estimated in two steps. In first step, a participation or selection equation is estimated by maximum likelihood probit regression, in which decision to work in labour market or not is used as response variable that depends on different explanatory factors. From the coefficients estimated from probit regression, Inverse Mills Ratio (IMR) is calculated. In second step, wage function is estimated with IMR as an additional regressor that will account for the bias due to non random nature of the sample of wage earners. A significant coefficient for IMR points at the presence of the sample selectivity. The two steps involved in the estimation of the sample selection model with the Heckman approach are described as follows,

Step-1: Estimation of selection equation by ML probit regression

$$ACTIVE = \begin{cases} 1 & if \ \ person \ is \ involved \ in \ waged \ \ work \ \ in \ \ labor \ \ market \\ 0 & Otherwise \end{cases}$$

$$\text{ACTIVE}_i = \alpha_0 + \sum \alpha_K (V_K)_i + \eta_i \qquad (2.4)$$

where V_K are different explanatory variables that affect likelihood of participation of individuals in to waged work and α_K are the respective coefficients associated to these contributory factors. This probit regression model tells which factors contribute significantly in enhancing or reducing the probability of a person to be a wage worker in the labour market. From the above probit regression (Eq. 2.4), the IMR is calculated by following relation,

$$IMR_i = \frac{\phi(V'\alpha)}{1 - \Phi(V'\alpha)} \ , \qquad (2.5)$$

where $\phi(.)$ and $\Phi(.)$ are the density function and distribution function of the standard normal distribution, respectively.

Step-2: Estimation of the wage regression

In this step, IMR is added as an additional explanatory variable in the Mincerian wage model (Eq. 2.1). By introducing the IMR as an additional explanatory variable, the wage model takes the following form,

$$\ln W_i = \beta_0 + \beta_1 (SCH1)_i + \sum \beta_K (X_K)_i + \beta_{IMR} (IMR)_i + \varepsilon_i \qquad (2.6)$$

In the above equation, a significant coefficient associated to IMR will hint at the presence of sample selection bias. Although some authors (like Cosslett, 1991; Heckman and Robb, 1985; Ichimura, 1993; Klien and Spady, 1993) have proposed semi parametric estimation of the sample selection model in the sense that they estimate selection equation in some non-parametric or semi-parametric way. We will concentrate on the parametric estimation of the selection equation as Vella (1998) applied Heckman (1979) parametric and Newey (1991) semi parametric two-step procedures on a sample of US female workers to correct for the possible sample selection bias. He found similar results from the two said correction

techniques but those results were substantially different from those obtained from the simple OLS estimation which ignores any bias due to sample selectivity.

García et al. (2001) while estimating returns to education for Spanish data used the Heckman (1979) two-step procedure to correct for the bias coming from the non randomness of wage earners' sample. Kim (2011) used Heckman's two-step estimation method to correct for possible sample selection bias in the Korean labour market data. He reported significant differences between the results obtained with and without sample selectivity correction. The said two-step estimation has also been used by Arabsheibani and Mussurov (2007) for Kazakh data set. This is one of the studies that corrected for both endogeneity and sample selection biases at the same time. After correcting for sample selectivity, their IV estimates for returns to schooling dropped by 3 percentage points. However they applied this correction to only female's sample. Coefficient associated with IMR found significant and negative implying that non-participants would have been better paid if they had opted to work in the labour market. Similarly, Chen and Hamori (2009) estimated the Mincerian sample selection corrected model using data from urban China. By applying Heckman (1979) two step estimation they found economic returns to education to be significantly different from the estimates not corrected for the sample selection bias. The coefficient related to IMR found positive and significant as 0.3568. It implies that labour market participants had an advantage over non-participants which they have also verified by showing difference between schooling levels of workers and non workers. Like Arabsheibani and Mussurov (2007), Chen and Hamori (2009) also applied this correction only to the female's sample. This may be due to the possibility of the lower participation rates of females for both countries. The said two-step estimation method has also been used by Zhang et al. (2005) for the separate Mincerian models for years 1988 to 2001 for China. They found no significant evidence of the sample selection bias except for the years after 1997 for women and the year 2001 for men.

This said procedure is also used by Arrazola & Hevia (2006) in order to correct bias due non random nature of the female sample for Spanish case. The Heckman (1979) two step procedure is also used by many other studies estimating the Mincerian earnings regression (like Kara, 2006; Agrawal, 2011; Horowitz and Schenzler, 1999 and many others)

There are some studies that reported statistically significant but very little impact of sample selection bias on the estimated coefficients of human capital earnings regression model. Like,

Bagheri and Kara (2005) used Heckman's two-step approach for Turkish data and found the bias due to the sample selectivity as small. However they found its effect as higher for females than for males which they attributed to lower labour force participation of females. Asadullah (2006) estimated the Heckman sample selection model to explain the differences in wage structure using data from Bangladesh. They did not find any significant effects due to sample selection bias. Flabbi (1999) as well, did not see any significant bias due to sample selection. They also used Heckman two-step method for sample selectivity correction in the Italian data.

Many studies employed and it is a general understanding that not correcting for the sample selectivity is more harmful for females' sample (Schultz, 1993). However, the problem of bias due to non random nature of the sample is not only limited to females but this problem may have an effect for the males as well (Hoffman and Link, 1984; Kimmel, 1997).

We prefer to include this correction in our analysis because we are also excluding self employed people and employers, from the wage regression estimation. Moreover, we also prefer to include this correction in our estimation strategy because even a small amount of this bias can have a significant impact on measures for the policy makers at macro level (Bagheri and Kara, 2005).

2.3 Parametric and Semi-parametric Estimation of the Mincerian Model

There are some studies that also estimated the Mincerian wage regression in a semi-parametric way. In the literature concerned, the semi-parametric estimation of the Mincerian model typically refers to the estimation of sample selection model in which first step participation equation is estimated by some non-parametric or semi-parametric way and estimation of the wage regression in second step by inserting first step selectivity correction term as additional regressor. They are semi-parametric in the sense that, they tackle the problem of sample selection bias by adding non-parametric or semi-parametric correction term in the wage regression. For example, Martins (2001) compared the parametric and semi-parametric results of the Mincerian wage regression for Portuguese data. Parametric approach

used the well known Heckman's (1976) maximum likelihood estimation of the selection or participation equation assuming joint normality of the errors. While the semi-parametric approach based on the Newey (1991) and Klein & Spady (1993) methods for the estimation of the participation equation. The main wage model becomes semi-parametric as it contains the semi-parametric components coming from the estimation of the selection or participation equation. Semi-parametric estimation produced stronger evidence for the presence of sample selectivity bias as compared to parametric method. Schafgans (1998) compared the wage differences between Malay and Chinese ethnic groups in western Malaysia. They applied both parametric and semi-parametric approaches to correct for sample selection bias in estimation of Mincer's earnings model. For parametric estimation of the sample selection models, they used Heckman's ML and Heckman's two-step procedures. Their semi-parametric approach used the estimator based on the work of Ichimura (1993), Robinson (1988) and Powell (1989). Another similar study using same data is conducted by Schafgans (2000) that differed from (Schafgans, 1998) in the sense that it was not focused to the ethnic wage differences. Lei (2005) used Canadian data to compare parametric and semi-parametric results from estimation of human capital earnings function. Like Schafgans (1998), they used Heckman ML and two step estimator to correct for selectivity bias in parametric framework. For semi-parametric estimation, they estimated first step selection equation by Robinson (1988) method and Newey (1991) method which uses Klein and Spady (1993) estimator. They reported presence of selectivity bias for Canadian females from both techniques. For males, selection bias found significant only from the Newey's (1991) semi-parametric approach. Using Hausman (1978) specification test they preferred semi-parametric approach for selection equation as well as for main wage equation. They also addressed the endogeneity problem of the Mincerian wage regression by applying Instrumental variables approach. But like other studies (Martins, 2001; Schafgans, 1998; Schafgans, 2000) they did not tackled the problem of endogeneity while correcting for sample selection in a single specification. Fernández & Rodríquez-Poo (1997) and Bhalotra & Sanhueza (2004) are among other studies that compared parametric and semi-parametric estimations in similar ways, as they also made use of Klein and Spady (1993) estimator for semi-parametric estimations. See Christofides et al. (2003) for a review of comparison of selectivity models with different parametric and semi-parametric techniques with particular reference to wage models.

Chapter 3: Estimation of the Mincerian Wage Model for the French Data

3.1 A Brief Review of the Mincerian Studies for the French Data

Inspiring from studies in other countries all over the world the relationship between earning/wage and its determinants has also been carried out for the French data by many authors covering different aspects of the French labour market.

Wage structures in the French labour market depend on law and other regulations from central government (minimum wage etc.), on negotiations with trade unions at branch level and on the firm based collective bargaining which is much decentralized particularly after 80s (Barrat et al., 2007). We only mentioned these institutional features to put some light on factors that may affect wages but not taken in the studies estimated in Mincerian framework. This may be a reason that why Mincerian studies have a certain portion of variation as unexplained. Here, we give short review of the French literature about estimation and use of the Mincerian wage regression in different situations.

3.1.1 Common data sources

The major data source for studies estimating Mincer's earnings function for the French labour market is Training and Qualification Survey (Enquête sur la Formation et Qualifications Professionnelles i.e.FQP). The data from different rounds of this survey have been used in majority of the studies focusing on the French labour market (for example, Jarousse & Mingat, 1986; Selz & Thélot, 2004; Boumahdi & Plassard, 1992; Goux & Maurin, 1994; Daly et.al, 2006; Fougère et.al, 2001; Sofer, 1990 all used FQP data coming from different years). Other than FQP, data from Labour Force Surveys (Enquête d'Emploi Continue) have also been used in different studies (Bargain & Melly, 2007 ; Abdelkarim & Skalli, 2005 and Selz & Thélot, 2004 for example). Some other data sources have been used as well. For example Meurs & Ponthieux (2000) used Longitudinal Survey on Careers of Youth (Enquête Jeaune et Carrières) and Simonnet (1996) used Survey on Careers and Mobility (Enquête Carrières et Mobilité).

3.1.2 Estimation and trends

In her doctoral thesis, Riboud (1974) used the French data for first time to estimate education-earnings relationship in the context of the Mincerian model. She used different specifications of human capital variables to explain the earnings differences in France. Returns were reported of approximately 10% for each additional year of schooling and about 8% for each additional year of experience. Introduction of quadratic term for schooling variable did not significantly increase the explanatory power of the model, as R^2 just increased from 0.30 to 0.305. Models estimated by Riboud were based on data on employed people coming from 1964 labour force survey[1].

Jarousse & Mingat (1986) using data from Enquête sur la Formation et Qualifications Professionnelles (FQP: English: Training and Qualification survey) 1977 wave, estimated the Mincerian regression model to explore its applicability for the French data. They defined education variable as number of years of schooling. The schooling years were calculated in the similar way that Mincer (1974) had used to calculate for potential experience i.e., years of schooling were computed by ending year as regular student minus birth year minus 6. Because of richness of data, they had the chance to use actual experience instead of potential experience. From a standard Mincerian approach using annual earning as response variable, they reported 9.9% increase in earnings resulting from each additional year of schooling attained and approximately 6% increase due to each additional year of experience while earnings get their peak with an experience of about 31 years. They have also controlled for full time and part time work by introducing two more explanatory variables representing number of months worked full-time and number of months worked part-time during the reference year. The effect of these measures found to be significant and their inclusion substantially increased the explanatory power of the earnings function. In another specification, the authors included the seniority in the current job in addition to the overall experience variable. This new measure which then became much common in the French studies applying human capital model (as we see later in the review) found to be significantly contributing to earnings but its inclusion did not increase as much the overall performance of the model judged by adjusted R^2, which may be due to its mixed effect with actual

[1] This was the first wave of Labour Force Survey in France

experience. Their results also revealed that non-certified years are not valued in the labour market. Finally, they favoured the robustness of the Mincerian earnings regression for the French data. However, the study was limited to the French born males only. A very detailed study to see the changes in the returns to schooling over the long period from 1964 to 1998 is conducted by Selz & Thélot (2004) who estimated the Mincerian model using 9 different data sets, collected in different years over the mentioned period. Their results showed a decrease in returns to schooling from 1964 to 1985 and after that the returns to schooling remained stable between 8.5% and 9% per additional year of schooling. They reported higher returns to men compared to women but the gap was shown to become narrower over the time. As far as the public and private wage differentials are concerned, they found returns to schooling as lower in public sector for all the data sets. Their results may suffer from biases due to unobserved ability, measurement error and non-random selection in to wage work. Moreover, their results may also be under or over estimated for not taking any variable other than education and experience as it is a common practice now to include some other factors affecting wages. Boumahdi & Plassard (1992) also worked on the relationship between earning and its determinants of human capital and other factors in the French labour market. For this, they applied human capital earnings model using data from FQP 1984. Excluding females, unemployed, associated to military and non-French people, their analysis was based on the French national males, aged between 16 and 65. Other than human capital indicators, the authors included firm size, indicator for private sector and region of residence at survey time to proxy for work location. Their results confirmed that all these factors exerted significant impacts on earnings in the labour market. The authors claimed novelty of their work in the French context as they tried to eliminate the bias arising due to endogeneity of schooling using IV2SLS approach. They used educational levels of parents and number of siblings as instruments for schooling. They reported higher returns to education by IV2SLS estimation (11.3%) compared to the OLS approach (8.7%). The returns to experience found to be similar with both of the estimation methods. The Hausman (1978) test provided evidence against the exogeneity of schooling variable.

Yet another study limited to the male workers was carried out by Goux & Maurin (1994) who used FQP 1993 data for estimation. They estimated a variant of Mincer's proposed human capital model in the sense that instead of common norm they entered schooling and experience both in cubic. The results from their cross-sectional analysis revealed about 7-9%

coefficient associated to the linear schooling term and 2% to the linear experience term. The amount of variance explained by these two variables found to be comparable with other studies. We have discussed only their results from the cross-sectional analysis but they also applied a pseudo panel analysis. Guillotin & Sevestre (1994) is another example of panel analysis.

Like in other countries, the Mincerian wage regression is also used to explain the gender wage gaps in France as well. Barnet-Verzat & Wolff, (2008) investigated the gender wage differentials. They also controlled for age, seniority in firm, location of work, and educational levels. They controlled for any possible sector or firm specific effects, by taking the sample from workers of a specific French firm working in private sector. Their results from the mean regression as well as from the quantile regression showed that females earned less than males. The educational levels, age (as proxy for experience) and job positions found to be significantly contributing in the wage determination. The seniority in job, found to be insignificant in the mean regression and also across the wage distribution as well. The gender wage gap in the French labour force is also studied in Sofer (1990) using data from FQP survey of 1977. They included the seniority in job, time passed out of labour force and of course well known human capital measures of schooling and experience. Experience is taken in linear form instead of general practice of quadratic form. They computed the experience variable, as survey year minus year of entrance in labour force minus the time spent out of labour force during the period between survey year and entrance year. They reported returns to schooling about 8.5% while returns to experience as 3.2%. Separate regression results revealed that men enjoy slightly higher returns to education and experience compared to women. The seniority and duration passed out of market found non-significant in separate regression while in joint estimation, these measures were although found to be significant but with t-values not much above the significance threshold. Finally, they also estimated earnings equations for sub samples based on proportion of females in an occupation. From this analysis, a decreasing trend in returns to human capital variables is found as the percentage of women increase in a particular occupation. This also points at female's concentration in low paying jobs. They did not tackled the potential biases due to problems involved in estimation of the Mincerian model, like bias due to unobserved innate ability , omitted variables bias, sample selection bias and measurement error bias. The gaps between men and women wages

for the French labour market are also studied by some others (like Daly et.al, 2006; Meurs and Ponthieux, 2000 among others).

Using the French labour market data, the Mincerian wage regression has been used to see the effects of different factors and phenomenon on the wages. For example, Simonnet (1996) estimated the Mincerian model while trying to unveil the effects of internal and external job mobility on wages. She used data from French Enquête Carrière et mobilité (Career and Mobility Survey). She considered 1748 individuals after restricting analysis to full time young workers aged between 30 and 35 years. She specified education in 7 dummy variables representing different educational levels and experience in linear instead of quadratic form. After controlling for education, experience, sector of firm's activity and regional effects, she reported that men earn more due to within firm mobility in job position while women may face penalty for within firm change. On the other hand case is reversed in favour of women for the mobility across firms. Further, she also pointed out that differences in male female wages are more resulting from differences in career paths than due to differences in returns associated with the explanatory factors considered. As far as the human capital variables are concerned, the returns to education found higher for men for the most of the educational levels while higher in favour of women for some educational levels that are more related to technical education. The returns to experience were found higher for men. A similar study is carried out by Fougère et al. (2001) who used an extended Mincer type equation to evaluate the effects of on the job training provided by firms to their employees, on the wages of workers and their mobility across firms. They took data from FQP 1993 wave and limited their analysis only to males working in private firms with permanent contract in 1988 and reported wages in year 1992. Wage rate was calculated using annual salaries and taking the periods worked full time and part time into account. This wage rate was then used to calculate wages for all workers assuming as they have worked full time. The schooling variable was specified as levels form which was found significantly affecting wages. By applying maximum likelihood method for joint estimation of selection equation for training, selection equation for mobility and wage equation and by controlling for many other factors (see Fougère et.al, 2001 for details) the authors concluded that training provided by employers lowers the possible negative effect of mobility across firms on wages.

In order to see the effect of different individual and firm related characteristics on wage differentials across different sectors of work, Araï et.al (1996) estimated earnings function using different sets of explanatory variables by the OLS method. In human capital measure they included 4 dummy variables used to capture the effects of different educational levels, age in intervals, and seniority with present firm. Other variables included were dummies for socio-professional category, region, family situation (married, unmarried, divorced etc), nationality, work schedule and type of contract. The results showed significance of human capital variables, also that people with permanent contract earn more than those with fixed term contract or working as trainees. Unmarried persons earn significantly less than married or divorced people, men get 13% higher wages and there exist no substantial effects of nationality on wages. We think that these factors like family situation are more related the selection in to the labour force than to wages but the authors did not estimate selection equation to correct for possible sample selectivity problem. Finally, they concluded that the people with similar endowments get different earnings and there exist significant effects of factors related to firms on wages. Another Mincerian application for the French data is Tahar & Plassard, (1990) that explores the impact of work nature in job and industry of work on the wages of individuals. They used data from FQP 1977 wave. They defined work nature by 6 dummy variables and employment industry by 9 dummy variables. Schooling variables is defined as number of years of schooling while experience in general and specific with firm indicators and duration of training during job were included in the set of explanatory variables. The explanatory power of the different specifications pointed that human capital variables explain the major portion of the variation in response variable followed by the dummies for nature of work in job due to which explanatory power increased by 10 percentage points. The coefficient related to schooling ranged from 4.4% to 9% depending on the specification used, and returns to an extra year of experience ranged from 3% to 3.8%, while returns to seniority or experience with current employer remained stable around 1.5%. Their results also confirmed that women get lower wages compared to men and effect of nationality found insignificant.

There are some studies that compared estimation results of the Mincerian regression for France with other countries, for example Rouault & Kaukewitsch (1998) and Bell et.al (1996) compared the French earnings function estimation results with those from West Germany and United Kingdom, respectively. Both these studies specified 5 levels of comparable

educational levels. Araï & Skalli (1996) compared the French Mincerian returns to schooling with those for the Swedish data and reported returns to schooling as 6% for France. From these 3 comparisons it comes out that returns to education for the French workers are higher than those from Sweden and United Kingdom but lower than those for the West German workers (see Guille & Skalli, 1999, for a more detailed review of these and some other French studies). Another review covering the French literature may be found in Guillotin & Tensaout (2004) who did a meta analysis based on 14 French studies of the Mincerian type, conducted at different times, using different methodologies and based on different data sources. Their Meta analysis produced returns to schooling of nearly 7.5% per additional year of schooling, from the OLS approach. Their results also confirmed international finding (Card, 1994 and Ashenfelter et al., 1999) that the schooling coefficients from IV2SLS estimation are higher than the corresponding OLS approach.

Human capital theory found admissible by Hanchane & Moullet (2000) for the French data when confronted with filtering theory. Their results also found higher IV estimates than those from the OLS method and endogeneity of schooling found evident with Hausman (1978) test. For IV2SLS estimation, they used different factors for tastes and financial potential for education like parental education, father's socio-professional category, rank of individual in his siblings and size of the residence as instruments. For the French data, the human capital theory is also supported by Gurgand & Maurin (2007) using data from different waves of Labour Force Surveys during 1990s. They also confirmed downward bias in the OLS estimates by comparing with IV estimates for returns to schooling. They used birth cohorts of people as source of exogenous variation in schooling attainments based on the idea that people would have different opportunities and costs for schooling depending on their birth cohort.

A relatively recent French analysis on wage determinants under the Mincerian framework is done by Viger (2007). He used data for 4 different time periods for estimation. Focusing on the more recent data of 2002, we see that contrary to some French studies limited to males only, he included both male and female workers aged 16-65 in the analysis. Like Jarousse & Mingat (1986), he computed years of schooling variable as ending year of schooling minus year of birth minus 6, while experience was defined as the difference between current age and the age when the individual left schooling. He also explored the effects of gender and region

on the salaries. But for region he made only distinction between Îl de France region (Paris region) and the rest of France. Results from his (Viger, 2007) doctoral thesis revealed a return to schooling of 7.5% and to experience approximately 3.7% from the standard Mincerian model having education and experience estimated through the OLS approach. The results from extended specification confirmed that men get higher wages than women and also people in Paris region are better remunerated than people in other regions. To eliminate the possibility of endogeneity bias, IV method was used. For IV2SLS estimation, month of birth (like Angrist & Krueger, 1991) and socio-professional category of father (also used by Steunou, 2003) were used as instruments for endogenous schooling. The validity of instruments, found evident from Sagan test (Sargan, 1964, as given in Gujarati & Sangeetha, 2007) while the exogeneity of schooling variable was strongly rejected by Hausman (1978) test. From comparison among two methods, the IV2SLS estimates of returns to schooling were found higher than the OLS ones with an amount of about 40%, while returns to experience were not much different from both techniques. He also corrected for sample selection bias that was found significant with a negative effect. But like general trend in France and all over the world, he did not tackle both of the biases in one specification to produce more robust coefficient estimates. Separate regressions in this study pointed at higher returns to schooling for females compared to males and reverse case with respect to the returns to experience.

As noted in Chapter 2, there are many studies that tried to control for bias due to unobserved ability through proxy variable and fixed effects estimates. But analysis that correct for the ability bias using IQ test scores as proxy for ability or using twins or sibling fixed effect models have not conducted in France due to non availability of such kind of data (Guillotin & Tensaout, 2004).

From the review of the relevant French literature, it is found that there exist different kinds of studies. For example, Jarousse & Mingat (1986), Boumahdi & Plassard (1992), Goux & Maurin (1994) and some others were limited to males only. Selz & Thélot (2004), Goux & Maurin (1994), Daly et al. (2006), Sofer (1990) and Simonnet (1996) were limited only to full time workers. Some studies that specified education as years of schooling (Jarousse & Mingat, 1986; Boumahdi & Plassard, 1992; Goux & Maurin, 1994; Sofer, 1990) and also some that specified education as different educational levels attained (Simonnet, 1996;

Fougère et al., 2001; Araï et al., 1996 and many others). We will not use the levels specification as we are more concerned about the econometric problems related to the estimation of the Mincerian wage regression. For example while addressing the endogeneity problem using IV2SLS approach; it is not very common in literature to take schooling in form of levels. We found some studies using wage rate (Guillotin & Sevestre, 1994; Daly et al., 2006), some using annual earnings (Jarousse & Mingat, 1986; Sofer, 1990; Boumahdi & Plassard, 1992), and some with monthly wages (Barnet-Verzat & Wolff, 2008; Meurs & Ponthieux, 2000 and some others), as the response variable in Mincer's semi-logarithmic human capital model. We see a trend in the French studies of using seniority in current job as an explanatory variable in addition to overall experience (Boumahdi & Plassard, 1992; Barnet-Verzat & Wolff, 2008; Daly et al., 2006; Sofer, 1990; Meurs & Ponthieux, 2000; Tahar & Plassard, 1990). So following the general pattern of the Mincerian approaches in France, for the purpose of comparability and taking advantage of availability of such measure we also chose to include experience with current employer in addition to overall experience for estimation for the French case in the present work. However, the seniority term is also found non-significant in some studies (Barnet-Verzat & Wolff, 2008), and also as we can see in Jarousse & Mingat (1986) that experience when entered alone in a specification gave almost similar coefficient to that obtained from adding experience and seniority coefficients when both entered at the same time in another specification. This kind of evidence can also be found by comparing results from relevant time periods in different studies (for example, comparing Viger, 2007 with Daly et al., 2006 and Meurs & Ponthieux, 2000). From the French results we note that there exist a general trend of higher men's wages and also higher returns for men associated to human capital variables in most of the cases. The type of contract that an individual working under seem a very important determinant of his economic gains from the job but we have found only a few studies that included type of contract in the set of explanatory variables and revealed its significance. Coming to the problems that lead to potential biases in coefficient estimates of the Mincerian wage regression, we find very few studies correcting for sample selectivity for the French case (for example, Meurs & Ponthieux, 2000; Viger, 2007) and also few studies that tackled the endogeneity of schooling variable (Boumahdi & Plassard, 1992; Hanchane & Moullet, 2000; Viger, 2007; Steunou, 2003). There is no study that tried to eliminate both these biases in a single specification (the case is similar to that in all over the globe as we found only a few such studies). For France, general pattern in the IV2SLS approach is to use parental education or socio-professional

category of father, or measure related to siblings as instruments. Size of the residence and month of birth has also been used as instruments by some authors.

From this review, we feel that there exists ample space for estimation of the Mincerian wage regression for the French workers using more recent data, with some new instruments, and eliminating potential biases from sample selectivity and endogeneity of education in one specification, and other robust estimation techniques, in order to provide more reliable estimates for different determinants of wages in the French labour market. Other than introducing new instruments, we also apply semi parametric estimation. Our semi-parametric strategy is different from others in the sense that instead of estimating selection equation in non parametric or semi parametric way, we focus on the possibility non-parametric relationship between endogenous explanatory variable (schooling) and instruments. So we estimate first stage schooling equation under non parametric framework.

3.2 The Data

For estimation of the Mincerian wage regression for the French labour market, we have taken data from Labour Force Survey (Enquête d'Emploi Continu) by National Institute of Statistics and Economic (INSEE: Institut National de la Statistique et des Etudes Économiques) conducted in 2007. The data were downloaded from INSEE website **www.insee.fr**. Data from Labour Force Survey is collected for a large number of measures related to different social, demographic, economic and labour market conditions. We have used some of these measures. We give a brief detail of the variables used in the present analysis for the French data.

3.3 Variables used in Estimation of the Mincerian Model for France

In his pioneering work, Mincer (1974) used linear education term and linear and quadratic terms for experience as explanatory factors for log earning function. So education and experience (or potential experience or ages as proxy) are integral parts of the design matrix of the Mincerian wage function but in addition to these human capital factors, now it is common

practice to add several other control variables that affect wages in the labour market. These other factors include social, demographic, regional, and economic measures.

3.3.1 Response variable for the French wage regression

Log Monthly Wage: Log of annual earnings was used as response variable in Mincer's earning function in pioneering work presented in his famous book 'Schooling, Experience, and Earnings'. But after that work, in following years, different earnings measures have been used in the literature as response variable in the human capital model like annual earning (Flabbi, 1999; Lorenz & Wagner, 1990; Uusitalo, 1999; Wahba, 2000; Jarousse & Mingat, 1986; Boumahdi & Plassard, 1992 and many others) monthly wages (Ismail, 2007; Bhalotra & Sanhueza, 2004; Tansel, 1994; Shabbir, 1994; Siddiqui & Siddiqui, 1998 and many others) and hourly wages (Brunello & Miniaci, 1999; Sanroman, 2006; Liu et.al, 2000; Trostel et al., 2002; Melly, 2005 and many others). And the coefficient associated to schooling (we may expect same for other coefficient estimates) found not to depend much on that which of these measures is used, as it is shown in a Meta analysis for the Portuguese data (Pereira & Martins, 2004). We prefer to include log of monthly wages as response variable because both data sets under consideration in the present work collect this measure directly from respondents.

3.3.2 Explanatory variables for the French wage regression

As it is common now to add other potential determinants of wages as well in the Mincerian wage regression in addition to standard human capital factors. For the French analysis, our set of explanatory or independent variables consists of the following variables:

Schooling: The major determinant of wages considered in the Mincerian wage model is individual's educational attainment. The effect of education on schooling was captured by defining education as number of years attended in school in the basic work of Mincer (1974). Most of the studies use number of years of completed schooling as indicator for educational attainment (Lorenz & Wagner, 1990). But in the available literature, we also found many studies that use schooling in the form of levels of education i.e. they introduce dummy variables representing different educational levels (for example, Tansel, 1994; Hawley, 2004; Falaris, 2004; Albrecht et al., 2003 and many others). In the literature concerning France, we also found both kinds of the studies that define education in levels form (Charnoz et al., 2011;

Barnet-Verzat & Wolff, 2008; Daly et al., 2006; Simonnet, 1996; Fougère et al., 2001, for example) and that define education as number of years of schooling (Jarousse & Mingat, 1986; Boumahdi & Plassard 1992; Viger, 2007; Goux & Maurin, 1994 and others). As noted earlier that we are more focused on the econometric issues related to estimation of the Mincerian model like endogeneity of schooling and sample selection bias. The problem of endogeneity of schooling is not tackled commonly in a specification that takes schooling in levels form as it seems difficult to use fitted values for dummies (for educational levels) in second stage regression. So we define schooling as number of years of schooling. For the French data, we could have followed Jarousse & Mingat (1986) or Viger (2007) who calculated years of schooling by ending year as regular student minus birth year minus 6. But it is not possible to calculate schooling duration in this way for the Pakistani data due to unavailability of measure on ending year of schooling process. So keeping in view the comparability of results from two countries, we have converted levels of schooling into number of years of schooling. The evidence for this kind of conversion from levels into years of schooling is found in many studies (these include Brunello & Miniaci, 1999; Chen & Hamori, 2009; Lassibille, 1998 and others). This type of conversion is also justifiable as Jarousse & Mingat (1986) concluded that non-certified years are not rewarded in the labour market, so certified years can be taken from levels of schooling attained. The brief description about conversion of educational levels into duration of education in years completed is given below.

The Labour Force Survey (Enquête d'Emploi Continu) by INSEE provides information on the highest level of education compatible with UNESCO levels[2]. We attribute 5 years of schooling to individual who reported as having level *«Primaire» (primary),* 9 year to those who reported *«Brevet des collèges ou professionnel court non diplômé»(junior secondary),* 11 years of education to those who reported as having *«Diplômes niveau CAP ou BEP» (basic vocational),* 12 years for those who have attained an educational level as *«Autres diplômes professionnels niveau bac»* or *«Bac pro, bac pro agricole»* or *«Bac général, bac technologique»* or *«Capacité en droit, DAEU» (as senior secondary, general or vocational tracks),* 14 years for *«Dut, BTS»* or *«CPGE,1er cycle universitaire disciplinaire»* diploma

[2] ISCED: International Standard Classification of Education. for details see :
http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx

holders (non university further education and bachelor level), 16 years for *«Université 2nd cycle, écoles niveau licence-maitrise»* (master level), 17 years for *«Université troisième cycle, grandes écoles»* or *«Ecoles d'ingénieur»* or *«Ecoles de commerce»* (Master in engineering, MBA etc.) and 20 years of education for *«Doctorats sauf santé»* or *«Doctorats de santé» (PhD, MD etc.)* and of course, 0 to those with no schooling.

Job Seniority: In the studies concerning France, we have noted a general trend of adding two measures for experience, i.e. general experience and seniority in the current job. We have noted that experience in current enterprise is also contributory in addition to general experience in some studies that used both measures in the same specification (Boumahdi & Plassard 1992; Daly et al., 2006 for both males and Females; Meurs & Ponthieux, 2000 for both males and females). So following these studies, we also include two measures for experience. These two measures are seniority or current job experience and labour market experience prior to current job, we call it as past experience. The job seniority variable is available in Labour Force Survey (Enquête d'Emploi Continu). It is collected in number of months of job seniority; we converted this measure into number of years in order to be comparable with other studies.

Past Experience: Like schooling, labour market experience is also a vital and virtually an integral part of every wage function estimated in the Mincerian setting. Due to unavailability of actual measure of experience in most of cases, it is a common custom to use potential experience. As in the basic work by Mincer (1974), potential experience is generally calculated as age minus years of schooling attained minus 6. Here, 6 is taken as assumed age when individual started schooling, but some also subtracted 5 (Chiswick, 1983a; Abbas & Foreman-Peck, 2007) or 7 (Korsun, 2010; Uusitalo, 1999) instead of 6 from the relation for potential experience. A number of studies used individual's age as a proxy for experience (Gwartney & Long, 1978; Sanroman, 2006; Hyder, 2007; Barnet-Verzat & Wolff, 2008 and many others). The only studies that used actual experience we found are by Palme & Wright (1998) and Callan & Harmon (1999).

We calculate potential experience as subtracting ending year of schooling process (available in Labour Force Survey) from year of survey (i.e. 2007). Then we calculate past experience (labour market experience prior to current job) as difference between potential experience and current job seniority. i.e,

$$\text{Potential EXP} = 2007 - (\text{Ending year of schooling process})$$
$$\text{Past Experience} = (\text{Potential EXP}) - (\text{Job Seniority})$$

So for experience we use two measures, 'past experience (experience before current job)' and 'current job experience or seniority'.

Hours Worked: There can be a tendency that more educated or more skilled workers have opportunities to work for more time. If we control for the number of hours worked in the estimation process, then results may be different (Schultz, 1988). The effect of hours worked on observed wages have been captured by some people (Gwartney & Long, 1978; Lassibille, 1998). As we are using monthly wages as response variable (in semi log wage function), so need to control for hours worked is further strengthened. This measure is available in the Labour Force Survey by INSEE. Therefore, we have added number of hours worked for the reported salary in the design matrix. In order to eliminate the possible effects of extreme values, we have excluded observations having less than 20 work hours per week or more than 40 work hours per week. Statistically, it is equivalent to exclusion of lower and upper 2.5^{th} percentiles of the distribution of hours worked.

Gender: The difference due to gender in labour market monetary gains for individuals is evident in many studies. It may be so because of different reasons, like market preferences in favour of a particular gender, lesser availability of educated and highly skilled women, women (men) may work less (more) due to their household responsibilities etc. Almost every study employing the Mincerian wage regression includes control for gender. We also control for gender wage differences by including a dummy variable in the set of explanatory variables. For the estimation process females will be taken as reference category.

Regional and Location Effects: The effects of working in different regions are found significantly causative in the wage determination process in many studies (García et al., 2001; Bagheri & Kara, 2005; Flabbi, 1999; Heckman & Hotz, 1986; Staneva et al., 2010 and Ismail, 2007 are some examples). Urban and rural differentials are also found considerable in many studies (like Behrman et al., 1985; Chiswick, 1983a; Arabsheibani & Mussurov, 2007; Korsun, 2010). These regional and urban-rural wage differences have also been controlled for in some Mincerian studies concerning France, for example Viger (2007) and Abdelkarim & Skalli (2005) controlled for working in Paris region or not. Similarly, Simonnet (1996) and

Araï et.al (1996) controlled regional effects by introducing dummy variables for 7 and 8 regions respectively. As in Sofer (1990), mostly regional dummies found non-significant so instead of using many regions we control these kind of effects by defining 3 categories i.e. Paris region, non-Paris urban areas and rural areas. Rural area will be used as reference category.

Professional or General Degree: The effect of technical training on earnings has been found considerable in many studies (Jimenez & Kugler, 1987; Gaag & Vijverberg, 1989; Khandker, 1990 and Tansel, 1994), so we feel it essential to capture these effects. We take this by defining a dummy variable for having professional educational degree or not. We have defined those as professional or technical diploma/degree holders who have obtained «Diplômes niveau CAP ou BEP» or «Autres diplômes professionnels niveau bac» or «Bac pro, bac pro agricole» or «Capacité en droit, DAEU» or «Dut, BTS» and those who have engineering degree «Ecoles d'ingénieur» or Commerce or «Ecoles de commerce» or medical «Doctorats de santé» degrees.

Work Sector: Another important factor that affects wages of individuals in the labour market is sector of work i.e. whether an individual is working in public sector or private sector of economy. There are a number of studies that focused on pay gaps between these two sectors of work and found such gaps as significant (like Lindauer & Sabot, 1983; Terrell, 1993; Al-Samarrai & Reilly, 2005; Nielsen & Rosholm, 2001). Many others have also controlled for these effects in estimation of the Mincerian wage model (for example Smith, 1976; Lassibille, 1998; Gardeazabal & Ugidos, 2005) and for France, we found Selz & Thélot (2004) and Simonnet (1996). Following Selz & Thélot (2004) and Simonnet (1996), we considered working in state organization or local authorities or public hospitals or public firms like «la poste» or «EDF-GDF» etc, as public sector workers and others as workers working in private sector of economy. For the estimation process, private sector will be taken as reference category.

Type of Contract: The type of contract under which an individual is working may have a sizeable effect on the wage he gets. There are not many studies that controlled for effects that contract statuses put on labour market wages. We have found one study (Gardeazabal & Ugidos, 2005) that controlled for contract type. For France, Araï et al. (1996) and Skalli (2007) are the only studies we found that controlled for contract types and reported this factor

as influencing wages extensively. This provides support for controlling for the contract type under which individuals are working. We control this by defining 3 categories i.e. temporary workers, those having fixed term contract, and those having permanent contract. For the estimation process, temporary workers will serve as reference category.

The following Table 3.1 gives a summary of variables used for estimation of the Mincerian model for the French data.

Table 3.1: Brief Description of Variables used in Estimation of the Mincerian Model for the French data

Response Variable	
LNWAGE	Natural logarithm of monthly wage of individual from main job
Explanatory Variables	
SCH1	Variable for education, measured in number of years of completed schooling
BEFEX2	Past experience (experience in labour market before current job), measured in number of years
BEFEX22	Past experience squared
EXP2	Current job seniority (labour market experience within present job), measured in number of years
EXP22	Current job seniority squared
HOURS3	Number of hours devoted to monthly salary (i.e. hours worked per month)
DGENDER4	Gender of the individual (Male =1 ; Female=0)
RDRURAL0	Dummy variable indicating whether individual is resident of rural area or not (Reference category)
DNPARIS5	Dummy variable indicating whether individual is resident of urban area other than Paris region or not. (non-Paris Urban=1; else=0)
DPARIS6	Dummy variable indicating whether individual is resident of Paris region or not. (Paris region =1; else=0)
DTYPDIP7	A dummy variable indicating whether individual have degree/diploma in professional education or general education (Professional Diploma=1; General=0)
DPUBLIC8	A dummy variable indicating whether individual is working in Public sector or private sector (Public =1; Private=0)
RDTMPCT0	Dummy variable indicating that individual is working temporarily (Reference category)
DFIXCT9	Dummy variable indicating that individual is working under Fixed Term Contract (Fixed Term Contract=1; else= 0)
DPERCT10	Dummy variable indicating that individual is working under Permanent Contract (Permanent Contract=1; else= 0)

We have used vector X_K (that contains explanatory variables other than schooling) while explaining different estimation techniques in different sections of Chapter 2. Now based on

the variables given in Table 3.1, we can define vector X_K which will be used for estimation of the Mincerian model for the French data as follows:

$$X_K = \begin{bmatrix} BEFEX2, EFEX22, EXP2, EXP22, \\ HOURS3, GENDER4, DNPARIS5, \\ DPARIS6, DTYPDIP7, DPUBLIC8, \\ DFIXCT9, DPERCT10 \end{bmatrix} \quad (3.1)$$

3.4 Variables used in Estimation of Participation Equation for France

We estimated selection equation by the maximum likelihood probit regression as a first step in order to apply Heckman's (1979) two-step approach for sample selection correction. There are different factors that affect the decision of individuals to participate or not as wage worker in the labour market. These factors include different personal, social, household and regional characteristics. Below, we give a brief description of the variables used in the probit estimation of the selection or participation equation for French labour market data.

3.4.1 Response variable for selection equation for France

ACTIVE: The response variable taken for the selection equation is being 'ACTIVE' in the labour market. This is a binary response taking 1 for active and 0 otherwise. By 'ACTIVE', we mean that a person who is engaged in some waged wok. The people involved in self employment are excluded from analysis as their income may be determined by other factors like capital they invest into the self employment activity.

3.4.2 Explanatory variables for selection equation for France

Different measures that affect participation decision of the people have been used as independent variables in the selection equation. The explanatory variables used in estimation of selection equation for France are briefly described in following lines:

Age: Age is considered as an important factor that affects decision of individuals to participate or not as wage worker in labour market. Inclusion of age in the explanatory variables' set for selection equation is reasonable as people's responsibilities grow with growing age which leads them to work in the labour market. Moreover, age is also related to some other factors that affect participation decision for example marriage or availability or non-availability of income from family etc. This is included in almost every study (Bhalotra & Sanhueza, 2004; García et al., 2001; Agrawal, 2011; Kozel & Alderman, 1990 and many others) that estimates selection equation. Keeping its importance in view, we also include age in the design matrix for participation or selection equation.

Education or Schooling: Education has a significant and positive role in the probability of being a wage worker in the labour market (Tansel, 1994). It is also reasonable to add schooling as independent variable in selection equation because people with more schooling have higher chances to get jobs in the work market. In developed countries education plays an important role in participation odds of people in the labour force as people with lower schooling have more chances to remain unemployed for more periods (Brunello & Miniaci, 1999). So like age, education is also an important determinant of decision concerning labour market participation and is used as regressor in many studies (see, for example, García et. al., 2001; Martins, 2001; Riboud, 1985; Burger, 2011; Asadullah, 2006). So we also include education measured in number of years of schooling in the set of explanatory variables for participation equation.

Gender: We add gender in the set of independent variables for selection equation as gender may have an influence on the decision to participate in labour market or not. The evidence for significance of gender exists in literature (for example Asadullah, 2006; Agrawal, 2011). Being female will be used as reference category for estimation.

Marital Status: Like gender, marital status also affects individual's decision to participate in the work market as wage worker or not. Viger (2007) and Lie (2005) reported the significance of the causal effect that marital status have on decision to work or not in the labour market. We control for this by defining dummy variable for married and single or widowed or divorced. Category representing single or widowed or divorced will serve as reference category for estimation process.

Regional Effects: The regional characteristics or conditions may also affect the probability of work activity. This is may be due to differences in opportunities in different areas and different labour market conditions or it may be due to differences in the supply and demand for the skilled and unskilled workers in different areas. We control regional effects by defining 3 regions in similar way as we did for the wage equation i.e., dummy variables for rural areas, non-Paris urban areas, and Paris region. Dummy for rural areas will be used as reference category in estimation of probit regression for selection equation.

Home Ownership: The home ownership may also have an effect on the participation decision of individuals regarding working in labour force as it is used as explanatory variable in selection equation (Burger, 2011). An indicator of land owned by individual (Alderman et al., 1996) is also used as explanatory factor for selection equation which provides a support for the inclusion of home ownership in the participation or selection equation. As we exclude self employed so instead of land owned by individual, home ownership seem more appropriate. Moreover, indicator for home ownership may serve as proxy for income from other sources used as contributory factor in many studies (for example Agrawal, 2011; Buchinsky, 2001; Lei, 2005; Palme & Wright, 1998). Indicator for home ownership may be good proxy for income from other sources in terms of forgone expenditures that would have been expended if did not own their home. French Labour Force Survey data collects the indicator for this measure and that lead us to include dummy variable for home ownership in the set of explanatory variables. People who do not own their home will be used as reference group in estimation.

Financial Allocation: Another measure which may have effects similar to income from other sources on participation decision, is any financial allocation received by individual without working for that. We may expect that people getting any kind of financial allocation or support may have slighter motivation to work as wage earner in the labour market compared to the people not getting any such support. This kind of measure, for example indicator for social grants is used in literature (Burger, 2011). We define people as getting financial allocation if they get any financial support from government like childcare allowance, isolated parents allowance, allowance for disability, widowhood allowance, or unemployment benefits and non receivers otherwise. People who do not get any financial aid will be our reference category for maximum likelihood probit estimation of participation equation.

Professional Degree: We believe that people with professional diploma have higher chances of being involved in waged work as they have increased odds to get a salaried job. We take this variable as it is defined for use in wage equation. People with non-professional degrees or diplomas will serve as reference category in estimation.

Younger Children: The presence of younger children in household may affect the participation decision of individuals. Particularly, presence of younger children in the household may reduce the probability of work participation of women. That may also affect male's decision because of increased needs of the household due to more children and non working of females due to childcare activities. Effects of younger children on participation decision has been controlled for and found significant in many studies estimating selection equation (like Bhalotra & Sanhueza, 2004; Wahba, 2000; Palme & Wright, 1998). Different studies take different age limits to define children as to be dependent children. Some take less than 5 years of age (Gustafsson & Jacobsson, 1985), some take as less than 6 years (Arabsheibani & Mussurov, 2007) while some take this limit as 7 years (Chen & Hamori, 2009) for children to be considered as younger or dependent. We define this limit to be as 6 years of age i.e., a child less than 6 years old will be considered as dependent child. We take this measure as the number of children under 6 years in the concerned household.

The following Table 3.2 gives a summary of variables used for estimation of the participation equation through maximum likelihood probit regression for the French data.

Table 3.2: Brief Description of Variables used in Estimation of Selection Equation for the French Data

Response Variable	
ACTIVE	A dummy variable indicating whether the person is ACTIVE in the Labour Market or not (AVTIVE =1; else=0)
Explanatory Variables	
AGE1	Age of the individual in completed years at the last day of reference week.
SCH2	Variable for education, measured in number of years of completed schooling.
DGENDER3	Gender of the individual (Male =1 ; Female=0)
DMSTAT4	Legal marital Status of the Person (Married=1; Single/Divorced/Widowed=0)
RDRURAL0	Dummy variable indicating whether individual is resident of rural area or not (Reference category)
DNPARIS5	Dummy variable indicating whether individual is resident of an urban area other than Paris region or not. (non-Paris Urban=1; else=0)
DPARIS6	Dummy variable indicating whether individual is resident of Paris region or not. (Paris region=1; else=0)
DHOMOWN7	A dummy variable indicating that whether individual owns his house or not (Owner=1 ; else=0)
DALLOC8	A dummy variable indicating that whether person is getting any kind of financial aid/allocation or not (Yes=1 ; No=0)
DTYPDIP9	A dummy variable indicating whether individual have degree/diploma in professional education or general education (Professional Diploma=1; General=0)
CH6Y10	Number of children under 6 years of age in the house hold to which an individual belongs

Based on the variables described in above table, the vector V_K used to denote explanatory variables for the probit regression (in Chapter 2, Eq. 2.4) takes the following form for the French case:

$$V_K = \begin{bmatrix} AGE1, SCH2, DGENDER3, DMSTAT4, \\ DPARIS5, DNPARIS6, DHOMOWN7, \\ DALLOC8, DTYPDIP9, CH6Y10 \end{bmatrix} \qquad (3.2)$$

3.5 Instrumental Variables to be used in the IV2SLS Estimation Approach

We propose two instrumental variables which we call Z1 (Instrument-1) and Z2 (Instrument-2). We give a brief overview of these two instruments in the following lines. We apply IV2SLS based on these instruments in different specifications and then based on the performance and relevance; we choose one that can be considered as most appropriate of these two instruments.

3.5.1 Definition of Instrument-1 (Z1)

Many studies have used father's or mother's or sibling's education (Sanroman, 2006; Girma & Kedir, 2003; Boumahdi & Plassard, 1992; Blackburn & Neumark, 1993; Maluccio, 1998 and many others) as instrument for endogenous schooling. So it is reasonable to believe that schooling of an individual may depend on the general trend and motivation for schooling in his family members. Based on these grounds we introduce Instrument-1 (Z1) as "the average educational years in the household". The idea is that in family, different persons may have different tastes for education and different levels of ability. So averaging over the schooling of family members will eliminate the bias due to ability. This instrument is calculated from the data. As it is a data generated instrument, so its benefit is that it can be used in any study which provides some information about house hold education. It will net out ability bias up to some extent.

The instrument-1 may be affected by some measures related to genetic effects about tastes for schooling or innate ability. As a family based instrument, it may also be affected by different social or demographic factors, as members of a family are expected to share these factors.

3.5.2 Definition of Instrument-2 (Z2)

The educational decisions of an individual may be affected by different factors at the time when he has to decide for work or in favour of more schooling. Different people used different instruments for correcting endogeneity of schooling. We propose Instrument-2 (Z2) as "the average schooling in the country in the year in which person entered labour force, of

the age group, at which he entered labour force, calculated by Gender". In other words "it is the average schooling in country, of particular age group, for particular gender, in that particular year when he joined the labour force".

Average schooling for years 1950-2010 were calculated by Barro & Lee (2010). These measure are available at (http://www.barrolee.com/) for many countries of the world. Fortunately these measures were also available for France and Pakistan. We interacted their calculated measures with gender, year in which person entered labour force and age group at which person entered in to labour force.

Our Instrument-2 (Z2) is suitable as it takes into account the effect of different cohorts who had different perceived future benefits for education, have faced different costs of education and other factors that may have changed over the cohorts and generations. As this instrument is computed in such a way that it is based on the factors at the time when an individual has decided to join the labour force, so these factors may be uncorrelated with the earnings/wages of present time. In this instrument ability bias may be eliminated as different people will come from different areas, different families (some with low ability and some with high ability) so it will work as a balance. Also any effect of school proximity will also be net out as it is an overall average. Therefore, our second instrument (Z2) is technically having many instruments of ability, school proximity, family background, school quality etc, in it.

3.6 Estimation Results Based on the French Data

In this section, we present the estimation results of the Mincerian wage regression for the French data using different estimation techniques.

3.6.1 Preliminary estimation

As a first step, we present the results from the OLS estimation of model given in Eq. 2.1 (Section 2.1). The model is estimated using variables described in Table 3.1 for the French

data. For further reference we name this model as **Model 3.1**[3]. In Table 3.3, we give results from the preliminary OLS estimation (Model 3.1) for the French labour force data.

Table 3.3: OLS Estimation of the French Wage Regression

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.5528	0.5525	2578.56	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	4.7342	1.81E-02	261.20	<0.0001
SCH1	0.0670	8.01E-04	83.60	<0.0001
BEFEX2	0.0133	6.45E-04	20.52	<0.0001
BEFEX22	-0.0002	2.01E-05	-11.18	<0.0001
EXP2	0.0184	6.96E-04	26.46	<0.0001
EXP22	-0.0001	2.06E-05	-6.66	<0.0001
HOURS3	0.0064	6.77E-05	94.07	<0.0001
DGENDER4	0.1687	4.19E-03	40.26	<0.0001
DNPARIS5	0.0117	4.86E-03	2.41	0.0159
DPARIS6	0.2212	6.78E-03	32.65	<0.0001
DTYPDIP7	-0.0498	4.18E-03	-11.91	<0.0001
DPUBLIC8	0.0049	6.89E-03	0.72	0.4736
DFIXCT9	0.3415	1.48E-02	23.15	<0.0001
DPERCT10	0.4963	1.36E-02	36.40	<0.0001

From the above results (for Model 3.1), we see that the model is overall significant and R^2 is reasonably large as compared to those reported in the other studies (in fact, higher to many other studies) of such relationship estimated in France and in the other countries of the world. Also we see that a 6.70% increase in monthly wage is associated with each additional year of schooling. The relationship between the past experience (by past experience we mean time spent in labour market before present job) and wages seems to be of concave type and a year of extra experience in labour market increases an individual's wage by an amount of 1.33% in start but this return to experience decreases by 0.02% with each additional year of experience. The experience-wage relationship gets to peak with an experience of approximately 29 years.

[3] see appendix A2 for algebraic form of Model 3.1

Another important measure, similar to labour market experience, is job seniority i.e. experience with the current employer or in current job. The coefficient related to job seniority suggests that each additional year that passed in the current job increases wages by 1.84% in start while this return decreases by 0.01% with every extra year worked in the current enterprise. Here, we see that present job seniority is just a little more rewarding as compared to past experience (1.84% versus 1.33%). The estimate for quadratic term reveals that like past experience, the relationship between wages and current experience (job seniority) is also of concave nature. The number of hours worked affects wages significantly as expected. Roughly, with an extra hour of work, people get an extra salary of 0.64%. Our results show that the significant effect of gender is also present in the French labour market which is in favour of men. Males enjoy a wage premium of about 16.87% over females. The direction of the coefficient of gender is similar to that found in literature (Viger, 2007; Barnet-Verzat & Wolff 2008; Araï et al. 1996 and some others) that males are better paid in France. But our results are in more closeness with those obtained in Barnet-Verzat & Wolff (2008) who reported a total gender wage gap of nearly 18%. The effects of regional labour market segmentation has been taken into account by using 3 categories for whether an individual is working in the rural areas, or in the non-Paris urban areas or in Paris region. We found that the people in the rural France earn less than their counterparts who live in the Paris region or in the other urban areas of France. People in the urban areas other than Paris region (non-Paris urban areas) get wage premium of about 1.2% (statistically significant) while workers from Paris region have wages which are about 22% higher compared to rural workers. So it also shows that Paris region (which is mostly urban area) workers are better paid than other urban areas in France. The finding regarding wage premium for Paris region workers over rural workers is in line to that obtained in Viger (2007). The higher wage premium associated to Paris region compared to that of other urban areas of France may be due to the availability of more job opportunities as Paris regions can be considered as the industrial and economic hub of France. On one hand the presence of higher job opportunities allow workers to debate for higher and better salaries and on the other hand the situation motivates employers to better pay their workers in order to avoid mobility across firms and enterprises. Comparing the type of diploma (whether professional or general), surprisingly, our results show that the people having professional degree get lesser wages by an amount of 5% than those with general educational diploma. This is against the previous evidence from the other studies (like Tansel, 1994; and Simonnet, 1996 for France). One possible explanation may lie in a way that how

we defined the variable. Because, we define it as having degree type which is considered as professional while the other studies defined it as having technical/vocational training. So it may be due to existence of a general wage differential in the jobs that typically relates to the professional kind of diploma holders. Moreover, there may be some individuals who are facing a wage penalty for not being working in the field relevant to the professional degree they hold. We also control public-private wage differences by introducing a dummy variable. The non- significance of the coefficient associated to that variable hints at the absence of any noticeable wage difference between two sectors of employment and this result is in line with that found in Simonnet (1996).

Finally, controlling for the type of contract, as expected, we find that the people working under fixed term contract have 34% higher wages while those who work under a permanent contract enjoy a wage premium of about 49%, with reference to temporary workers. The directions of such findings are consistent with those reported in Araï et al. (1996) but the magnitude of penalty for temporary contract workers is much larger in our investigation.

3.6.2 Instrumental variables 2SLS estimation using Instrument-1

Now keeping in view the problem of endogeneity of schooling discussed in Section (2.2.1), we apply the IV2SLS estimation approach (explained in Section 2.2.1.1) using Z1 as instrumental variable as defined in Section (3.5) for possibly endogenous schooling. For further reference, we name this model that uses Z1 as instrument for endogenous schooling as **Model 3.2**[4]. The results from first stage and second stage regressions of Model 3.2 are given in Table 3.4-A and Table 3.4-B, respectively.

[4] see appendix A2 for algebraic form of Model 3.2

Table 3.4-A: First Stage of IV2SLS Estimation for France using Z1 as Instrument

Global Measures of Model				
R-Square	**Adj R-Sq**	**F Value**	**P-value**	**N**
0.7459	0.7458	6123.63	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	0.9943	8.34E-02	11.92	<0.0001
BEFEX2	-0.0494	2.89E-03	-17.11	<0.0001
BEFEX22	0.0005	9.00E-05	5.51	<0.0001
EXP2	-0.0121	3.13E-03	-3.87	0.0001
EXP22	-0.0003	9.20E-05	-3.58	0.0003
HOURS3	0.0012	3.04E-04	4.05	<0.0001
DGENDER4	-0.2163	1.88E-02	-11.50	<0.0001
DNPARIS5	0.0621	2.18E-02	2.84	0.0045
DPARIS6	0.2631	3.05E-02	8.62	<0.0001
DTYPDIP7	0.6557	1.84E-02	35.57	<0.0001
DPUBLIC8	0.0594	3.10E-02	1.91	0.0555
DFIXCT9	0.6011	6.62E-02	9.08	<0.0001
DPERCT10	0.4792	6.13E-02	7.82	<0.0001
Z1	0.8916	3.99E-03	223.51	<0.0001

From IV2SLS estimation approach with Instrumeent-1 (Z1), Table 3.4-A presents the results from the first stage regression of schooling on Instrument-1 and other exogenous variables. The model is overall significant as indicated by the significance of the F-Statistic and high R^2. The value of R^2 is larger than that found in many other studies. The results show that instrumental variable (Z1) is significantly related to endogenous schooling which points out that instrument we are using is relevant.

Now we present the results from second stage of IV2SLS estimation. The predicted values for schooling variable $\hat{SCH}1_Z1$ from first stage regression will replace the schooling variable SCH1which is suspected to be endogenous.

Table 3.4-B: Second Stage of IV2SLS Estimation for France using Z1 as Instrument

Global Measures of Model				
R-Square	**Adj R-Sq**	**F Value**	**P-value**	**N**
0.5435	0.5433	2483.71	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	4.6386	1.91E-02	242.89	<0.0001
SCH1_Z1	0.0766	9.98E-04	76.79	<0.0001
BEFEX2	0.0145	6.52E-04	22.31	<0.0001
BEFEX22	-0.0002	2.00E-05	-11.47	<0.0001
EXP2	0.0189	6.99E-04	27.09	<0.0001
EXP22	-0.0001	2.10E-05	-6.17	<0.0001
HOURS3	0.0063	6.80E-05	93.03	<0.0001
DGENDER4	0.1734	4.21E-03	41.17	<0.0001
DNPARIS5	0.0082	4.87E-03	1.69	0.0915
DPARIS6	0.2056	6.86E-03	29.97	<0.0001
DTYPDIP7	-0.0619	4.26E-03	-14.53	<0.0001
DPUBLIC8	0.0014	6.92E-03	0.20	0.8423
DFIXCT9	0.3231	1.48E-02	21.79	<0.0001
DPERCT10	0.4760	1.37E-02	34.67	<0.0001

Table 3.4-B shows the estimation results from second stage of IV2SLS approach (i.e. Model 3.2) that uses Z1 as instrument for schooling in the first stage regression. The results show that returns to education has increased by using IV2SLS method (7.67% as compared to 6.70% from the OLS method) which is in accordance with general finding in the literature (Bhalotra & Sanhueza, 2004; Conneely & Uusitalo, 1997; Angrist and Kruger, 1992; Ashenfelter et al., 1999). Our schooling coefficient from the IV2SLS estimation is roughly 14% higher to that found from the OLS estimation which is common in literature as after reviewing many studies Card (1994) found range of this difference to be between 10% and 30%. The effect and concave relationship of past experience as well as current job seniority with wages remained roughly similar. From the above IV2SLS estimation, return to past experience is 1.45% (compared to 1.33% when using the OLS) while each additional year that spent in current job increases wages by 1.89% (compared to 1.84% from the OLS). The effects of gender and number of hours devoted to work also remained similar to that found in the OLS estimation (Model 3.1). For effects concerning geographical regions, we see that

lower premium of being in the non-Paris urban areas over rural areas is even lesser from the IV2SLS estimation and it becomes statistically insignificant, while wage premium for Paris region becomes 20.56% (decreased by roughly 1.5 percentage points to that in the OLS Model 3.1). The penalty for professional education increased by about 1 percentage point while effect of being a public sector worker compared to private sector worker is non-significant as it was before (in Model 3.1). The wage gains for holders of fixed term and permanent contracts over those who work temporarily, is present in the IV2SLS estimation as well. The amount of wage premiums for fixed term and permanent contract workers over temporary workers decreased by nearly 2 percentage points for both types of contract holders.

3.6.2.1 Hausman test for endogeneity of schooling using Instrument-1

Table 3.4-C: Endogeneity Test for Schooling with Z1 as Instrument

Efficient under H0	Consistent under H1	Statistic	P-value
OLS	2SLS	262.58	<0.0001

In order to test that schooling is actually endogenous or not for the present sample, we have applied the Hausman (1978) test. Table 3.4-C shows the results from this test. The substantially low *p*-value suggests that null hypothesis of no difference between IV2SLS and OLS estimates (and hence the OLS should be preferred as being more efficient than the IV2SLS) is rejected. So it means that the schooling variable (SCH1) is not exogenous for the present sample which leads to conclude that the IV2SLS estimation is more robust compared to the OLS one.

3.6.3 Instrumental variables 2SLS estimation using Instrument-2

Now we have re-analyzed the relationship between wage and its determinants by applying IV2SLS approach using Z2 as an instrumental variable for endogenous schooling. For

convenience we name this model as **Model 3.3**[5]. The results from first stage and second stage regressions of Model 3.3 are given in Tables 3.5-A and 3.5-B below.

Table 3.5-A: First Stage of IV2SLS Estimation for France using Z2 as Instrument

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-Value**	**P-value**	**N**
0.2856	0.2852	833.83	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	6.9604	2.12E-01	32.89	<0.0001
BEFEX2	-0.1482	4.86E-03	-30.48	<0.0001
BEFEX22	0.0014	1.58E-04	9.17	<0.0001
EXP2	-0.0089	5.82E-03	-1.52	0.1277
EXP22	-0.0008	1.56E-04	-4.87	<0.0001
HOURS3	0.0047	5.10E-04	9.28	<0.0001
DGENDER4	-0.4695	3.15E-02	-14.92	<0.0001
DNPARIS5	0.3552	3.66E-02	9.72	<0.0001
DPARIS6	1.5910	5.02E-02	31.72	<0.0001
DTYPDIP7	1.2536	3.06E-02	40.99	<0.0001
DPUBLIC8	0.3518	5.19E-02	6.77	<0.0001
DFIXCT9	1.4442	1.14E-01	12.69	<0.0001
DPERCT10	1.5706	1.07E-01	14.73	<0.0001
Z2	0.2922	1.70E-02	17.14	<0.0001

From the IV2SLS approach, using Instrument-2 (Z2) to correct for endogeneity of schooling (i.e. Model 3.3), the above Table 3.5-A shows the results from estimation of first stage schooling equation. These results show that model is over all significant, however the explanatory power of the model is less than that when used Z1 as instrument for education or schooling (i.e. first stage of Model 3.2). But such values of first stage R^2 are common in literature for example Boumahdi & Plassard (1992) and Pons & Gonzalo (2003) reported lesser first stage R^2 than ours. The highly significant coefficient associated with Instrument-2 (Z2) hints at the relevance of Instrument-2 used in the schooling equation. We computed

[5] see appendix A2 for algebraic form of Model 3.3

fitted values ($\hat{SCH}1_Z2$) for schooling variable (SCH1) which are to be used as explanatory variable replacing SCH1 in the main wage equation estimated in the second stage of IV2SLS estimation.

Table 3.5-B: Second Stage of IV2SLS Estimation for France using Z2 as Instrument

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.4355	0.4855	1609.69	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	4.0385	8.88E-02	45.47	<0.0001
SCH1_Z2	0.1371	8.76E-03	15.65	<0.0001
BEFEX2	0.0227	1.38E-03	16.41	<0.0001
BEFEX22	-0.0003	2.30E-05	-11.56	<0.0001
EXP2	0.0221	9.10E-04	24.26	<0.0001
EXP22	-0.0001	2.50E-05	-2.63	0.0086
HOURS3	0.0060	8.80E-05	68.52	<0.0001
DGENDER4	0.2027	6.35E-03	31.91	<0.0001
DNPARIS5	-0.0137	6.34E-03	-2.15	0.0312
DPARIS6	0.1077	1.61E-02	6.71	<0.0001
DTYPDIP7	-0.1378	1.19E-02	-11.57	<0.0001
DPUBLIC8	-0.0210	8.45E-03	-2.49	0.0129
DFIXCT9	0.2079	2.35E-02	8.83	<0.0001
DPERCT10	0.3487	2.40E-02	14.55	<0.0001

Table 3.5-B gives the results from second stage of Model 3.3, which is based on the IV2SLS approach using Z2 (Instrument-2) as an instrument for endogenous schooling in first stage schooling equation. The results point at the overall significance of the model. The explanatory power of the model is well comparable to other studies applying the Mincerian model to data from different countries. The returns to schooling are about 13.7% for each additional years of schooling. Comparing results from above Table 3.5-B with those from the OLS estimation (Model 3.1, given in Table 3.3), we see that returns to schooling are about 2 times than returns obtained from the OLS. General finding in the literature concerning difference between schooling coefficient from the IV2SLS and OLS estimation is that endogeneity

corrected (IV2SLS) estimate are 10-30 % or approximately 2 percentage points higher compared to the OLS estimates (Card, 1994; Ashenfelter et al., 1999). However, there are some studies that reported the IV2SLS schooling coefficient as nearly double than that obtained in specification taking education as exogenous (for example, Harmon & Walker, 1995; Card, 1993; Butcher & Case, 1994). The returns to labour market experience found to be higher than the both of previous models, Model 3.1 (OLS) and Model 3.2 (IV2SLS with Z1 as instrument) and this is robust to both kinds of experience terms (past experience and current job seniority), we controlled for. People get 2.27% increment for every more year of past experience. Similarly, every extra year of current job seniority increases wages of individuals by approximately 2.21%. The impact of work duration is almost identical to those found in previous two models. As before, Model 3.3 (Table 3.5-B) also provides evidence of gender wage differential in favour of men for the French labour market. But the magnitude of this differential is increased by 3 percentage points when compared to that found in preceding models. The above results reveal that women get 20% lesser wages compared to their counterparts of opposite gender. The wage difference due to regions, changed considerably in the above model when compared to results from the Model 3.1 and Model 3.2. The wage differential between non-Paris urban areas and rural workers turned out to be in favour of rural workers by a magnitude of 1.4% while this was non-significant in Model 3.2 and in favour of the non-Paris urban area workers in the OLS Model 3.1. The wage gain for workers in the Paris region over rural workers becomes 10.77% after decreasing by 11-12 percentage points than those found in previous two models. Coefficient of dummy variable for professional degree indicates that professional diploma holders earn less than people with general educational degree. Comparison between 3 models suggests that wage penalty for professional or a technical degree holder is more serious from the above Model 3.3 (Table 3-B). From the Model 3.3, professional degree holders get 13.8% lesser wages compared to those who possess a diploma in general education while this loss was about 5% and 6% from the Model 3.1 (OLS) and Model 3.2, respectively. Contrary to earlier specifications used, difference in the wages between public and private sectors of work is found significant in the above specification. The coefficient associated to dummy variable for public-private wage differential hints that individuals working in private sector of economy get roughly 2% higher wages as compared to those who are working in public sector. Concerning the impact of contracts statuses, we note that wage premium for people working under fixed term contract and permanent contract over temporary workers has decreased considerably in comparison

with two prior models. Precisely, the results from above specification (Model 3.3), it is noted that individuals with fixed term and permanent term contracts get 21% and 35% higher wages respectively than their counterparts working as temporary workers. The discrepancy between wage premiums for contract types between the IV2SLS and OLS estimation is higher when we have used Z2 as instrument in IV2SLS estimation. Numerically, we found differences in wage premiums associated with contract statuses of approximately 15 percentage points between Model 3.3 and Model 3.1while the similar difference of 2 percentage points is found between Model 3.2 and Model 3.1.

3.6.3.1 Hausman test for endogeneity of schooling using Instrument-2

Table 3.5-C: Endogeneity Test for Schooling with Z2 as Instrument

Efficient under H0	Consistent under H1	Statistic	P-value
OLS	2SLS	64.66	<0.0001

Table 3.5-C shows the output of Hausman (1978) test for exogeneity of schooling using Z2 as instrument. Similar to Table 3.4-C, the above results also reject the exogeneity hypothesis for schooling. These results suggest that for having unbiased estimates of the effects of education on the wage determination process, endogeneity problem of education must be addressed.

3.6.4 Choice between two instruments (Z1, Z2) for the French data

From the IV2SLS estimations based on Instrument-1(Z1) and Instrument-2 (Z2), it is clear that the schooling coefficient (returns to education) from OLS estimations are downward biased. The magnitude of difference between the OLS and IV2SLS estimates for effect of education (we are more concerned about effect of education as it is the suspected endogenous variable) depends on the instrument used for schooling in first stage schooling equation. The explanatory power of Instrument-1 (Z1) is more than that of Instrument-2 (Z2) as judged by first stage R^2 (See Table 3.4-A and Table 3.5-A).

In order to proceed further by choosing more appropriate instrument, we look at the nature and behaviour of the both instruments.

Instrument-1, as already discussed in Section 3.5, it is the average educational years in the household. But it may face the problems like there can be some family effects in schooling i.e. Instrument-1 may be affected by the same problems as schooling variable itself. For example, ability may be related to family situation and also there may be a genetic effect concerning ability and taste for education. Moreover, members of a household have more chances to face similar costs of schooling, to have similar opportunities for schooling (in terms of distance to or availability of educational institutions etc.), there may be some common demographic and geographical issues affecting schooling attainments because members of same household are more likely to share a common social and community environment.

Instrument 2 is the average schooling in the country, of particular age group, for particular gender, in that particular year when an individual entered in the labour market for monetary gains. So problems like similar demographic, geographical, schooling costs and opportunities, genetic ability effects with Instrument-1 may be eliminated in instrument-2, because it is an overall average. Moreover, as we noted in Section 3.5 that our second instrument (Z2) is such that it technically has many instruments of ability, school proximity/availability, family background, social environment, regional and demographic characteristics, school quality etc, in it.

For a further glance into the matter, we look in to correlation matrix of, LNWAGE (Response variable), SCH1 (endogenous explanatory variable), Instrument-1 (Z1) and Instrument-2 (Z2). Table 3.6 presents correlation matrix of LNWAGE, SCH1, Z1, and Z2.

Table 3.6: Correlation Matrix among Response, Endogenous Variable and Instruments (Z1, Z2), for the French Data

	LNWAGE	SCH1	Instrument-1 (Z1)	Instrument-2 (Z2)
LNWAGE	1.0000	0.3399	0.3360	-0.1360
SCH1		1.0000	0.8481	0.2728
Instrument-1 (Z1)			1.0000	0.2633
Instrument-2 (Z2)				1.0000

From the above correlation matrix, we see that correlation between Instrument-1 (Z1) and schooling is higher than the correlation between Instrument-2 (Z2) and schooling. But Instrument-1 (Z1) is also highly correlated with response 'LNWAGE' (virtually, as highly as schooling itself), so it violates the requirement of instrumental variables that instrument should not be directly correlated with response, with severity. While Instrument-2 (Z2), although have low correlation with schooling but it has also very low correlation with response variable (LNWAGE).

Finally, keeping in view the correlation matrix and the more adequate definition of Instrument-2 (Z2), we prefer this instrument to be used for further analysis. So Model 3.3 (IV2SLS model with Z2 as an instrument for endogenous schooling) will be the reference model for IV2SLS estimation for further analysis regarding specification of the model.

3.6.5 Sample selection model: Step-1: Estimation of participation equation for the French data

Now in order to correct for impact of sample selection bias on the estimated coefficients related to different wage determinants, we have estimated a sample selection model given in Section 2.2.3 (Eq. 2.6). We have used the Heckman (1979) two-step procedure to correct for the possible bias due to non random selection of individuals into wage earners' sample. For this purpose, we first estimated a participation or selection equation (as given in Section 2.2.3, Eq. 2.4) in which we regress binary response variable indicating positive or negative outcome of whether an individual decides to participate in the labour market or not. The details of explanatory variables are provided in the Section 3.4 (or Table 3.2 for a brief description). The maximum likelihood estimation of selection or participation equation is carried out by

Logistic Procedure in SAS 9.3 that uses Fisher's scoring as optimization technique. The estimation of selection equation is based on 81742 observations. The results from estimation of participation equation by maximum likelihood probit regression are given in the Table 3.7 below.

Table 3.7: Estimation of Selection Equation for France

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	P-value
INTERCEPT	-1.0298	2.63E-02	1537.34	<0.0001
AGE1	0.0104	4.08E-04	645.04	<0.0001
SCH2	0.0859	1.60E-03	2861.48	<0.0001
DGENDER3	0.2192	9.57E-03	524.70	<0.0001
DMSTAT4	0.1411	1.17E-02	145.61	<0.0001
DNPARIS5	-0.1525	1.22E-02	156.60	<0.0001
DPARIS6	-0.0685	1.65E-02	17.16	<0.0001
DHOMOWN7	-0.1914	1.05E-02	334.96	<0.0001
DALLOC8	-1.1516	1.67E-02	4774.45	<0.0001
DTYPDIP9	0.3754	1.05E-02	1273.09	<0.0001
CH6Y10	0.2571	1.08E-02	562.96	<0.0001

Table 3.7 gives the results of participation equation estimated for the French labour force data. The estimation reveals that all the explanatory factors contribute significantly to the decision of an individual regarding participation into labour market as wage worker or not. Table 3.7 shows that age, schooling, being male and being married significantly enhance the chances of the individual's positive decision regarding participation in the labour force. The results were expected so, as it is clear that increase in age motivates one more and more to work for financial gains because with increasing age individuals become more independent and have lesser aids from other sources like family etc. The positive effect of age concerning participation into labour market is consistent with literature (for example Arabsheibani & Mussurov, 2007; Chen & Hamori, 2009; Hoffman & Link, 1984; and Meurs & Ponthieux, 2000; Viger, 2007 for the French data). The effect of schooling is also positive as expected because the level of education has a significant and positive role in the probability of being a wage worker in the labour market (Tansel, 1994; Coelho et al., 2008) and also for France (Meurs & Ponthieux, 2000; Viger, 2007). The enhanced chance for being engaged in work due to rise in schooling attainments for the French data is also evident in Riboud (1985). The

males have more chances to participate in the labour force. This is like found in literature (for example Asadullah, 2006) and is as expected because it is also evident from other studies that marital status influences the decision to work positively for males and negatively for females (Lei, 2005) and also husband's income affects negatively women's participation (Martins, 2001; Riboud, 1985; Iglesias & Riboud, 1985). Moreover, females are assumed as having more odds of inactivity in labour force due to their family responsibilities. Marital status affects positively the chances of work involvement and it is consistent with existing evidence (Viger, 2007, for example). Opposite to expectation, people who are in the rural areas have higher chances of being in employment, as our coefficient suggests that individuals in urban area (Paris or non-Paris) have lower chances of being included in the labour force. One possible explanation may be that in urban areas in general and in the Paris region in particular, there exist a more tough competition for the jobs due to increased availability of educated and skilled workers. Secondly it may be due to more population in these areas. These two possibilities are further strengthened from the frequencies of these 3 categories in our cleaned sample that is used for maximum likelihood probit estimation of selection equation. Table 3.8 shows that 78.3% people are from urban areas including the Paris region while only 21.7% from rural areas.

Table 3.8: Proportion of Individuals According to Area of Residence

Area	Frequency	Percentage
Rural	17979	21.7
Non-Paris Urban	52329	63.1
Paris Region	12589	15.2
Total	82897	100.0

We have used dummy variable for homeownership in order to see the effect of unearned income in form of assets on participation decision, because we expected that it may represent other un-earned incomes (may be in form of lesser expenditures). The estimated coefficients from our probit regression on the French data (Table 3.7) confirm this expectation by suggesting a negative impact of being a house owner on work decision. It is in well coherence with the results from other studies like Asadullah (2006) that provided negative impacts for being owner of land capturing similar effects. We have also captured the effects of any non wage income measure by dummy variable for getting any type of financial support. Results

show that this variable significantly affects the participation decision of people and the individual who get any financial grant or support have lesser odds to be active in the labour force. This is in accordance with the existing evidence reported in different studies, as Burger (2011, for South Africa) reported a significantly negative impact of grants for child support and grants for old age people, on the decision to get involved in work activity and Asadullah, (2006, for Bangladesh) and Lei (2005, for Canada) that amount of unearned income of individual or household decreases the positive attitude towards work. Finally, concerning number of young children, we report a coefficient contrasting with expectation and with that from many other studies (Viger, 2007; Arabsheibani & Mussurov, 2007; Chen & Hamori, 2009; Coelho et al., 2008; Riboud, 1985 and some others). Our coefficient suggests that the number of children under 6 years of age in the household enhance the chances of work participation, while the studies mentioned above reported negative effect of younger children on work participation, but all these studies (except Viger, 2007, the French one) were focused on the behaviour of such measure for women's sample. It is found in the literature that married men have higher tendency to work and we can justifiably expect that after having children men will even feel more need to participate in monetary gain activities. As our sample consists of both males and females, so we may expect that negative effects of younger children on females participation is netted out by a more strong but positive impact of younger children on work participation of men.

3.6.5.1 Marginal effects of explanatory variables on the probability of selection into the wage earners' sample in France

The coefficients obtained from maximum likelihood estimation of probit regression, just tell about the direction of the effects that explanatory variables put on the binary response variable. They are so because of the structure of the probit regression model. But from a regression point of view, it is natural to think about marginal effects that explanatory variables have on the response variable. Usually the marginal effect of a particular variable is defined as the derivative of the predicted response with respect to that particular variable. In case of probit model,

$$P(y_i = 1) = cdf\left(\beta_0 + \sum_k \beta_k X_{ki} + \varepsilon_i\right)$$

or

$$P(y_i = 1) = cdf\left(\hat{\beta}_0 + \sum_k \hat{\beta}_k X_{ki}\right)$$

$$\frac{\partial P(y_i = 1)}{\partial X_1} = \frac{\partial}{\partial X_1}\left[cdf\left(\hat{\beta}_0 + \sum_k \hat{\beta}_k X_k\right)\right]$$

$$= pdf\left(\hat{\beta}_0 + \sum_k \hat{\beta}_k X_k\right) * \frac{\partial}{\partial X_1}\left(\hat{\beta}_0 + \sum_k \hat{\beta}_k X_k\right)$$

$$= pdf\left(\hat{\beta}_0 + \sum_k \hat{\beta}_k X_k\right) * \hat{\beta}_1 \qquad (3.3)$$

So in probit model, marginal effects of X_1 (for example) on probability of $Y_i = 1$ i.e. $P(Y_i = 1)$ depend on the values of independent variables and unlike to linear regression model, it is unique for every observation. In order to ease representation, generally average marginal effects are used. This average marginal effect is based on evaluation of function (Eq. 3.3) at the means of the explanatory variables. For example average marginal effect of a variable say X_1 on $P(Y_i = 1)$ will be,

$$marginal_Effect_X_1 = pdf\left(\hat{\beta}_0 + \sum_k \hat{\beta}_k * \overline{X}_k\right) * \hat{\beta}_1 \qquad (3.4)$$

Using above relation we have calculated the marginal effects for all explanatory variables in the selection equation which are given in Table 3.9.

Table 3.9: Marginal Effects of Different Explanatory Variables on the Probability of Selection into Wage Earner's Sample

Variable	MARGINAL EFFECT on P(ACTIVE=1)
AGE1	0.0039
SCH2	0.0321
DGENDER3	0.0820
DMSTAT4	0.0528
DNPARIS5	-0.0570
DPARIS6	-0.0256
DHOMOWN7	-0.0716
DALLOC8	-0.4306
DTYPDIP9	0.1404
CH6Y10	0.0961

From Table 3.9, we see that on average, with an increase of one year in age, the probability of participation into waged work increases by 0.0039 percentage points. Each additional year of schooling increases the chances of being active in the labour market by approximately 0.032 percentage points. The effect of gender can be interpreted as being male have about 0.082 points higher probability to work as compared to probability of being in waged work for females. Similarly, married people have 0.0528 points higher chances of being active in the labour market for monetary gains, than people living as single. The probability of being working for people in rural France is higher by 0.057 points and 0.0256 points than people from non-Paris urban areas and Paris region, respectively. Concerning effects of homeownership, the marginal effects suggest that those who own a house have 0.0716 points lesser chances to decide in favour of work in labour market compared to those who do not own a house. As described earlier that people getting any kind of financial support have lesser tendency to work. In terms of numbers, such people have 0.43 points fewer probability of being active in work market compared to those who do not get any kind of financial aid. Coming to the degree effects, the Table 3.9 reveals that on average, French individuals with a professional diploma have 0.14 points higher probability to get a job compared to the individuals with a general educational degree. Finally, on average, an additional child of less than 6 years of age, in the household raises the likelihood of being active by 0.096 percentage points, for the people of working age in that household.

3.6.6 Sample selection model: Step-2: Estimation of wage regression

The second step of the sample selection model, consists of estimation of wage regression after adding IMR (calculated from first step probit regression) as an additional explanatory variable (Eq. 2.6, Section 2.2.3). The results from step-2 of the sample selection model are given in Table 3.10. We name sample selection model as **Model 3.4**[6] for further reference.

Table 3.10: Heckman (1979) Sample Selection Model for the French Data

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.5551	0.5549	2417.41	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	4.7394	1.81E-02	262.10	<0.0001
SCH1	0.0576	1.12E-03	51.58	<0.0001
BEFEX2	0.0119	6.53E-04	18.21	<0.0001
BEFEX22	-0.0002	2.00E-05	-10.41	<0.0001
EXP2	0.0175	6.99E-04	24.98	<0.0001
EXP22	-0.0001	2.05E-05	-6.37	<0.0001
HOURS3	0.0063	6.77E-05	93.03	<0.0001
DGENDER4	0.1416	4.75E-03	29.83	<0.0001
DNPARIS5	0.0248	4.96E-03	5.00	<0.0001
DPARIS6	0.2233	6.76E-03	33.03	<0.0001
DTYPDIP7	-0.0872	5.20E-03	-16.77	<0.0001
DPUBLIC8	0.0028	6.88E-03	0.40	0.6862
DFIXCT9	0.3370	1.47E-02	22.90	<0.0001
DPERCT10	0.4857	1.36E-02	35.64	<0.0001
IMR	0.1405	1.17E-02	12.03	<0.0001

Table 3.10 presents the estimates corrected for the sample selection bias using the Heckman (1979) two-step approach. The IMR was calculated by the relation (2.5) given in Section (2.2.3) based on the coefficients from probit regression for the French data given in Table 3.7 in previous Section 3.6.5. The results signify the presence of significant sample selection bias

[6] see appendix A2 for algebraic form of Model 3.4

in the OLS estimates. The significant positive coefficient associated with IMR indicates that individuals who decided not to participate would have earned less than participants, if they had decided to participate in the labour market work activity. By comparing results from the above Model 3.4 with those from OLS Model 3.1(which ignores possibility of sample selection bias), it is also clear from coefficients for other variables that they generally change with the inclusion of sample selectivity term. Returns to schooling dropped from 6.70% (OLS, Model 3.1) to 5.76% (above Model 3.4) which indicates an upward bias in the OLS estimates resulting from bias due to non randomness of the sample. The returns to labour market experience decreased but not much. Particularly, returns associated to past experience decrease from 1.33% to 1.19% for each extra year of the past experience. The returns to experience in current job or seniority also decreased from 1.84% to 1.75%. Comparing with OLS Model 3.1, the impact of hours of work did not found affected by sample selectivity, while the impact of being male decreased by a little more than 2.7 percentage points. Sample selectivity corrected Model 3.4 suggests that men get 14% higher wages compared to women while this measure was nearly 17% in Model 3.1 (uncorrected for sample selection bias). It means that gender wage gap would have been lesser if non participants had decided to include themselves in labour market's waged work activity. Wage premium for urban workers over rural workers also exists after sample selection correction. Comparing with Model 3.1, this wage premium over rural workers is increased by 1 percentage point for workers in the non-Paris urban areas while remained almost identical for the Paris region workers when sample selectivity controlled (i.e. Model 3.4). The wage penalty for professional diploma holders almost doubled when we corrected the model for possible selectivity bias. The wage loss for professional diploma holders is 8.7% in sample selection model (Model 3.4) compared to approximately 5% in the OLS (Model 3.1, Table 3.3) that did not corrected for possible sample selection bias. The absence of public sector wage differential over private sector reported in the OLS model (Table 3.3) is also found to be robust with the Heckman (1979) correction. In Model 3.4 (corrected for sample selectivity), the wage premiums for people working with a fixed term or permanent contract over temporary contract workers remained significant with very small changes in coefficients compared to the OLS results. So it means that presence of sample selection bias does not affect much the impacts of contracts under which individuals work in the labour market.

3.6.7 Dealing with endogeneity and sample selection bias, simultaneously

So far we have addressed problems of endogeneity of schooling and of sample selection bias in separate specifications. Comparing the direction of biases in the OLS estimation due to both these issues, we note that changes in coefficients (small or large) linked to hours of work, dummy for diploma type, public-private differential, and premiums for type of contract, are in similar directions with both corrections (magnitudes may differ with both corrections). While for coefficients related to human capital variables of schooling and experience and gender gap, tend to move in different directions from the two bias correcting specifications implemented separately. For example in comparison to the uncorrected OLS specification (Model 3.1, Table 3.3) the impact of schooling and experience increases when endogeneity is addressed (robust to choice of different instruments i.e. Model 3.2 and Model 3.3) but it decreases when sample selection problem is addressed (Model 3.4). Similarly, concerning gender variable, coefficient increases a little in the IV2SLS approaches while decreases with the Heckman sample selection correction.

Moreover, in the literature concerning estimation of the Mincerian wage regression, the studies that tackle the problems of endogeneity bias and sample selection bias at the same time in a single specification are relatively few (for example, see Maluccio, 1998; Arrazola and Hevia, 2006; García et al., 2001; Arabsheibani and Mussurov, 2007). Out of these, in the last two studies mentioned, the correction for sample selectivity was limited only to female samples. Moreover, we have seen that for France too, there exist no study that addressed biases due to endogeneity of schooling and sample selection in single specification.

So keeping in view the need of the estimated effects of different wage determinants corrected for both kind of biases (endogeneity and sample selection biases) and the scarcity of literature that addresses both said problems simultaneously in a single model, we estimated a model that tackled both these potential sources of bias simultaneously. For this we apply the IV2SLS estimation based on most appropriate instrument (we selected Z2 to be more appropriate instrument) of the two proposed instruments and that also includes sample selection correction term coming from probit estimation of selection or participation equation (Step-1 of sample selection model, Section 3.6.5).

More specifically, this model replaces schooling by its fitted values from first stage schooling regression (based on Z2 as instrument, Table 3.5-A) and IMR as an additional explanatory variable.

Algebraically, this model takes the form,

$$\ln W_i = \beta_0 + \beta_1 (\hat{SCH}1_Z2)_i + \sum \beta_K (X_K)_i + \beta_{IMR} (IMR)_i + \varepsilon_i \qquad (3.5)$$

We believe that estimated impact of different explanatory factors from this model (Eq. 3.5) will be more reliable and representative as this model tackles both sources of bias, simultaneously.

For the French data, the results from specification in Equation 3.5 that corrects for both sample selectivity and endogeneity are presented in Table 3.11. We name it as **Model 3.5**[7] for differentiation and further reference.

[7] see appendix A2 for algebraic form of Model 3.5

Table 3.11: Addressing Endogeneity and Sample Selection Biases Simultaneously for the French Data (Model 3.5)

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.4903	0.4900	1863.38	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	4.2426	8.57E-02	49.49	<0.0001
SCH1_Z2	0.0818	8.50E-03	9.62	<0.0001
BEFEX2	0.0149	1.34E-03	11.12	<0.0001
BEFEX22	-0.0002	2.30E-05	-8.72	<0.0001
EXP2	0.0173	8.82E-04	19.65	<0.0001
EXP22	-0.0001	2.40E-05	-2.46	0.0139
HOURS3	0.0059	8.50E-05	69.03	<0.0001
DGENDER4	0.0873	6.42E-03	13.60	<0.0001
DNPARIS5	0.0445	6.19E-03	7.19	<0.0001
DPARIS6	0.1459	1.55E-02	9.42	<0.0001
DTYPDIP7	-0.2617	1.17E-02	-22.42	<0.0001
DPUBLIC8	-0.0226	8.15E-03	-2.78	0.0055
DFIXCT9	0.2256	2.27E-02	9.93	<0.0001
DPERCT10	0.3462	2.31E-02	14.98	<0.0001
IMR	0.5513	9.20E-03	59.91	<0.0001

No we interpret the results from the above model which is possibly free from both types of biases. From Table 3.11, we see that each additional year of schooling causes an increase of about 8.2% in wages of individuals. The past experience influence wages positively. With each additional year spent in labour market prior to the present job, wage increases by 1.5%. Every extra year of seniority in present job enhances the wages of individuals by 1.73%. The experience (both past and current job experience) wage relationship is concave which means increasing years of experience cause wage increase with a decreasing rate. In other words, marginal gains associated to experience tend to decrease with increase in the experience (robust to both past as well as current experience measures). Work hours affect wages positive with magnitude of 0.59% associated with every additional hour worked. The raw gender wage gap is approximately 8.7% in favour of men i.e. women get 8.7% lesser wages compared to men. This gender gap is lesser in magnitude than the uncorrected OLS (Model 3.1), endogeneity corrected separately (Model 3.3) and sample selection corrected separately

(Model 3.4) specifications. So if we had not corrected for both kinds of biases in the same specification (i.e. Model 3.5) than we would have reported higher gender gap which is lower actually. This fact points out the gains in accuracy of raw gender wage gap due to correcting for both biases in a single model. Workers in the non-Paris urban areas get 4.45% and workers in Paris region get about 14.6%, higher wages compared to the rural workers, respectively. The negative effect of professional degree on wages increases substantially compared to the uncorrected OLS or models that correct for endogeneity or sample selectivity separately. It had a magnitude of 4.9% in the OLS specification (Model 3.1), 13.7% in endogeneity corrected (Model 3.3), and 8.7% in sample selection corrected (Model 3.4) specifications but from Model 3.5, we note that workers having professional diploma get 26% fewer wages compared to people with a general educational degree.

The wage differential between public and private sectors from Model 3.5 is closer to that found in Model 3.3 as compared to the OLS Model 3.1 or Model 3.4. The workers in public sector earn 2.26% lower, compared to workers in private sector. The bias free wage premium for fixed term contract workers over temporary contract workers found as 22.6% while similar premium associated to permanent contract is 34.6%. These two coefficients linked to contract statuses wage premiums are more in line with those found from Model 3.3 (endogeneity corrected model), so we may say that sample selection bias have limited effect on the coefficients related to nature of contract. The overall impact of sample selection bias is positive and significant even with a higher coefficient than that found when only sample selectivity was corrected (i.e. Model 3.4).

We believe that coefficients from Model 3.5 to be more reliable and representative as they are computed from a specification that tackled endogeneity bias and sample selection bias at the same time.

3.6.8 Heterogeneity of the error term

From the literature review concerning the Mincerian studies in the world and in France too, we see very few studies that addressed the problem of heteroscedasticity of the error term of the wage model. Violation of this assumption may lead to serious problems in estimation and can lead to draw incorrect or misleading conclusions about the impacts of different contributing factors. Only some studies in the literature are found that used heteroscedasticity

corrected standard error to judge the significance of the coefficients related to different explanatory factors (like Johnson & Chow, 1997; Martins, 2001; Hawley, 2004; Uusitalo, 1999; Dickson, 2009; Broekhuizen, 2011; Flabbi, 1999 and some others). Most of such studies used White's (1980) heteroscedasticity robust standard errors. To our knowledge, the only study that formally tested the presence of heteroscedasticity is Ewing (2000) in which assumption of homoscedasticity was found violated. Focusing on the studies based on the data from the French labour markets, we found that this issue has rarely been addressed. From our review of the French studies, we found no study that formally tested the validity of homoscedasticity of error variances. Only one study that applied heteroscedasticity robust estimation for France is carried out by Meurs & Ponthieux (2000). So we feel it as necessary to verify the homoscedasticity assumption and possibly remedy it if found violated. If the errors terms are found to be heteroscedastic, the remedial measures for heteroscedasticity will be taken that lead towards more efficient coefficient estimates.

3.6.8.1 Testing heteroscedasticity of the error term

As we have mentioned that there are only a few studies that tested or corrected for the possibility of heteroscedasticity of errors. In order to test the validity of homoscedasticity assumption, we applied White's (1980) test on the errors calculated from Model 3.5. The output of the test is shown in Table 3.12.

Table 3.12: White's Test for Heteroscedasticity on Errors from Model 3.5

Test Statistic	DF	P-value	Variables
3628.08	81	<0.0001	Cross of all variables

From Table 3.12, the violation of homoscedasticity of errors assumption is evident. The highly significant value of the test statistic signifies that error variance is not same across observations and remedial measures must be taken to have efficient estimates.

3.6.9 Adaptive estimation of simultaneous model

From Table 3.12, it is clear that the error terms of Model 3.5 are heteroscedastic i.e. they do not possess a constant variance and $\text{var}(\varepsilon_i)=\sigma_i^2$. It has been well established that in the presence of heteroscedasticity, the OLS estimator (OLSE) becomes inefficient and its usual covariance matrix estimator becomes biased and inconsistent. If σ_i^2's are known, the simplest solution to estimate Model 3.5, efficiently, is the use of weighted least squares (WLS). While using the WLS, the model given in (Eq. 3.5) is divided by σ_i^2 that makes the random errors homoscedastic. But, usually, σ_i^2's are unknown and their estimates are used. In this case, the estimated weighted least squares (EWLS) may be used that can result in more efficient estimates as those by the OLS (see Fuller & Rao, 1978). The other popular method to have more efficient estimates is the adaptive estimation procedure. This method makes use of a variance-stabilizing transformation such that the assumption of homoscedasticity is satisfied, without knowing the functional form of the error variance; see, e.g., Carroll (1982), Carroll & Ruppert (1982), Robinson (1987) and Aslam (2006) etc. Recently, Ahmed et al. (2011) and Aslam et al. (2012) make a convincing use of the adaptive estimator as proposed by Carroll (1982).

On the lines of Carroll (1982), we assume the variance of error terms ε_i to be a smooth function of the mean values as

$$\text{var}(\varepsilon_i) = \sigma_i^2 = g(d_i),$$

where g is unknown and d_i can be estimated by

$$d_i = \ln \hat{W}_i,$$

where $\ln \hat{W}_i$ is the OLS estimate of $\ln W_i$

Like Carroll, we present a kernel estimator of σ_i^2 as a form of the Nadaraya-Watson (1964) estimator,

$$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{n} K\left(\frac{d_j - d_i}{h}\right)\hat{\varepsilon}_j^2}{\sum_{j=1}^{n} K\left(\frac{d_j - d_i}{h}\right)} \quad , \qquad (3.6)$$

where $k(\cdot)$ is the kernel function with h as the smoothing parameter and $\hat{\varepsilon}_j$ are the OLS residuals.

Now we divide the Model 3.5 given in Eq. 3.5 by $\hat{\sigma}_i$ and then apply the OLS for estimation. This procedure of estimation is known as adaptive estimation and results in efficient coefficient estimates.

For estimation of error variances (3.6), we use the normal kernel,

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Following Li & Stengos (1994) and Ahmed et al. (2011), we use a simple approach to compute the optimum value of the smoothing parameter, h as $h = c s_x n^{-0.2}$ for c = 0.8 with s_x as the standard deviation of x.

The results from adaptive estimation of the Model 3.5 are given in Table 3.13. We name adaptive version of Model 3.5 as **Model 3.5W**[8].

[8] see appendix A2 for algebraic form of Model 3.5W

Table 3.13: Model 3.5W (Adaptive Estimation of Model 3.5)

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.4741	0.4738	1746.24	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	4.2615	9.02E-02	47.27	<0.0001
SCH1_Z2	0.0772	8.59E-03	8.98	<0.0001
BEFEX2	0.0143	1.37E-03	10.47	<0.0001
BEFEX22	-0.0002	2.30E-05	-8.65	<0.0001
EXP2	0.0172	8.89E-04	19.37	<0.0001
EXP22	-0.0001	2.30E-05	-2.84	0.0045
HOURS3	0.0060	8.00E-05	75.06	<0.0001
DGENDER4	0.0860	6.61E-03	13.02	<0.0001
DNPARIS5	0.0473	6.12E-03	7.73	<0.0001
DPARIS6	0.1573	1.50E-02	10.50	<0.0001
DTYPDIP7	-0.2545	1.12E-02	-22.63	<0.0001
DPUBLIC8	-0.0188	8.02E-03	-2.34	0.0193
DFIXCT9	0.2296	2.35E-02	9.77	<0.0001
DPERCT10	0.3587	2.37E-02	15.16	<0.0001
IMR	0.5528	9.22E-03	59.94	<0.0001

Table 3.13 shows the estimation results of the Model 3.5W (adaptive version of Model 3.5). From these results, a similar (to Model 3.5) pattern for the impacts of explanatory variables on log wages can be seen. However, the magnitudes of these impacts change a little. Comparing with Model 3.5, in Model 3.5W returns to education decreased from 8.17% to 7.72% per extra year of schooling. Returns to past experience and current experience remained almost similar (negligible changes in coefficients) from simple (Model 3.5) and adaptive (Model 3.5W) estimations. The impact of hours worked and raw male-female gaps also remained almost similar to that obtained in the previous Model 3.5. The premiums for the non-Paris urban area workers (over rural workers) remain similar while the premium for Paris region workers increased by 1 parentage point. From Model 3.5W, the wage penalty for people with professional degrees is found as 25% which is nearly 1 percentage point less than that from Model 3.5. Concerning the public-private wage differences, results from above Model 3.5W show that private sector workers earn 1.88% (statistically significant) more than workers working in public sector. The wage gain that fixed term contract workers enjoy over

temporary workers is 22% which is same as found in simple estimation of the model (Model 3.5, Table 3.11). While the wage premium for permanent contract workers (over temporary workers), increased by about 1 percentage point. From adaptive estimation (Model 3.5W) results, permanent contract worker get salaries which are 36% higher than people working on temporary basis.

Comparing with simple estimation (Model 3.5), the global significance and predictability of the adaptive estimation (Model 3.5W) is a little less but it is well above than such measures found in the literature concerning estimations of the Mincerian wage regressions for other countries and data sets.

3.6.10 Semi-parametric estimation of the Mincerian wage regression

As we have described briefly in Section 2.3 that the Mincerian wage regression is also estimated semi-parametrically by many researchers and it is semi-parametric because of containing non-parametric component coming from non-parametric estimation of the selection or participation equation. But here in the present thesis, instead of using non parametric and semi parametric techniques for correction of the sample selection bias, we use the parametric two-step Heckman (1979) estimation procedure to correct for the possibility of this kind of bias. However, we explore the possibility that the relationship between instruments and schooling may not be linear or parametric. We relax parametric assumptions on the first stage schooling equation while correcting for the endogeneity of education by instrumental variables technique. We estimate the first stage schooling equation non-parametrically by using LOESS regression. Hence our second stage wage regression will be semi-parametric as we will insert non-parametrically predicted values of schooling coming from first stage. So our study is unique in the sense that it is semi-parametric due to having non-parametric component from first stage schooling equation and not due to the non-parametric or semi-parametric estimation of the participation equation. For this we estimate first stage schooling equation non-parametrically. Then our semi parametric estimation will be based on two steps that can be described as below.

Step-1 Non-parametric estimation of first stage schooling equation

$$SCH1 = f(X\ , Z2) + \nu \tag{3.7}$$

Where, X is the vector of all exogenous (except schooling SCH1) variables and $Z2$ is Instrumental variable. We use Z2 as instrument as we have chosen it to be more robust compared to other proposed instrument (Section 3.6.4). The non-parametric estimation of the first stage schooling regression (Eq. 3.7) will be carried out by LOESS regression.

Get $\hat{SCH1_NP_i}$ (Non-parametrically predicted value of $SCH1_i$ based on above Eq. 3.7, the first-stage Regression)

Step-2 Semi-parametric estimation of the Mincerian wage model

In semi-parametric estimation the non-parametrically fitted values for schooling ($\hat{SCH1_NP}$) by LOESS regression will be used in palace of endogenous schooling (SCH1) in the second stage wage equation which takes the following form,

$$\ln W_i = \beta_0 + \beta_1(\hat{SCH1_NP})_i + \sum \beta_K (X_K)_i + \beta_{IMR}(IMR)_i + \varepsilon_i \qquad (3.8)$$

The above function (3.8) is semi parametric as it contains fitted values of schooling from a non-parametrically estimated first stage schooling equation.

3.6.10.1 Non-parametric estimation of first-stage schooling equation

For the semi parametric estimation of the model, first we have to estimate first stage schooling equation based on instrument (Z2) and all exogenous variables in a non parametric way. For this, as we have more than one explanatory variables in the schooling equation, we choose to apply the Locally Weighted Scatter plot Smoothing (LOWESS) or LOESS regression.

LOESS regression was originally proposed by Cleveland (1979) and further developed by Cleveland et al. (1988). It is a non-parametric regression method and denotes a method that is also known as locally weighted polynomial regression.

For the simplicity, let $SCH1_i$ be the ith observation of the response vector $SCH1$ corresponding to the ith vector x'_i of X with p predictors (all explanatory variables and Instrument Z2, in our case) as follows

$$SCH1_i = f(x'_i) + \nu_i \tag{3.9}$$

where ν_i is the random error and $f(x'_i)$ is the unknown regression function. Then according to Härdle (1994), "the basic idea of LOESS is to start with a local polynomial least squares fit and then to robustify it. " Local" means here a *k*-NN type neighbourhood. The procedure starts from a k-NN pilot estimate and iteratively defines robustness weights and re-smoothes several times". Further details along with algorithm for computation can be found in Härdle (1994).

But this LOESS regression is not recommended in a situation when we have more than 4 or 5 explanatory variables. Also SAS 9.3 did not work for the estimation by LOESS method with explanatory variables more than 7 while we have 13 explanatory variables for the first-stage schooling regression.

In this situation we decided to make use of principal component analysis. So for non parametric estimation of the endogenous schooling, we applied the LOESS regression on Instrument-2 (Z2) and the number of principal components that explain major part of variation in the all exogenous explanatory variables (of course excluding schooling). For this purpose we first computed principal components based on explanatory variables except endogenous schooling. The results from the principal component analysis based on **BEFEX2, BEFEX22, EXP2, EXP22, HOURS3, DGENDER4, DNPARIS5, DPARIS6, DTYPDIP7, DPUBLIC8, DFIXCT9, DPERCT10, IMR**, for the French data are given in Table 3.14 below.

Table 3.14: Principal Component Analysis on all Exogenous Variables

Principal Component	Eigen value	Difference	Proportion	Cumulative
1	129823.228	52331.581	0.6233	0.6233
2	77491.647	76532.618	0.3720	0.9953
3	959.029	947.550	0.0046	0.9999
4	11.479	3.214	0.0001	1.0000
5	8.265	7.955	0.0000	1.0000
6	0.309	0.020	0.0000	1.0000
7	0.289	0.068	0.0000	1.0000
8	0.221	0.047	0.0000	1.0000
9	0.174	0.095	0.0000	1.0000
10	0.079	0.007	0.0000	1.0000
11	0.071	0.024	0.0000	1.0000
12	0.047	0.037	0.0000	1.0000
13	0.011	---	0.0000	1.0000

These results revealed that first 3 principal components explain almost 100% variability (as shown is Table 3.14) in the set of these explanatory variables. So, first 3 principal components can be used to capture the total variability in all exogenous explanatory variables.

So, now we can apply non-parametric LOESS regression for the first stage schooling equation on first 3 principal components and Instrument-2 (Z2). Based on first 3 principal components and Z2, the Equation for LOESS regression (Eq. 3.7) will take the following form.

$$SCH1_i = f\left(PC1_i, PC2_i, PC3_i, Z2_i\right) + \nu_i, \qquad (3.10)$$

where PC1, PC2, PC2 are respectively first 3 principal components and Z2 is Instrument-2. We use SAS procedure "PROC LOESS" for LOESS regression estimation. LOESS method applied in SAS 9.3 has produced a data driven value of smoothing parameter as 0.05917.

The advantage of this exercise is that we have taken into account the effects of instrument and all other exogenous variables (through their principal components) and also computation is made feasible due to reduced dimensionality.

3.6.10.2 Semi-parametric estimation of second-stage wage regression

Now we estimate the main wage model as given in Eq. 3.8 (in Section 3.6.10). This model incorporates non-parametrically fitted values for schooling from first stage schooling equation (Eq. 3.10). The results of semi parametric estimation of the Mincerian wage regression are presented in Table 3.15 (which we named it as **Model 3.6**[9] for convenience). This model replaces endogenous schooling (SCH1) by its non-parametrically predicted ($\hat{SCH}1_NP$) values from the LOESS regression.

Table 3.15: Semi-Parametric Estimation of the French Wage Regression (Model 3.6)

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.4627	0.4624	1667.97	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	4.1732	2.15E-02	193.889	<0.0001
SCH1_NP	0.0750	1.33E-03	56.444	<0.0001
BEFEX2	0.0038	7.78E-04	4.822	<0.0001
BEFEX22	0.0001	2.40E-05	5.399	<0.0001
EXP2	0.0164	8.32E-04	19.726	<0.0001
EXP22	-0.0001	2.40E-05	-3.085	0.00204
HOURS3	0.0059	8.10E-05	73.680	<0.0001
DGENDER4	0.0506	5.65E-03	8.954	<0.0001
DNPARIS5	0.0734	5.91E-03	12.427	<0.0001
DPARIS6	0.2762	8.05E-03	34.330	<0.0001
DTYPDIP7	-0.1584	6.19E-03	-25.575	<0.0001
DPUBLIC8	0.0051	8.19E-03	0.623	0.53329
DFIXCT9	0.3342	1.75E-02	19.078	<0.0001
DPERCT10	0.4654	1.62E-02	28.689	<0.0001
IMR	0.5486	1.39E-02	39.484	<0.0001

Table 3.15 shows the results from semi parametric estimation of the Mincerian wage regression for the French data. From the semi parametric results, the global significance of the

[9] see appendix A2 for algebraic form of Model 3.6

model is a little less than that obtained in its parametric version i.e. Model 3.5. However, the coefficients related to independent variables faced some changes and needs to be interpreted. The schooling coefficient decreased by 0.68 percentage points with reference to parametric model. From this semi parametric estimation about 7.5% increase in wages is associated with an increase of one year in length of educational period. The returns associated to an additional year of past experience changed substantially. Wages increase by just 0.38% with every extra year of past experience while this measure was 1.49% from parametric Model 3.5. An important difference is that in semi-parametric estimation (above Model 3.6) the wage and past experience relationship in not concave. The wage gain related to additional year of present experience (seniority with current job), is found as 1.64% which is close to that found from parametric model (Model 3.5). The relationship of job sonority and wages is of concave nature from semi-parametric model as well. The impact of hours worked on wages is similar to that found in parametric model.

The penalty for women in wages compared to men has decrease from 8.7% to 5.1%. Semi parametric regression generated coefficient suggests wage premium of about 7.3% for workers from non-Paris urban areas and 27% premium for workers from Paris region over their rural counterparts respectively. These premiums are higher to those found from parametric model (Model 3.5) by approximately 3 and 13 percentage points for non-Paris urban areas and Paris region workers respectively. Similar to the parametric Model 3.5, the effect of professional degree on wages is negative, but wage penalty has decreased from 22% to about 16%. The public-private wage differential found non-significant in the semi parametric model which was significant in the parametric Model 3.5.

The wage differences among workers working under different contract types found higher than those found from parametric Model 3.5. Precisely, from above semi-parametric Model 3.6, fixed term and permanent contract holders enjoy wage premium of 33% and 47% respectively over those working as temporary workers. These two effects are increased by 10 and 12 percentage points as compared to parametric Model 3.5. Finally, the coefficient related to term for sample selection correction is also found significantly positive and its magnitude is virtually identical to that found in parametric Model 3.5.

3.6.10.3 Testing heteroscedasticity of errors from semi-parametric model

As we did for the parametric model, we test the validity of equal error variances assumption for semi parametric model also. The results from White's (1980) test for presence of heteroscedasticity in error from Model 3.6 (semi-parametric model) are presented in Table 3.16 below.

Table 3.16: White's Test for heteroscedasticity on Errors from Model 3.6

Test Statistic	DF	P-value	Variables
3486.97	81	<0.0001	Cross of all variables

From Table 3.16, results clearly reject the homoscedasticity assumption. So like in parametric Model 3.5, adaptive estimation is also justifiable for semi parametric Model 3.6 in order to minimize the effects of homoscedasticity violation.

3.6.11 Adaptive estimation of semi-parametric model

The adaptive estimation of semi-parametric Model 3.6 is carried out with the same strategy that was used in adaptive estimation of Model 3.5 (the parametric model). The adaptive estimation results of Model 3.6 are given in Table 3.17 and we name it as **Model 3.6W**[10].

[10] see appendix A2 for algebraic form of Model 3.6W

Table 3.17: Model 3.6W (Adaptive Estimation of Model 3.6)

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.4616	0.4614	1661.14	<0.0001	27136
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	4.2575	2.12E-02	201.247	<0.0001
SCH1_NP	0.0679	1.31E-03	51.952	<0.0001
BEFEX2	0.0037	7.64E-04	4.801	<0.0001
BEFEX22	0.0001	2.30E-05	4.561	<0.0001
EXP2	0.0159	8.18E-04	19.445	<0.0001
EXP22	-0.0001	2.40E-05	-3.254	0.00114
HOURS3	0.0060	7.90E-05	75.087	<0.0001
DGENDER4	0.0475	5.56E-03	8.557	<0.0001
DNPARIS5	0.0754	5.81E-03	12.980	<0.0001
DPARIS6	0.2745	7.91E-03	34.708	<0.0001
DTYPDIP7	-0.1639	6.09E-03	-26.918	<0.0001
DPUBLIC8	0.0069	8.05E-03	0.859	0.39051
DFIXCT9	0.3357	1.72E-02	19.502	<0.0001
DPERCT10	0.4722	1.59E-02	29.615	<0.0001
IMR	0.5539	1.37E-02	40.561	<0.0001

The results from Model 3.6W (adaptive version of semi-parametric model) given in Table 3.17 point at the overall significance of the model. The returns associated to an additional year of schooling are about 6.79% which are less by about 1 percentage point than those found in simple semi-parametric estimation (Model 3.6). The impact of, current job seniority on wages is 1.59% per additional year which was 1.64% in Model 3.6. Coefficients linked to past experience and hours worked, wage premium for fixed term contract and permanent contract workers, are similar as those found from Model 3.6. Similarly, the coefficients related to other variables are approximately same as the coefficients estimated in Model 3.6. So they need not be interpreted again. The raw gap between wages of males and females is decreased from 5.1% (Model 3.6) to 4.7% (Model 3.6W). The estimated coefficient for sample selection correction term increased a little in adaptive estimation of semi parametric model.

3.6.12 Comparison of parametric and semi-parametric models for France

For a comparison among parametric and semi-parametric models and simple and adaptive estimation of these two models, Table 3.18 presents the coefficients estimated from parametric model (simple and adaptive estimations) and from semi parametric model (simple and adaptive estimations).

Table 3.18 : Comparison of Parametric and Semi-Parametric Mode Estimated for the French Data

Variable	Model 3.5 Parametric (Simple)	Model 3.5W Parametric (Adaptive)	Model 3.6 Semi-parametric (Simple)	Model 3.6W Semi-parametric (Adaptive)
INTERCEPT	4.2426	4.2615	4.1732	4.2575
SCH1	0.0818	0.0772	0.0750	0.0679
BEFEX2	0.0149	0.0143	0.0038	0.0037
BEFEX22	-0.0002	-0.0002	0.0001	0.0001
EXP2	0.0173	0.0172	0.0164	0.0159
EXP22	-0.0001	-0.0001	-0.0001	-0.0001
HOURS3	0.0059	0.0060	0.0059	0.0060
DGENDER4	0.0873	0.0860	0.0506	0.0475
DNPARIS5	0.0445	0.0473	0.0734	0.0754
DPARIS6	0.1459	0.1573	0.2762	0.2745
DTYPDIP7	-0.2617	-0.2545	-0.1584	-0.1639
DPUBLIC8	-0.0226	-0.0188	0.0051	0.0069
DFIXCT9	0.2256	0.2296	0.3342	0.3357
DPERCT10	0.3462	0.3587	0.4654	0.4722
IMR	0.5513	0.5528	0.5486	0.5539
R-Square	**0.4903**	**0.4741**	**0.4627**	**0.4616**

3.6.12.1 Choice between parametric and semi-parametric models

From the comparison of the parametric and semi-parametric models (in both simple and adaptive versions), we note that global predictive performance of both models is not notably different from each other in both their versions. However, we choose parametric model to be more appropriate for French data as its predictive performance is a little higher than semi-

parametric one. Moreover, this choice is also justifiable due to simplicity and easiness in estimation of parametric model provided that both models give quite similar global performances. So we may conclude that parametric model is appropriate for the estimation.

3.6.12.2 Choice between simple and adaptive versions of parametric model

Concerning the choice between simple and adaptive estimations of parametric model, its adaptive version (Model 3.5W) is preferred model over simple estimation, as it tackles issues of endogeneity, sample selection bias and heterogeneity of error term at the same time and its global performance is also not substantially different from that of Model 3.5 (simple estimation of the parametric model).

So from Section 3.6.12.1 and Section 3.6.12.2, it is clear that adaptive version of the parametric model (i.e. Model 3.5W) is the appropriate and robust specification for estimation of the Mincerian wage model for the French data.

3.6.13 Interval estimation of selected model for the French data

As in the preceding section we have chosen Model 3.5W (adaptive version of parametric model) to be more appropriate and preferred for the estimation of the Mincerian wage model for the French labour market data, So here we present interval estimates of the coefficients linked to different explanatory factors estimated in Model 3.5W to have more clearer picture about the impacts that different variables put on the wage determination process in the French labour market. The interval estimates are given in the Table 3.19. The lower and upper limits of the coefficients are calculated by following relations.

For lower limit of the coefficient estimate we have,

$$\beta_{Lower\ Limit} = \hat{\beta} - 1.96 * S.E\left(\hat{\beta}\right) \qquad (3.11a)$$

and for upper limit, the relation used is as follows,

$$\beta_{Upper\ Limit} = \hat{\beta} + 1.96 * S.E\left(\hat{\beta}\right) \qquad (3.11b)$$

Table 3.19: Interval Estimation of Model 3.5W selected for France

Variable	Parameter Estimate	Standard Error	$\beta_{Lower\ Limit}$	$\beta_{Upper\ Limit}$
INTERCEPT	4.2615	9.02E-02	4.0848	4.4382
SCH1_Z2	0.0772	8.59E-03	0.0603	0.0940
BEFEX2	0.0143	1.37E-03	0.0116	0.0170
BEFEX22	-0.0002	2.30E-05	-0.0002	-0.0002
EXP2	0.0172	8.89E-04	0.0155	0.0190
EXP22	-0.0001	2.30E-05	-0.0001	0.0000
HOURS3	0.0060	8.00E-05	0.0058	0.0061
DGENDER4	0.0860	6.61E-03	0.0730	0.0989
DNPARIS5	0.0473	6.12E-03	0.0353	0.0593
DPARIS6	0.1573	1.50E-02	0.1280	0.1867
DTYPDIP7	-0.2545	1.12E-02	-0.2765	-0.2324
DPUBLIC8	-0.0188	8.02E-03	-0.0345	-0.0030
DFIXCT9	0.2296	2.35E-02	0.1835	0.2757
DPERCT10	0.3587	2.37E-02	0.3123	0.4051
IMR	0.5528	9.22E-03	0.5347	0.5709

Table 3.19 gives the confidence intervals for coefficients related to different explanatory variables estimated from adaptive estimation of the parametric model (Model 3.5W) which is our preferred model for the French analysis. Table 3.19 signifies that based on the most appropriate model chosen in the present work, the increase in wages linked to an additional year spent in schooling process ranges from 6.03% to 9.40%. Similarly, every extra year that an individual passed in labour market before his current job increases his wage by a magnitude ranging from 1.16% to 1.70% while these measure are 1.55% and 1.89% for seniority in the current job of the individuals. The positive influence of each additional hour worked is found to lie between 0.58% and 0.61% increase in wages. The wage difference between male and female workers is found to be between 7.3% and 9.89% in favour of men. From above results, the ranges of wage premiums for workers in the non-Paris urban areas and workers in the Paris region over their counterparts working in rural areas are 3.53% to 5.92% and 12.80% to 18.67%, respectively. Concerning wage difference between professional and general degree holders, the 95% confidence interval estimation shows that professional degree holders face a minimum wage penalty of 23.24% while magnitude of the maximum wage penalty is found as 27.65%. It is already noted that people working in public

sector of employment get lesser wages compared to those working in private sector. Table 3.19 shows that this wage gain for private sector workers ranges from 0.3% to 3.45% in favour of private sector workers. In the French labour market the selected model shows that people working under fixed term contracts enjoy at least 18.35% and at most 27.57% more wages over workers who work temporarily. Similar wage premium is found to lie between 31.23% and 40.51% for permanent contract workers.

Chapter 4: Estimation of the Mincerian Wage Model for the Pakistani Data

4.1 A Brief Review of the Mincerian Studies for the Pakistani Data

Following the world wide trend of investigating the relationship between acquired human capital and gains in the labour market, many studies has also used Mincer's proposed human capital earnings regression to explore the effects that human capital and other social, demographic, regional etc factors exert on the wages/earnings of individuals working in labour market of Pakistan. Here, we briefly review the literature devoted to estimation of the Mincerian wage regression in the Pakistani context.

4.1.1 Common data sources

For estimation of the Mincerian Wage regression for Pakistan's labour market, data from different surveys have been used for analysis in many previous studies. Most of the studies used data from Pakistan Integrated Household Survey (PIHS is used by Aslam, 2009; Nasir, 2002; Khan & Toor, 2003), Pakistan Social and Living Standards Measurement (PSLM, used by Farooq & Sulaiman, 2009; Abbas & Foreman-Peck, 2007; Aslam & Kingdon, 2009; Qureshi, 2012) and Labour Force Survey (used by Hyder & Reilly, 2005; Hyder, 2007; Nasir, 1998; Nasir, 2000; and some others). Other data sources included Population, Labour Force and Migration Survey (PLMS, used by Shabbir, 1994) and Pakistan Socio Economic Survey (used by Nazli, 2004).

4.1.2 Estimation and trends

Aslam et al. (2008) estimated the Mincerian wage regression for the Pakistani data. They corrected for potential bias due endogeneity of schooling and other family shared factors like ability or environment, by using the IV and fixed effects approaches, but found no significant evidence for such biases. For estimation through the IV2SLS approach they used parental education and distance to nearest school as instruments. They reported higher returns to schooling for females (8.3% for each additional year) compared to 4.5% for males. The generalization of their data was limited as they have collected data from only 1000 households covering only nine districts of two provinces. While exploring the possibility that differences in the returns to education are causing low female education in Pakistan, Aslam

(2009) reported that different factors have effects on the wage determination process in the Pakistani labour market. In addition to the standard human capital variables, controls were used to capture the regional and urban-rural effects on wages. Using a nationally representative sample (Pakistan Integrated Household Survey, PIHS) and excluding self employed people, Aslam (2009) estimated Mincer's human capital earnings function using different estimation methods and specifications. Higher returns to schooling for females reported in this study, which was found robust to different estimation methods. Estimates for returns to education from fixed effects model which was used to correct bias due to unobserved ability, found lower than those from the OLS. This may possibly be due to not correcting for measurement error. For the correction of endogeneity due to unobservable factors and possible measurement error in schooling, the IV2SLS estimation was applied which produced higher returns to schooling estimates than the OLS method. Parental education and spouse's education were used as instruments for schooling. Two-step Heckit method used to correct for sample selection bias which was found significant, but like general trend in the literature, study did not corrected both endogeneity and sample selection biases in one single specification. Farooq & Sulaiman (2009) is another study that estimated earnings regression for the Pakistani data. They used data from Pakistan Social and Living Standard (PSLM) survey 2004-05. Estimating the Mincerian model separately for both genders, they found higher returns to education for women. Schooling variable used as levels of schooling and individual's age was used as proxy for experience. They also found significant differences in wages across provinces. This differs from our study in the sense that, they did not restricted to wage workers only but included self employed people as well. They did not correct any problem faced in estimation of the Mincerian wage model like endogeneity and sample selection etc, that may bias the OLS estimates. Abbas & Foreman-Peck (2007) investigated the impact of education on earnings in Pakistan for wage workers and 2 groups of self employed people. They applied Mincer's semi-log model using data from PSLM 2004-05. They estimated wage functions in both ways, by specifying education as year of schooling and levels of schooling. They tried to correct for endogeneity bias by applying the IV2SLS estimation, using different indicators for obstacles in schooling attainment like expenses, school distance, parent's unwillingness, individual not willing, or due to involved in help at home or at work. Their results revealed higher IV estimates for returns to schooling compared to the OLS estimates while returns to experience were similar in both specifications. We think that the instruments they used for endogenous schooling may be collinear. For example

'education too expensive' may be correlated with 'parent's unwillingness' , 'had to help at home' and 'had to help at work' may affect schooling in similar way. Similarly, instrument 'school is far away' may be correlated with 'education too expensive' as second includes schooling cost etc. The Heckman two-step method used to correct for possible bias arising from sample selectivity which was found positive and significant. But they did not estimate the model correcting for both biases at the same time. Alternative to the IV2SLS approach they estimated the family fixed effects model to correct for bias due to ability and other shared family traits. Like Aslam et al. (2008), they found lower estimates from the fixed effects model than OLS which may be due to possibility of measurement error as Hertz (2003) reported an increase in the fixed effects estimates after correcting for measurement error on 13% of the original sample. The fixed effects estimates may also be lower due to not controlling for unobservable factors coming from other than family background. These two possible reasons for lower estimates in fixed effects model can be removed by using the IV2SLS approach as it corrects for any biases due to unobservables and also due to measurement error. This notion is further strengthen by Abbas & Foreman-Peck (2007) because they found similar results from the standard IV2SLS approach and fixed effects approach using the IVs. Female's returns to education found higher in all estimation methods. Nasir (2002) also estimated the Mincerian returns to education for the Pakistani labour market using data from PIHS 1995-96. They reported schooling coefficient as 7.9%, returns to experience as 4.7% and wages get at peak with 30 years of experience. Their results showed significant differences across provinces and significant positive effects on wages for being from urban areas, being a male, having technical training, being studied at private school. We think that the effect of technical training may also be captured from having professional degree or not. However, instead of belonging from urban or rural area, we believe that location of work is more suitable candidate to be included in the wage equation. They did not correct any possible biases related to estimation of the Mincerian wage regression.

Shabbir (1994) using nationally representative data collected in 1979, estimated the Mincerian wage model for males only. The results showed that each additional year of schooling increase monthly earnings by 7-8%, while returns to experience were about 6%. They controlled for urban-rural origin, provincial effects and occupational categories using different specifications. They found that earnings in Punjab are relatively lower than other provinces and there exist a significant premium for being from urban origin in all provinces except for

Khyber Pakhtunkhwa (NWFP at that time) province. Separate specifications revealed that returns to schooling were similar across provinces except for Balochistan where schooling coefficient found lower. In one specification, the study controlled for occupational categories, but the way groups made, we think that against the spirit of Mincer's model, because being a professional/technician or clerks or in agriculture may be strongly correlated with schooling levels attained. The better way would have been to capture these effects through indicator for any professional education. Similarly, in the presence of the urban-rural measure, the controlling for occupation in such a way may cause collinearity problem. For example keeping in view the Pakistani society, it seems very rare for a person from urban origin to be involved in agricultural activity. Moreover, their results may be biased as they ignored the possibility of any biases due to measurement error or endogeneity or sample selectivity problems. Another study reporting schooling returns about 7.4% is conducted by Guisinger et al. (1984). They reported earnings to be at their peak with 38 years of experience. But this study may not be representative due to small sample size and limited to only males from one urban district of one province. Moreover, they did not try to correct for any problem related to estimation of the wage regression. Similar data has been used for some other similar studies. For example, Haque (1977) explored the contributing factors of earnings for people working in formal and informal sectors.

In order to the see changes in the returns to schooling and other wage determinants over the period 1990-91 to 2001-02 in Pakistan, Khan & Toor (2003) applied the Mincerian regression by specifying schooling as the levels of education attained. Their results confirmed the increase in returns for education over the period considered. They also, confirmed the features of the Pakistani labour market of higher female educational returns but lower earnings compared to males. Their results demonstrated significant differences in earnings among provinces and between urban and rural areas which provides support for adding these measures into design matrix for estimation of the Pakistan based wage models.

Like in other countries of the globe, in Pakistani literature concerning estimation of the Mincerian wage regression there are also some studies focused to see the gaps between public and private sector wages and returns to different factors. For example Aslam & Kingdon (2009) investigated such wage gaps, and reported a large wage gap in favour of public sector employees. This wage premium for public sector workers found even larger for female

workers than male workers. Nasir (2000) is another study that investigated the wage differentials between public and private sectors in the Pakistani labour market. They specified schooling variable as levels of schooling and used age as the proxy for labour market experience. Although they reported that public sector have higher wages but returns to human capital measures found higher in private sector. This supports the human capital theory as private sector gives more importance to productivity and efficiency as compared to public sector. Similar direction of wage gaps between public and private sectors is also reported by some other studies for Pakistan (for example, Hyder & Reilly, 2005 and Hyder, 2007). All these studies used the Mincerian regression in which schooling variable defined as levels of education attained and all these did not correct for any kind of bias. This significant premium in favour of public sector workers provides support for us to include a dummy variable indicating that whether a person is working in public sector or private sector in the set of explanatory variables.

Similar to the studies on wage gaps between public and private sectors, some studies also used the Mincerian model to determine the gender wage gaps present in the Pakistani labour market. Like, Ashraf & Ashraf (1993) who studied the gaps between men and women using data collected in two time periods 1979 and 1985. They reported that gap existed in favour of men. Controlling for occupational categories their results showed significant inter-provincial effects while urban workers get 18% higher wages compared to rural workers. Although the earnings gap found in favour of men but returns to education were reported higher for females compared to males. Another study limited to Punjab province also used the Mincerian model, while studying gender wage differentials is conducted by Yasin et al. (2010) and produced results showing that men earn more despite having lower returns to educational attainments. Demonstrating wage differentials, Siddiqui & Siddiqui (1998) also revealed similar results like Ashraf & Ashraf (1993) and Yasin et al. (2010) that women have higher schooling coefficient but lower earnings. They attributed this situation as due to women's concentration in low paying jobs.

Nasir (1998) used the Mincerian model in order to judge the contribution of human capital and non human capital factors in the process of wage determination of waged and salaried worker in the Pakistani labour market. He used data from the Labour Force Survey (LFS) 1993-94 wave. He included schooling in the form dummies for different levels of education

achieved and age for experience. He found significantly different labour market conditions for men and women. Men earn higher than woman, and also men working in urban area enjoy a premium over their rural counterparts but such premium was not found for women. While endogeneity ignored, he did control for sample selection bias by using the Heckman two-step procedure and sample selectivity found significantly negative for females only. Finally, he concluded that human capital factors are significant explanatory factors in the determination of wages, as he found that only these factors explained 38% variation in response variable while with the addition of all other factors, the model explained 45% of such variation. This highlights the importance of the accumulated human capital for economic rewards in the labour market. The importance of human capital measures in the determination of wages in the Pakistani work market is also signified in Siddiqui & Siddiqui (1998) by showing that model with just schooling and age (experience proxy) accounted for 37% of variation in log earnings function while the full extended model accounted for 47% of the variation in the dependent variable.

Alderman et al. (1996) also highlighted the significance of the human capital variables with similar reasons that these accounted for most of the variation. Although they tried to cope with possibility of biases due to sample selection and endogeneity but inclusion of cognitive attainments and schooling in the same specification may have caused bias in the estimated coefficients due to possible collinearity as cognitive skills measured by literacy and numeracy may themselves be products of schooling. Further, their study cannot be generalized even for the rural areas as being relying on very small sample size (890 for selection equation and just 195 for the wage equation) and limited to males from only 4 districts of the 3 provinces ignoring Balochistan which may have entirely different labour market conditions as being the province with largest area but lowest population.

From the above brief review of the Pakistani studies that estimated the Mincerian regression, we see a general trend showing significant differences among wages in different regions, higher earnings of the people from urban origin, men get higher wages but lower returns to education compared to women, and better public sector remuneration. Further most of the studies did not correct for the any problems related to estimation of the Mincerian model like measurement error or endogeneity of schooling or sample selection bias (For example, Qureshi, 2012; Farooq & Sulaiman, 2009; Aslam & Kingdon, 2009; Hyder & Reilly, 2005;

Nasir, 2002; Siddiqui & Siddiqui, 1998; Khan & Irfan, 1985; Shabbir, 1994 and many others). The problem of bias arising due to non random selection into labour force has tackled by a few studies (Aslam, 2009; Abbas & Foreman-Peck, 2007; Khan & Toor, 2003; Nasir, 1998; Ashraf & Ashraf, 1993) using the Heckman two-step approach. The number of studies that correct for bias due to endogenous schooling is even lesser (only Aslam, 2009; Aslam et al., 2008 and Abbas & Foreman-Peck, 2007). This is may be because of the fact that most of the studies in the Pakistani context take schooling variable in the form of dummies representing different levels of schooling attained (Aslam & Kingdon, 2009; Farooq & Sulaiman, 2009; Jaffry et al., 2007; Hyder, 2007; Hyder & Reilly, 2005; Nasir, 1998 and 2000; Ashraf & Ashraf, 1993 and many others). This seems to be true, as Abbas & Foreman-Peck (2007) did not applied the IV approach in the specification using educational levels attained by stating to be so due to not having the enough instruments for each category of educational level. Similarly, although corrected for endogeneity in years of schooling's specification Aslam (2009) also not corrected it when level's specification used.

Now as we are more concerned to the problems related to estimation of the Mincerian regression. We must take the issue of endogenous schooling while estimating the model, and we are using average years of schooling as instruments. Instrumenting levels of schooling by average years of schooling does not seem plausible, so we use the schooling variable as the number of years of schooling completed.

For Pakistan, according to our knowledge no study has tried to eliminate the both problems (endogenous schooling and sample selection bias) in a single specification (that would have produced the estimates potentially free from both kinds of biases) except Alderman et al. (1996) that corrected the both biases in a single specification. But as we already mentioned that inclusion of cognitive skills in the presence of schooling, having very low sample size and using 20 years old data, their results may not be representative of the today's labour market conditions and the wage determination process in Pakistan.

So we believe that, this strengthens, a need for the estimation of the Mincerian wage regression, by using most recent available, larger in size and nationally representative data, and by eliminating the biases in order to give the most suitable model for the waged and salaried workers in the Pakistani labour market.

4.2 The Data

For estimation of the Mincerian wage regression for the Pakistani labour market, we have taken data from Labour Force Survey (LFS) by Federal Bureau of Statistics (FBS) conducted in 2008-09. The LFS data collects information about different social, economic, demographic and household characteristics of individuals. LFS is especially formulated in order to explore the characteristics of the Pakistani labour market. It has been conducted in different years since 1963. The advantage of LFS data is that it is collected over 4 quarters of year and hence avoids any seasonal effects. Survey coverage is for 4 provinces of Pakistan from both urban and rural areas. This survey excludes tribal areas and military restricted areas, but excluded area covers only 2% of national population (as given in the methodology section in report on LFS 2008-09 published by FBS (http://www.pbs.gov.pk/content/labour-force-survey-2008-09).

4.3 Variables used in Estimation of the Mincerian Model for Pakistan

Variables in the wage equation for the Pakistani data are very similar to those used for the French analysis but some variables are different in both analyses. Here, in this section, we briefly describe the variables involved in the Pakistani analysis.

4.3.1 Response variable for the Pakistani wage regression

Log Monthly Wage: Like for France, we have used monthly wages for the Pakistani analysis as well. Monthly wages in Pakistani rupees (PKR) are available for wage workers in LFS by FBS. The previous Pakistani studies used monthly (for example Shabbir, 1994; Nasir, 1998; Siddiqui & Siddiqui, 1998) and hourly (Guisinger et.al, 1984; Hyder, 2007 and others) wages as response in Mincer's semi logarithmic earnings function. As we are controlling for hours worked, so use of monthly wages is justifiable.

4.3.2 Explanatory variables for the Pakistani wage regression

Similar to that we did for the French data analysis in Chapter 3, in addition to human capital variables; we will add some other explanatory variables for the Pakistani analysis which we believe to be relevant and available for Pakistan's labour market. Here we give a description and importance of explanatory variables used for estimation of wage regression for the Pakistani data.

Schooling: In the studies applying the Mincerian model, schooling is taken in two forms as number of years of schooling and levels of education. Most of the previous Pakistani studies defined education in levels form (for example Hyder, 2007; Abbas & Foreman-Peck,2007; Nasir, 1998; Nasir, 2000; Ashraf & Ashraf, 1993) but as stated before that we are focused to correct for endogeneity bias which is not commonly tackled with IV2SLS approach if education is taken in levels form. So we take education as number of years of completed schooling. Direct measure for years of schooling is not available in Pakistan's LFS, so following many studies (Warunsir & Mcnown, 2010; Uusitalo, 1999; Arabsheibani & Mussurov, 2007 and references given in Section 3.1.1), we converted levels of schooling attained into number of years of schooling. The way how we converted different levels of education into number of years of schooling is given in what follows.

From the questionnaire designed for LFS by FBS 2008-09, we assign 1 year of schooling for those who went to school up to 'Nursery' but left below 'Kinder Garden', 2 year for those who went up to 'Kinder Garden' but did not pass 'primary' level. Similarly, we give 5 years for 'primary school completers', 8 year for 'middle school completers', 10 year to those who got 'secondary school certificate', 12 year for 'higher secondary school certificate holders', 14 year to holders of 'bachelor's degree', 16 year to those who got a 'masters degree' or educational degree awarded after 16 years of education. Finally, 18 years for schooling assigned for 'M.Phil or Ph.D'. We are unable to assign different years for M. Phil and Ph.D degrees because survey questionnaire does not separate these two educational levels.

Experience: Contrary to the French data, we do not have any measure about job seniority and year or age when individual left schooling process available in Pakistan's LFS data. So in order to capture the influence that wages face due to passing time in labour force work activity, we use only potential experience following many studies (Tansel, 1994; Uusitalo,

1999; Zhang et al., 2005; Korsun, 2010 and Lorenz and Wagner, 1990 are just few examples). Following the pioneering work of Mincer (1974) and many others we calculated potential experience by following relation.

$$Potential \quad Experience = Age_{survey\ year} - schooling - 6$$

For those who have no schooling we replaced 6 by 14 in the above relationship. As it is general practice, experience is used in both linear and quadratic terms for the estimation.

Hours Worked: Number of hours worked is controlled in the Pakistani estimation also. The LFS by FBS provides information about the number of hours worked during week preceding the data collection day. As data provides information about wages on monthly basis and working hours on weekly basis, so following Sanroman (2006) and Heckman & Hotz (1986), we multiply weekly hours worked by 4 in order to get number of hours worked during the month. Like for the French data, we also exclude observation falling in lower and upper 2.5^{th} percentiles to avoid outliers' effect. Numerically these values are 20 hours and 72 hours worked per week.

Gender: The significant effect of gender has been reported for Pakistan by many studies (Hyder & Reilly, 2005; Hyder, 2007; Khan & Toor, 2003; Nasir, 2002 and many others). The gender may have an effect on wages due to discrimination in labour market against a particular gender or due differences in skill levels possessed by men and women. Gender control is part of virtually every study that estimates the Mincerian wage regression jointly for both genders. We also add a dummy variable for gender. Females will serve as reference category in the estimation practice.

Work Location: The location at which an individual is working may have a notable effect on wages. The significant difference in wages for workers in urban and rural area has been reported for Pakistan (for instance, see Aslam, 2009; Nasir, 2002; Khan & Irfan, 1985; Siddiqui & Siddiqui, 1998) and other countries of the world (like Ismail, 2007; Arabsheibani & Mussurov, 2007; Korsun, 2010 and many others). Due to richness of the Pakistani labour market data, we are able to control for location of work urban or rural instead to proxy it with place of birth or household location (as Aslam, 2009 and Shabbir, 1994 did). We control the

urban-rural wage differential by adding a dummy variable. The rural location of work will be reference category for estimation of the wage equation.

Professional Degree: Having a technical or professional degree affects the wages of individuals in the labour market. For Pakistan, Hyder (2007) reported higher wages for people who have obtained technical training. Similar to the French analysis, we control such type of training by defining whether an individual got a degree in professional or technical education. We take individuals with degrees in engineering, medical, computer sciences and agriculture as professional degree holders and people with other degrees as general (or non professional) degree holders. The effect of the type of degree on wages is taken into account by adding a dummy variable in the set of explanatory variables. For estimation, non-professional degree holders will be reference category.

Work Sector: As we described in Section (3.1.1) that working in public or private sector of economy may substantially affect wages of individuals. This is so because the process of wage determination is different in public and private sectors of work (Hyder & Reilly, 2005; Lall & Sakellariou, 2010) as the private sector is more centred or rewarding for productivity and efficiency. Many studies that worked on public and private sector wage differentials found these effects as significant (Smith, 1976; Gunderson, 1979; Mann & Kapoor, 1988). Similar to other countries of the world, the differences between public and private sector wages found significant for Pakistan as well (for instance Hyder & Reilly, 2005; Hyder, 2007; Nasir, 2000; Qureshi, 2012). Keeping in view the significance of these differences, we defined a dummy variable indicating whether a person is working in public sector or private sector to control for these effects. The public sector is defined as working in an organization under federal government, provincial government, local bodies; or public corporation or public limited companies while others are considered as working in private sector. Individuals working in private sector will be our reference category for estimation.

Type of Contract: As it is stated in few studies (Gardeazabal & Ugidos, 2005 for instance) that contract status under which people work in labour market significantly influence wages. Similar to that for the French analysis we control for type of contract in the Pakistani analysis. We define 3 categories of workers i.e. permanent contract workers, fixed term contract workers and people who work temporarily or without any contract. Individual who work temporarily or without any contract will be reference category in estimation process.

Provincial Effects: Working in a particular region or area can affect wages in positive or negative way. It is so because of differences in local labour market conditions, differences in levels of economic activities, differences in educational or job opportunities. In Pakistan these effects can be taken into account by differentiating in individuals working in the labour markets of different provinces. Many Pakistan based studies controlled for provincial effects and found them as extensively affecting wages (Shabbir, 1994; Khan & Irfan, 1985; Ashraf & Ashraf, 1993; Yasin et al., 2010; Siddiqui & Siddiqui, 1998). We also control for 4 provinces by defining dummy variables indicating that whether individual is working in a particular province. For the estimation process, people working in Punjab province as being the most populous province, will serve as reference category.

The following Table 4.1 gives a summary of variables used for estimation of wage regression for the Pakistani data.

Table 4.1: Description of Variables used in Estimation of the Mincerian Wage Model for the Pakistani Data

Response Variable	
LNWAGE	Natural logarithm of monthly wage of individual from main job
Explanatory Variables	
SCH1	Variable for education, measured in number of years of completed schooling
EXP2	Potential experience measured in years (AGE-6-SCH1)
EXP22	Potential experience squared
HOURS3	Number of hours devoted to monthly salary (i.e. hours worked per month)
DGENDER4	Gender of the individual (Male =1 ; Female=0)
DWORKLOC5	Dummy variable indicating work location of individual (Urban=1 ; Rural=0)
DTYPDIP6	A dummy variable indicating whether individual have degree/diploma in professional education or general education (Professional Diploma=1; General=0)
DPUBLIC7	A dummy variable indicating whether individual is working in Public sector or private sector (Public =1; Private=0)
RDTMPCT0	Dummy variable indicating that individual is working temporarily without any contract (Reference category)
DFIXCT8	Dummy variable indicating that individual is working under Fixed Term Contract (Fixed Term Contract=1; else= 0)
DPERCT9	Dummy variable indicating that individual is working under Permanent Contract (Permanent Contract=1; else= 0)
RDPUNJAB0	Dummy variable indicating that individual's household is located in Punjab province (Reference category)
DSINDH10	Dummy variable indicating that individual's household is located in Sindh province (Yes=1 ; else=0)
DKPK11	Dummy variable indicating that individual's household is located in Khyber Pakhtunkhwa province (Yes=1 ; else=0)
DBALO12	Dummy variable indicating that individual's household is located in Balochistan province (Yes=1 ; else=0)

Based on the variables given in above Table 4.1, the vector X_K will take the following form for estimation of the Mincerian model in the Pakistani context,

$$X_K = \begin{bmatrix} EXP2, EXP22, HOURS3, DGENDER4, \\ DWORKLOC5, DTYPDIP6, DPUBLIC7, \\ DFIXCT8, DPERCT9, DSINDH10, \\ DKPK11, DBALO12 \end{bmatrix} \quad (4.1)$$

4.4 Variables used in Estimation of Participation Equation for Pakistan

While selecting variables for estimation of selection or participation equation based on the Pakistani data, we have taken into account different factors used in international studies as well as measures typically relevant to the Pakistani context. This section gives an overview of the variables used in analysis of selection process into wage earners sample depending on their relevance and availability.

4.4.1 Response variable for selection equation for Pakistan

Active: The response variable for the selection equation is dichotomous variable indicating that whether an individual is a wage worker or not in the labour market. This binary response variable takes 1 for waged worker and 0 otherwise. The Pakistan LFS provided information on wages only for those who are in waged or salaried job, so self employed people were excluded from the analysis. It is very common practice in studies estimating the Mincerian wage regression to exclude self employed people from their analysis (for example Chen & Hamori, 2009; Zhang et.al, 2005; Sanroman, 2006; Heckman & Hotz, 1986; Liu et al., 2000).

4.4.2 Explanatory variables for selection equation for Pakistan

The explanatory variables that we have used for the maximum likelihood estimation of participation equation are:

Age: Age is included as contributory factor in the selection or participation equation. With growing age people tend to be more willing to be employed. In addition to references given in Section (3.2.2), some Pakistani studies (Aslam, 2009; Nasir, 1998) also found age to be a

significant contributory factor in the participation decision of individuals concerning waged work.

Education or Schooling: Education measured in number of years of schooling is also used as an explanatory factor. Other than references given in Section (3.2.2), evidence from the previous Pakistani studies (like Aslam, 2009; Kozel & Alderman, 1990) also supports the inclusion of this variable. In a developing country it is more important to include education in selection equation's design matrix because people with more schooling have higher odds to attract employers and get selected for job in the market due to relative scarcity of educated and skilled workers.

Younger Children: The effect on participation decision due to presence of younger children in household is controlled for by taking number of children less than 6 years of age in the household.

Actives from Household: As many studies (Tansel, 1994; Kozel & Alderman, 1990) used income from other sources to affect the motivation to participate in the labour market. We control these effects by following García et al., (2001) who used number of income earners from the household. It is reasonable to include such measure because if more family members are active in the labour market then this leads to increase in family income which in turn may lead to lesser tastes for work for some other members of the family. Unlike for the French data, the measure on financial allocation or homeownership was not available for the Pakistani data.

Size of Household: We believe that size of the household may have an effect on the participation decision of individuals concerning labour market. The significance of the household size in contributing towards participation decision is documented by some researchers estimating selection equation (for example Bhalotra & Sanhueza, 2004; Burger, 2011; Lei, 2005). The measure is also reported as a significant factor affecting participation decision for Pakistan by Qureshi (2012). Therefore, we add size of the individual's household in the set of explanatory variables. By size we mean number of family members.

Gender: Gender may have a stronger effect in participation decision for the Pakistani labour force as compared to that of the French one. It is so because of the structure of the Pakistani

society where males are generally considered more liable for income generating activities. The importance of gender in participation decision is reported in Aslam (2009) and Aslam & Kingdon (2009) for Pakistan. We control for effects of gender with a dummy variable and females will be reference category for the probit regression estimation by ML method.

Marital Status: A dummy variable is defined to control for effect of marital status which separates married people from those who are single or widowed or divorced. It seems reasonable to believe that marital status affects wages because marriage increases the needs for the household. This effect is controlled and reported as significantly affecting participation decision of individuals for the Pakistani data (Aslam, 2009; Hyder, 2007; Nasir, 1998).

Location of Household: The difference in the educational levels, tastes for waged work and access to information in urban and rural areas may cause differences in chances of being in salaried work or not. People from rural areas may be more inclined to be involved in agricultural activities as compared to people in urban areas. Due to the location of household (urban or rural), considerable differences have been reported in likelihood for a person to be a waged worker in some previous Pakistan based studies (Aslam, 2009; Aslam & Kingdon, 2009; Nasir, 1998). It is also noted to be significant for other countries (like Asadullah, 2006; Arabsheibani & Mussurov, 2007; Palme & Wright, 1998). Hence a dummy variable is included that indicates urban or rural location of household to which an individual belongs. For the estimation, people from household located in rural areas will serve as reference category.

Head of Household: A dummy variable is also added in the independent variables to take the effect of being head of the household on participation decision. It seems logical to believe that being head of the household increases the chances of positive decision regarding waged work participation because head is regarded as more liable for providing resources to run the household as compared to other members of the family, particularly in eastern societies like Pakistan. This variable is found significantly contributory in some studies (Aslam, 2009; Aslam & Kingdon, 2009; Hyder, 2007) estimating selection equation for Pakistan.

Professional Degree: Like increase in schooling increases odds of being selected for a job, similar is the case for having a professional degree. Difference in demand and supply of workers with professional or non professional degrees may cause differences in their chances

for getting a salaried work. We expect that people with professional degree may get their jobs in a period shorter than that required for non-professional degree holders to get a job. Dummy variable for professional degree is defined identically as we did for wage equation estimation for Pakistani data.

Provincial Effects: The effect of being a resident of a particular province has found to influence significantly the decision of individual to work or not in the labour market as wage worker in some previous studies about Pakistan (for instance, see Aslam & Kingdon, 2009; Hyder, 2007). We also control for these provincial effects on participation decisions of individuals. We define dummy variable indicating province to which a particular individual belongs. Similar to wage function, Punjab will serve as reference category in estimation of the probit regression via ML method.

Following Table 4.2 gives a summary of variables used for estimation of participation equation through the ML probit regression for the Pakistani data.

Table 4.2: Description of Variables used in Estimation of Selection Equation for the Pakistani Data

Response Variable	
ACTIVE	A dummy variable indicating whether the person is ACTIVE in the Labour Market or not (AVTIVE =1; else=0)
Explanatory Variables	
AGE1	Age of the individual in completed years at the last day of reference week.
SCH2	Variable for education, measured in number of years of completed schooling
CH6Y3	Number of children under 6 years of age in the household to which an individual belongs
NACTIVE4	Number of persons economically active in labour market from individual's household.
SIZEHH5	Size of household (Number of persons in the household)
DGENDER6	Gender of the individual (Male =1 ; Female=0)
DMSTAT7	Legal marital Status of the Person (Married=1; Single/Divorced/Widowed=0)
DLOCHH8	Location of the individual's household (Urban=1 ; Rural=0)
DHEADHH9	Dummy variable indicating whether individual is head of the household or not (Head=1 ; else=0)
DTYPDIP10	A dummy variable indicating whether individual have degree/diploma in professional education or general education (Professional Diploma=1; General=0)
RDPUNJAB0	Dummy variable indicating that individual's household is located in Punjab province (Reference category)
DSINDH11	Dummy variable indicating that individual's household is located in Sindh province (Yes=1 ; else=0)
DKPK12	Dummy variable indicating that individual's household is located in Khyber Pakhtunkhwa province (Yes=1 ; else=0)
DBALO13	Dummy variable indicating that individual's household is located in Balochistan province (Yes=1 ; else=0)

Based on the variables described in above table, the vector V_K (refer to Chapter 2, Eq. 2.4) takes the following form for estimation of the probit regression for the Pakistani data:

$$V_K = \begin{bmatrix} AGE1, SCH2, CH6Y3, NACTIVE4, \\ SIZEHH5, DGENDER6, DMSTAT7, \\ DLOCHH8, DHEADHH9, DTYPDIP10, \\ DSINDH11, DKPK12, DBALO13 \end{bmatrix} \quad (4.2)$$

4.5 Estimation Results Based on the Pakistani Data

The present section presents the results for the Mincerian wage regression based on the Pakistani data estimated by different estimation methods. Mainly the analysis is done in a similar way as we did for the French case in Chapter 3 but it differs in the set of variables used in the estimation process.

4.5.1 Preliminary estimation

First we present the results of the preliminary estimation of the model given in Eq. 2.1 which is estimated via OLS approach. The variables used for estimation process for the Pakistani data are given in Table 4.1. We name this model as **Model 4.1**[11]. The results from Model 4.1 are given in following Table 4.3.

[11] see appendix B2 for algebraic form of Model 4.1

Table 4.3: OLS Estimation of the Pakistani Wage Regression

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.3968	0.3964	989.58	<0.0001	19574
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	7.9780	2.76E-02	289.50	<0.0001
SCH1	0.0479	8.35E-04	57.36	<0.0001
EXP2	0.0305	1.09E-03	27.94	<0.0001
EXP22	-0.0004	2.24E-05	-16.63	<0.0001
HOURS3	0.0019	3.93E-04	4.83	<0.0001
DGENDER4	0.1044	1.35E-02	7.72	<0.0001
DWORKLOC5	0.0694	7.29E-03	9.52	<0.0001
DTYPDIP6	0.4725	2.41E-02	19.60	<0.0001
DPUBLIC7	-0.0848	5.45E-03	-15.56	<0.0001
DFIXCT8	0.0120	1.30E-02	0.92	0.3559
DPERCT9	0.2305	9.58E-03	24.07	<0.0001
DSINDH10	-0.0430	8.32E-03	-5.17	<0.0001
DKPK11	-0.0475	9.66E-03	-4.91	<0.0001
DBALO12	0.0114	1.07E-02	1.07	0.2865

Table 4.3 shows the results from preliminary (OLS i.e. Model 4.1) estimation of the Mincerian wage regression for the Pakistani data. The *p*-value associated with F-statistic speaks at overall significance of the model. The explanatory power of the model judged by R^2 is 39.68% which is close to the other Pakistani studies employing the Mincerian model (Shabbir, 1994; Abbas & Foreman-Peck, 2007; Khan & Toor, 2003 and Guisinger et al., 1984, for example) and also near to the R^2 value in the pioneering work of Mincer (1974). First we discuss the human capital variables. We see that for the Pakistani labour market, education plays a significant role in the process of wage determination. Wages increased by about 4.80% with each additional year of schooling. The impact of education found in Table 4.3 seems less as compared to the OLS specification for some other studies working with Pakistani data (Nasir, 2002 and Guisinger et al., 1984 that reported schooling coefficient as above 7% with the OLS specification). One possible explanation for this difference may be the fact that our OLS specification differs from those in terms of the explanatory variables. For example, both of the studies mentioned did not control for the urban-rural differences,

public-private differences and also they did not take into account the types of contract under which individuals are working. Our schooling coefficient found also less than that reported in Aslam (2009) (7.2% for males and 16.6% for females) and Abbas & Foreman-Peck (2007) (9.2% for males and 14% for females). The above stated explanation may also be true for these two studies as well, as they did not control for the said factors in their OLS specification. The returns to experience are found to be 3.045% with each additional year spent as wage worker in labour market. The negative coefficient related to quadratic term for experience reveals the concavity of the experience–wage relationship which is in confirmatory in almost all Mincer based studies. From coefficients related to linear and quadratic terms of experience, we inferred that wages get at peak with an experience of about 41 years. As expected the hours worked contribute significantly in explaining wages in the Pakistani labour market. An additional hour of work increases wage by 0.19%. Consistent with many other country's results (Johnson & Chow, 1997 for China; Lassibille, 1998 for Spain; Asadullah, 2006 for Bangladesh; Ismail, 2007 for Malaysia; Korsun, 2010 for Ukraine and many others) we found significant raw gender wage differentials in favour of men. The wage penalty for women is also found evident in previous Pakistani studies (Ashraf & Ashraf, 1993; Nasir, 2002; Khan & Toor, 2003; Nazli, 2004; and Hyder & Reilly, 2005). However we found a raw wage gap of a magnitude near to 10% between two genders, a gap which is lower than the one estimated in all the studies quoted above about Pakistan. The explanation for this difference with previous studies may lie in the increased schooling of women in recent years as we use more recent data and secondly it may have dropped because of extra control variables that we used. Keeping in view the significant wage differences between workers from the urban and rural labour markets reported by many researchers (for example Ismail, 2007; Arabsheibani & Mussurov, 2007; Korsun, 2010), we also controlled for the urban and rural effects on wages. Instead of using household location or place of birth (as used by Aslam, 2009 and Shabbir, 1994) to have an idea about the work location, the LFS 2008-09 data permitted us to use the precise dummy variable indicating work location as urban or rural. Consistent with some of the other studies employed in the Pakistani context, we also report wage premium of nearly 7% for people working in urban areas over their rural counterparts. The finding about direction of the wage differences between urban and rural workers is same as in some other studies (Khan & Irfan, 1985; Farooq & Sulaiman, 2009). But the magnitude of that gap is a little lower from some other studies (for example Nasir, 1998 and Khan & Toor, 2003 reported with an amount 12% and 21%, respectively). From

above results (Table 4.3, Model 4.1), we note that technical or professional degree has significant positive effect on wages of individuals in Pakistan's labour market. The coefficient associated with dummy variable for professional degree shows that people possessing a professional degree get about 47% higher wages with reference to similar workers having a non professional degree. The wage premium for people with technical training or education is also found by some others (Nasir, 2002; Khan & Toor, 2003; Hyder, 2007) for the Pakistani data. Concerning wage differences between public and private sectors of work, our results show that people working in private sector of economy are better paid as compared to workers in public sector. The results suggest that individuals working in private jobs get 8.4% more wages than those working in public sector. This is in contrast with the previous findings (Hyder & Reilly, 2005; Hyder, 2007) that reported higher wages for public sector employees in Pakistan. Related to differences in wages due to the contract under which individuals work, our results show that there are no significant differences in wages of individuals working under fixed term contract and working temporarily in the Pakistani labour market. However, the impact of working under a permanent contract substantially influences labour market wages. The coefficient related to dummy variable for permanent contract revealed that people working under permanent contract get 23% and 22% higher salaries as compared to those working under fixed term contract and who work temporarily, respectively.

Finally, we focus on the provincial effects. We have taken Punjab province as reference category. The results show that people working in the Sindh and Khyber Pakhtunkhwa provinces face significantly lower wages as compared to the Punjab province with a magnitude of about 4.3% and 4.7%, respectively, while workers in the Balochistan province have wages which are not statistically significantly different from workers in the Punjab province. So we may conclude that individuals working in Punjab (the province with largest population) and Balochistan (the province with the largest area) get higher wages compared to those working in two other provinces. The finding of higher wages for workers in Balochistan province is in line with the findings from some other studies (like Khan & Irfan, 1985; Ashraf & Ashraf, 1993; Siddiqui & Siddiqui, 1998; and Khan & Toor, 2003; Farooq & Sulaiman, 2009) but the finding of higher wages for workers in Punjab province compared to the Sindh and Khyber Pakhtunkhwa provinces is in contrast with the previous findings (for example, Khan & Irfan, 1985; Ashraf & Ashraf, 1993; Shabbir, 1994; Khan & Toor, 2003; Kozel &

Alderman, 1990). This contrast hints at the changes in the labour market conditions in different regions in recent past years in Pakistan.

4.5.2 Instrumental variables 2SLS estimation using Instrument-1

The problem of endogeneity is tackled by the IV2SLS approach. For the Pakistani data, we employ the IV2SLS technique using the similar instruments as we used for the French data. Definitions of the instruments are given in the Section (3.5). For the Pakistani analysis, the IV2SLS estimation is applied by putting explanatory variables from Table 4.1 and using the instruments (as defined in Section 3.5) in the model explained in section 2.2.1.1.

First we used Z1 as an instrument for endogenous schooling. We name this first IV2SLS model for Pakistan as **Model 4.2**[12]. Results from the IV2SLS estimation for the Pakistani data with Z1 as instrument for schooling (i.e. Model 4.2) are given in Table 4.4-A and Table 4.4-B for first stage and second stage regressions, respectively as below.

[12] see appendix B2 for algebraic form of Model 4.4

Table 4.4-A: First Stage of IV2SLS Estimation for Pakistan using Z1 as Instrument

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.7646	0.7645	4888.02	<0.0001	19574
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	3.8999	1.53E-01	25.52	<0.0001
EXP2	-0.0882	6.07E-03	-14.53	<0.0001
EXP22	-0.0003	1.26E-04	-2.62	0.0088
HOURS3	-0.0318	2.19E-03	-14.49	<0.0001
DGENDER4	1.1798	7.65E-02	15.42	<0.0001
DWORKLOC5	-0.4573	4.15E-02	-11.02	<0.0001
DTYPDIP6	1.4916	1.35E-01	11.06	<0.0001
DPUBLIC7	-0.1506	3.06E-02	-4.92	<0.0001
DFIXCT8	0.4903	7.25E-02	6.77	<0.0001
DPERCT9	1.7197	5.22E-02	32.92	<0.0001
DSINDH10	0.5846	4.65E-02	12.56	<0.0001
DKPK11	0.6737	5.43E-02	12.41	<0.0001
DBALO12	0.8570	6.05E-02	14.16	<0.0001
Z1	0.9273	5.74E-03	161.54	<0.0001

The results from first stage schooling regression are presented in Table 4.4-A. From these results we see that model is over all significant with reasonably higher value of R^2. The substantially lower p-value associated with instrument (Z1) means that instrument used is significantly contributory in explaining endogenous schooling which proves its relevance.

Table 4.4-B: Second Stage of IV2SLS Estimation for Pakistan using Z1 as Instrument

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.3945	0.3941	980.42	<0.0001	19574
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	7.7539	2.95E-02	262.95	<0.0001
SCH1_Z1	0.0648	1.12E-03	58.02	<0.0001
EXP2	0.0340	1.11E-03	30.57	<0.0001
EXP22	-0.0004	2.30E-05	-17.12	<0.0001
HOURS3	0.0029	3.99E-04	7.14	<0.0001
DGENDER4	0.1151	1.37E-02	8.42	<0.0001
DWORKLOC5	0.0539	7.39E-03	7.30	<0.0001
DTYPDIP6	0.3820	2.47E-02	15.49	<0.0001
DPUBLIC7	-0.0668	5.56E-03	-12.03	<0.0001
DFIXCT8	-0.0128	1.31E-02	-0.98	0.3282
DPERCT9	0.1563	1.02E-02	15.33	<0.0001
DSINDH10	-0.0533	8.42E-03	-6.33	<0.0001
DKPK11	-0.0467	9.76E-03	-4.79	<0.0001
DBALO12	0.0240	1.08E-02	2.22	0.0266

We are mainly interested in estimation of wage regression. The results from second stage of IV2SLS estimation which uses schooling variable replaced by its fitted values from the first stage schooling equation (Table 4.4-A), are shown in above Table 4.4-B. The significant F-statistic in the above table shows over all significance of the model. Goodness of fit of the above model (Model 4.2) judged by R^2 (0.3945) is similar to that found in the OLS estimation (Model 4.1) and also well comparable to concerned literature. Coming to the contributory factors, we note that impact of education is increased by nearly 2 percentage points from that obtained in OLS estimation. Downward bias in the OLS estimates for schooling coefficient is in line with the relevant international (Butcher & Case, 1994; Maluccio, 1998; Harmon & Walker, 1999; Chen & Hamori, 2009 and many others) as well as Pakistani (Abbas & Foreman-Peck, 2007; Aslam, 2009) literature. From the IV2SLS estimation with Z1 as instrument (Model 4.2) results, wages increase by 6.48% with increase of one year in the schooling attainment. The returns associated to other human capital variable i.e. experience, increased by from 3.05% (in OLS Model 4.1) to 3.4% (above Model 4.2) for each additional

year spent in labour market. Similar to that noted in OLS Model 4.1, concavity of the wage experience relationship holds in Model 4.2 as well. Wages get at their peak with roughly 43 years of experience. The impact that hours worked have on wages is significant and increased a little in Model 4.2 compared to the uncorrected OLS (Model 4.1) results. The raw gender wage gap is found in favour of men with a magnitude of 11.5%. It means that everything remaining equal, on average females get wages which are approximately 11.5% lesser as compared to their male counterparts. The effect of working in urban areas is decreased a little in Model 4.2 to 5.4% compared to Model 4.1 that produced such effect to be approximately 7%. The effect of technical or professional degree decreased by 9 percentage points as compared to the results in Table 4.3 (OLS estimation or Model 4.1). Here, from Model 4.2, the coefficient suggests a wage gain of 38% for people with a professional or technical educational degree over those who possess a non vocational or general degree. The wage penalty for being a worker in public sector is found to be about 6.68% with a little decrease from the OLS (8.47%). Similar to the OLS results, we find no evidence of significant wage differences between temporary workers and workers working under fixed term contracts. However, people working under permanent contracts enjoy significantly higher wages compared to temporary and fixed term contract workers. The magnitude of the wage premium for permanent contract workers is found 15.6% after a decrease of about 7-8 percentage points from that reported in the OLS (Model 4.1) results. The wage differentials due to working in different provinces are similar to those found in Model 4.1. But the wage premium for individuals in Balochistan province compared to Punjab province turns out to be significant which was non-significant in previous specification. From above Model 4.2, the people working in Balochistan get higher wages followed by workers in the Punjab, Khyber Pakhtunkhwa and Sindh provinces, respectively.

4.5.2.1 Hausman test for endogeneity of schooling using Instrument-1

Table 4.4-C: Endogeneity Test for Schooling with Z1 as Instrument

Efficient under H0	Consistent under H1	Statistic	P-value
OLS	2SLS	518.10	<0.0001

In order to test the exogeneity of schooling for the Pakistani data, we apply the Hausman (1978) test. The substantially low p-value associated to Hausman test statistic in Table 4.4-C clearly rejected the exogeneity assumption regarding schooling variable. So like for France, for Pakistan too, due to existence of endogeneity in the schooling variable, IV2SLS estimation is more preferred than the OLS estimation which ignores the presence of endogeneity bias.

4.5.3 Instrumental variables 2SLS estimation using Instrument-2

Now we repeat the IV2SLS estimation process with Z2 (defined in Section 3.5) as an instrument for endogenous schooling. We name this model as **Model 4.3**[13]. The results from this IV2SLS estimation are given in Table 4.5-A (first stage schooling regression) and Table 4.5-B (second stage wage regression).

[13] see appendix B2 for algebraic form of Model 4.3

Table 4.5-A: First Stage of IV2SLS Estimation for Pakistan using Z2 as Instrument

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.4668	0.4664	1317.24	<0.0001	19574
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	3.4398	4.57E-01	7.53	<0.0001
EXP2	0.0432	1.38E-02	3.14	0.0017
EXP22	-0.0006	1.99E-04	-3.03	0.0024
HOURS3	-0.0508	3.30E-03	-15.39	<0.0001
DGENDER4	-4.5419	1.97E-01	-23.09	<0.0001
DWORKLOC5	0.8868	6.11E-02	14.51	<0.0001
DTYPDIP6	4.6796	2.02E-01	23.20	<0.0001
DPUBLIC7	-1.0095	4.54E-02	-22.25	<0.0001
DFIXCT8	1.3593	1.09E-01	12.50	<0.0001
DPERCT9	4.1972	7.50E-02	55.94	<0.0001
DSINDH10	0.5535	7.01E-02	7.90	<0.0001
DKPK11	-0.0573	8.15E-02	-0.70	0.4820
DBALO12	-0.6930	8.99E-02	-7.71	<0.0001
Z2	1.9483	8.00E-02	24.36	<0.0001

Results from the first stage of IV2SLS estimation using Z2 as an instrument are given in Table 4.5-A. The above table shows the significance of the model in global sense. R^2 in the first stage regression is lower compared to that obtained in first stage regression of IV2SLS estimation with Z1 as instrument (Model 4.2, Table 4.4-A) but even then it is well above to the first stage R^2 found in different other (Boumahdi & Plassard,1992; Pons & Gonzalo, 2003; Abbas & Foreman-Peck,2007; Aslam ,2009) studies. The significance of the coefficient associated to instrument Z2 in the above first stage regression explaining suspected endogenous schooling means that Z2 can be considered as a relevant instrument for the endogenous schooling.

Table 4.5-B: Second Stage of IV2SLS Estimation for Pakistan using Z2 as Instrument

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.2922	0.2917	621.06	<0.0001	19574
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	7.2108	7.75E-02	93.09	<0.0001
SCH1_Z2	0.1056	5.43E-03	19.46	<0.0001
EXP2	0.0425	1.65E-03	25.73	<0.0001
EXP22	-0.0004	2.50E-05	-16.68	<0.0001
HOURS3	0.0052	5.32E-04	9.69	<0.0001
DGENDER4	0.1410	1.55E-02	9.13	<0.0001
DWORKLOC5	0.0166	9.49E-03	1.75	0.0800
DTYPDIP6	0.1626	3.93E-02	4.13	<0.0001
DPUBLIC7	-0.0234	8.32E-03	-2.81	0.0050
DFIXCT8	-0.0729	1.64E-02	-4.43	<0.0001
DPERCT9	-0.0236	2.59E-02	-0.91	0.3618
DSINDH10	-0.0783	9.84E-03	-7.95	<0.0001
DKPK11	-0.0450	1.08E-02	-4.17	<0.0001
DBALO12	0.0544	1.26E-02	4.33	<0.0001

Table 4.5-B present results for wage regression from second stage of IV2SLS estimation that uses Z2 as instrument in first stage schooling equation. The model's global performance is less than both of the previous two models namely Model 4.1 (OLS) and Model 4.2 (IV2SLS with Z1 as instrument for schooling) as R^2 is 0.29 decreasing by approximately 10 percentage points (compared to previous two models). The returns to schooling differ significantly from the models estimated previously. From the IV2SLS estimation using Z2 as instrument (i.e. Model 4.3), we found that each additional year of schooling increases wages of individual by 10.56%. The returns to schooling from above model are more than double than those found from the OLS specification (Model 4.1) which assumes schooling as exogenous. The higher schooling coefficient from the IV2SLS approach is quite common in the literature devoted to returns to schooling. But vast difference (as we found in above Model 4.3) between the schooling coefficients produced by the IV2SLS and OLS is rare but reported in some studies (Card, 1993; Butcher & Case, 1994). From above model we find returns related with an extra year of labour market experience as 4.2% which are higher as compared to those from the

OLS Model 4.1 (3.05%) and IV2SLS with Z1 as instrument (3.4%). The impact of hours worked found to be 0.52% which is higher compared to that found in previous two specifications. The gender wage gap is significant in favour of men as it was before but in this specification its magnitude is higher. The wage gains related to being a male worker is found to be 14% from above IV2SLS specification using Z2 as an instrument (Model 4.3). A substantial difference that comes out with this model is non-significance (at 5% significance level) of the effect of work location. In contrast to the findings in the OLS (Model 4.1) and IV2SLS with Z1 as instrument (Model 4.2) models, the above Model 4.3 shows that wage penalty for a rural worker is less severe. The effect of professional or vocational degree on wages is minimized in the above estimation given in Table 4.5-B. From these results, the workers with general educational degrees earn 16% less compared to those possessing a professional or technical educational degree. This magnitude of wage gains to professional degree holders is more close to that found in Hyder (2007), than those found in other models of the present work (Model 4.1 and Model 4.2). The wage differential between public and private sector employees seems to be narrowed in the Model 4.3. Form above Table 4.5-B, we found that wage differential in favour of private sector workers is about 2% which is significant at significance level of 5%. This wage gap is substantially lower in terms of magnitude from both the previous estimations. Another interesting result obtained from Model 4.3 is that permanent contract workers do not have any extra wages due to their contract status over temporary workers while people with fixed term contract face a statistically significant wage penalty of about 7% compared to temporary workers. These results concerning contract statuses are in contrast with expectations and results from previous specifications. But this may be due to more concentration of temporary workers in private jobs and as we have seen the wage gains for private sector narrowed in this specification so we may suspect that this is may be due to correlation between being a private sector worker and in temporary work. Lastly, focusing on provincial effects, our results from Model 4.3 show that people in the Balochistan province get 5.4% better wages as compared to workers in the Punjab province and difference is statistically highly significant. While the people from Khyber Pakhtunkhwa and Sindh provinces get respectively 4.4% and 7.8% lesser wages as compared to individuals working in the Punjab province. From the above estimation (Model 4.3), the amounts of wage premiums or penalties concerning provinces change but the order of wages remains same as found in previous estimations, that is, people working in the

Balochistan province get higher wages followed by workers from Punjab, Khyber Pakhtunkhwa and Sindh provinces, respectively.

4.5.3.1 Hausman test for endogeneity of schooling using Instrument-2

Table 4.5-C: Endogeneity Test for Schooling with Z2 as instrument

Efficient under H0	Consistent under H1	Statistic	P-value
OLS	2SLS	115.80	<0.0001

In order to test the presence of endogeneity bias in schooling coefficient after using Z2 as instrument for schooling, the Hausman (1978) test is applied. The significance of the Hausman test statistic in Table 4.5-C points out the presence of endogeneity bias in the OLS estimated schooling coefficient. Due to existence of endogeneity in schooling variable, the IV2SLS estimation is more preferred (in case of Z2 as instrument as well) than OLS estimation which assumes schooling to be exogenous.

4.5.4 Choice between two instruments for the Pakistani data

From both the IV2SLS estimations (with Z1and Z2 as instrument separately i.e. in Model 4.2 and Model 4.3, respectively), we have seen that the OLS estimates for schooling coefficient are downward biased. In order to choose the most appropriate instrument for further analysis, we proceed in similar way as we did for the French analysis (Section 3.6.4). We calculated correlation matrix among response variable (LNWAGE), endogenous explanatory variable (SCH1), Instrument-1 (Z1), and Instrument-2 (Z2). The correlation matrix is presented in Table 4.6 below.

Table 4.6: Correlation Matrix among Response, Endogenous Variable and Instruments (Z1, Z2), for the Pakistani Data

Variable	LNWAGE	SCH1	Instrument-1 (Z1)	Instrument-2 (Z2)
LNWAGE	1.0000	0.4762	0.4993	-0.0531
SCH1		1.0000	0.8275	0.3458
Instrument-1 (Z1)			1.0000	0.1845
Instrument-2 (Z2)				1.0000

Similar to that we have seen from correlation matrix of these variables for the French data, for Pakistani data (Table 4.6) too, we find that correlation between Instrument-1 (Z1) and endogenous schooling (SCH1) is stronger than the correlation between Instrument-2 (Z2) and endogenous schooling (SCH1). On this hand Instrument-1 (Z1) seems to be more relevant but on the other hand (similar to the French data case, Table 3.6) correlation of Instrument-1 (Z1) with response variable (LNWAGE) is also stronger than the correlation of Instrument-2 (Z2) with response (LNWAGE). The higher correlation between instrument and response variable violates the assumption of an instrument that it should affect response only through its effect on the endogenous variable and otherwise should be uncorrelated with response variable.

Therefore, we decide between these two instruments in a similar way as we did for the French case, that is, based on more robust definition (definition of both instrument are same for the French and Pakistani analyses) of Instrument-2 (i.e. Z2) and lesser correlation between response variable and Instrument-2 (Z2). So based on this criterion, we choose Instrument-2 (Z2) to be more appropriate and it will be used as instrument for further analysis of the Pakistani labour market data.

4.5.5 Sample selection model: Step-1: Estimation of participation equation for the Pakistani data

As of now we have addressed the problem of endogeneity bias. Now we go for to address the bias due to non randomness of the wage earners sample. For this, we estimate (as we did for the French case in Section 3.6.5) sample selection model proposed by Heckman (1979) which is based on two step estimation (first step probit regression for selection equation and second step wage regression having IMR as an additional regressor).

In this section we present the results from Step-1 of the sample selection model i.e. maximum likelihood estimation of the probit regression for selection equation, for the Pakistani data. In the selection equation estimation, indicator for salaried worker is binary response variable and the description of the explanatory variables is given in details in Section 4.4.2 (see Table 4.2 for a brief summary about variables used for estimation of selection equation for the Pakistani data). For the Pakistani data, estimation of selection equation is based on 66245 observations. The results from the probit estimation of the selection equation for the Pakistani data are given in Table 4.7 below.

Table 4.7: Estimation of Selection Equation for Pakistan

Variable	Parameter Estimate	Standard Error	Wald Chi-Square	P-value
INTERCEPT	-0.3235	2.63E-02	150.89	<0.0001
AGE1	-0.0096	6.10E-04	249.85	<0.0001
SCH2	0.0366	1.11E-03	1077.79	<0.0001
CH6Y3	-0.0127	5.35E-03	5.631	0.0176
NACTIVE4	-0.0659	4.51E-03	213.62	<0.0001
SIZEHH5	-0.0084	2.63E-03	10.25	0.0014
DGENDER6	0.2883	1.67E-02	298.40	<0.0001
DMSTAT7	-0.0351	1.55E-02	5.10	0.0239
DLOCHH8	0.5063	1.12E-02	2051.45	<0.0001
DHEADHH9	0.1273	1.66E-02	58.85	<0.0001
DTYPDIP10	0.3226	5.80E-02	30.98	<0.0001
DSINDH11	-0.1501	1.26E-02	141.11	<0.0001
DKPK12	0.0609	1.56E-02	15.28	<0.0001
DBALO13	-0.2077	1.63E-02	162.06	<0.0001

Table 4.7 presents the results from estimation of the probit regression for selection equation (first step of the sample selection model). First, our results show that increase in age exerts a negative effect on the probability of being salaried in the labour market. This is in contrast to results found by many other people (Meurs & Ponthieux, 2000; Arabsheibani & Mussurov, 2007; Chen & Hamori, 2009; Burger, 2011 for other countries and Nasir, 1998 and Aslam, 2009 for Pakistan) who found the effect of age as positive. However, Aslam & Kingdon (2009) reported similar impact of age on the selection into wage earners sample. One possible explanation for this difference could lie in the way we defined the response variable in the selection equation. We defined response variable as being salaried or not instead of being

active in the labour market. Furthermore, the negative coefficient is justifiable as employers tend to recruit fresh graduates so chances of being selected in salaried work decreases with increasing age. As expected, education plays a significant positive role in the chances of being a salaried or waged worker. This is in line with the findings from other studies for Pakistan (like Kozel & Alderman, 1990; Nasir, 1998; Aslam, 2009; Aslam & Kingdon, 2009) and for other countries of the world (Riboud, 1985; Lei, 2005; Martins, 2001, for example). More children under 6 years of age reduce the chances of being involved in waged work. As we are applying a joint regression for males and females so this negative effect seems to be present mainly because of women as it is reasonable to believe that woman's chances of monetary work activity will be lesser if she has to look after for younger children at home. The negative impact of younger children on selection equation is found significant in other countries for females (Riboud, 1985; Arabsheibani & Mussurov, 2007; Chen & Hamori, 2009) but our results are more in line with Aslam & Kingdon (2009) and Aslam (2009), which are Pakistan based studies that found negative effects of younger children on selection into wage earners sample for both genders. To capture the effects of other sources of income on the probability of selection into salaried work, we have used number of other persons active (salaried or self employed) in the labour market from the concerned house hold. Our results show that likelihood of being in waged work decreases with increasing number of persons working in the labour market from household. It is reasonable to believe that when more people from house hold are working then over all house hold's income will be increased which leads to lesser motivation for work for some of the household members. Negative effects for this kind of variable has also been found in other studies that used unearned income from other sources (Kozel & Alderman, 1990; Tansel, 1994; Lei, 2005; Agrawal, 2011) and that used spouse income/wage to capture for a similar effect (Riboud, 1985; Coelho et al., 2010). The size of household has negative effect of being in waged job. This effect of household size is found mixed in literature. It is found negative for both genders in some studies (Agrawal, 2011 for example) and positive for males while negative for females in some (Lei, 2005) studies. The positive effect of gender points out that being a male enhances the likelihood of selection in to waged work. This is reasonable as participation rates for men are much higher than for women in Pakistan. Also keeping in view the nature of the Pakistani society males are considered more responsible for the income generating activities. The more odds of being in waged work for males are also manifested in other countries from south Asia (Asadullah, 2006 for Bangladesh and Agrawal, 2011 for India). The effect of being married on the

probability of being a salaried worker is negative. Like negative effect of younger children, we think that this is also because of joint estimation for both genders. This belief seems reasonable because some other studies reported positive effect of being married for males and negative for females (Lei, 2005; Aslam, 2009 for example). Another important factor that we controlled for is the effect that location of household to which individual belongs puts on the probability of his/her selection in to wage earners sample. According to expectation and similar to that found in literature (Asadullah, 2006; Arabsheibani & Mussurov, 2007; Chen & Hamori, 2009 and others) we also found that people belonging to urban areas have higher probability to get into the salaried work as compared to their rural counterparts. People in rural areas have more tendencies towards working in agricultural activities and other professions of self employment as compared to urban workers. As we have excluded persons who are self employed or involved in agricultural activities from analysis, so it is easily explainable that people from urban households have higher odds for being a wage worker. The positive effect of belonging to urban area is also found for Pakistani data by Aslam & Kingdon (2009). Another well expected result is the positive effect of being head of household on the likelihood of participation in salaried activity. It was expected so because head of household is more responsible for providing income to run the household in general as well as in particular in the Pakistani context. This is in confirmatory with results from Aslam et al. (2008) and Aslam (2009) which are only studies we found that controlled for effects of being head of household on the probability of labour market participation. The impact of having a professional or technical educational degree significantly increases the chances of getting involved into waged work. This positive impact is similar to results found for France in the present thesis and also in line with findings by some other international (Tansel, 1994 for Turkey) and Pakistani (Hyder, 2007; Nasir, 1998 for example) researches. Finally, taking the provincial effects, we see that individuals who belong to the Sindh and Balochistan provinces have significantly lower chances of involved in salaried work compared to individuals from Punjab province while people from Khyber Pakhtunkhwa have significantly higher odds of being in salaried work than their counterparts from the Punjab province. Our results regarding comparison of Sindh and Balochistan with Punjab are similar to those found in Aslam (2009) but in opposition to those found by Aslam & Kingdon (2009). The results regarding comparison between Khyber Pakhtunkhwa and Punjab are different in our study in the sense that both these studies (Aslam & Kingdon, 2009 and Aslam, 2009) found non-significant differences in the likelihood of being in salaried work due to belonging

from these two provinces. In summary, we may conclude that the people from Balochistan province have lower probability of getting a job compared to other provinces. This may be due to relative scarcity of availability of highly educated and skilled workers required for salaried jobs in this province. This may also be due to more tendencies of the people from the Balochistan province towards self-employment activities.

4.5.5.1 Marginal effects of explanatory variables on the probability of selection into the wage earners' sample in Pakistan

As we have discussed in Chapter 3 that estimated coefficients from the Maximum likelihood probit regression tell only the significance and direction of effect that different explanatory variables exert on the binary response variable. But they do not indicate the marginal effects on the probability of being '1' in binary response. Table 4.8 presents the marginal effects that different contributory factors put on the probability of being selected in wage earners sample in Pakistan's labour market. These effects are calculated in similar way as we have explained for the French data case (Section 3.6.5.1; Eq. 3.4).

Table 4.8: Marginal Effects of Different Explanatory Variables on the Probability of Selection into Wage Earner's Sample

Variable	MARGINAL EFFECT on P(ACTIVE=1)
AGE1	-0.0036
SCH2	0.0138
CH6Y3	-0.0048
NACTIVE4	-0.0249
SIZEHH5	-0.0032
DGENDER6	0.1090
DMSTAT7	-0.0133
DLOCHH8	0.1913
DHEADHH9	0.0481
DTYPDIP10	0.1219
DSINDH11	-0.0567
DKPK12	0.0230
DBALO13	-0.0785

From above Table 4.8, we see that with an increase of one year in an individual's age, the probability of getting into waged or salaried work in labour market decreases by

approximately 0.0036 percentage points. The increase in education enhances the odds of to be a wage worker in such a way that likelihood of being a salaried workers in the labour market increases by 0.0138 percentage points with an increase of one more year in individual's schooling. The probability of waged work participation reduces by 0.0048 percentage points for every more child of less than 6 years of age in the household. Similarly, if one more person from household joins labour market for monetary gains then individual's chances for to be a salaried worker diminishes by 0.0249 percentage points. Size of the household also has a negative effect on the likelihood of being in salaried work participation. From the marginal effects (Table 4.8) we see that if size of the household increased by one person, then this reduces the odds of being in waged job by 0.0032 percentage points. The other two measures related to personal characteristics are gender and marital status. We note that being a female and being married lessens the probability of holding a salaried job by approximately 11 and 0.0133 percentage points compared to being a male and living as single, respectively. Location of household is another significant contributory factor related to binary response of being in job or not. People who belong to a family living in urban areas have 0.19 points more probability of getting a job when compared to those who belong to family from rural areas. Being head of family have 0.048 points more odds to be a salaried or waged worker than other members of the household. Comparison between professional and general educational degrees shows that individuals who have a professional degree have a 0.12 points higher probability of getting a job compared to those people who possess a general educational degree. Finally, concerning marginal effects of provinces, we note that people from the Sindh and Balochistan provinces have 0.056 and 0.0785 points lower probability while individuals from Khyber Pakhtunkhwa province have 0.023 points higher probability for getting a salaried job respectively with reference to the Punjab province. From Table 4.8, although marginal effects of age, younger children and household size look small in magnitude but these are shown to be statistically significant in Table 4.7.

4.5.6 Sample selection model: Step-2: Estimation of wage regression

This section presents the results from Step-2 of the sample selection model (as given in Eq. 2.6). For the Pakistani data, this Step-2 is based on explanatory variables given in Table 4.1 and by adding the IMR calculated from the Step-1 probit coefficients given in Section 4.5.5

(Table 4.7) as an additional regressor. We name this second step of sample selection model as **Model 4.4**[14] whose results are given in Table 4.9 below.

Table 4.9: Heckman (1979) Sample Selection Model for the Pakistani Data

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.4007	0.4003	934.15	<0.0001	19574
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	7.8843	2.87E-02	274.94	<0.0001
SCH1	0.0419	9.89E-04	42.33	<0.0001
EXP2	0.0300	1.09E-03	27.56	<0.0001
EXP22	-0.0003	2.24E-05	-15.60	<0.0001
HOURS3	0.0018	3.92E-04	4.58	<0.0001
DGENDER4	0.0578	1.41E-02	4.11	<0.0001
DWORKLOC5	0.0164	8.63E-03	1.89	0.0581
DTYPDIP6	0.3904	2.51E-02	15.56	<0.0001
DPUBLIC7	-0.0813	5.44E-03	-14.95	<0.0001
DFIXCT8	0.0099	1.29E-02	0.76	0.4447
DPERCT9	0.2310	9.54E-03	24.20	<0.0001
DSINDH10	-0.0238	8.47E-03	-2.81	0.0049
DKPK11	-0.0552	9.65E-03	-5.72	<0.0001
DBALO12	0.0429	1.10E-02	3.90	<0.0001
IMR	0.2702	2.38E-02	11.37	<0.0001

Table 4.9 presents the estimation results from the second step of Heckman's (1979) sample selection model. IMR computed by the relation given in Section 2.2.3 (Eq. 2.5) based on the probit estimates of selection equation (given in Table 4.7) is added as an additional regressor in the wage regression. This Model 4.4 corrects for sample selection bias, however, education is treated as exogenous in this model in order to see the effect of selectivity bias separately. From the results in Table 4.9, we see that Model 4.4 seems to be over all significant with predictive ability quite similar to that found in the OLS specification (Model 4.1) which ignores the possibility of sample selection bias. The coefficient related to schooling drops

[14] see appendix B2 for algebraic form of Model 4.4

down from 4.79% to 4.19% after the correction for sample selection bias. The impact of experience remains almost similar to that found in uncorrected (OLS, Model 4.1) estimation. The concavity also remains similar from both corrected and uncorrected estimations. The impact of hours worked on wages is found pretty similar in uncorrected OLS Model 4.1 (0.19%) and in sample selection corrected Model 4.4 (0.18%). The raw gender wage differential changes substantially due to addition of selectivity correction term. The gender wage premium in favour of men drops down to almost half from 10.44% (Model 4.1) to 5.78% (Model 4.4). So we can say that the half of the raw gender gap produced in OLS specification was only due to the presence of sample selection bias. This may question the findings from studies working on gender wage differentials that do not address the problem of non randomness of the waged earner's sample. Another difference that is noted due to correcting for possible sample selection bias is related to the reduction in differential effect of urban over rural work location on wages. The wage loss due to working in a rural area shrinks to 1.6% in above Model 4.4 from nearly 7% in uncorrected estimation (OLS, Model 4.1). This effect in sample selection model (Model 4.4) is just near the significance threshold while it was highly significant in the OLS model. So similar to gender gaps, sample selection bias affects urban rural wage differences severely as well. In the above sample selection Model 4.4, the wage differential between professional and general educational degrees also reduced to 39% premium for professional degree holders while it was about 47% when sample selection bias ignored (Model 4.1). The wage loss linked with working in public sector establishments remain almost identical from both sample selectivity corrected and uncorrected estimations. The non-significance of the wage premiums of fixed term contract holders over temporary workers is also evident as it was in the OLS specification. Similarly wage gains for permanent contract workers over temporary workers (reference category) found with an identical magnitude of 23% that was found in Model 4.1 (OLS or uncorrected specification). But these results concerning contract statuses are substantially different from those found in Model 4.3 when only endogeneity was corrected and sample selectivity was ignored. The negative effect of being from the Sindh province reduces to 2.3% (from 4.3% in Model 4.1) and Khyber Pakhtunkhwa province increased to 5.5% (from 4.7% in Model 4.1) compared to wages from Punjab province. The wages for people working in Balochistan province which were found similar to that in Punjab province in the OLS estimation, now turn out to be significantly different from Punjab province. The coefficients from sample selection Model 4.4 suggest that workers working in the Balochistan province enjoy approximately

4.3%, 6.7%, and 9.8% higher wages compared to the Punjab, Sindh and Khyber Pakhtunkhwa provinces, respectively. The coefficient related to IMR (selectivity correction term) found to be positive and significant. The positive significant coefficient hints at the fact that people who are working in salaried jobs in the Pakistani labour market currently have higher wages compared to the expected wages of non-participants if they had decided to work. This finding is similar to that we found for the French case. The previous evidence from the Pakistani studies is somewhat mixed. Some reported sample selection bias to be positive and significant (Abbas & Foreman-Peck, 2007), some reported significantly negative (Qureshi, 2012; Aslam & Kingdon, 2009 and Aslam, 2009) while there exist some researches that reported respectively positive and negative (Nasir, 1998; Hyder, 2007) but statistically non-significant coefficient associated to sample selection correction term.

4.5.7 Addressing endogeneity and sample selection bias, simultaneously for the Pakistani data

Like for the French data, for Pakistani data by comparing the results from models without any correction (Model 4.1) and the two models that correct for endogeneity (Model 4.3) and sample selectivity (Model 4.4), separately, we come to know that both these corrections put different effects (in magnitude as well as in direction) on the uncorrected estimates (from the OLS or Model 4.1) for most of the explanatory factors. For example the returns to schooling and experience increase when corrected for endogeneity of schooling (endogeneity correction refers to the IV2SLS model with Z2 as instrument, as we choose Z2 to be more appropriate instrument in Section 4.5.4 i.e. Model 4.3) while these effects tend to decrease when we addressed the bias due to sample selection (i.e. Model 4.4). Similarly wage differential due to gender goes up to 14% (from 10.43% in the OLS) in endogeneity corrected specification but drops down to 5.78% in the specification than accounts for only sample selection bias and wage premium for private sector workers reduces to 2.3% in endogeneity corrected Model 4.3 while remains virtually unaltered in sample selection Model 4.4 compared to nearly 8% from the uncorrected OLS (Model 4.1) specification. Moreover, the wage premiums for technical or professional degree holders drop by considerably different amounts in the models that tackled endogeneity (Model 4.3) and sample selection (Model 4.4) biases separately, compared to the OLS Model 4.1 which ignores both these issues. The wage gains for professional degree holders drop by 8 percentage points in sample selection Model 4.4 while

it decreases by approximately 31 percentage points in endogeneity corrected Model 4.3 when both these models compared to the simple OLS (Model 4.1) specification. The results concerning effect of contract statuses remain unchanged with sample selection correction but change markedly in the specification that takes endogeneity of schooling into account. These effects changed so much from the OLS that the OLS speciation produced results that showed that permanent contract holders enjoy a wage premium over temporary workers while wage gaps between fixed term contract holders and temporary workers reported to be non-significant. But endogeneity corrected Model 4.3 produced results that are totally reversed in this regard. Specifically, endogeneity corrected Model 4.3 showed that there exist no difference in wages of temporary workers and workers that work under permanent contracts while fixed term contracts holders face significant wage penalty against their counterparts who work temporarily. Similarly, when we tackled these biases separately, different changes have been noted in the differences in wages due to working in different provinces.

The above discussion accentuates the need for the correction of both these biases at the same time for the Pakistani data too. Correcting both biases in single model will enable us to produce more accurate, reliable and more representative effects that different explanatory variables put on the wage determination process in waged or salaried sector of the Pakistani labour market.

So we estimated a model that tackled both these biases at the same time. We estimated such model in similar fashion as we did for the French data (see Section 3.6.7), that is by estimating the model given by Eq. 3.5 using set of explanatory variables given in Table 4.1 for the Pakistani labour market. We name this model as **Model 4.5**[15] for convenience and further reference. The results from Model 4.5 that incorporated both kinds of possible biases are given in Table 4.10 below.

[15] see appendix B2 for algebraic form of Model 4.5

Table 4.10: Addressing Endogeneity and Sample Selection Biases Simultaneously for the Pakistani data (Model 4.5)

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.32145	0.32096	661.83	<0.0001	19574
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	6.9690	7.78E-02	89.62	<0.0001
SCH1_Z2	0.0862	5.46E-03	15.79	<0.0001
EXP2	0.0407	1.65E-03	24.62	<0.0001
EXP22	-0.0004	2.60E-05	-13.99	<0.0001
HOURS3	0.0047	5.32E-04	8.89	<0.0001
DGENDER4	0.0059	1.59E-02	0.37	0.7106
DWORKLOC5	-0.1338	1.04E-02	-12.81	<0.0001
DTYPDIP6	-0.0621	3.99E-02	-1.56	0.1192
DPUBLIC7	-0.0155	8.32E-03	-1.86	0.0627
DFIXCT8	-0.0759	1.64E-02	-4.62	<0.0001
DPERCT9	-0.0133	2.59E-02	-0.51	0.6076
DSINDH10	-0.0218	9.98E-03	-2.19	0.0286
DKPK11	-0.0673	1.08E-02	-6.23	<0.0001
DBALO12	0.1435	1.28E-02	11.19	<0.0001
IMR	0.7762	2.25E-02	34.48	<0.0001

In Table 4.10, we present estimation results for the Pakistani data from Model 4.5 that combines both sample selection correction and correction for endogeneity of schooling. From the results of Model 4.5, we see that the overall significance of the model in terms of explained variation is lower as compared to the OLS Model 4.1 and sample selection corrected Model 4.4 but a little higher than Model 4.3 which is an endogeneity correcting specification. After addressing both biases concurrently, we see that each additional year of education increases wages of workers by approximately 8.6% in Pakistan's labour market. The schooling coefficient from this model is decreased compared to endogeneity corrected Model 4.3 (10.43%) and increased compared to the sample selectivity corrected Model 4.4 (4.19%). Schooling coefficient in Model 4.5 is also higher than that found in the simple OLS Model 4.1 (4.79%) that ignores possibility of endogeneity and sample selectivity. The experience wage relationship shows that individuals get 4.07% gains in their wages with an extra year that they pass in labour market as wage worker. The evident concavity of wage

experience relationship means that this rate of increase in wages due to experience decreases as they get more and more experience. The impact of hours worked for reported salary found significant in this Model 4.5 as well. However its magnitude is closer to that found in the endogeneity corrected Model 4.3 which means that effect of sample selection bias is smaller on this coefficient. From above results an extra hour of work increases wages by 0.47%. The most substantial and noticeable difference that we note from above Model 4.5 is regarding gender wage gaps. The above model which is possibly free (or at least with minimum magnitudes) of both major sources of biases, suggests that there are no significant differences between wages of females and males in the Pakistani labour market's waged or salaried sector. This is different from previous 3 specifications (Model 4.1, Model 4.3, and Model 4.4) which reported significant raw gender wage differentials in favour of men. So this outcome questions the finding of studies that report significant gender wage differentials without addressing possibility of biases arising due to endogeneity of schooling and non random nature of the salaried worker's sample or findings of those studies that address one of these problems. Another difference that comes out form Model 4.5 is in urban-rural wage differential. This differential turns its direction in favour of workers in rural areas with 13.4% higher wages for rural workers. This is contrasting with the expectations and existing evidence from literature but it is explainable to some extent as people working in rural areas may get some remote area allowance etc. The wage premium reported as significant in previous models in favour of professional or vocational degree holders over general degree holders turns out to be non-significant in the above Model 4.5. This result also emphasizes the need to correct for both biases in same specification. Similar to dummies of gender and professional degree, the coefficient related to public-private wage differential against public sector workers considerably changed from uncorrected estimates. The coefficient related to public-private wage differences shows 1.5% wage gain for private sector workers, and this coefficient is not significant with 5% significance level. This finding is more close to the endogeneity corrected Model 4.3 than uncorrected Model 4.1 or Model 4.4 that corrects for sample selection bias. Results from Model 4.5 concerning contracts under which individuals work in labour market, are also more close to results found from endogeneity corrected Model 4.3 for these measures. Similar to Model 4.3, Model 4.5 also reveals that the workers having fixed term contracts have 7.5% lesser wages compared to those working on temporary basis while there are no significant gaps between wages of temporary workers and those who work under permanent contracts. Model 4.5 produces findings regarding provincial effects on

wages which are in more similarity with such measure obtained in Model 4.4. People working in the Sindh and Khyber Pakhtunkhwa provinces get respectively 2% and 6.7% lesser wages compared to those who work in Punjab province. The individuals working in the Balochistan province get 14% higher salaries compared to wages of workers in the Punjab province while this positive effect was found lesser in other specifications. Selectivity correction term found significantly positive in presence of endogeneity correction as well, in fact with relatively higher coefficient compared to that in Model 4.4 when only sample selection bias was addressed.

4.5.7.1 Testing heteroscedasticity of errors from simultaneous model

As we already noted in Section 3.6.8, that from review of literature, we found only few studies that tested the homoscedasticity assumption of the linear model in the context of the Mincerian literature. This scarcity holds for Pakistan as well because we found no study that formally tested the validity of vital assumption of homoscedasticity of error terms. However, Hyder (2007) is the only study that used heteroscedasticity robust standard errors for testing. The homoscedasticity assumption is a must to hold for the inference from linear model to be valid and reliable. So just like we did for the French analysis (Section 3.6.8.1), for the Pakistani analysis too, we applied White's (1980) test to see any possible violation of this assumption regarding equality of variances of error terms computed from Model 4.5 that addresses both vital issues of endogeneity and sample selection bias. The outcome from White's (1980) test of heteroscedasticity for Pakistani data (Model 4.5) is presented in Table 4.11 below.

Table 4.11: Test for Heteroscedasticity on Error from Model 4.5

Statistic	DF	P-value	Variables
912.15	90	<0.0001	Cross of all variables

From Table 4.11, it is clear that the homoscedasticity assumption about error variances is violated. To eliminate or curtail the adverse effects of heteroscedasticity of errors on the

estimation process, we proceed in a similar way that we used for the French data analysis i.e. we estimate the Model 4.5 by adaptive estimation.

4.5.8 Adaptive estimation of simultaneous model (Pakistan)

Due to presence of heteroscedasticity in errors computed from Model 4.5, we go for the adaptive estimation of the Model 4.5. The procedure applied for the adaptive estimation is same as we applied for adaptive estimation of the French model. This procedure is explained in the Section 3.6.9. We directly present results from the adaptive estimation of Model 4.5 (we call it as **Model 4.5W**[16] for differentiation). These results are presented in Table 4.12.

Table 4.12: Model 4.5W (Adaptive Estimation of Model 4.5)

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.3029	0.3024	607.00	<0.0001	19574
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	6.9553	8.03E-02	86.67	<0.0001
SCH1_Z2	0.0867	5.60E-03	15.49	<0.0001
EXP2	0.0413	1.72E-03	23.97	<0.0001
EXP22	-0.0004	2.50E-05	-14.26	<0.0001
HOURS3	0.0049	5.31E-04	9.28	<0.0001
DGENDER4	0.0190	1.59E-02	1.20	0.2320
DWORKLOC5	-0.1282	1.03E-02	-12.50	<0.0001
DTYPDIP6	-0.0433	4.20E-02	-1.03	0.3022
DPUBLIC7	-0.0178	8.03E-03	-2.21	0.0268
DFIXCT8	-0.0749	1.63E-02	-4.61	<0.0001
DPERCT9	-0.0060	2.56E-02	-0.24	0.8140
DSINDH10	-0.0275	9.97E-03	-2.76	0.0058
DKPK11	-0.0658	1.07E-02	-6.16	<0.0001
DBALO12	0.1449	1.27E-02	11.38	<0.0001
IMR	0.7449	2.26E-02	32.99	<0.0001

[16] see appendix B2 for algebraic form of Model 4.5W

The results from of Model 4.5W show that global performance of model is nearly similar to that found in its simple estimation (Model 4.5).

From results of the adaptive version of the simultaneous model (Model 4.5W), we note that impact of human capital variables on earning changes just a little i.e. from 8.62% to 8.67% per year for education and 4.07% to 4.13% for every year of potential labour market experience. Wage gains associated to each additional hour of work is found to be 0.49% in above Model 4.5W (compared to 0.47% in Model 4.5). Similar to its simple estimation, the wage differential due to gender and type of diploma are found non-significant in adaptive estimation as well. Comparing coefficients for other variables estimated from simple estimation (Model 4.5) and the adaptive estimation (Model 4.5W) of simultaneous model (Table 4.10 and Table 4.12, respectively), we come to know that both of these estimations give the results that are virtually almost indistinguishable. The impacts related to different wage determinants are almost similar in terms of magnitude as well as in direction. So we can say that we did not get any gains from the adaptive estimation of the model in terms of changes in coefficients associated with different explanatory factors. Due to this strong similarity between results from these two models, we do not interpret results from Model 4.5W in details. However, in addition to sample selection and endogeneity biases, Model 4.5W tackles issue of heteroscedasticity as well whose presence is found evident by White's (1980) test (see Table 4.11). Therefore, despite of quite similar results from the simple and adaptive estimations of the model, we consider adaptive version (Model 4.5W) to be more robust among the two.

4.5.9 Semi-parametric estimation of the Mincerian model (Pakistan)

For the Pakistani labour market data, so far we have estimated the Mincerian model parametrically. For the French data, we have explored the plausibility of semi-parametric estimation of the Mincerian model. As it is already explained in Section 3.6.10 (and Section 3.6.10.1) that our semi-parametric estimation consists of the IV2SLS estimation with the first stage schooling equation estimated non-parametrically. While the correction for sample selection is done by adding IMR which is computed from parametric estimates of selection equation (Step-1 of sample selection model, Table 4.7). On the similar lines, as we did for French case, we have estimated the Mincerian model semi-parametrically for the Pakistani data.

4.5.9.1 Non-parametric estimation of first stage schooling equation

For the Non-parametric estimation of first stage schooling equation, we faced similar problem (as we faced in the case of French data, Section 3.6.10.1) that the LOESS regression is not recommended with more than 4 explanatory variables. Due to similar constraints we applied similar strategy in order to estimate first stage schooling equation non-parametrically by the LOESS regression. So we computed principal components based on all exogenous variables in first stage schooling equation, and then LOESS regression is applied on Instrument-2 (Z2) and the number of principal components that explain major portion of variation.

The results from the principal component analysis based on variables **EXP2, EXP22, HOURS3, DGENDER4, DWORKLOC5, DTYPDIP6, DPUBLIC7, DFIXCT8, DPERCT9, DSINDH10, DKPK11, DBALO12, IMR,** are shown in Table 4.13.

Table 4.13: Principal Component Analysis on all Exogenous Variables

Principal Component	Eigen value	Difference	Proportion	Cumulative
1	324589.268	324501.623	0.9997	0.9997
2	87.645	77.739	0.0003	1.0000
3	9.906	9.344	0.0000	1.0000
4	0.562	0.311	0.0000	1.0000
5	0.250	0.048	0.0000	1.0000
6	0.203	0.044	0.0000	1.0000
7	0.159	0.008	0.0000	1.0000
8	0.151	0.076	0.0000	1.0000
9	0.076	0.012	0.0000	1.0000
10	0.064	0.005	0.0000	1.0000
11	0.059	0.028	0.0000	1.0000
12	0.031	0.016	0.0000	1.0000
13	0.015		0.0000	1.0000

From Table 4.13, we see that first 2 principal components explain 100% variability in the set of schooling equation explanatory variables.

Hence our non-parametric estimation of the schooling equation will be the LOESS regression estimation based on first 2 principal components and Instrument-2 (Z2). So for the Pakistani data Eq. 3.7 (for the LOESS regression) becomes,

$$SCH1_i = f\left(PC1_i, PC2_i, Z2_i\right) + \nu_i, \qquad (4.3)$$

4.5.9.2 Semi-parametric estimation of the wage regression

Now in our semi-parametric model the fitted values for schooling (SCH1_NP) by the LOESS regression will replace the endogenous schooling (SCH1) in the second stage estimation of wage equation as explained in Section 3.6.10. For the Pakistani data, the LOESS method applied in SAS 9.3 has produced a data driven value of smoothing parameter as 0.02567. Then we estimated wage model semi-parametrically as given in Eq. 3.10 using explanatory variables relevant to Pakistani labour marker given Table 4.1. We name this semi-parametric model as **Model 4.6**[17] and the results of this semi-parametric estimation of the Mincerian wage regression are presented in Table 4.14 given below.

[17] see appendix B2 for algebraic form of Model 4.6

Table 4.14: Semi-Parametric Estimation of the Pakistani Wage Model (Model 4.6)

Global Measures of Model				
R-Square	**Adj R-Sq**	**F-value**	**P-value**	**N**
0.2834	0.2829	552.50	<0.0001	19574
Parameter Estimates				
Variable	**Parameter Estimate**	**Standard Error**	**t-value**	**P-value**
INTERCEPT	8.0321	2.96E-02	271.581	<0.0001
SCH1_NP	0.0087	1.02E-03	8.541	<0.0001
EXP2	0.0227	1.12E-03	20.215	<0.0001
EXP22	-0.0003	2.30E-05	-11.922	<0.0001
HOURS3	-0.0002	4.04E-04	-0.450	0.65272
DGENDER4	-0.0527	1.45E-02	-3.630	0.0003
DWORKLOC5	-0.0613	8.90E-03	-6.880	<0.0001
DTYPDIP6	0.3852	2.59E-02	14.888	<0.0001
DPUBLIC7	-0.1060	5.61E-03	-18.903	<0.0001
DFIXCT8	0.0519	1.33E-02	3.903	0.0001
DPERCT9	0.3623	9.84E-03	36.807	<0.0001
DSINDH10	0.0247	8.73E-03	2.823	0.0048
DKPK11	-0.0772	9.96E-03	-7.754	<0.0001
DBALO12	0.0727	1.13E-02	6.412	<0.0001
IMR	0.8108	2.45E-02	33.071	<0.0001

The results from semi-parametric Model 4.6 show that our semi-parametric model is although globally significant but have lesser explanatory power as compared to the parametric model (Model 4.5). From Model 4.6, the coefficients related to different independent variables are quite different and surprising in some of the cases. First, the coefficient associated to schooling shows that each additional year of schooling increases wages by just less than 1% which is contrary to results from not only our all previous models but also hundreds of studies all over the globe. The impact of experience is also less than all other specifications although it is in bounds that found generally in international literature. From semi-parametric model, an extra year of experience enhances wages by an amount of 2.26% in the beginning and this return decreases with increasing years of experience due to concavity of the wage experience relationship. The impact of hours worked on wages becomes extremely less and statistically insignificant which was found to be significant in all the previous models (Model 4.1 through Model 4.5W). Another major difference that comes out with this semi-parametric Model 4.6

is wage differential in favour of women. This finding is against the results of all previous models, conclusion from many other Pakistani studies and also against the general consensus in global literature concerning gender wage differentials that reports raw gender wage differential in favour of men in heavy majority of the studies. The impact of work location is in favour of rural workers as it was in simple and adaptive estimations of parametric model. But the premium for rural workers is less than that found in parametric model (6.1% compared to 13.3% and 12.8% in the simple and adaptive estimation of parametric model, respectively). Coefficient linked with dummy variable for professional or technical degree is closer to that found in Model 4.4 (that only corrected for sample selection bias and not endogeneity) that suggested a 39% wage gains due to possessing a professional or technical degree. The wage penalty for working in public sector is negative as in all other specifications estimated but once again size of that wage penalty produced by semi-parametric Model 4.6 is closer to that found in the sample selection Model 4.4 and OLS Model 4.1 as compared to its closeness to coefficients from other models estimated for the Pakistani data in the present work. Regarding contract statuses, the semi-parametric Model 4.6 indicates that temporary workers get significantly lesser wages as compared to those who work under a fixed term or permanent contract. The results demonstrate a 5.19% and 36.2% wages gains for fixed term contract workers and permanent contract workers respectively, over those who work temporarily. Concerning provincial effects, the direction of differences in wages of workers in the Balochistan and Khyber Pakhtunkhwa provinces compared to Punjab remains same as found in parametric model. However, the wage premium for those who are working in the Balochistan province is found less than that found in parametric model that correct for both biases (Model 4.5) and greater than that found in the parametric specifications that address sample selectivity (Model 4.4) or endogeneity of education (Model 4.3), separately. The above model gives different results concerning comparison between wages of workers in the Punjab and Sindh provinces. Contrary to previous models, semi parametric model shows that workers in the Singh province have 2.47% higher wages relative to their counterparts in the Punjab province. Finally, we see that selectivity correction term is significant and positive in this semi-parametric Model 4.6 as well which is consistent with the parametric Model 4.5.

4.5.10 Comparison of parametric and semi-parametric models

For a comparison among the parametric and semi-parametric models and simple and adaptive estimation of parametric model, Table 4.15 presents the coefficients estimated from the parametric model (simple Model 4.5 and adaptive Model 4.5W) and from the semi-parametric Model 4.6. We do not present results from the adaptive estimation of the semi-parametric model as it performed worst than all the other models in terms of global significance and also as we noted in previous section that semi-parametric model produces results which are quite unreasonable and against the general consensus in Pakistani as well as in international literature concerning schooling coefficient.

Table 4.15: Comparison of Parametric and Semi-Parametric Models Estimated for the Pakistani Data

Variable	Model 4.5 Parametric (Simple)	Model 4.5W Parametric (Adaptive)	Model 4.6 Semi-parametric (Simple)
INTERCEPT	6.9690	6.9553	8.0321
SCH1	0.0862	0.0867	0.0087
EXP2	0.0407	0.0413	0.0227
EXP22	-0.0004	-0.0004	-0.0003
HOURS3	0.0047	0.0049	-0.0002
DGENDER4	0.0059	0.0190	-0.0527
DWORKLOC5	-0.1338	-0.1282	-0.0613
DTYPDIP6	-0.0621	-0.0433	0.3852
DPUBLIC7	-0.0155	-0.0178	-0.1060
DFIXCT8	-0.0759	-0.0749	0.0519
DPERCT9	-0.0133	-0.0060	0.3623
DSINDH10	-0.0218	-0.0275	0.0247
DKPK11	-0.0673	-0.0658	-0.0772
DBALO12	0.1435	0.1449	0.0727
IMR	0.7762	0.7449	0.8108
R-SQUARE	**0.3215**	**0.3029**	**0.2834**

4.5.10.1 Choice between models estimated for the Pakistani data

From the Table 4.15 that compares results from the simple and adaptive version of the parametric model (Model 4.5 and Model 4.5W, respectively) and simple estimation of semi-

parametric model (Model 4.6) for the Pakistani labour market data. We see that global performance of simple estimation of parametric model (Model 4.5) as given by R^2 is a little higher compared to that found in adaptive estimation of parametric model (Model 4.5W) or semi-parametric estimation (Model 4.6). However, as we already noted that coefficients from semi-parametric Model 4.6 for some of the factors seem unreasonable and inconsistent with general consensus in literature. So on the basis of some unreasonable coefficient estimates and better performance in terms of R^2 we prefer parametric model. We prefer its adaptive version i.e. Model 4.5W for the Pakistani data as this model takes issues of endogenous schooling, sample selection bias, and heterogeneity of error terms simultaneously. Moreover, as it uses IV2SLS estimation for correcting endogeneity of schooling, so it accounts for measurement error as well.

4.5.11 Interval estimation of selected model for the Pakistan data

In this section we give the results from interval estimation of the most appropriate model that we selected for estimation of the Mincerian wage model for the Pakistani data (i.e. Model 4.5W). These results are given in the following Table 4.16 which details the lower and upper limits for the impacts that different explanatory factors put on the wages in the Pakistani labour market. These confidence intervals are computed by the relations given in Chapter 3 (Section 3.6.13; Eq. 3.11a and 3.11b) by using the results from most appropriate model chosen for the Pakistani data.

Table 4.16: Interval Estimation of Model 4.5W selected for Pakistan

Variable	Parameter Estimate	Standard Error	$\beta_{Lower\ Limit}$	$\beta_{Upper\ Limit}$
INTERCEPT	6.9553	8.03E-02	6.7980	7.1126
SCH1_Z2	0.0867	5.60E-03	0.0757	0.0976
EXP2	0.0413	1.72E-03	0.0379	0.0447
EXP22	-0.0004	2.50E-05	-0.0004	-0.0003
HOURS3	0.0049	5.31E-04	0.0039	0.0060
DGENDER4	0.0190	1.59E-02	-0.0122	0.0503
DWORKLOC5	-0.1282	1.03E-02	-0.1483	-0.1081
DTYPDIP6	-0.0433	4.20E-02	-0.1256	0.0390
DPUBLIC7	-0.0178	8.03E-03	-0.0335	-0.0020
DFIXCT8	-0.0749	1.63E-02	-0.1068	-0.0431
DPERCT9	-0.0060	2.56E-02	-0.0561	0.0441
DSINDH10	-0.0275	9.97E-03	-0.0471	-0.0080
DKPK11	-0.0658	1.07E-02	-0.0867	-0.0448
DBALO12	0.1449	1.27E-02	0.1200	0.1699
IMR	0.7449	2.26E-02	0.7007	0.7892

From the interval estimation results of the Model 4.5W selected for the Pakistani data, we note that schooling coefficient ranges between7.57% and 9.76%. So based on the 95% confidence level, we may conclude that each additional year of education enhances wages of individual at least by 7.57% and at most by 9.76%. The returns to experience lie between 3.79% and 4.47%. The range of impact of hours worked is found to be 0.39% to 0.60% meaning that every additional hour devoted to work causes an increase of this range in wages of individuals. The impact of work location is significant ranging from 10.81% to 14.83% in favour of workers working in rural areas. The wage gains that private sector workers have over those who work in public sector jobs varies from 0.2% to 3.35% for Pakistan's labour market. The wage difference between temporary workers and worker working under fixed term contract fluctuate from 4.31% to 10.68 against the fixed term contract workers. However such a difference is found as non significant between permanent contract workers and temporary workers. The gender wage gap and wage differential due to type of diploma are also found as non-significant. Concerning the ranges in differences in the wages between workers from different provinces, we see that wage losses for workers in the Sindh and Khyber Pakhtunkhwa provinces compared to Punjab province (reference category) oscillate

from 0.8% to 4.71% and from 4.48% to 8.67% for the both provinces, respectively. It shows that wage loss for people of Khyber Pakhtunkhwa compared to workers from Punjab is more compared to the similar loss for workers from the Sindh province as lower limit for Sindh and upper limit for Khyber Pakhtunkhwa are close to each other. The people of Balochistan get higher wages compared to all other provinces. For example they enjoy a wage premium that varies from 12% to 17% over worker in the Punjab province. Similarly workers in Balochistan get about 16.19% to 16.71 higher wages than workers in the Sindh province and about 16.48% to 20.67% than workers in Khyber Pakhtunkhwa province. A similar wage gap ranges from 3.7% to about 4% between workers form the Sindh and Khyber Pakhtunkhwa provinces in favour workers from the Sindh province.

Chapter 5: The Mincerian Model as a Quantile Regression Model

5.1 The Mincerian Wage Model in the Context of Quantile Regression

An OLS regression is based on the mean of the conditional distribution of the regression's dependent variable. This approach is used because one implicitly assumes that possible differences in the impacts of the independent variables along the conditional distribution are unimportant. However, this may prove inadequate in some research agendas. If explanatory variables influence parameters of the conditional distribution of the response variable other than the mean, then an analysis that disregards this possibility will be severely weakened (see Koenker and Bassett, 1978). In such a situation, the solution lies in the use of quantile regression. Unlike OLS, quantile regression models allow for a full characterization of the conditional distribution of the dependent variable. Quantile regression as introduced may be viewed as a natural extension of classical least squares estimation of conditional mean models to the estimation of an ensemble of models for conditional quantile functions. With help of quantile regression approach, we explore the possibility that whether different explanatory factors affect uniformly over the conditional distribution of the response variable or they exert different effects on response variable in different quantiles of the conditional distribution of response variable. Quantile regression captures heterogeneity in effects of different factors that mean regression would not have captured. One advantage of quantile regression is its lesser sensitivity to outliers (Koenker & Bassett, 1978; Mwabu & Schultz, 1996; Falaris, 2004) and it may perform better than OLS in case when the homoscedasticity assumption is violated (Deaton, 1997; Lee & Lee, 2006). Moreover, quantile regression is also less affected by the deviations of error term from normality (Buchinsky, 1998a).

To briefly recall the ordinary quantile, consider a real valued random variable Y characterized by the following distribution function

$$F(y) = \text{Prob}\ (Y \leq y),$$

the τth quantile of Y is defined as the inverse function

$$Q\ (\tau) = \inf\ \{y: f(y) \geq \tau\},$$

where $0 < \tau < 1$. In particular, the median is $Q(1/2)$.

The τth sample quantile $\hat{\xi}(\tau)$, which is an analogue of $Q(\tau)$, may be formulated as the solution of the optimization problem

$$\min_{\xi \in R} \sum_{i=1}^{n} \rho_{\tau}(y_i - \xi),$$

where $\rho_{\tau}(z) = z(\tau - I(z < 0)), 0 < \tau < 1,$ is usually called the check function.

When covariates X are considered, the linear conditional quantile function, $Q(\tau \mid X = x) = x'\beta(\tau)$, can be estimated by solving

$$\hat{\beta}(\tau) = \arg\min \sum_{i=1}^{n} \rho_{\tau}(y_i - x'_i \beta), \qquad (5.1)$$

for any $\tau \in (0,1)$. The quantity $\hat{\beta}(\tau)$ is called the regression quantile. The case $\tau = 1/2$, which minimizes the sum of absolute residual, is usually known as median regression. Thus, a least squares of the mean regression model would be considered with the dependence of the conditional mean of Y on the covariates X. While on the other hand, the quantile regression estimator tackles this issue at each quantile of the conditional distribution, providing thus a more complete description of how the conditional distribution of Y given $X = x$ depends on x.

For more details about quantile regression, see Fitzenberger (1997), Koenker & Hallock, (2001), Buhai (2004), Martins & Pereira (2004) and Chen & Wei (2005).

In our case, for running the quantile regression according to (5.1), we set the dependent variable $y_i = \ln W$, ith row vector x'_i includes intercept and explanatory variables relevant for each country i.e. France and Pakistan.

5.2 A Short Review of the Mincerian Wage Model under Framework of Quantile Regression

Like its estimation in mean regression setting, the Mincerian wage model has also been estimated in the quartile regression structure in many research studies covering many countries and regions. Such studies mainly focused on exploring the behaviour of wage gaps due to different reasons (race, gender, public/private sector, ethnic belonging) in different parts of the conditional wage distribution.

One advantage of the quantile regression with particular reference to the Mincerian model is that, with this applied, we can also control for some unobservable factors due to which individuals have been located in particular quantile of the wage distribution (Staneva et al., 2010). Here, we briefly review such studies that applied quantile regression for the Mincerian wage model.

Buchinsky (2001) studied the changes in the returns to education for the US female workers over the period of 1967—1990. For this, they applied the Mincerian model with quantile regression estimation. They controlled for 9 different age groups for each of the 5 data sets they used for the specified period. For mean regression they corrected for sample selectivity using the Heckman (1979) two step method and semi-parametric correction as well. While for the quantile estimation, they corrected for possible bias due to non-randomness of the sample with Buchinsky (1998b) approach using Ichimura (1993) semi-parametric estimator for the participation equation. Results from this study showed that returns to education differ significantly over time and across age groups and these changes are not uniform over the quantiles of the conditional wage distribution of US female workers over the period considered.

Mwabu & Schultz (1996) used quantile regression for estimating the Mincerian relationship in order to compare the racial differences in returns to educational attainments between the White and African males, evaluated at different quantiles. The results from their investigation revealed different patterns of educational returns for the White and African males in different segments of the wage distributions. For example, for higher education, returns to education for African people decrease along with deciles of the wage distribution while this pattern is

reverse in case of the White people. However, their results may be subject to sample selection bias as they took only waged people and did not take any corrective measure for possibility of sample selectivity. Girma & Kedir (2003) estimated human capital earnings function by applying the quantile regression in order to see the heterogeneity of schooling coefficient at different points of the wage distribution in Ethiopia. They also addressed the endogeneity issue using parent's schooling as instruments. They reported that returns to schooling in Ethiopia decrease along with higher quantiles of wages as they found coefficient associated to schooling as 28% at 10^{th} quantile while only 9% for 90^{th} quantile. Moreover, the wage penalty for being female is not same at all points of wage distribution as it increases initially but decreases in upper half of the wage distribution. This study serves as another evidence for higher returns to schooling via the IV2SLS method than OLS one. Coelho et al. (2008) estimated returns to education by using the sample selection corrected quantile regression for married women data coming from Brazil. For the sample selection correction they used Buchinsky (1998b) approach which is based on Newey (1991) using Klein & Spady (1993) estimator for selection equation. They reported that returns to schooling were not uniform over the quantiles but higher for upper quantiles of the wage distribution. Their results showed that the White women enjoy wage premiums and such wage premiums are higher in upper tail of the wage distribution compared to lower tail. Comparing the parametric and semi-parametric correction for sample selectivity, they preferred the one from semi-parametric approach by using Horowitz & Härdle (1994) specification tests. However, from returns to schooling reported, we see that returns estimated with no correction and with semi-parametric correction as similar to each other but substantially different from those obtained via parametric correction for possible sample selection bias. Another study that applied quantile regression to assess the returns to education in 4 countries with transition economies is Staneva et al. (2010). They conducted separate analyses for Bulgaria, Russia, Kazakhstan, and Serbia. They corrected for endogeneity of schooling by Lee (2007) control function approach while sample selection by Newey (1991) based on the Buchinsky (1998b) approach. For Bulgaria and Kazakhstan, they found increasing returns to schooling in the higher quantiles of wage distribution while opposite trend for Russia where higher returns to schooling were found at lower ends of the wages. These results are robust for both of the genders. For Serbia, their estimation exposed that estimates at different points of the wage distribution were not considerably different from those obtained through the OLS approach and this fact is more evident for men than for women. Sample selectivity found to affect

significantly for Russian and Kazakh estimations and correction for endogenous schooling lead to higher returns to schooling. However, effect of endogeneity seemed to be stronger for female estimates. Patrinos et al. (2009) for different Asian and Latin American countries and Budría & Pereira, (2005) for some European countries, applied quantile regression on the Mincerian function and documented higher returns to education in higher quantiles of the relevant wage distributions.

For the French data Barnet-Verzat & Wolff (2008), found returns associated to different levels of education that increase with moving towards upper qauntiles. Also for the Pakistani data Hyder & Reilly (2005) applied the quantile regression and results showed that differences between public and private sector wages vary in different quantiles, and also explained portion of such gaps differ with quantiles. Similar is the case for higher educational levels and other determinants of wages they controlled for. They also reported that gender wage gaps decrease as we move up in the conditional wage distribution.

Similar to mean regression in the Mincerian context, some authors applied quantile regression to explore the public and private sector wage differentials at different points of the distribution of wages. Like Mueller (1998) and Poterba & Rueben (1994) worked on the public sector wage premiums against private sector workers over the entire wage distribution applying the quantile regression for the Canadian and US data, respectively. Estimation results for both studies revealed a general trend that wage differentials differ with the choice of quantile. Similar evidence seen for the West Germany for both men and women, that wage differential between public and private sector is different across the conditional wage distribution. Both raw wage gap and its discriminatory or unexplained part showed a decreasing trend with moving up along with wage distribution (Melly, 2005). Another work that estimated the Mincerian model in quantile regression setting is Lucifora & Meurs (2006) while working on the differences between public and private sector wages in France, Italy and Great Britain. Similar to many others, their results come with the finding that the OLS approach cannot capture the wage gaps that are representative for the entire distribution of wages. For all 3 countries they found that public-private wage differentials tend to decrease in the upper quantiles. Contrary to general trend in studies (Staneva et al., 2010; Buchinsky, 2001; Melly, 2005; Mueller, 1998 and many others) that evaluated relationship at 5 quantiles (10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th}), they evaluated at each decile of the wage distribution. For

United Kingdom, similar findings were also reported in Blackaby et al. (1999) and Disney & Gosling (1998). For the distribution of wages in Panama, Falaris (2004) compared the effects that different determinants exercise on the labour market wages at the selected quantiles. Opposite to common norm, the study included self employed people also in the analysis. They conclude with the finding that in Panama, working in public sector has more positive effect in lower end of the wage distribution and similar finding found for working in large size firms. Being a member of union, have positive effects that are higher in lower end of the distribution for men while women's wages remained unaffected by unionization. Working in the famous Canal Zone affects wages in a uniform way over the entire distribution. Returns to higher education and experience found higher in the upper quantiles of the distribution of wages for men. While for women returns to education and experience were statistically indifferent when evaluated at different quantiles. So we may say that as for as the human capital factors concerned, in case of female workers in Panama, quantile regression proved to be less extra informative over the regression at means. For France, Bargain & Melly (2007) evaluated the wage gaps between public and sector workers using mean as well as quantile regression. Their results indicated that men get higher wages in private sector while women get higher wages in public sector. Quantile regression pointed out that for both genders, wage gaps in favour or against public sector workers reduce by going along with higher points of the wage distribution.

There are many people who applied quantile regression for the Mincerian model to have an insight about the amount of gender wage gaps in different parts of the wage distribution. These include, Ajwad & Kurukulasuriya (2002) who estimated Mincer's model with both the OLS and quantile regression approaches to observe the gender and ethnic differences in wages in the labour market of Sri Lanka. Surprisingly, in Sri Lankan context, they found ethnicity to be non significant contributor of variation in wages and these results were valid for both mean as well as quantile regression estimations. The gender wage gap was significant which found to increase with going towards higher parts of the wage distribution and it is in accordance with a hypothesis of glass ceiling effect. Focusing on human capital factors, returns to most of the educational levels were higher for low paid workers while the converse is true for returns associated to the labour market experience. Penalty for people in rural areas found to be more severe in the higher quantiles. From their results we see an important fact captured by quantile regression but not by mean regression, that is, mean regression predicted

a penalty for being in public sector of work but quantile based regression revealed it as not to be true for the upper portion of the distribution where it turns to be a premium in fact. Using the Swedish data, Albrecht et al. (2003) applied the quantile regression based on the Mincerian model to expose the wage gaps between men and women at different levels of wage distribution. Limiting to the full time Swedish workers, they pointed out that the gender wage differences were more pronounced at the top of the wage distribution which is known as glass ceiling effect. However the unexplained portions of these gaps were lower in the upper tail suggesting more gender equality in high paying jobs. For both men and women returns to education rise with the rising of quantiles for all educational levels, while a U-shape relationship found between returns to experience and quantiles. Comparing the effects of being immigrant they concluded that penalty for being an immigrant is more prominent at the lower parts of the distribution. Their results were generally robust with different data sets used. García et al. (2001) also applied quantile regression to see the gender gaps across the wage distribution for the Spanish data. Stressing the need for availability of bias free estimates they corrected for possible biases due to endogeneity and sample selection in case of quantile regression as well. Their estimation and decomposition results showed that the wage gap between men and women vary across the quantiles and the gap is more prominent at higher quantiles. The gap which is not attributable to differences in endowments is also not same over the entire wage distribution which means that different explanatory factors affect the wages of individuals differently in different parts of the wage distribution. So in such cases quantile regression estimates can lead to clearer picture of the relationship between wage and its determinants compared to insights provided by the mean regression.

Like Albrecht et al. (2003) did for Sweden; the existance of glass ceiling effect was also investigated for the 11 European countries by Arulampalam et al. (2007). They also concluded that gender wage gaps were not same in the entire wage distribution. For most of the countries, these gaps were higher in the upper tail while for Spain and Austria these gaps were higher at lower tails of the relevant wage distributions. For France and Italy, they reported gender pay gaps to be higher at the both ends but lower in the middle of the wage distributions. Separate analysis for public and private sectors come out with the finding that wage gaps were present in almost all quantiles, for all countries except Netherland and Finland. After controlling for age, educational level, experience, health status and some other factors, their study gave a general conclusion that men get higher wages across the wage

distribution, across the countries included and across the both major sectors of employment. The quantile regression in the Mincerian setting is also used to discover the glass ceiling effect in some other countries. For Spain, the glass ceiling effect also found evident for the people with higher educational levels in De la Rica et al. (2008) who applied quantile regression for human capital wage model. They found higher wage gaps in higher quantiles for people with higher education but higher wage gaps in lower quantiles for low educated people. Now as clear from their sample composition, people were more concentrated in low educational groups where gaps are higher at bottom, so this provides a reason why Arulampalam et al. (2007) found higher gaps at lower ends of the Spanish wage distribution. The similar results that gender wage gaps and their unexplained portion is not same in different parts of the wages distribution, also found from Spain (Gardeazabal & Ugidos, 2005), for France (Jellal et. al, 2008) and for Switzerland (Bonjour & Gerfin, 2001). All these studies used the Mincerian approach via quantile regression controlling for different individual specific and firm specific measures.

From the above review of the literature concerning the Mincerian wage model in the framework of quantile regression, we see that effects of different determinants of wages are not same across the conditional wage distribution in almost all the cases coming throughout the globe. We see that a vast majority of the studies evaluated the relationship between wage and its determinants at 5 different points of the wage distribution (10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} quantiles) with few exceptions that evaluated the relationship at 7 different quantiles (Albrecht et al., 2003; Falaris, 2004; Jellal et al., 2008) and some evaluated at each of the deciles (for example, Albrecht et al., 2004; Lucifora & Meurs, 2006). We also noticed only a few studies that corrected for bias due to endogeneity of schooling (Girma & Kedir, 2003; Jellal et al., 2008; Staneva et al., 2010). Also from this review it is clear that endogeneity in quantile regression framework is tackled (where tackled) with similar approach as commonly used in mean regression, that is by the IV2SLS estimation. In such cases IV2SLS is applied as estimating first stage schooling equation by the mean regression while second stage wage equation is estimated with quantile regression (see García et al., 2001 for example). As far as the problem of sample selection bias is concerned, many authors tackled the sample selection bias problem in the quantile regression by applying Buchinsky (1998b) approach which is based on Newey (1991) series approximation using Ichimura (1993) estimator for selection equation (like Buchinsky, 2001 ; Albrecht et al., 2004; Nicodemo, 2009 ; Staneva et al., 2010)

but some others (like Coelho et al., 2008 ; Huber & Melly, 2011) used Klein and Spady (1993) estimator instead of Ichimura (1993) for estimation of the selection equation. The only study that corrected both kinds of biases in the same specification for quantile regression is García et al. (2001) that applied Buchinsky (1998b) approach but using parametric probit estimates of the selection equation used for power series approximation instead of using semi parametric (Ichimura, 1993 or Klein & Spady,1993 estimators) for the estimation of the selection equation. Some studies stated that traditional bias correction term cannot be used in quantile regression (Buchinsky, 1998b; Buchinsky, 2001). However we found such studies (Hyder & Reilly, 2005; Coelho et al., 2008) that did apply Heckman (1979) two-step method to eliminate the possibility of sample selection bias in quantile regression setting too so we may follow these studies.

5.3 Estimation of the Mincerian Quantile Regression Model for France

We have estimated the Mincerian wage regression as quantile regression model as explained in Section 5.1 (Eq. 5.1). We have preferred parametric model compared to semi-parametric model and adaptive estimation over simple estimation within the framework of parametric model based on estimation of the mean regression model for the French data (Chapter 3, Section 3.6.12.1 and Section 3.6.12.2). But as noted in previous Section 5.1 that quantile regression performs well in presence of heteroscedasticity (Deaton, 1997; Lee & Lee, 2006), so quantile regression is applied on the simple version of the parametric model which is preferred over semi-parametric model.

For the French case we recall parametric model (Model 3.5, Eq. 3.5) and for running the quantile regression according to (5.1), we set the dependent variable $y_i = \ln W_i$, *i*th row vector x'_i includes schooling variable $\hat{SCH}1_Z2$ fitted from first stage regression based on instrument Z2 (from first stage of Model 3.3 Table 3.5-A), variables included in X_K (Eq. 3.1), and *IMR* (based on probit coefficient in Table 3.7).

We followed García et al. (2001) and Ribeiro (1997) as they have sued $(\hat{SCH}1_Z2)_i$ as the fitted values from first stage schooling equation calculated in mean regression analysis in quantile regression too. Moreover, the vector β in (5.1) contains all the coefficients associated to all these variables included in X'_i. For sample selection correction we have followed Hyder & Reilly (2005) and Coelho et al. (2008) by using Heckman two-step method in quantile regression.

Finally, quantile regression estimates are found for the quantiles (τ = 0.05, 0.1, 0.2, 0.25, 0.50, 0.75, 0.80, 0.90 and 0.95) i.e., 5^{th} , 10^{th} , 20^{th} , 25^{th} , 50^{th} , 75^{th} , 80^{th} , 90^{th} , and 95^{th} percentiles of the French wage distribution.

The following Table 5.1 presents the results from quantile regression analysis of Mincer's specification for the French labour force data.

Table 5.1: Parameter Estimates of the Mincerian Wage Model at Different Quantiles of the French Wage Distribution

Parameter	Q_0.05	Q_0.10	Q_0.20	Q_0.25	Q_0.50	Q_0.75	Q_0.80	Q_0.90	Q_0.95
INTERCEPT	3.2301	3.6558	3.9098	4.0074	4.1445	4.5454	4.6250	4.8791	4.8681
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
SCH1_Z2	0.0800	0.0672	0.0663	0.0671	0.0832	0.0892	0.0915	0.0843	0.1034
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
BEFEX2	0.0118	0.0084	0.0085	0.0093	0.0125	0.0177	0.0195	0.0216	0.0301
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
BEFEX22	-0.0002	-0.0001	-0.0001	-0.0001	-0.0001	-0.0002	-0.0003	-0.0003	-0.0004
	<0.0001	0.0199	0.0003	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
EXP2	0.0135	0.0125	0.0137	0.0141	0.0173	0.0192	0.0201	0.0207	0.0232
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
EXP22	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0001	-0.0001	-0.0001	-0.0001
	0.3100	0.4572	0.1223	0.1518	0.0791	0.0233	0.0080	0.0143	0.0057
HOURS3	0.0093	0.0089	0.0085	0.0082	0.0067	0.0044	0.0038	0.0032	0.0027
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DGENDER4	0.0513	0.0512	0.0573	0.0609	0.0761	0.0862	0.0896	0.0851	0.1017
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DNPARIS5	0.0076	0.0102	0.0166	0.0201	0.0373	0.0574	0.0570	0.0781	0.0723
	0.3519	0.1562	0.0026	0.0004	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DPARIS6	0.0104	0.0291	0.0646	0.0794	0.1376	0.1779	0.1786	0.2285	0.1962
	0.6128	0.1302	0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DTYPDIP7	-0.1823	-0.1588	-0.1786	-0.1822	-0.2501	-0.3058	-0.3164	-0.3293	-0.3669
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
DPUBLIC8	0.0025	0.0095	0.0143	0.0137	-0.0063	-0.0281	-0.0346	-0.0459	-0.0825
	0.8033	0.4361	0.0684	0.0836	0.4305	0.0055	0.0032	0.0020	<0.0001
DFIXCT9	0.5305	0.4931	0.4237	0.3750	0.2386	0.1363	0.1262	0.1762	0.1176
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0113
DPERCT10	0.7030	0.6111	0.5060	0.4540	0.3273	0.2390	0.2380	0.3039	0.2388
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
IMR	0.3918	0.3574	0.3668	0.3792	0.4983	0.6366	0.6577	0.6768	0.6879
	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Table 5.1 shows the results from quantile regression of the parametric model for the French data. The above results, point at the importance of the use of quantile regression as they clearly suggest that impacts that most explanatory factors exert on wages are not same over the entire wage distribution in the French labour market. For example, the returns to education are higher at the both ends of the wage distribution. Returns to schooling initially decrease as

we move up along with wage distribution, attains its minimum at about 25th quantile then increases onwards from 50th quantile with only exception at 90th quantile. The returns to education for lowest paid 5% group is found as 8% while for highest paid 5% group as 10.34% for each addition year of schooling. The lowest returns to education are found for 20th quantile (i.e. 6.63%). These results concerning returns to education are consistent with the literature from other studies like Martins & Pereira (2004) and Flabbi et al. (2008) that reported lower returns to education at the lower ends of the earnings distribution. The higher schooling returns at higher quantiles are also reported by Falaris (2004) and Coelho et al., (2008). The results are also consistent with the French literature which is more relevant as Barnet-Verzat & Wolff (2008) found higher returns to different educational levels in the upper part of the wage distribution. But our results differ a little in that we report higher returns for extreme lower paid group as compared to other people in the lower half of the wage distribution. A similar pattern has been noted for returns to past experience, that it is higher for upper part of the wage distribution. Mainly returns to past experience fluctuate from nearly 1.18% (at 5th quantile) to 3.01% (at 95th quantile). The returns to past experience is less than 1% for 10th, 20th, and 25th quantiles but statistically significant. The impact of current job seniority is also not uniform across the wage distribution. Like past experience, the increase in wage with additional year of current job experience also increases as we move up along with wage distribution. From 5th quantile, the returns associated with job seniority increases from 1.35% to 2.32 % at the 95th quantile of the wage distribution. The degree of concavity of the experience (both past and current) wage relationship is also different in different parts of the wage distribution. The effect of working hours on wages decrease as we go in the upper part of the wage distribution. For the 5th quantile wage increase due to an additional hour of work is found near to 1% and at the 95th quantile this effects becomes as 0.27% after consistently decreasing along with quantiles of the wage distribution. Similar to many others (for instance Kuhn, 1987; García et al., 2001) our results also suggest the total wage gap between two genders is not uniform across the wage distribution. Although significant in favour of men over the entire range of wage distribution but like Albrecht et al. (2003), our results reveal that gender wage gap is more pronounced in the upper quantiles of the wage distribution. The higher wage premiums for males at higher quantiles of the wage distribution roughly hints at the presence of glass ceiling effect against women in the French labour market which is found in De la Rica et al. (2008, for Spain) and Arulampalam et al. (2007, for some other European countries). Concerning France, our results differ with those of

Arulampalam et al. (2007) in one direction as they reported higher gender wage gaps in the middle of the distribution for France which our results suggest to be not so. The lower raw gender wage gaps at the lower parts of wage distribution may be due to the possibility that women are more concentrated in low paying jobs in France which is documented in different studies for France (for instance, Jellal et al., 2008; Barnet-Verzat & Wolff, 2008).

Another important fact exposed by the quantile regression analysis is non uniformity of the regional effects on wages. The wage gains that urban workers have over rural workers differ substantially across different parts of the wage distribution. The quantile regression reveals that there are no significant differences in wages of individuals working in different areas for the lower paid 10% workers. After 10^{th} quantile wage premiums for workers in non-Paris urban areas over workers in rural areas increase as we move to higher quantiles of the wage distribution. These wage gains fluctuate between 1.66% (at 20^{th} quantile) and 7.81% (at 90^{th} quantile). The wage premium for workers in the Paris region over rural workers is also found higher in the upper quantiles of the wage distribution and lower in lower parts (insignificant for lowest paid 10% individuals). These results suggest that highly paid people also enjoy higher wage premiums due to working in a particular location. The more widening wage gaps in upper parts of the wage distribution opens the possibility that factors affecting an individual's placement at different locations in wage distribution may be major causes of urban-rural wage gaps in France. Like in mean regression the impact of professional degree on wages of individuals is negative and significant in the entire distribution of wages. However, the magnitude of this impact is different in different quantiles and cost of professional degree in terms of lesser wage is higher at upper quantiles. This wage penalty fluctuates around 15-18% up to 25^{th} quantile but then it goes to 25% in middle and growing rapidly reaches 36.69% at the 95^{th} quantile of the conditional wage distribution. Another major advantage that quantile regression revealed over mean regression based models is that public-private wage differential comes out to be significant at some points of the wage distribution which was reported as non-significant in some of mean regression models estimated in the present work. The wage differential between public and private sector workers is found non-significant in the lower half of the wage distribution, while it is in favour of private sector workers in upper half of the wage distribution. Within that upper half, the wage loss for public sector workers is also not homogeneous and this penalty increases as we move up within the upper part of the wage distribution. The wage premiums enjoyed by the fixed term or permanent contract holders over the people working temporarily or without

any contracts found significant over the entire wage distribution but amount of such benefits depends on the choice of quantile. Wage gains that fixed term contract holders have over temporary workers decrease as we move up along with conditional wage distribution. The gains for fixed term contract workers over temporary workers found 53% for 5^{th} quantile and reached 11.76% at 95^{th} quantile. Similarly, the wage premiums for permanent contract holder are also higher in the lower parts of the wage distribution and this effect decreases as we move to upper quantiles. These results highlight the fact that contract statues are more important for low paying jobs. Finally the impact of sample selection bias is not uniform as well across the wage distribution. The coefficients for sample selection correction term in the different quantiles of the wage distribution can serve as an indication that sample selection bias may be more substantial in the highly paid groups as compared to low paid groups.

The similar patterns for the coefficients associated to all explanatory variables in different parts of the conditional wage distribution can also be seen in the plots of quantile regression coefficients given in the following figures (5.1a, 5.1b, 5.1c and 5.1d).

Figure 5.1a

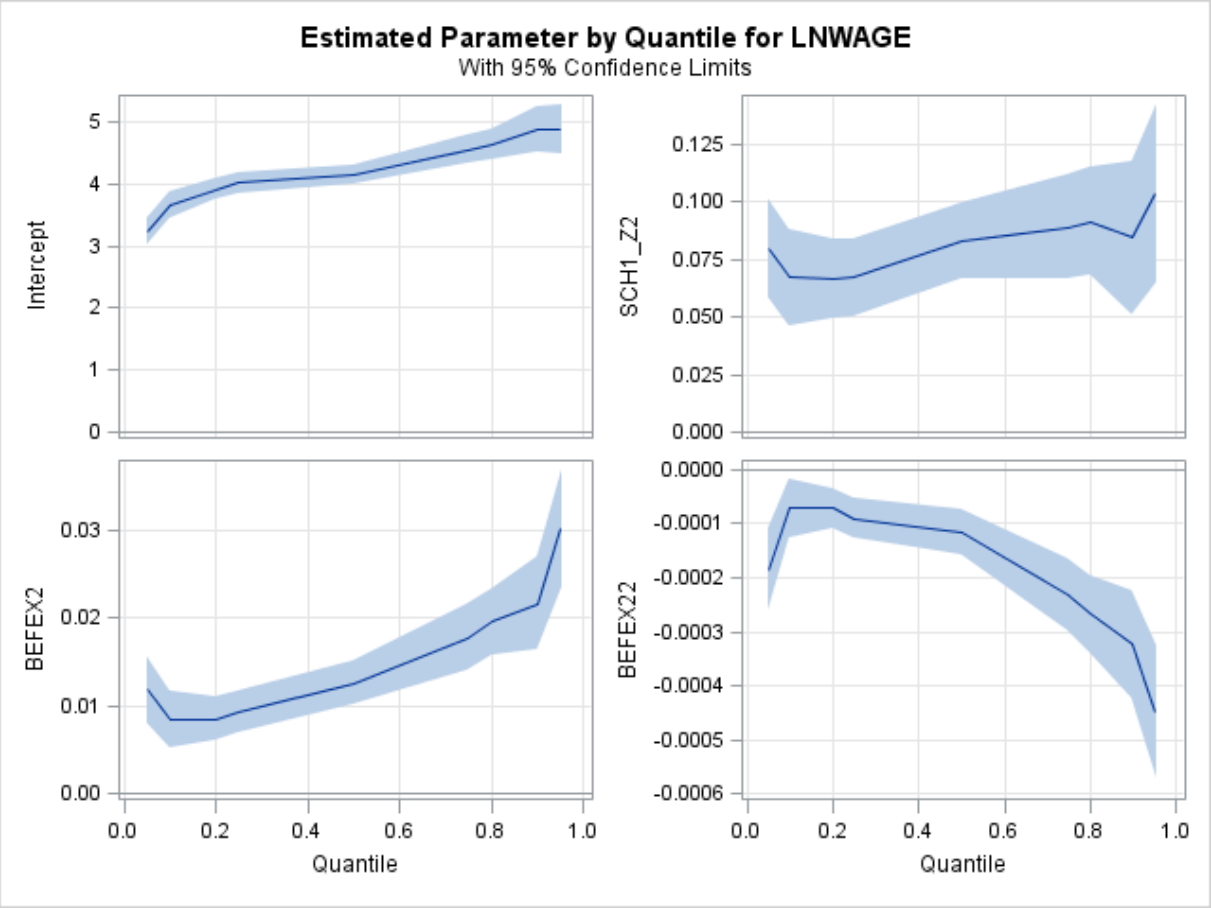

Estimated Parameter by Quantile for LNWAGE
With 95% Confidence Limits
Intercept
SCH1_Z2
BEFEX2
BEFEX22
Quantile
Quantile

Figure 5.1b

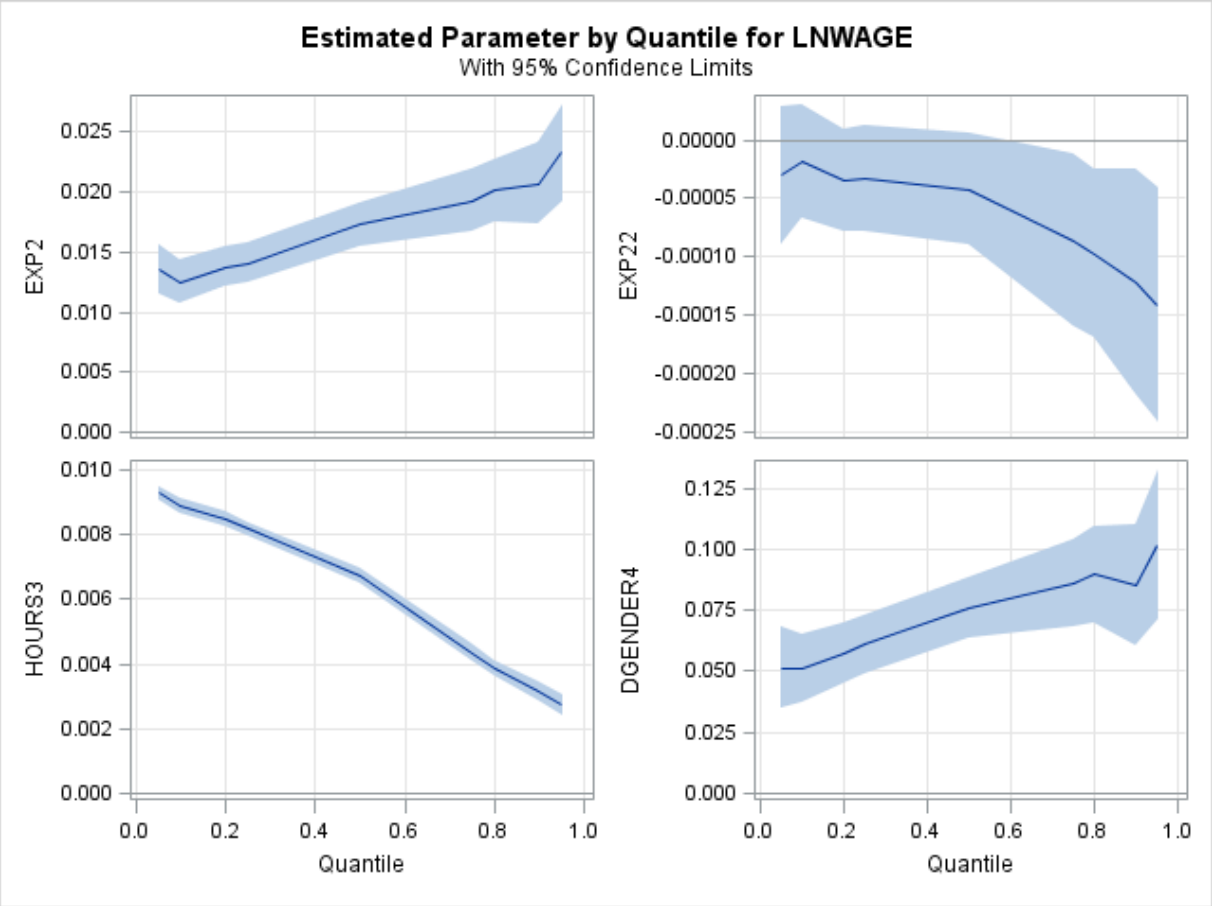

Estimated Parameter by Quantile for LNWAGE
With 95% Confidence Limits
EXP2
0.025
0.020
0.015
0.010
0.005
0.000
EXP22
0.00000
-0.00005
-0.00010
-0.00015
-0.00020
-0.00025
HOURS3
0.010
0.008
0.006
0.004
0.002
0.000
DGENDER4
0.125
0.100
0.075
0.050
0.025
0.000
0.0
0.2
0.4
0.6
0.8
1.0
Quantile
Quantile

Figure 5.1c

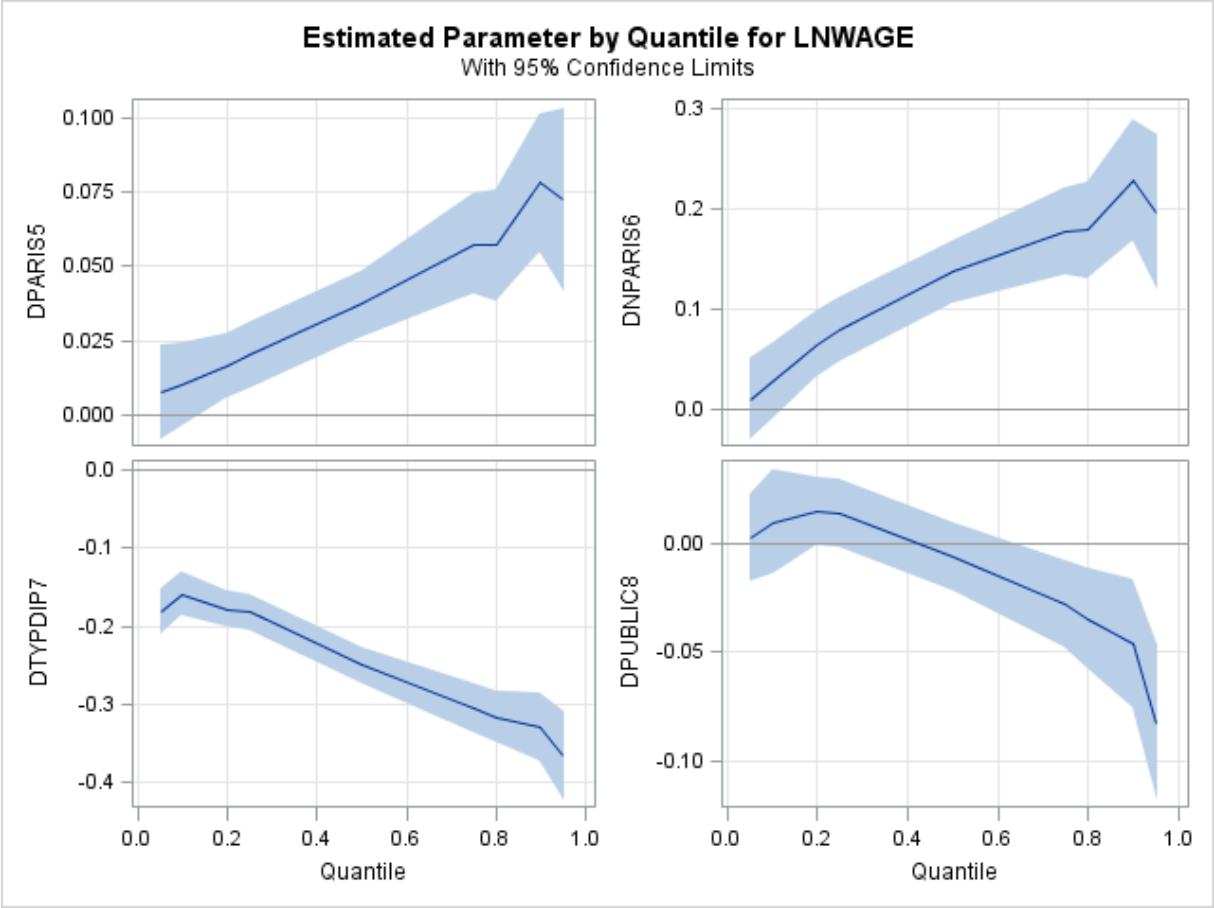
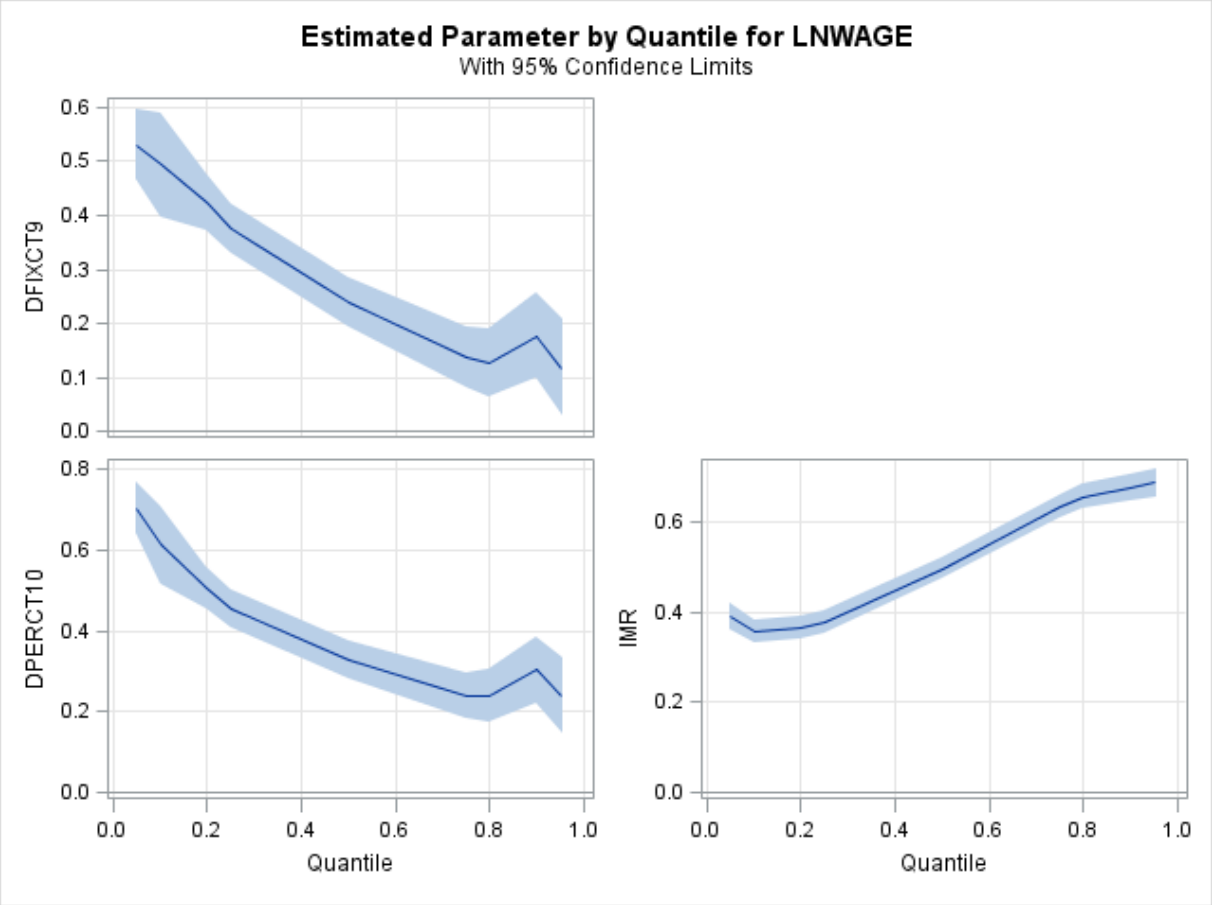

Estimated Parameter by Quantile for LNWAGE
With 95% Confidence Limits
DPARIS5
0.100
0.075
0.050
0.025
0.000
DNPARIS6
0.3
0.2
0.1
0.0
DTYPDIP7
0.0
-0.1
-0.2
-0.3
-0.4
DPUBLIC8
0.00
-0.05
-0.10
0.0
0.2
0.4
0.6
0.8
1.0
Quantile

Figure 5.1d



5.4 Estimation of the Mincerian Quantile Regression Model for Pakistan

We have preferred parametric model over semi-parametric model in general based on mean regression model estimation for Pakistani data as well (Chapter 4, Section 4.5.10.1), so for Pakistani data too, quantile regression is applied on the simple version of parametric model in a similar way as we applied on the French data in previous section.

For Pakistani case we recall parametric Model 4.5 and for running the quantile regression according to (5.1), we set the dependent variable $y_i = \ln W_i$, ith row vector x'_i includes $\hat{SCH}1_Z2$ fitted from first stage regression based on instrument Z2 (first stage of Model 4.3

i.e. Table 4.5-A), all variables included in X_K (Eq. 4.1) for Pakistan, and *IMR* (based on probit coefficients in Table 4.7).

Moreover, the vector β in (5.1) contains all the coefficients related to variables relevant for the Pakistani data

Similar to that we did for France in the previous Section 5.3, for Pakistan as well, the Mincerian model is also estimated at similar quantiles i.e. for 5^{th} , 10^{th} , 20^{th} , 25^{th} , 50^{th} , 75^{th} , 80^{th} , 90^{th} , and 95^{th} percentiles of the Pakistani wage distribution.

The following Table 5.2 presents the results from quantile regression analysis of the Mincerian model for the Pakistani labour force data.

Table 5.2: Parameter Estimates of the Mincerian Wage Model at Different Quantiles of the Pakistani Wage Distribution

Parameter	Q_0.05	Q_0.10	Q_0.20	Q_0.25	Q_0.50	Q_0.75	Q_0.80	Q_0.90	Q_0.95
INTERCEPT	7.2413 (<0.0001)	7.1770 (<0.0001)	7.1142 (<0.0001)	7.0476 (<0.0001)	6.9463 (<0.0001)	6.9515 (<0.0001)	6.9123 (<0.0001)	6.8401 (<0.0001)	7.0086 (<0.0001)
SCH1_Z2	0.0340 (0.0002)	0.0397 (<0.0001)	0.0516 (<0.0001)	0.0588 (<0.0001)	0.0779 (<0.0001)	0.1059 (<0.0001)	0.1161 (<0.0001)	0.1472 (<0.0001)	0.1521 (<0.0001)
EXP2	0.0192 (<0.0001)	0.0228 (<0.0001)	0.0289 (<0.0001)	0.0318 (<0.0001)	0.0404 (<0.0001)	0.0454 (<0.0001)	0.0482 (<0.0001)	0.0562 (<0.0001)	0.0585 (<0.0001)
EXP22	-0.0002 (<0.0001)	-0.0002 (<0.0001)	-0.0003 (<0.0001)	-0.0003 (<0.0001)	-0.0004 (<0.0001)	-0.0004 (<0.0001)	-0.0004 (<0.0001)	-0.0004 (<0.0001)	-0.0004 (<0.0001)
HOURS3	0.0029 (<0.0001)	0.0033 (<0.0001)	0.0037 (<0.0001)	0.0039 (<0.0001)	0.0054 (<0.0001)	0.0062 (<0.0001)	0.0060 (<0.0001)	0.0068 (<0.0001)	0.0063 (0.0003)
DGENDER4	0.1407 (<0.0001)	0.1259 (<0.0001)	0.1008 (<0.0001)	0.1090 (<0.0001)	0.0207 (0.2494)	-0.0829 (<0.0001)	-0.0780 (0.0002)	-0.1579 (<0.0001)	-0.2056 (<0.0001)
DWORKLOC5	-0.1019 (<0.0001)	-0.1072 (<0.0001)	-0.1031 (<0.0001)	-0.1060 (<0.0001)	-0.1283 (<0.0001)	-0.1629 (<0.0001)	-0.1568 (<0.0001)	-0.1516 (<0.0001)	-0.1412 (<0.0001)
DTYPDIP6	-0.1650 (0.1467)	0.0229 (0.7962)	0.1058 (0.1631)	0.0915 (0.1145)	0.1544 (0.0035)	0.0117 (0.8238)	-0.0867 (0.0859)	-0.3197 (<0.0001)	-0.5290 (<0.0001)
DPUBLIC7	-0.0249 (0.0109)	-0.0341 (0.0003)	-0.0301 (0.0002)	-0.0213 (0.0031)	-0.0007 (0.9320)	-0.0077 (0.5240)	-0.0070 (0.5913)	-0.0104 (0.5915)	-0.0258 (0.4211)
DFIXCT8	-0.0910 (<0.0001)	-0.0849 (<0.0001)	-0.0712 (<0.0001)	-0.0743 (<0.0001)	-0.0801 (<0.0001)	-0.0900 (<0.0001)	-0.1004 (<0.0001)	-0.1222 (0.0003)	-0.1044 (0.0984)
DPERCT9	0.1916 (<0.0001)	0.2219 (<0.0001)	0.1757 (<0.0001)	0.1466 (<0.0001)	0.0539 (0.0463)	-0.1020 (0.0024)	-0.1628 (<0.0001)	-0.3449 (<0.0001)	-0.4452 (<0.0001)
DSINDH10	-0.0156 (0.2068)	-0.0010 (0.9354)	0.0074 (0.4200)	0.0025 (0.7835)	-0.0158 (0.0535)	-0.0619 (<0.0001)	-0.0685 (<0.0001)	-0.1184 (<0.0001)	-0.1196 (<0.0001)
DKPK11	-0.0619 (<0.0001)	-0.0539 (<0.0001)	-0.0467 (0.0001)	-0.0423 (<0.0001)	-0.0500 (<0.0001)	-0.0550 (<0.0001)	-0.0654 (<0.0001)	-0.0915 (<0.0001)	-0.1431 (<0.0001)
DBALO12	0.1521 (<0.0001)	0.1639 (<0.0001)	0.1584 (<0.0001)	0.1563 (<0.0001)	0.1375 (<0.0001)	0.1281 (<0.0001)	0.1219 (<0.0001)	0.0963 (<0.0001)	0.0820 (0.0212)
IMR	0.3354 (<0.0001)	0.4461 (<0.0001)	0.5218 (<0.0001)	0.5472 (<0.0001)	0.7196 (<0.0001)	0.8934 (<0.0001)	0.9264 (<0.0001)	1.0317 (<0.0001)	1.1843 (<0.0001)

Table 5.2 shows the results from quantile regression applied on parametric model (Model 4.5) that we chose as more appropriate compared to semi-parametric model for analysis based on data from Pakistan's labour market. From above results it is clear that coefficients associated with different explanatory factors are different in magnitude (and in direction in some cases) in different quantiles of the Pakistani labour market's wage distribution. This signifies the gains in insights brought up by the application of quantile regression.

Commenting on contributory factors one by one, we see that impact of education on wages of individuals increase as we move up with quantiles of the conditional wage distribution. Each additional year of schooling increases wages by roughly 3.4% in the 5^{th} quantile and this effect gets to 15.21% for each additional year in 95^{th} quantile after consistently increasing over different middle quantiles. The results showing higher returns to schooling in higher quantiles of the wage distribution are consistent with our French quantile regression results and also with many other studies, like Martins & Pereira (2004) and Flabbi et al. (2008). Similar findings have been reported in Staneva et al. (2010) for Bulgaria and Kazakhstan. The finding of higher schooling coefficient in upper part and lower schooling coefficient in lower part is also in line with the quantile regression results previously found for the Pakistani data (Hyder & Reilly, 2005) but our finding is different in an important aspect that our results show a consistent increase in returns to schooling by moving along with higher quantiles while Hyder & Reilly (2005) reported that returns to schooling drop in the middle parts of the wage distribution. Similar finding of higher schooling coefficient in upper parts of the wage distribution for Pakistan is also documented by Fasih et at. (2012) when they estimated the Mincerian returns in different quantiles of wage distribution for different countries. The results indicate that education plays a vital role in the wage determination process in the Pakistani labour market, as returns to schooling are lowest for lowest paid group and highest for highest paid group. The magnitude of influence of experience on wages is also found as not uniform in different parts of the wage distribution. The returns to experience found higher in upper parts of the wage distribution. The returns to experience are just 1.9% for every extra year in 5^{th} quantile of the wage distribution which gets at peak (5.85%) in 95^{th} quantile after continuously increasing in all the quantiles between these two most lower and most upper quantiles of the conditional wage distribution. The non uniformity of returns to experience is similar to that we noted for the French labour force data but the amount of differences among returns in different quantiles are higher for Pakistani data. The coefficient linked to quadratic term of experience exposed that degree of concavity in experience wage relationship is also not homogeneous in different parts of the conditional wage distribution. The returns to experience are different from those found in Hyder & Reilly (2005) for Pakistan as they reported lower coefficients in higher quantiles but instead of actual or potential experience, they used age as an experience proxy. Internationally, these results are in line with Ajwad & Kurukulasuriya (2002) who reported lower experience returns in lower parts of the wage distribution. The effect of hours worked on wages is found statistically significant in all part

of the wage distribution with having higher effect in the upper half of the wage distribution as compared to the lower half. An important fact exposed by quantile regression is related to gender wage gaps. The raw gender wage gap is not only heterogeneous in different quantiles of the Pakistani wage distribution but also changes its direction in upper part of the distribution. In lower half of the wage distribution gender wage differential is found to be in favour of men but in upper half this differential is found as in favour of women. Consistent to results from mean regression (parametric Model 4.5), the gender wage gap is not significant at 50^{th} quantile of the wage distribution. The finding regarding gender wage gap is in contrast with findings from several other studies that reported higher wage gaps in favour of men in upper tail of the wage distribution compared to its lower tail (for example Ajwad & Kurukulasuriya, 2002 for Srilanka ; Albrecht et al., 2003 for Sweden ; García et al., 2001 for Spain). Decreasing wage gaps with moving up with quantiles of the wage distribution in favour of men has also been reported in literature (Arulampalam et.al, 2007 for Spain). For Pakistan, Hyder & Reilly (2005) reported lower gender gaps in upper parts of the wage distribution but contrary to ours their reported gender wage differential always remained in favour of males. One possible explanation for the wage gaps in favour of women in upper half of the wage distribution may lie in the relative scarcity of skilled and highly educated women to be included in high paying jobs. The lower supply of such women in highly paid job positions may increase their wages in upper part of the conditional wage distribution which lead wage differential in favour of women in that part of distribution. Similar to that we found in mean regression, quantile regression also revealed that wages are considerably higher for workers working in rural areas as compared their urban counterparts but the amount of wage premium for rural workers in not even across the wage distribution. The wage penalty for urban workers is minimum with a magnitude of 10.19% at 5^{th} quantile which increase up to 75^{th} quantile (16.29%) and then decreases a little in quantiles 80^{th} and above but always remains higher than penalty that urban workers face in lower half of the conditional wage distribution for the Pakistani labour market data. The differences between wages due to having a professional and general educational degree are found significant at 50^{th}, 90^{th} and 95^{th} quantiles. Evaluated at 50^{th} quantile, individuals with professional degree get 15.44% higher wages over others while professional degree holders face a wage penalty of 32% and 53% in the 90^{th} and 95^{th} quantiles respectively. The wage penalties or premiums for professional or general degree holders are found as non-significant at all other selected points of the conditional distribution of wages. These finding are opposite to those from the French

data based quantile regression results as there we found significant wage penalty for professional degree holders in all parts of the wage distribution. The wage loss for the people working under fixed term contracts in comparison to those who work temporarily is although not identical at different points of wage distribution but it mainly moves around 7-9 % up to 75th quantile. While this wage loss for fixed term contract workers is non-significant for the best paid 5% workers. The quantile regression proves its better performance by providing a deep insight into wage gaps between temporary workers and permanent contract holders. As we know that this kind of wage difference was found insignificant in mean regression (refer to parametric Model 4.5), but quantile regression exposed the fact that permanent workers enjoy wage premium over temporary workers which (wage premium) decreases while moving up along the wage distribution and turns out as a penalty (for permanent contract workers) in the upper half of wage distribution. So we can say that permanent contract workers in lower half of the wage distribution enjoy wage gain while face a wage loss in upper half of the wage distribution. The maximum wage gain that permanent contract workers get is 22.19% noted for 10th quantile while the maximum wage penalty they face is of magnitude 44.52% noted for highest paid 5% workers. So we may conclude that these premiums in lower half and penalties in upper half of the distribution for permanent contract holders with respect to temporary workers caused non-significance of such measure in the mean regression model. Looking at results from quantile regression regarding wage differences among people working in different provinces, we note that these differences are not alike at different points selected for evaluation of wage distribution of the Pakistani labour market. Taking most populous Punjab province as reference, we see that wages for workers in Punjab and Sindh province are similar in the lower half while workers in Sindh province face wage penalty in upper half which is increasing in magnitude. For example workers in Sindh get 6.2% lesser wages than workers in Punjab at 75th quantile but this amount gets to 11.96% wage loss at 95th quantile. Contrary to non-significance between wages of Sindh and Punjab provinces in lower part of the wage distribution, the wages differ significantly between workers in the Punjab and Khyber Pakhtunkhwa provinces in all parts of the conditional wage distribution. In all quantiles people working in Khyber Pakhtunkhwa province get fewer wages compared to those in Punjab province. This wage premium for workers in Punjab over workers in Khyber Pakhtunkhwa province decreases initially from 5th quantile (6.19%) up to 75th quantile (5.5%) and then increase in upper quantiles and gets at 14.31% at 95th quantile. Comparing wage of workers in the Punjab and Balochistan provinces, we see that workers in

Balochistan province get significantly higher wages in all parts of the wage distribution. The amount of wage gains of workers in Balochistan is not even across the wage distribution, in fact it increases as we move to the upper portion of wage distribution (from 33.54% at 5^{th} quantile to 118% at 95^{th} quantile). The wage premium for people working in Balochistan province in general and higher premium in higher quantiles in particular may be due to lesser availability of highly educated and skilled workers as this province is smallest with regard to population and also is generally considered as with lowest educational opportunities. Comparing the wage gaps among the Sindh, Balochistan and Khyber Pakhtunkhwa provinces, we see that people working in Balochistan province get higher wages, then followed by workers in Sindh province followed by those who are working in Khyber Pakhtunkhwa province and this finding seems robust to all quantiles of the wage distribution. The coefficient associated to IMR is also significant in all parts of the wage distribution which increases by moving along with higher quantiles. So it means that if we had ignored sample selection bias then it would have more severe effects on coefficient estimates for highly paid groups of individuals compared to the effects that it would have exerted on coefficient estimates for low paid groups. These results concerning sample selection bias in different parts are similar to those found in Hyder & Reilly (2005) for the Pakistani data and those that we found for the French data quantile regression in the present thesis.

The differences in impacts of different explanatory variables in different quantiles of the Pakistani wage distribution that we interpreted in detail in the above discussion are also clear and evident in the following figures (5.2a, 5.2b, 5.2c and 5.2d). These figures show that in the Pakistani labour market, how impacts of different independent variables on wages change in different quantiles of the conditional wage distribution in Pakistan.

Figure 5.2a

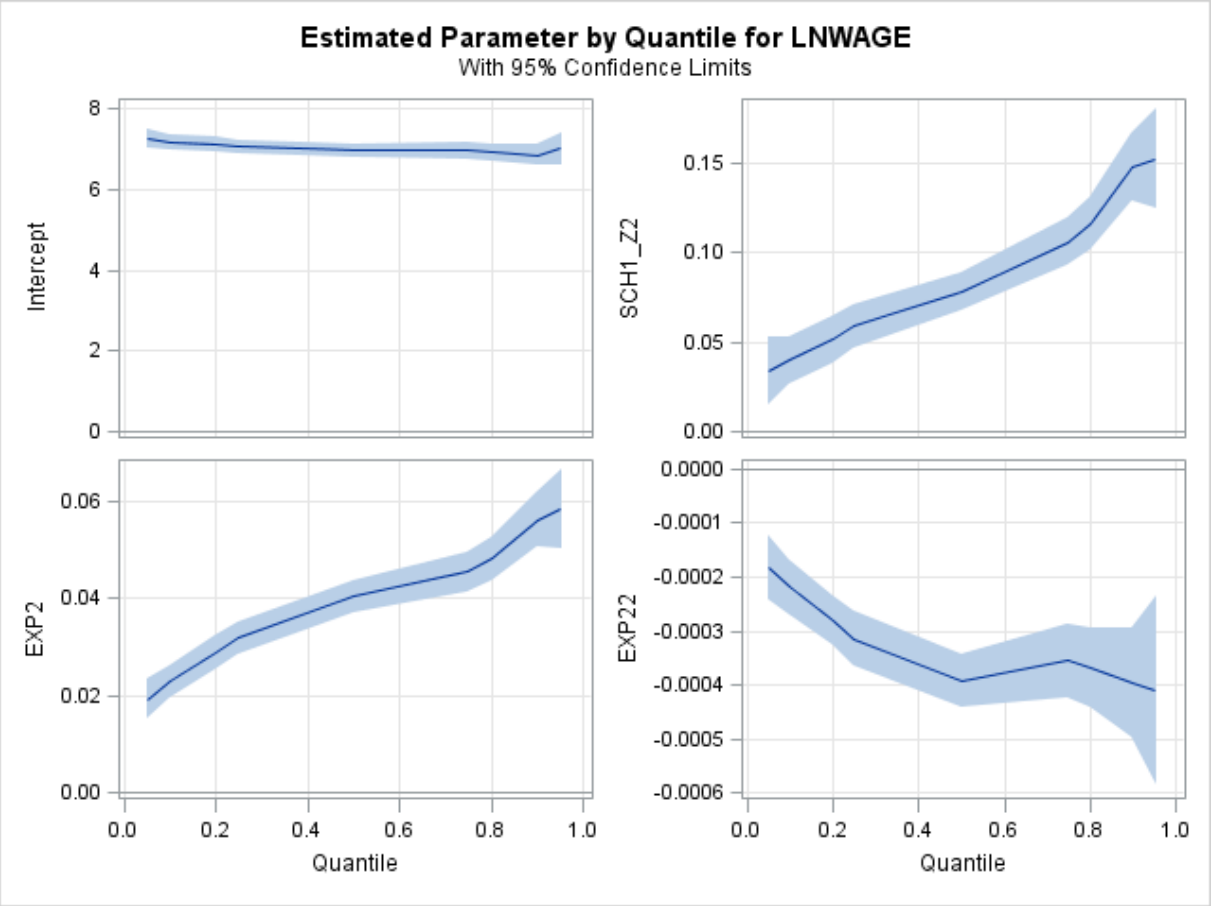

Estimated Parameter by Quantile for LNWAGE
With 95% Confidence Limits
Intercept
SCH1_Z2
EXP2
EXP22
Quantile
Quantile

Figure 5.2b

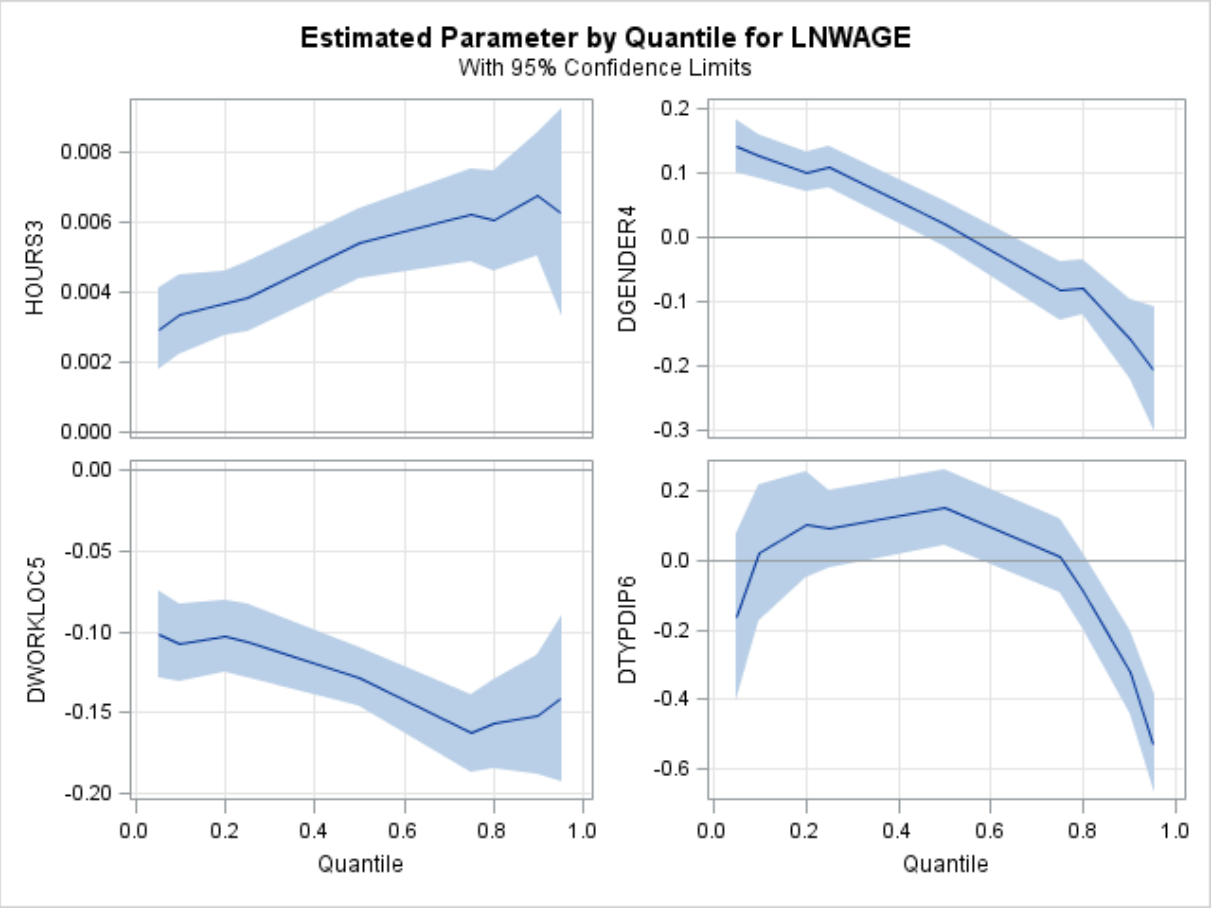

Estimated Parameter by Quantile for LNWAGE
With 95% Confidence Limits
HOURS3
0.008
0.006
0.004
0.002
0.000
DGENDER4
0.2
0.1
0.0
-0.1
-0.2
-0.3
DWORKLOC5
0.00
-0.05
-0.10
-0.15
-0.20
DTYPDIP6
0.2
0.0
-0.2
-0.4
-0.6
0.0
0.2
0.4
0.6
0.8
1.0
Quantile
Quantile

Figure 5.2c

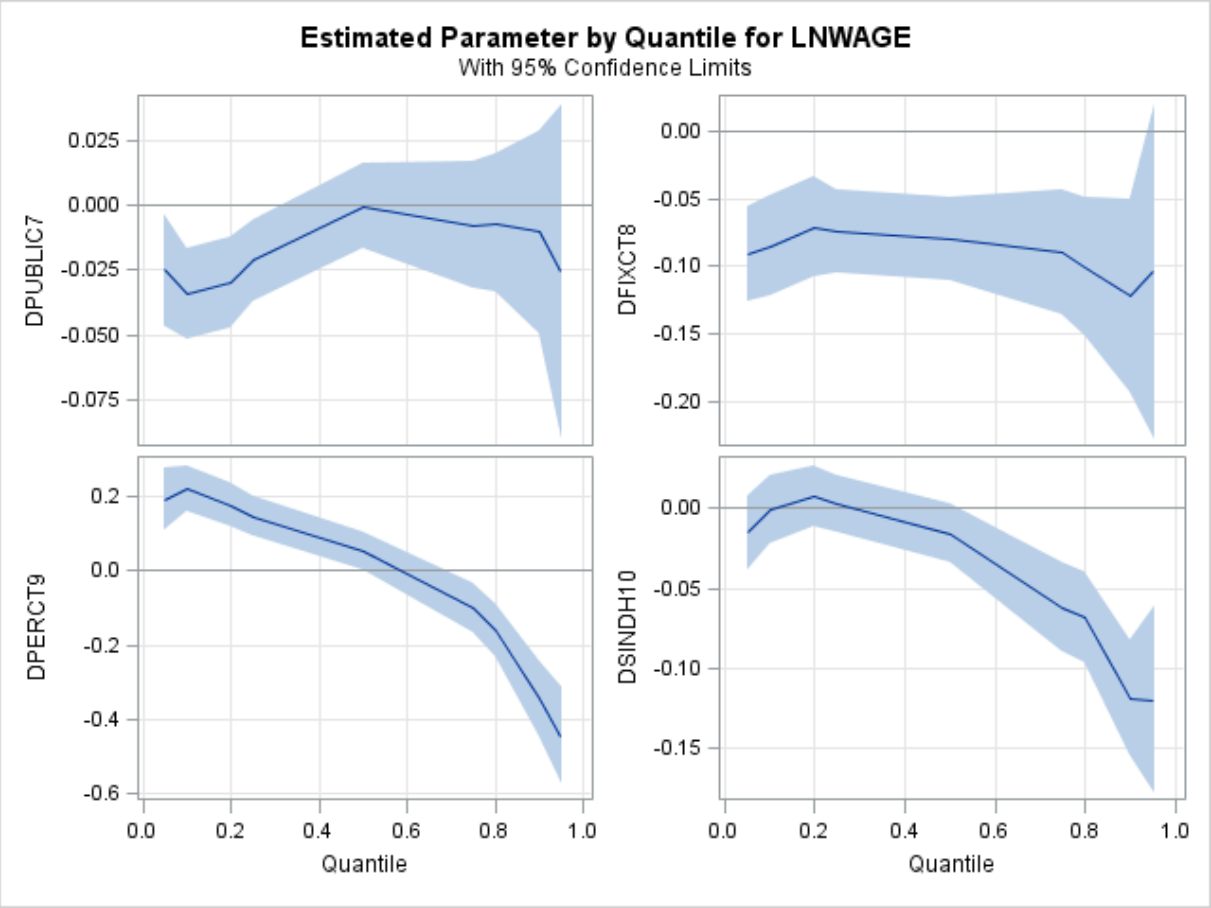
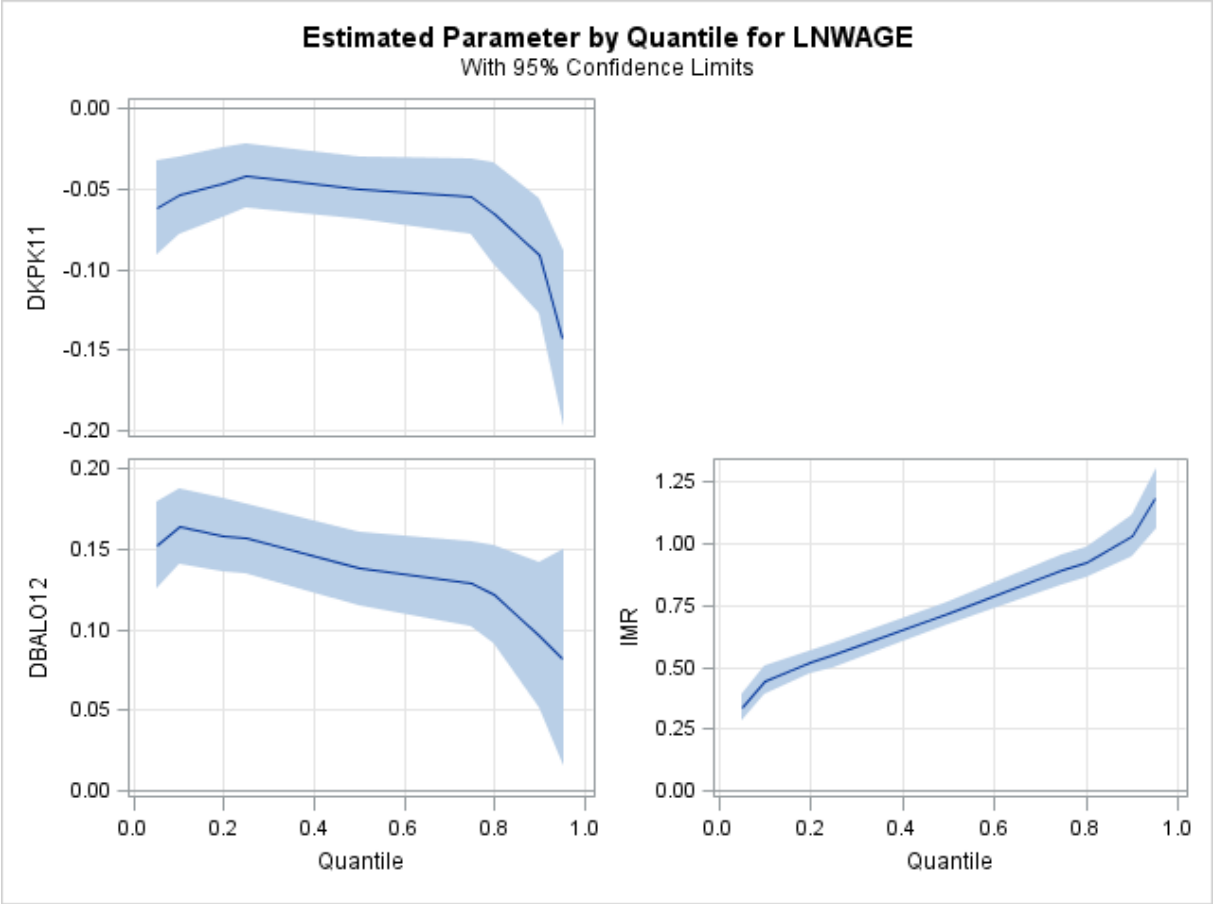

Estimated Parameter by Quantile for LNWAGE
With 95% Confidence Limits
DPUBLIC7
DFIXCT8
DPERCT9
DSINDH10
Quantile
Quantile

Figure 5.2d


Estimated Parameter by Quantile for LNWAGE
With 95% Confidence Limits
DKPK11
0.00
-0.05
-0.10
-0.15
-0.20
DBALO12
0.20
0.15
0.10
0.05
0.00
IMR
1.25
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0
Quantile
Quantile

Chapter 6: Summary and Conclusions

In the present doctoral thesis, we have estimated the Mincerian wage regression for data from the French and the Pakistani labour markets. The major sources of bias like endogeneity of schooling, measurement error and sample selection have been addressed. In order to tackle the bias arising due to endogeneity of schooling, and measurement error, we applied well known IV2SLS approach which is considered as a standard solution to counter these problems. For the IV2SLS estimation, we have proposed two new instrumental variables for endogenous schooling. The first instrument is defined as "the average schooling in the household to which an individual belongs". The other instrument is defined as "the average schooling in country, of the year, of the age group, of gender, at the time when an individual joined the labour market as wage worker". Based on the significance in the first stage schooling regressions, both instrumental variables were found relevant for both countries; France as well as for Pakistan. For both countries and using both instrumental variables, the Hausman (1978) test suggested schooling variable as endogenous. However, test for validity of instruments was not possible because we have used only one instrument at a time while validity test (Sargan, 1964) requires number instruments to be greater than endogenous explanatory variables. Due to high possible correlation between two proposed instruments, we have not used them in the same specification. Confirmatory with general findings reported in literature, higher schooling coefficients were found in the IV2SLS estimation compared to the OLS one that considers education to be exogenous.

In order to choose one instrument which is most suitable of the two proposed instruments, we have analyzed the correlation matrix among response variable (log of monthly wages), suspected endogenous variable (schooling), and both instrumental variables (Instrument-1 and Instrument-2). For both countries, we found a similar behaviour of instruments. Due to the fact that Instrument-1 has more severe direct effect on response variable compared to that of Instrument-2, and more robust definition of Instrument-2, we have selected Instrument-2 to be more suitable for the IV2SLS approach for both the French and the Pakistani data sets.

To address the other potential source of bias in the estimation of the Mincerian wage model if estimated with the simple OLS estimation i.e. bias due to sample selectivity, we have employed the Heckman two-step procedure which is well known and most used solution to problem arising due to non random selection of people into wage earners sample. The correction term for sample selectivity was found positive and statistically significant for both

of the labour market's data. It means in both countries' labour markets, those who did not decide to participate in the labour market waged work activities would have earned lesser wages than participants if they (non-participants) had decided to join labour force as wage worker. But an important feature we noted in both countries is that the coefficient related to gender dropped whenever we corrected for sample selectivity which means that the effect of sample selection bias is different for males and females in both countries. Now as the gender coefficient drops and we have taken females as reference category so we may conclude that women who did not joined labour force a wage worker would have earned higher than those women who joined labour force if they (non-participating women) had joined the labour market as wage worker while this is reverse for men i.e non-participants would have earned lesser wages compared to participants if they (non-participating men) had joined labour force. While estimating Heckman (1979) sample selection model, we had to estimate a participation equation as a first step. The estimation results from selection or participation equation provided information about how different factors affect decision or chances of individuals to be involved in waged or salaried work in the labour market. Schooling attainments and being a male found to increase the chances of being a wage worker in both countries. This finding is in lines with similar findings in other studies. The results pointed out that age exerts a different effect on chances of being involved in waged work in France and Pakistan. In France, age found to augment the probability of being wage worker while situation is contrary in Pakistan. One possible explanation may lie in relatively higher unemployment in Pakistan than in France, as in an economy with higher chances of employment, people decide to join labour market as they grew up and get jobs more easily as compared to an economy with relatively lower employment opportunities where job search takes more time. On the other hand, in an economy with relatively higher unemployment, individuals decide to join the labour force with growing age but a portion of them don't get salaried work due to lower employment opportunities and tendency among employers to recruit fresh and young graduates. This decreases the chances of getting a job with growing age for those who did not find jobs when they were young and fresh graduates. Presence of younger dependent children in the household found to increase odds of work activity in France and to decrease these odds in Pakistan. These differences may be due to the joint estimation of selection equation for both genders, as in France it is more expected that husband and wife share financial responsibilities of household, so after a certain period of child bearing for women, number of younger children increases chances of both genders to work in order to meet the increased

requirements of household due to younger children; the large development in France of "crèches" and early childhood schooling reinforces this effect. While in the context of the Pakistani society, males are considered to be more responsible for household resources while females as less responsible. We suspect that the negative coefficient related to presence of younger children in the Pakistani selection equation as may be due to more strong negative effect on women's participation contrary to France where we expected that coefficient to be positive for both genders. One other explanation for difference in sign of coefficient associated with presence of younger children in the household may lie in differences in chances of re-employment of women after breaks in work activity due to maternity. Positive impact of marital status in France and negative for Pakistan may be due to similar reasons as we noted for such coefficient related to positive and negative impacts of younger children on odds of being a waged worker in France and Pakistan, respectively. As expected, in both countries professional degree found to increase the likelihood of getting a job position in labour force. This is according to the common understanding or expectation that people having professional degrees get jobs more easily. Other sources of income controlled by indicators of homeownership and any financial aid for France, and number of people active in labour force for Pakistan, showed a negative impact on probability of waged work for individuals. In Pakistan, people in urban areas found to have higher probability to participate in labour force compared to workers from rural areas while this effect is reverse for France, where people in rural area shown to have higher odds of being involved in waged work compared to people in Paris region or other non-Paris urban areas. For Pakistan, size of household and being head of family discovered to decrease and increase likelihood for being a salaried worker, respectively. Significant differences have been noted due to be resident of different provinces on work participation for the Pakistani data.

From review of literature devoted to estimation of the Mincerian wage function, we found scarcity of the studies in general and absence of such studies for France and Pakistan in particular that correct for both problems (endogeneity and sample selection bias) in a single specification. Moreover, in the specifications in which we have tackled the issues of endogenous schooling and sample selection bias separately, we found different impacts of these corrections on uncorrected estimates for most of the variables and this finding is robust for both of the countries under consideration. The changes in the coefficients of most explanatory factors were in different directions due to endogeneity and sample selection

corrections. The magnitudes of differences due to both corrections were also different generally. Therefore keeping in view the different impacts for correcting endogeneity bias and sample selection bias, relative scarcity of literature that corrects for both these biases in a single specification, and absence of studies correcting for both problems at the same time for France as well as for Pakistan, we have estimated a model that took both issues into account together. This model tackled both issues simultaneously, i.e. corrected endogeneity bias by IV2SLS using Instrument-2 and the Heckman (1979) approach for sample selection correction.

We have also noted from concerning literature that assumptions of linear model in the context of the Mincerian wage model have rarely been tested. Some of the studies used heteroscedasticity consistent standard errors but studies that formally tested the presence of heteroscedasticity of the Mincerian model's error terms were found very rare and even do not exist for the two countries we are dealing with. So we have tested the validity of homoscedasticity assumption by applying White's (1980) test. For both of the countries, errors found to be heteroscedastic. In order to avoid the adverse effects of heteroscedasticity of errors on the estimation process, we have done the adaptive estimation of the simultaneous model (simultaneous model that corrects for endogeneity and sample selectivity in a single specification). The advantage of adaptive estimation over just using heteroscedasticity corrected standard errors is that adaptive estimation provides not only efficient parameter estimates but also consistent standard errors.

In addition to the parametric estimation, we have also estimated the Mincerian wage regression model semi-parametrically for both countries, France and Pakistan. Generally, in the context of the Mincerian model, semi-parametric estimation known to mean that selection equation is estimated non-parametrically and hence final wage model is semi-parametric because it contains non-parametric component for selectivity correction term. But our semi-parametric estimation is different in the sense that it contains non-parametric component coming from first stage schooling equation in the IV2SLS estimation i.e we estimated first stage schooling equation non-parametrically by LOESS regression. To our knowledge this is first study that focused on the non-parametric estimation of first stage schooling equation. Based on the global performance of the parametric and semi-parametric models, we found that for France both parametric and semi-parametric models performed almost equally well.

Therefore, due to ease in estimation and more familiarity in applied literature we select parametric model to be more appropriate for the French data. Concerning simple and adaptive estimations of the parametric model, we noted that the global performances of the both models as quite similar. Although having similar global performances, we selected adaptive specification of the parametric model to be more robust and appropriate for the French data because in addition to possible biases due to measurement error, endogeneity and sample selectivity, it tackles problem arising due to heteroscedasticity of errors as well which were found to be so by application of White's (1980) test. Similarly, for the Pakistani analysis, we explored that overall performance of parametric and semi-parametric models are not too much different. But contrary to the French analysis, for the Pakistani analysis semi-parametric model produced unreasonable values for some coefficients which were in huge difference with general consensus of the relevant literature. After comparing simple and adaptive versions of the parametric model with semi-parametric model, for Pakistan too, we found parametric model in its adaptive specification as most apt for estimation of impacts that different contributory factors exert on the wage determination process in the labour market of Pakistan.

Concerning results, it is noted that generally, models estimated for the French labour market data have better global performance in terms of variation explained in the response variable as compared to models estimated for Pakistan's labour market data. From the most appropriate models selected for both countries, we found that education is a little more rewarded in Pakistani labour market than in French labour market. A more year spent in schooling process found to increase wages by 7.72% and 8.67% in France and Pakistan, respectively. Average schooling in France is higher than average schooling in Pakistan so this is according to general economic behaviour and relationship between price and supply which is also found in Selz & Thélot (2004) and Palme & Wright (1998) that returns to schooling decrease with increase in schooling attainments in the society. This difference between schooling coefficients of the two countries can also be attributed with economic development levels of both countries as schooling coefficient is found generally lower for developed countries than developing countries (Fasih et at., 2012). The effects of labour market experience were captured by potential experience in Pakistani analysis while by current job seniority and past experience in French analysis, so this effect is not directly comparable. However, in both countries measure for experience increase wages at a decreasing rate (i.e. experience–wage is

of concave nature). In the final selected model gender wage gap is insignificant in Pakistan (however this gap was significant in uncorrected specification or specifications that corrected for endogeneity bias or sample selection bias separately) while men enjoy a wage premium of nearly 8.6% in France. From the whole estimation process, we have seen that magnitude of raw gender gaps in wages varied greatly among the uncorrected OLS specification, endogeneity corrected specification, and sample selectivity corrected specification and specification in which both problems addressed simultaneously. This situation magnifies the importance of estimating a model that simultaneously corrects for both these sources of bias and also this finding may question the findings of the studies working on gender wage gaps but ignore both or one of these two potential sources of bias. Concerning the impact of work location in France, we explored that as expected people in rural area earn less than workers in the Paris or non-Paris urban areas. Workers in non-Paris urban areas and Paris region have roughly 4.7% and 15.7% higher wages respectively, compared to workers in rural areas. For Pakistan, contrary to expectation and evidence from existing literature, we found that people working in urban areas get significantly lesser wages as compared to workers working in rural areas of Pakistan. From the model selected to be most appropriate, we noted that rural area workers enjoy a wage premium of approximately 12% over their counterparts in urban areas. In both countries, professional degree has a negative effect on wages. However, this negative effect is more pronounced and statistically significant in France (25% lesser wages for professional degree holders compared to those with general degree), while this negative effect is statistically insignificant for the Pakistani labour market data. Raw wage gaps between public and private sectors of work are in favour of private sector employees but the magnitude and significance of such wage gaps both are not much higher and it is almost similar for both countries. Keeping in view the importance of contract statutes' impacts, we also controlled for these effects. The impacts of contract types on wages were found different in the French and the Pakistani labour markets. For French data, we found that people working under fixed term and permanent contracts get 23% and 36% higher salaries, respectively than those working temporarily. But these effects found different for workers working in Pakistan's labour market. In Pakistan people working under fixed term contracts get approximately 7.5% lesser wages compared to people who work temporarily, while interestingly we found no significant differences between wages of people working temporarily and working under permanent contracts. For Pakistan, we have also controlled for the provincial effects. We found significant differences in wages of individuals working in different provinces of Pakistan. Our

results showed that people working in the Balochistan province get higher wages followed by those working in the Punjab, Sindh and Khyber Pakhtunkhwa provinces, respectively. This finding once again justifies the inclusion of provincial effects in almost all studies estimating the Mincerian wage regression for Pakistan.

We also applied the interval estimation on the selected models for the French and the Pakistani data in order to give a range of coefficients linked to different factors affecting wages in both countries with a certain level of confidence.

After finding parametric model to be better than semi-parametric one from mean regression for both countries' data, we have applied quantile regression on the parametric model. Similar to the scarcity in mean regression estimations, the model dealing with sample selection and endogeneity problems simultaneously are rarely found in literature dedicated to estimation of the Mincerian wage function as a quantile regression model. Results from quantile regression proved the worth of application of quantile regression by showing its superiority over mean regression for both of the countries. It is proved in the way that we found impacts of explanatory factors on wages as substantially different in different quantiles of the wage distributions. The returns to education were found higher in upper parts compared to lower parts of the wage distribution for France as well as for Pakistan. However, variation in the schooling coefficient in different quantiles of wage distributions found lower in France (8% at 5^{th} quantile to 10.3% at 95^{th} quantile) than that variation in Pakistan's labour market (3.4% at 5^{th} quantile to 15.2% at 95^{th} quantile). So returns to schooling can be considered relatively more uniform in different parts of the wage distribution in France than in Pakistan. Similarly returns to job seniority and experience for the French case were found to be less variable as compared to returns to potential experience for the Pakistani case. The impact of hours devoted to work activity found also different in different quantiles of wage distribution but this effect decreased in the French case and increased in the Pakistani case as we move up along with the conditional wage distributions of the two countries. From quantile regression results it is also exposed that mean regression may not have correctly captured the raw gender wage gaps. It is so because wage gaps between men and women were found to differ largely across the wage distributions of the both countries. But there is a difference in both countries results that for France gender wage gap continuously increases as we move towards upper quantiles of the wage distribution and the gap always remained in favour of men. It means

that in the French labour market raw gender wage gap is more evident in the highly paid groups compared to lowest paid groups of workers. While for Pakistan, raw gender wage gap found in favour of men in the lower half of the wage distribution, decreased with moving towards upper quantiles in the first half and then became in favour of women in the second half of the wage distribution, increased with moving up in the upper half of the wage distribution. So men get a wage premium in first half and face wage penalty in second half of the wage distribution in the Pakistani labour market. The impact of work location found as non-significant for the lowest paid 10% of workers in France. For the other quantiles, people in rural areas earn less than their counterparts in Paris and other urban areas and this wage loss noted as increasing in higher parts of the wage distribution of the French workers. For the Pakistani labour market focusing on the impact of work location, we found that wage premium for people working in rural areas is not homogeneous in all quantiles of the wage distribution; in fact this wage premium in favour of rural workers looked like an inverse u-shaped curve depending on the position of individual in the wage distribution. The wage penalty associated to professional degree gets larger while moving across the quantiles of the wage distribution in France (18% for lowest paid 5% workers to 37% for highest paid 5% workers) and it is significant in all quantiles of the French wage distribution. As far as the quantile regression is concerned, from results for the Pakistani data, difference in professional and non-professional degrees comes out to be non significant for most of the quantiles of wage distribution while found significant in favour of non-professional degree holders for highest paid 10% of workers. Quantile regression model have also exposed that wage gains for private sector workers over their counterparts in public sector are only significant in first half of the wage distribution in Pakistan while the case is reverse in the French labour market where these differences were found to be significant only in second half of the wage distribution. Mean regression would not have captured these differences in wage premiums or losses associated with private or public sector employees. Quantile regression results concerning contract statuses showed that wage gains associated to fixed term or permanent contracts over temporary workers were lower in the upper parts and higher in the lower parts of the wage distribution in France. It means that holding a contract (fixed term or permanent) is more beneficial for those working in low paying jobs. For Pakistan we noted that fixed term contract workers earn less than temporary workers and the amount of difference is not only different in direction but also less variable compared to comparison of similar effects in France. For the French data, an interesting feature revealed by quantile regression is that

individuals working under permanents contracts get higher wage premiums in lower half and lower wage premiums in upper half of the wage distribution as compared to individuals who work temporarily. In Pakistan, it is similar to France in the sense that permanent contracts found to be more beneficial and rewarding in low paying jobs in Pakistan as well but in upper half of the wage distribution, people working under permanent contracts face a wage penalty compared to temporary workers. This explains why mean regression produced this effect as to be non-significant for the Pakistani data. For Pakistan, like effects of other explanatory variables, wage differentials related to working in different provinces were found substantially different in various parts of the wage distribution.

Although highly significant in all parts of the wage distribution in both countries, however the coefficients linked to sample selection correction term hinted at more strong effects of sample selectivity on estimated coefficients of different contributory factors in the upper parts of the wage distribution for both France and Pakistan. It means, we may conclude that, people who would have been placed in higher quantiles of the wage distribution out of those who decided not to join labour force as wage worker, would have faced more severe penalties (in form of lower wages) compared to possible penalties of those who decided not to participate in labour force and would have been placed in lower parts of the wage distribution, in comparison to participants in relevant quantiles.

The present study is an addition in literature devoted to estimation of the Mincerian wage regression for the French and the Pakistani labour market. It has advantages of using more recent data from both countries, adding some explanatory factors like hours of work, type of diploma and type of contract which have been rarely used in existing literature concerning to the both countries under consideration in particular, and in general for other countries as well. We have proposed two new instrumental variables in order to tackle the bias due to endogeneity of schooling variable. According to our knowledge these instruments have never been used before in the estimation process of the Mincerian model. The work is a new addition to the French and the Pakistani literature as a model that tackled both problems of endogeneity and sample selection simultaneously and such studies are also very rare in the overall literature dedicated to the Mincerian wage model for other countries of the world. The differences in results due to these corrections signified the importance of such simultaneous correction for estimation of the Mincerian wage model. Moreover, for France and Pakistan we

have first time formally tested the presence of heterogeneity of error term and applied adaptive estimation to avoid adverse effect of heteroscedasticity on estimated coefficients. We believe that addressing sample selection bias, endogeneity of education in same specification and eliminating effects of heteroscedasticity of error term through adaptive estimation helped to produce more efficient, more reliable, more accurate and more representative impacts that different covariates put on wage determination processes in the labour markets of these two countries.

We found that gaps in coefficients from models correcting endogeneity or sample selection bias separately and the OLS model were different in magnitude or in direction or in both ways. Similarly, gaps among coefficients from uncorrected model, endogeneity corrected model, sample selectivity corrected model and simultaneous model were also found different. This finding questioned the findings of many studies that work on wage differentials due to different factors like gender, sector of work (public-private), ethnicity, race, urban-rural etc but ignore one or both of these estimation problems or do not correct them in one specification. So studies that intend to get reliable and bias free estimates or tend to work on different kinds of wage differentials must address these issues together in order to get clearer picture.

Semi-parametric estimation containing non-parametric component for sample selectivity term has been applied in many studies for different countries. For the first time, we have estimated the Mincerian model in a semi-parametric way that contained non-parametric component from first stage schooling equation and explored the fact that for the first stage schooling equation, parametric estimation is more appropriate for both countries.

Finally, we estimated the parametric Mincerian wage model considered as more appropriate than the semi-parametric model from mean regression analysis, as a quantile regression model. For France and Pakistan, it is also first study that addressed both major sources of bias (sample selection and endogeneity of education) in same specification under the framework of quantile regression. Quantile regression revealed that most of the explanatory variables influence the labour market wages differently in different parts of the wage distributions of both countries.

From a futuristic point of view, we are planned to introduce one or two more instruments for endogenous schooling which will be similar to the two used in the present study. Those instruments will probably be defined as average schooling of an individual's age group interacted with gender, region, urban-rural location, or average education attainment in some other grouping based on some common characteristics. Like Instrument-1 in present study, these will be data generated instruments. In a future study, comparison between parametric and semi-parametric model that contain non-parametric component from probit selection equation in addition to non-parametric component from first stage schooling equation for French as well as Pakistani data is possible. We also plan to work on raw and decomposed wage differentials between men and women, public and private sectors, urban and rural areas and due to regional or provincial differences by applying a model that tackles problems of endogeneity, sample selection, and heteroscedasticity of errors concurrently.

Appendices

Appendix A1: Summary Statistics of the French Sample

The following Table A1.1 gives the summary statistics for the variables used for the estimation of the Mincerian wage model the French labour market data. These measures are based on the final sample used for estimation of the wage regression.

Table A1.1: Summary Statistics for the French Sample

Variable	Mean	Standard Deviation	Q1	Q3
LNWAGE	7.2551	0.4907	7.0031	7.5496
SCH1	11.6613	2.9272	11.0000	14.0000
BEFEX2	9.5091	9.7573	1.5000	15.3000
BEFEX22	185.6	305.9	2.2500	234.1
EXP2	10.2265	10.1859	2.0000	16.0000
EXP22	208.3	336.9	4.0000	256.0
HOURS3	144.6	31.2090	140.0	160.0
DGENDER4	0.5423	—	—	—
DNPARIS5	0.6160	—	—	—
DPARIS6	0.1492	—	—	—
DTYPDIP7	0.5280	—	—	—
DPUBLIC8	0.1094	—	—	—
DFIXCT9	0.1014	—	—	—
DPERCT10	0.8746	—	—	—
N = 27136				

Appendix A2: Models used for the French Data

The appendix A2 gives the algebraic forms of different models used in estimation of the Mincerian wage regression for the French data.

Model 3.1: Simple OLS model that do not address any possible biases

$$\begin{aligned}\ln W_i = {} & \beta_0 + \beta_1 (SCH1)_i + \beta_2 (\text{BEFEX2})_i + \beta_3 (\text{BEFEX22})_i + \beta_4 (\text{EXP2})_i + \beta_5 (\text{EXP22})_i \\ & + \beta_6 (\text{HOURS3})_i + \beta_7 (\text{DGENDER4})_i + \beta_8 (\text{DNPARIS5})_i \\ & + \beta_9 (\text{DPARIS6})_i + \beta_{10} (\text{DTYPDIP7})_i + \beta_{11} (\text{DPUBLIC8})_i \\ & + \beta_{12} (\text{DFIXCT9})_i + \beta_{13} (\text{DPERCT10})_i + \varepsilon_i\end{aligned}$$

Model 3.2: IV2SLS estimation with Instrument-1 (Z1) to correct for endogeneity

First Stage : schooling equation

$$\begin{aligned}SCH1_i = {} & \beta_0 + \beta_1 (\text{BEFEX2})_i + \beta_2 (\text{BEFEX22})_i + \beta_3 (\text{EXP2})_i + \beta_4 (\text{EXP22})_i \\ & + \beta_5 (\text{HOURS3})_i + \beta_6 (\text{DGENDER4})_i + \beta_7 (\text{DNPARIS5})_i \\ & + \beta_8 (\text{DPARIS6})_i + \beta_9 (\text{DTYPDIP7})_i + \beta_{10} (\text{DPUBLIC8})_i \\ & + \beta_{11} (\text{DFIXCT9})_i + \beta_{12} (\text{DPERCT10})_i + \beta_{13} (Z1) + \varepsilon_i \quad ,\end{aligned}$$

Z1 is the instrumental variable for endogenous schooling.

Second Stage : wage regression

$$\begin{aligned}\ln W_i = {} & \beta_0 + \beta_1 (\hat{SCH}1_Z1)_i + \beta_2 (\text{BEFEX2})_i + \beta_3 (\text{BEFEX22})_i + \beta_4 (\text{EXP2})_i \\ & + \beta_5 (\text{EXP22})_i + \beta_6 (\text{HOURS3})_i + \beta_7 (\text{DGENDER4})_i + \beta_8 (\text{DNPARIS5})_i \\ & + \beta_9 (\text{DPARIS6})_i + \beta_{10} (\text{DTYPDIP7})_i + \beta_{11} (\text{DPUBLIC8})_i \\ & + \beta_{12} (\text{DFIXCT9})_i + \beta_{13} (\text{DPERCT10})_i + \varepsilon_i\end{aligned}$$

$\hat{SCH}1_Z1$ are fitted values for endogenous schooling computed from first stage regression.

Model 3.3: IV2SLS estimation with Instrument-2 (Z2) to correct for endogeneity

First Stage : schooling equation

$$SCH1_i = \beta_0 + \beta_1(\text{BEFEX2})_i + \beta_2(\text{BEFEX22})_i + \beta_3(\text{EXP2})_i + \beta_4(\text{EXP22})_i + \beta_5(\text{HOURS3})_i + \beta_6(\text{DGENDER4})_i + \beta_7(\text{DNPARIS5})_i + \beta_8(\text{DPARIS6})_i + \beta_9(\text{DTYPDIP7})_i + \beta_{10}(\text{DPUBLIC8})_i + \beta_{11}(\text{DFIXCT9})_i + \beta_{12}(\text{DPERCT10})_i + \beta_{13}(Z2) + \varepsilon_i \quad ,$$

$Z2$ is the instrumental variable for endogenous schooling.

Second Stage : wage regression

$$\ln W_i = \beta_0 + \beta_1(\hat{SCH}1_Z2)_i + \beta_2(\text{BEFEX2})_i + \beta_3(\text{BEFEX22})_i + \beta_4(\text{EXP2})_i + \beta_5(\text{EXP22})_i + \beta_6(\text{HOURS3})_i + \beta_7(\text{DGENDER4})_i + \beta_8(\text{DNPARIS5})_i + \beta_9(\text{DPARIS6})_i + \beta_{10}(\text{DTYPDIP7})_i + \beta_{11}(\text{DPUBLIC8})_i + \beta_{12}(\text{DFIXCT9})_i + \beta_{13}(\text{DPERCT10})_i + \varepsilon_i$$

$\hat{SCH}1_Z2$ are fitted values for endogenous schooling computed from first stage regression.

Model 3.4: Heckman sample selection model to correct for sample selectivity

$$\ln W_i = \beta_0 + \beta_1(SCH1)_i + \beta_2(\text{BEFEX2})_i + \beta_3(\text{BEFEX22})_i + \beta_4(\text{EXP2})_i + \beta_5(\text{EXP22})_i + \beta_6(\text{HOURS3})_i + \beta_7(\text{DGENDER4})_i + \beta_8(\text{DNPARIS5})_i + \beta_9(\text{DPARIS6})_i + \beta_{10}(\text{DTYPDIP7})_i + \beta_{11}(\text{DPUBLIC8})_i + \beta_{12}(\text{DFIXCT9})_i + \beta_{13}(\text{DPERCT10})_i + \beta_{14}(\text{IMR})_i + \varepsilon_i$$

IMR is computed from the coefficient estimates of the probit regression for the French data.

Model 3.5: Model that addresses endogeneity and sample selection bias simultaneously

$$\begin{aligned}\ln W_i = {} & \beta_0 + \beta_1(\hat{SCH}1_Z2)_i + \beta_2(\text{BEFEX2})_i + \beta_3(\text{BEFEX22})_i + \beta_4(\text{EXP2})_i \\ & + \beta_5(\text{EXP22})_i + \beta_6(\text{HOURS3})_i + \beta_7(\text{DGENDER4})_i + \beta_8(\text{DNPARIS5})_i \\ & + \beta_9(\text{DPARIS6})_i + \beta_{10}(\text{DTYPDIP7})_i + \beta_{11}(\text{DPUBLIC8})_i \\ & + \beta_{12}(\text{DFIXCT9})_i + \beta_{13}(\text{DPERCT10})_i + \beta_{14}(\text{IMR})_i + \varepsilon_i\end{aligned}$$

Model 3.5W: Adaptive estimation of the Model 3.5

$$\frac{\ln W_i}{\hat{\sigma}} = \frac{1}{\hat{\sigma}}\left[\begin{aligned} & \beta_0 + \beta_1(\hat{SCH}1_Z2)_i + \beta_2(\text{BEFEX2})_i + \beta_3(\text{BEFEX22})_i + \beta_4(\text{EXP2})_i \\ & + \beta_5(\text{EXP22})_i + \beta_6(\text{HOURS3})_i + \beta_7(\text{DGENDER4})_i + \beta_8(\text{DNPARIS5})_i \\ & + \beta_9(\text{DPARIS6})_i + \beta_{10}(\text{DTYPDIP7})_i + \beta_{11}(\text{DPUBLIC8})_i \\ & + \beta_{12}(\text{DFIXCT9})_i + \beta_{13}(\text{DPERCT10})_i + \beta_{14}(\text{IMR})_i + \varepsilon_i\end{aligned}\right]$$

$\hat{\sigma}$ is computed from the Nadaraya-Watson (1964) approach using errors from Model 3.5

Model 3.6: Semi-parametric estimation of the Mincerian wage model

$$\begin{aligned}\ln W_i = {} & \beta_0 + \beta_1(\hat{SCH}1_NP)_i + \beta_2(\text{BEFEX2})_i + \beta_3(\text{BEFEX22})_i + \beta_4(\text{EXP2})_i \\ & + \beta_5(\text{EXP22})_i + \beta_6(\text{HOURS3})_i + \beta_7(\text{DGENDER4})_i + \beta_8(\text{DNPARIS5})_i \\ & + \beta_9(\text{DPARIS6})_i + \beta_{10}(\text{DTYPDIP7})_i + \beta_{11}(\text{DPUBLIC8})_i \\ & + \beta_{12}(\text{DFIXCT9})_i + \beta_{13}(\text{DPERCT10})_i + \beta_{14}(\text{IMR})_i + \varepsilon_i\end{aligned}$$

$\hat{SCH}1_NP$ are non-parametrically (through LOESS regression) predicted values for schooling variable.

Model 3.6W: Adaptive estimation of the Model 3.6

$$\frac{\ln W_i}{\hat{\sigma}} = \frac{1}{\hat{\sigma}}\left[\begin{array}{l} \beta_0 + \beta_1(\hat{SCH}1_NP)_i + \beta_2(\text{BEFEX2})_i + \beta_3(\text{BEFEX22})_i + \beta_4(\text{EXP2})_i \\ \quad + \beta_5(\text{EXP22})_i + \beta_6(\text{HOURS3})_i + \beta_7(\text{DGENDER4})_i + \beta_8(\text{DNPARIS5})_i \\ \quad + \beta_9(\text{DPARIS6})_i + \beta_{10}(\text{DTYPDIP7})_i + \beta_{11}(\text{DPUBLIC8})_i \\ \quad + \beta_{12}(\text{DFIXCT9})_i + \beta_{13}(\text{DPERCT10})_i + \beta_{14}(\text{IMR})_i + \varepsilon_i \end{array}\right]$$

$\hat{\sigma}$ is computed from the Nadaraya-Watson (1964) approach using errors from Model 3.6.

Appendix B1: Summary Statistics of the Pakistani Sample

The following Table B1.1 gives the summary statistics for the variables used for the estimation of the Mincerian wage model for Pakistan's labour market data. These measures are based on the final sample sued for estimation of the wage regression.

Table B1.1: Summary Statistics for the Pakistani Sample

Variable	Mean	Standard Deviation	Q1	Q3
LNWAGE	8.9337	0.5976	8.5172	9.2103
SCH1	7.3039	5.3599	2.0000	12.0000
EXP2	20.2193	11.9467	10.0000	29.0000
EXP22	551.5	569.6	100.0	841.0
HOURS3	48.1337	9.3584	44.0000	54.0000
DGENDER4	0.9306	—	—	—
DWORKLOC5	0.6726	—	—	—
DTYPDIP6	0.0205	—	—	—
DPUBLIC7	1.3701	—	—	—
DFIXCT8	0.0810	—	—	—
DPERCT9	0.3656	—	—	—
DSINDH10	0.2745	—	—	—
DKPK11	0.1693	—	—	—
DBALO12	0.1351	—	—	—
N = 19574				

Appendix B2: Models used for the Pakistani Data

The appendix B2 gives the algebraic forms of different models used in estimation of the Mincerian wage regression for the Pakistani data.

Model 4.1: Simple OLS model that do not address any possible biases

$$\begin{aligned}\ln W_i = {} & \beta_0 + \beta_1 (SCH1)_i + \beta_2 (\text{EXP2})_i + \beta_3 (\text{EXP22})_i + \beta_4 (\text{HOURS3})_i \\ & + \beta_5 (\text{DGENDER4})_i + \beta_6 (\text{DWORKLOC5})_i + \beta_7 (\text{DTYPDIP6})_i \\ & + \beta_8 (\text{DPUBLIC7})_i + \beta_9 (\text{DFIXCT8})_i + \beta_{10} (\text{DPERCT9})_i \\ & + \beta_{11} (\text{DSINDH10})_i + \beta_{12} (\text{DKPK11})_i + \beta_{13} (\text{DBALO12})_i + \varepsilon_i\end{aligned}$$

Model 4.2: IV2SLS estimation with Instrument-1 (Z1) to correct for endogeneity

First Stage : schooling equation

$$\begin{aligned}SCH1_i = {} & \beta_0 + \beta_1 (\text{EXP2})_i + \beta_2 (\text{EXP22})_i + \beta_3 (\text{HOURS3})_i + \beta_4 (\text{DGENDER4})_i \\ & + \beta_5 (\text{DWORKLOC5})_i + \beta_6 (\text{DTYPDIP6})_i + \beta_7 (\text{DPUBLIC7})_i \\ & + \beta_8 (\text{DFIXCT8})_i + \beta_9 (\text{DPERCT9})_i + \beta_{10} (\text{DSINDH10})_i \\ & + \beta_{11} (\text{DKPK11})_i + \beta_{12} (\text{DBALO12})_i + \beta_{13} (Z1) + \varepsilon_i\end{aligned}$$

Z1 is the instrumental variable for endogenous schooling.

Second Stage : wage regression

$$\begin{aligned}\ln W_i = {} & \beta_0 + \beta_1 (\hat{SCH}1_Z1)_i + \beta_2 (\text{EXP2})_i + \beta_3 (\text{EXP22})_i + \beta_4 (\text{HOURS3})_i \\ & + \beta_5 (\text{DGENDER4})_i + \beta_6 (\text{DWORKLOC5})_i + \beta_7 (\text{DTYPDIP6})_i \\ & + \beta_8 (\text{DPUBLIC7})_i + \beta_9 (\text{DFIXCT8})_i + \beta_{10} (\text{DPERCT9})_i \\ & + \beta_{11} (\text{DSINDH10})_i + \beta_{12} (\text{DKPK11})_i + \beta_{13} (\text{DBALO12})_i + \varepsilon_i\end{aligned}$$

$\hat{SCH}1_Z1$ are fitted values for endogenous schooling computed from first stage schooling regression.

Model 4.3: IV2SLS estimation with Instrument-2 (Z2) to correct for endogeneity

$$\text{First Stage : schooling equation}$$

$$\begin{aligned} SCH1_i = {} & \beta_0 + \beta_1(\text{EXP2})_i + \beta_2(\text{EXP22})_i + \beta_3(\text{HOURS3})_i + \beta_4(\text{DGENDER4})_i \\ & + \beta_5(\text{DWORKLOC5})_i + \beta_6(\text{DTYPDIP6})_i + \beta_7(\text{DPUBLIC7})_i \\ & + \beta_8(\text{DFIXCT8})_i + \beta_9(\text{DPERCT9})_i + \beta_{10}(\text{DSINDH10})_i \\ & + \beta_{11}(\text{DKPK11})_i + \beta_{12}(\text{DBALO12})_i + \beta_{13}(Z2) + \varepsilon_i \end{aligned}$$

$Z2$ is the instrumental variable for endogenous schooling.

$$\text{Second Stage : wage regression}$$

$$\begin{aligned} \ln W_i = {} & \beta_0 + \beta_1(\hat{SCH}1_Z2_i + \beta_2(\text{EXP2})_i + \beta_3(\text{EXP22})_i + \beta_4(\text{HOURS3})_i \\ & + \beta_5(\text{DGENDER4})_i + \beta_6(\text{DWORKLOC5})_i + \beta_7(\text{DTYPDIP6})_i \\ & + \beta_8(\text{DPUBLIC7})_i + \beta_9(\text{DFIXCT8})_i + \beta_{10}(\text{DPERCT9})_i \\ & + \beta_{11}(\text{DSINDH10})_i + \beta_{12}(\text{DKPK11})_i + \beta_{13}(\text{DBALO12})_i + \varepsilon_i \end{aligned}$$

$\hat{SCH}1_Z2$ are fitted values for endogenous schooling computed from first stage schooling regression.

Model 4.4: Heckman sample selection model to correct for sample selectivity

$$\begin{aligned} \ln W_i = {} & \beta_0 + \beta_1(SCH1)_i + \beta_2(\text{EXP2})_i + \beta_3(\text{EXP22})_i + \beta_4(\text{HOURS3})_i + \beta_5(\text{DGENDER4})_i \\ & + \beta_6(\text{DWORKLOC5})_i + \beta_7(\text{DTYPDIP6})_i + \beta_8(\text{DPUBLIC7})_i \\ & + \beta_9(\text{DFIXCT8})_i + \beta_{10}(\text{DPERCT9})_i + \beta_{11}(\text{DSINDH10})_i \\ & + \beta_{12}(\text{DKPK11})_i + \beta_{13}(\text{DBALO12})_i + \beta_{14}(\text{IMR})_i + \varepsilon_i \end{aligned}$$

IMR is computed from the coefficient estimates of the probit regression for the Pakistani data.

Model 4.5: Model that addresses endogeneity and sample selection bias simultaneously

$$\begin{aligned}\ln W_i = {} & \beta_0 + \beta_1(\hat{SCH}1_Z2)_i + \beta_2(\text{EXP2})_i + \beta_3(\text{EXP22})_i + \beta_4(\text{HOURS3})_i \\ & + \beta_5(\text{DGENDER4})_i + \beta_6(\text{DWORKLOC5})_i + \beta_7(\text{DTYPDIP6})_i \\ & + \beta_8(\text{DPUBLIC7})_i + \beta_9(\text{DFIXCT8})_i + \beta_{10}(\text{DPERCT9})_i \\ & + \beta_{11}(\text{DSINDH10})_i + \beta_{12}(\text{DKPK11})_i + \beta_{13}(\text{DBALO12})_i + \beta_{14}(\text{IMR})_i + \varepsilon_i\end{aligned}$$

Model 4.5W: Adaptive estimation of the Model 4.5

$$\frac{\ln W_i}{\hat{\sigma}} = \frac{1}{\hat{\sigma}}\left[\begin{aligned} & \beta_0 + \beta_1(\hat{SCH}1_Z2)_i + \beta_2(\text{EXP2})_i + \beta_3(\text{EXP22})_i + \beta_4(\text{HOURS3})_i \\ & + \beta_5(\text{DGENDER4})_i + \beta_6(\text{DWORKLOC5})_i + \beta_7(\text{DTYPDIP6})_i \\ & + \beta_8(\text{DPUBLIC7})_i + \beta_9(\text{DFIXCT8})_i + \beta_{10}(\text{DPERCT9})_i \\ & + \beta_{11}(\text{DSINDH10})_i + \beta_{12}(\text{DKPK11})_i + \beta_{13}(\text{DBALO12})_i + \beta_{14}(\text{IMR})_i + \varepsilon_i\end{aligned}\right]$$

$\hat{\sigma}$ is computed from the Nadaraya-Watson (1964) approach using errors from Model 4.5

Model 4.6: Semi-parametric estimation of the Mincerian wage model

$$\begin{aligned}\ln W_i = {} & \beta_0 + \beta_1(\hat{SCH}1_NP)_i + \beta_2(\text{EXP2})_i + \beta_3(\text{EXP22})_i + \beta_4(\text{HOURS3})_i \\ & + \beta_5(\text{DGENDER4})_i + \beta_6(\text{DWORKLOC5})_i + \beta_7(\text{DTYPDIP6})_i \\ & + \beta_8(\text{DPUBLIC7})_i + \beta_9(\text{DFIXCT8})_i + \beta_{10}(\text{DPERCT9})_i \\ & + \beta_{11}(\text{DSINDH10})_i + \beta_{12}(\text{DKPK11})_i + \beta_{13}(\text{DBALO12})_i + \beta_{14}(\text{IMR})_i + \varepsilon_i\end{aligned}$$

$\hat{SCH}1_NP$ are non-parametrically (through LOESS regression) predicted values for schooling variable.

Bibliography

Bibliography

1) Aakvik, A., Salvanes, K. G., & Vaage, K. (2003). Measuring Heterogeneity in the Returns to Education in Norway Using Educational Reforms. *IZA Discussion Paper No: 815* .

2) Abbas, Q., & Foreman-Peck, J. (2007). The Mincer Human Capital Model in Pakistan: Implications for Education Policy. *Cardiff Business School Working Paper Series. Working Paper No: E2007/24* .

3) Abdelkarim, O. B., & Skalli, A. (2005). Wage inequality in France. *Working Paper ERMES, Université Panthéon-Assas Paris II, April,2005* .

4) Agrawal, T. (2011). Returns to Education in India: Some Recent Evidence. *Indira Gandhi Institute of Development Research (IGIDR). Working Paper No. 2011-017* .

5) Ahmed , M., Aslam , M., & Pasha, G. R. (2011). Inference under Heteroscedasticity of Unknown Form Using an Adaptive Estimator. *Communications in Statistics - Theory and Methods , 40* (24), 4431-4457.

6) Ajwad, M. I., & Kurukulasuriya, P. (2002). Ethnic and Gender Wage Disparities in Sri Lanka. *World Bank Policy Research Working Paper No. 2859* .

7) Albrecht, J., Björklund, A., & Vroman, S. (2003). Is There a Glass Ceiling in Sweden? *Journal of Labor Economics , 21* (1), 145-177.

8) Albrecht, J., Vuuren, A. v., & Vroman, S. (2004). Decomposing the Gender Wage Gap in the Netherlands with Sample Selection Adjustments. *IZA Discussion Paper No. 1400* .

9) Alderman, H., Behrman, J. R., Ross, D. R., & Sabot, R. (1996). The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market. *Oxford Bulletin of Economics and Statistics , 58* (1), 29-55.

10) Al-Qudsi, S. S. (1989). Returns to Education, Sectoral Pay Differentials and Determinants in Kuwait. *Economics of Education Review , 8* (3), 263-276.

11) Al-Samarrai, S., & Reilly, B. (2005). Education, Employment and Earnings of Secondary School-Leavers in Tanzania: Evidence from a Tracer Study. *Poverty Research Unit at Sussex (PRUS).University of Sussex. Working Paper No.31* .

12) Angrist, J. D., & Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics , 106* (4).

13) Angrist, J. D., & Krueger, A. B. (1992). *Estimating The Payoff To Schooling Using The Vietnam Era Draft Lottery.* Cambridge, MA: National Bureau of Economic Research. Working Paper No. 4067.

14) Arabsheibani, G. R., & Mussurov, A. (2007). Returns to schooling in Kazakhstan OLS and instrumental variables approach. *Economics of Transition , 15* (2), 341-364.

15) Araï, M., & Skalli, A. (1996). The Strucure of Wages: A French-Swedish Comparison. *INSEE-DARES Conference Proceedings, International Comparisons of Wages, on 1-2 February, 1996* .

16) Araï, M., Ballot, G., & Skalli, A. (1996). Différentiels intersectoriels de salaire et caractéristiques des employeurs en France. *Économie et Statistique , 299* (299), 37-58.

17) Armitage, J., & Sabot, R. (1987). Socioeconomic Background and the Returns to Schooling in Two Low-Income Economies. *Economica , 54* (213), 103-108.

18) Arrazola, M., & Hevia, J. D. (2006). Gender Differentials in Returns to Education in Spain. *Education Economics , 14* (4), 469-486.

19) Arulampalam, W., Booth, A. L., & Bryan, M. L. (2007). Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. *Industrial and Labor Relations Review , 60* (2), 163-186.

20) Asadullah, M. N. (2006). Returns to Education in Bangladesh. *Education Economics , 14* (4), 453-468.

21) Ashenfelter, O., & Krueger, A. (1994). Estimates of the Economic Return to Schooling from a New Sample of Twins. *The American Economic Review , 84* (5), 1157-1173.

22) Ashenfelter, O., & Rouse, C. (1998). Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins. *The Quarterly Journal of Economics , 113* (1), 253-284.

23) Ashenfelter, O., Harmon, C., & Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship , with tests for publication bias. *Labour Economics , 6* (4), 453-470.

24) Ashraf, J., & Ashraf, B. (1993). An Analysis of the Male-Female Earnings Differential in Pakistan. *The Pakistan Development Review , 32* (4), 895-904.

25) Aslam, M. (2006). Adaptive procedures for estimation of linear regression models with known and unknown heteroscedastic errors. *Ph.D. Thesis, Bahauddin Zakariya University, Multan, Pakistan* .

26) Aslam, M. (2009). Education Gender Gaps in Pakistan: Is the Labor Market to Blame? *Economic Development and Cultural Change , 57* (4), 747-784.

27) Aslam, M., & Kingdon, G. (2009). Public–private sector segmentation in the Pakistani labour market. *Journal of Asian Economics , 20* (1), 34-49.

28) Aslam, M., Bari, F., & Kingdon, G. (2008). Returns to Schooling, Ability and Cognitive Skills in Pakistan. *Consortium on Educational Outcomes and Poverty (RECOUP).Working Paper No. 20* .

29) Aslam, M., Riaz, T., & Altaf, S. (2012). Efficient Estimation and Robust Inference of Linear Regression Models in the Presence of Heteroscedastic Errors and High Leverage Points. *Communications in Statistics- Simulation and Computation (to appear)* .

30) Assaad, R. (1997). The Effects of Public Sector Hiring and Compensation Policies on the Egyptian Labor Market. *The World Bank Economic Review , 11* (1), 85-118.

31) **B**agheri, F., & Kara, O. (2005). Rate of Return on Education and Sample Selection Bias. *Atlantic Economic Journal , 33*, 359-360.

32) Bargain, O., & Melly, B. (2007). Public Sector Pay Gap in France: New Evidence Using Panel Data. *IZA Discussion Paper No. 3427* .

33) Barnet-Verzat, C., & Wolff, F. C. (2008). Gender wage gap and the glass ceiling effect: A firm-level investigation. *International Journal of Manpower , 29* (6), 486-502.

34) Barrat, O., Chaput, H., Naboulet, A., & Wolff, L. (2007). Wage Bargaining and Compensation Practices in France : An Overview. *Actes du colloque Négociations salariales, emploi, politique économique et monétaire, http://tinyurl.com/wagebf (accessed on 03-06-2012)* .

35) Barro, R. J., & Lee, J. W. (2010). A New Data Set of Educational Attainment in the World, 1950–2010. *National Bureau of Economic Research (NBER). Working Paper No. 15902* .

36) Basmann, R. L. (1957). A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. *Econometrica , 25* (1), 77-83.

37) Becker, G. S. (1991). *A Treatise on the Family.* Cambridge, MA: Harvard University Press.

38) Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* Chicago: National Bureau of Economic Research. University of Chicago Press.

39) Behrman, J. R., & Rosenzweig, M. R. (1999). "Ability" biases in schooling returns and twins: a test and new estimates. *Economics of Education Review* (2), 159-167.

40) Behrman, J. R., Wolfe, B. L., & Blau, D. M. (1985). Human Capital and Earnings Distribution in a Developing Country: The Case of Prerevolutionary Nicaragua. *Economic Development and Cultural Change , 34* (1), 1-29.

41) Bell, D., Elliott, R. F., & Skalli, A. (1996). Wage Structure in Britain and in France : An Analysis of the Returns to Education and Age. *INSEE-DARES Conference Proceedings, International Comparisons of Wages, on 1-2 February, 1996* .

42) Belzil, C. (2006). *Testing the Specification of the Mincer Wage Equation.* Centre National de Recherche Scientifique, working paper GATE, #06-08, retrieved from ftp.gate.cnrs.fr/RePEc/2006/0608.pdf

43) Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy , 75* (4), 352-365.

44) Bhalotra, S., & Sanhueza, C. (2004). Parametric and Semi-parametric Estimations of the Return to Schooling in South Africa. *Econometric Society 2004 Latin American Meetings. Meeting No. 294.*

45) Blackaby, D., Murphy, P., & O'Leary, N. (1999). The Payment of Public Sector Workers in the UK: Reconciliation with North-American Findings. *Economics Letters , 65* (2), 239-243.

46) Blackburn, M. L., & Neumark, D. (1993). Omitted-Ability Bias and the Increase in the Return to Schooling. *Journal of Labor Economics , 11* (3), 521-544.

47) Bonjour, D., & Gerfin, M. (2001). The unequal distribution of unequal pay – An empirical analysis of the gender wage gap in Switzerland. *Empirical Economics , 26* (2), 407-427.

48) Bono, E. D., & Galindo-Rueda, F. (2004). Do a Few Months of Compulsory Schooling Matter? The Education and Labour Market Impact of School Leaving Rules. *IZA Discussion Paper No. 1233* .

49) Boumahdi, R., & Plassard, J. M. (1992). Note à propos du caractère endogène de la variable éducation dans la fonction de gains. *Revue économique , 43* (1), 145-156.

50) Bound, J., & Jaeger, D. A. (1996). *On the Validity of Season of Birth as an Instrument in Wage Equations: A Comment on Angrist & Krueger's "Does Compulsory School Attendance Affect Schooling and Earnings?* Cambridge, MA: National Bureau of Economic Research. Working Paper No.5835.

51) Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American Statistical Association , 90* (430), 443-450.

52) Brim, O. G. (1958). Family Structure and Sex Role Learning By Children: A Further Analysis of Helen Koch's Data. *Sociometry , 21*, 1-16.

53) Broekhuizen, H. V. (2011). Labour Market Returns to Educational Attainment,School Quality, and Numeracy in South Africa. *Maeter's Dissertation. Department of Economics. Stellenbosch University.South Africa* .

54) Brunello, G., & Miniaci, R. (1999). The economic returns to schooling for Italian men. An evaluation based on instrumental variables. *Labour Economics , 6* (4), 509-519.

55) Buchinsky, M. (2001). Quantile regression with sample selection: Estimating women's return to education in the U.S. *Empirical Economics , 26* (1), 87-113.

56) Buchinsky, M. (1998a). Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research. *The Journal of Human Resources , 33* (1), 88-126.

57) Buchinsky, M. (1998b). The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach. *Journal of Applied Econometrics , 13* (1), 1-30.

58) Budría, S., & Pereira, P. T. (2005). Educational Qualifications and Wage Inequality: Evidence for Europe. *IZA Discussion Paper No. 1763* .

59) Buhai, S. (2004). Quantile Regression: Overview and Selected Applications. *Tinbergen Institute and Erasmus University Rotterdam.*

60) Burger, R. (2011). Endogeneity, heterogeneity and sample selection: Estimating the shape of the schooling-earnings profile in the South African labour market. *Centre for the Study of African Economies (CSAE). 25th Anniversary Conference 2011: Economic Development in Africa. 20th - 22nd March 2011, St Catherine's College, Oxford. Paper No. 528.*

61) Butcher, K. F., & Case, A. (1994). The Effect of Sibling Sex Composition on Women's Education and Earnings. *The Quarterly Journal of Economics , 109* (3), 531-563.

62) Callan, T., & Harmon, C. (1999). The economic return to schooling in Ireland. *Labour Economics , 6* (4), 543–550.

63) Card, D. (1994). Eraning,schooling, and Ability Revisited. *National Bureau of Economic Research. Working Paper No. 4832* .

64) Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica , 69* (5), 1127-1160.

65) Card, D. (1999). The Causal Effect of Education on Earnings. *Chapter 30 in Ashenfelter and Card eds. Handbook of Labor Economics (Vol 3-A, p.1801-1863)* .

66) Card, D. (1993). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. *National Bureau of Economic Research (NBER). Working Paper No. 4833* .

67) Carroll, R. J. (1982). Adapting for Heteroscedasticity in Linear Models. *The Annals of Statistics , 10* (4), 1224-1233.

68) Carroll, R. J., & Ruppert, D. (1982). Robust Estimation in Heteroscedastic Linear Models. *The Annals of Statistics , 10* (2), 429-441.

69) Charnoz, P., Coudin, É., & Gaini, M. (2011). Wage inequalities in France 1976-2004: a quantile regression analysis. *Institut National de la Statistique et des Études Économiques (INSEE).Directorate of Economic Studies and Syntheses. Series of working papers . Working Paper No. G 2011 / 06* .

70) Chase, R. S. (1998). Markets for Communist Human Capital: Returns to Education and Experience in the Czech Republic and Slovakia. *Industrial and Labor Relations Review , 51* (3), 401-423.

71) Chen, C., & Wei, Y. (2005). Computational Issues for Quantile Regression. *Sankhyā: The Indian Journal of Statistics.QuantileRegression and Related Methods , 67* (2), 399-417.

72) Chen, G., & Hamori, S. (2009). Economic returns to schooling in urban China: OLS and the instrumental variables approach. *China Economic Review , 20* (2), 143-152.

73) Chevalier, A., & Walker, I. (1999). Further Results on the Returns to Education in the UK. *Mimeo, University of Warwick* .

74) Chiswick, B. R. (1983a). An Analysis of the Earnings and Employment of Asian-american Men. *Journal of Labor Economics , 1* (2), 197-214.

75) Chiswick, B. R. (1983b). The Earnings and Human Capital of American Jews. *The Journal of Human Resources , 18* (3), 313-336.

76) Chiswick, B. R. (2003). Jacob Mincer, Experience and the Distribution of Earnings. *IZA Dicussion Paper No. 847* .

77) Christofides, L. N., Li, Q., Liu, Z., & Min, I. (2003). Recent Two-Stage Sample Selection Procedures with an Application to the Gender Wage Gap. *Journal of Business & Economic Statistics , 21* (3), 396-405.

78) Cleveland , W. S., Devlin , S. J., & Grosse, E. (1988). Regression by local fitting: Methods, properties, and computational algorithms. *Journal of Econometrics , 37* (1), 87–114.

79) Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association , 74* (368), 829-836.

80) Coelho, D., Veszteg, R., & Soares, F. V. (2008). Quantile Regression with Sample Selection: Estimating Married Women´s Return of Education and Racial Wage Differential in Brazil. *Working Paper Retrieved at 18-03-2012 from (http://virtualbib.fgv.br/ocs/index.php/sbe/EBE08/paper/download/504/36) Its portugais version Published in Pesquisa e Planejamento Econômico (2010). Vol.40, No. 1,pp: 85-102* .

81) Conneely, K., & Uusitalo, R. (1997). *Estimating Heterogenous Treatment effects in the Becker schooling Model.* Unpublished Discussion Paper (Industrial relations Section, Princeton University).

82) Cosslett, S. (1991). *Semiparametric Estimation of a Regression Model with Sample Selectivity.* (W. A.Barnett, J. Powell, & G. Tauchen, Éds.) Cambridge, Cambridge

University Press: Nonparametric and Semiparametric Methods in Econometrics and Statistics.

83) **D**aly, A., Kawaguchi, A., Meng, X., & Mumford, K. (2006). The Gender Wage Gap in Four Countries. *Economic Development and Cultural Change , 47* (2), 165-176.

84) De la Rica, S., Dolado, J. J., & Llorens, V. (2008). Ceilings or floors? Gender wage gaps by education in Spain. *Journal of Population Economics , 21* (3), 751–776.

85) Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* The Johns Hopkins University Press.

86) Dickson, M. (2009). The Causal Effect of Education on Wages Revisited. *IZA Discussion Paper No. 4419* .

87) Dilmaghani, M. (2011). Religiosity, human capital return and earnings in Canada. *International Journal of Social Economics , 39* (1/2), 55-80.

88) Disney, R., & Gosling, A. (1998). Does it pay to work in the public sector? *Fiscal Studies , 19* (4), 347-374.

89) Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence froman Unusual Policy Experiment. *The American Economic Review , 91* (4), 795-813.

90) **E**vans, W. N., & Montgomery, E. (1994). Education and health:where there's smoke there's an instrument. *National Bureau of Economic Research. working Paper No. 4949* .

91) Ewing, B. T. (2000). The Wage Effects of Being Raised in the Catholic Religion: Does Religion Matter? *American Journal of Economics and Sociology , 59* (3), 419-432.

92) **F**alaris, E. M. (2004). A Quantile Regression Analysis of Wages in Panama. *Department of Economics. Alfred Lerner College of Business & Economics. University of Delaware. Working Paper No. 2004-01* .

93) Farooq, M., & Sulaiman, D. J. (2009). Gender Earnings Inequality and Discrimination in the Pakistani Labor Market. *The Dialogue , 4* (3), 373-385.

94) Fasih, T., Kingdon, G., Patrinos, H. A., Sakellariou, C., & Soderbom, M. (2012). Heterogeneous Returns to Education in the Labor Market. *The World Bank. Human Development Network. Education Team. Policy Research Working Paper No. 6170* .

95) Fernández, A. I., & Rodríquez-Poo, J. M. (1997). Estimation and specification testing in female labor participation models: parametric and semiparametric methods. *Econometric Reviews , 16* (2), 229-247.

96) Fersterer, J., & Winter-Ebmer, R. (2003). Smoking, discount rates, and returns to education. *Economics of Education Review , 22* (6), 561-566.

97) Filer, R. K., Jurajda, Š., & Plánovský, J. (1999). Education and wages in the Czech and Slovak Republics during transition. *Labour Economics , 6* (4), 581-593.

98) Fitzenberger, B. (1997). A guide to censored quantile regressions. *In Handbook of Statistics , 15 ,* 405–437.

99) Flabbi, L. (1999). Returns to Schooling in Italy :OLS, IV and Gender Differences. *Bocconi University. Econometrics and applied Economics Series. Working Paper No. 1*.

100) Flabbi, L., Paternostro, S., & Tiongson, E. R. (2008). Returns to education in the economic transition: A systematic assessment using comparable data. *Economics of Education Review , 27* (6), 724-740.

101) Flores-Lagunes, A., & Light, A. (2006). Measurement Error in Schooling: Evidence from Samples of Siblings and Identical Twins. *B.E. Journal of Economic Analysis & Policy , 5* (1: Article No. 14).

102) Fougère, D., Goux, D., & Maurin, É. (2001). Formation continue et carrières salariales: Une évaluation sur données individuelles. *Annales d'Économie et de Statistique , 62*, 49-69.

103) Fuchs, V. R. (1982). *Time preference and health: an exploratory study.* Economic aspects of health. Chicago: University of Chicago Press.

104) Fuller, W. A., & Rao, J. N. (1978). Estimation for a Linear Regression Model with Unknown Diagonal Covariance Matrix. *The Annals of Statistics, 6* (5), 1149-1158.

105) **G**aag, J. V., & Vijverberg, W. (1989). Wage Determinants in Côte d'Ivoire: Experience, Credentials, and Human Capital. *Economic Development and Cultural Change, 37* (2), 371-381.

106) García, J., Hernández, P. J., & López-Nicolás, A. (2001). How wide is the gap? An investigation of gender wage differences using quantile regression. *Empirical Economics, 26* (1), 149-167.

107) Gardeazabal, J., & Ugidos, A. (2005). Gender wage discrimination at quantiles. *Journal of Population Economics, 18* (1), 165-179.

108) Girma, S., & Kedir, A. (2003). Is education more beneficial to the less able? Econometric evidence from Ethiopia. *Department of Economics. University of Leicester. Discussion Paper No. 03/1* .

109) Goux, D., & Maurin, E. (1994). Éducation, expérience et salaire. *Économie & prévision , 116* (116), 155-178.

110) Greene, W. H. (2002). *Econometric Analysis (5th ed.).* Prentice Hall.

111) Griliches, Z. (1977). Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica , 45* (1), 1-22.

112) Griliches, Z. (1979). Sibling Models and Data in Economics: Beginnings of a Survey. *Journal of Political Economy , 87* (5, Part 2), S37-S64.

113) Gronau, R. (1974). Wage comparison--A Selectivity Bias. *Journal of Political Economy , 82* (6), 1119-1143.

114) Guille, M., & Skalli, A. (1999). Returns to Human Capital : A Review of the French Empirical Literature. *in Asplund, R.and P. Pereira Eds., Returns to Human Capital in Europe : A literature Review ETLA* .

115) Guillotin, Y., & Sevestre, P. (1994). Estimations de fonctions de gains sur données de panel : endogéneité du capital humain et effets de la sélection. *Économie & prévision , 116* (116), 119-135.

116) Guillotin, Y., & Tensaout, M. (2004). Meta-analysis of the return of schooling: comparing French and US results to introduce short term vs long term returns. *Retrieved from http://perso.univ-lemans.fr/~guillot/papiers/EALE-SOLE.pdf (Accessed on 19-03-2012)* .

117) Guisinger, S. E., Henderson, J. W., & Scully, G. W. (1984). Earnings, rates of return to education and the earnings distribution in Pakistan. *Economics of Education Review , 3* (4), 257-267.

118) Gujarati, D. N., & Sangeetha, S. (2007). *Basic Econometrics (4th ed.).* New Dehli: ATA McGraw-Hill Publishing Company Limited.

119) Gunderson, M. (1979). Earnings Differentials between the Public and Private Sectors. *The Canadian Journal of Economics / Revue canadienne d'Economique , 12* (2), 228-242.

120) Gurgand, M., & Maurin, E. (2007). A large scale experiment: Wages and educational expansion in France. *Paris School of Economics. Working Paper No. 2007-21* .

121) Gustafsson, S., & Jacobsson, R. (1985). Trends in Female Labor Force Participation in Sweden. *Journal of Labor Economics , 3* (1-II), S256-S274.

122) Gwartney, J. D., & Long, J. E. (1978). The Relative Earnings of Blacks and Other Minorities. *Industrial and Labor Relations Review , 31* (3), 336-346.

123) **H**anchane, S., & Moullet, S. (2000). Les rendements éducatifs privés : fondements, bilan et évaluation des nouvelles alternatives. *Revue de l'Institut d'Économie Publique. Numéro: 05–2000/1* .

124) Haque, N. U. (1977). Economic Analysis of Personal Earnings in Rawalpindi City. *The Pakistan Development Review , 16* (4), 353-382.

125) Härdle, W. (1994). Applied Nonparametric Regression. *Cambridge University Press* .

126) Harmon, C., & Walker, I. (1995). Estimates of the Return to Schooling for the United Kingdom. *The American Economic Review , 85* (5), 1278-1286.

127) Harmon, C., & Walker, I. (1999). The marginal and average returns to schooling in the UK. *European Economic Review , 43* (4–6), 879-887.

128) Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica , 46* (6), 1251-1271.

129) Hausman, J. (2001). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *The Journal of Economic Perspectives , 15* (4), 57-67.

130) Hawley, J. D. (2004). Changing returns to education in times of prosperity and crisis, Thailand 1985–1998. *Economics of Education Review , 23* (3), 273-286.

131) Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica , 47* (1), 153-161.

132) 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* (4), 475-492.

133) Heckman, J. J., & Hotz, V. J. (1986). An Investigation of the Labor Market Earnings of Panamanian Males Evaluating the Sources of Inequality. *The Journal of Human Resources , 21* (4), 507-542.

134) Heckman, J. J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics , 30* (1-2), 239–267.

135) Hertz, T. (2003). Upward Bias in the Estimated Returns to Education: Evidence from South Africa. *The American Economic Review , 93* (4), 1354-1368.

136) Hoffman, S. D., & Link, C. R. (1984). Selectivity Bias in Male Wage Equations: Black-White Comparisons. *The Review of Economics and Statistics , 66* (2), 320-324.

137) Horowitz, A., & Schenzler, C. (1999). Returns to General, Technical and Vocational Education in Developing Countries:recent evidence from Suriname. *Education Economics , 7* (1), 5-20.

138) Horowitz, J. L., & Härdle, W. (1994). Testing a Parametric Model against a Semiparametric Alternative. *Econometric Theory , 10* (5), 821-848.

139) Huber, M., & Melly, B. (2011). Quantile Regression in the Presence of Sample Selection. *School of Economics and Political Science. University of St. Gallen. Economics Working Paper No. 2011-09* .

140) Hyder, A. (2007). Wage Differentials, Rate of Return to Education, and Occupational Wage Share in the Labour Market of Pakistan. *Pakistan Institute of Development Economics (PIDE) Working Paper No. 2007:17* .

141) Hyder, A., & Reilly, B. (2005). The Public Sector Pay Gap in Pakistan: A Quantile Regression Analysis. *Poverty Research Unit at Sussex (PRUS) Working Paper No. 33* .

142) **I**chimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics , 58* (1-2), 71-120.

143) Iglesias, F. H., & Riboud, M. (1985). Trends in Labor Force Participation of Spanish Women: An Interpretive Essay. *Journal of Labor Economics , 3* (1-II), S201-S217.

144) Isacsson, G. (1999). Estimates of the return to schooling in Sweden from a large sample of twins. *Labour Economics , 6* (4), 471-489.

145) Ismail, R. (2007). The Impact of Schooling Reform on Returns to Education in Malaysia. *Munich Personal RePEc Archive (MPRA). University Library of Munich. Paper No. 15021* .

146) **J**affry, S., Ghulam, Y., & Shah, V. (2007). Returns to Education in Pakistan. *The Pakistan Development Review , 46* (4-II), 833–852.

147) Jarousse, J.-P., & Mingat, A. (1986). Un réexamen du modèle de gains de Mincer. *Revue économique , 37* (6), 999-1031.

148) Jellal, M., Nordman, C. J., & Wolff, F. C. (2008). Evidence on the glass ceiling effect in France using matched worker-firm data. *Applied Economics , 40* (24), 3233-3250.

149) Jimenez, E., & Kugler, B. (1987). The Earnings Impact of Training Duration in a Developing Country: An Ordered Probit Selection Model of Colombia's Servicio Nacional de Aprendizaje (SENA). *Journal of Human Resources , 22* (2), 228-247.

150) Johnson, E. N., & Chow, G. C. (1997). Returns to Schooling in China. *Pacific Economic Review , 2* (2), 101-113.

151) Jolliffe, D., & Campos, N. F. (2005). Does market liberalisation reduce genderdiscrimination? Econometric evidence from Hungary, 1986–1998. *Labour Economics , 12* (1), 1-22.

152) Jones, D. C., & Simon, K. I. (2005). Wage determination under plan and early transition:Bulgarian evidence using matched employer–employee data. *Journal of Comparative Economics , 33* (2), 227-243.

153) **K**ahn, L. M. (1998). Collective Bargaining and the Interindustry Wage Structure: International Evidence. *Economica, New Series , 65* (260), 507-534.

154) Kane, T. J., & Rouse, C. E. (1993). *Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?* Cambridge, MA: National Bureau of Economic Research. Working Paper No. 4268.

155) Kara, O. (2006). Occupational gender wage discrimination in Turkey. *Journal of Economic Studies , 33* (2), 130-143.

156) Kenny, L. W., Lee, L. F., Maddala, G. S., & Trost, R. P. (1979). UniversityReturns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data. *International Economic Review , 20* (3), 775-789.

157) Khan, F. S., & Toor, I. A. (2003). Changes in Returns to Education in Pakistan: 1990-2002. *The Lahore Journal of Economics , 8* (2).

158) Khan, S. R., & Irfan, M. (1985). Rates of returns to education and the determinants of earnings in Pakistan. *The Pakistan Development Review , , 24* (3-4), 671-683.

159) Khandker, S. R. (1990). Labor market participation, returns to education, and male - female wage differences in Peru. *The World Bank Policy Research Working Paper No. 461* .

160) Kim, H.-K. (2011). Does sample selection bias affect the effect of family background on the returns to schooling? Evidence from Korea. *Applied Economics , 43* (8), 963-972.

161) Kimmel, J. (1997). Rural Wages and Returns to Education:Differences Between Whites, Blacks, and American Indians. *Economics of Education Review , 16* (1), 81-96.

162) Klein, R. W., & Spady, R. H. (1993). An Efficient Semiparametric Estimator for Binary Response Models. *Econometrica , 61* (2), 387-421.

163) Koch, H. (1955). Some Personality Correlates of Sex, Sibling Position, and Sex of Sibling Among Five and Six Year Children. *Genetic Psychology Monographs , 52*, 3-50.

164) Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica , 46* (1), 33-50.

165) Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *The Journal of Economic Perspectives , 15* (4), 143-156.

166) Korsun, V. (2010). Wage Determination in Ukraine: Does Religiosity Matter? *Masters Dissertation. Kyiv School of Economics.*

167) Koutsoyiannis, A. (1977). Theory of Econometrics (2nd ed.). *London: Macmillan Education LTD* .

168) Kozel, V., & Alderman, H. (1990). Factors Determining Work Participation and Labour Supply Decisions in Pakistan's Urban Areas. *The Pakistan Development Review , 29* (1), 1-18.

169) Kuch, P., & Haessel, W. (1979). *An Analysis of Earnings in Canada.* Ottawa: Statistics Canada.

170) Kuhn, P. (1987). Sex Discrimination in Labor Markets: The Role of Statistical Evidence. *The American Economic Review , 77* (4), 567-583.

171) Lall, A., & Sakellariou, C. (2010). Evolution of Education Premiums in Cambodia: 1997–2007. *Asian Economic Journal , 24* (4), 333-354.

172) Lam, D., & Schoeni, R. F. (1993). Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil. *Journal of Political Economy , 101* (4), 710-740.

173) Lassibille, G. (1998). Wage Gaps Between the Public and Private Sectors in Spain. *Economics of Education Review , 17* (1), 83–92.

174) Lee, B.-J., & Lee, M. J. (2006). Quantile Regression Analysis of Wage Determinants in the Korean Labor Market. *The Journal of the Korean Economy , 7* (1), 1-31.

175) Lee, S. (2007). Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics , 141* (2), 1131-1158.

176) Lei, J. J. (2005). Parametric and Semiparametric Estimations of the Return to Schooling of Wage Workers in Canada. *Masters Dissertation. Department of Economics. Simon Fraser University* .

177) Leigh, A., & Ryan, C. (2008). Estimating returns to education using different natural experiment techniques. *Economics of Education Review , 27* (2), 149-160.

178) Li, Q., & Stengos, T. (1994). Adaptive Estimation in the Panel Data Error Component Model with Heteroskedasticity ofUnknown Form. *International Economic Review , 35* (4), 981-1000.

179) Lindauer, D. L., & Sabot, R. H. (1983). The public/private wage differential in a poor urban economy. *Journal of Development Economics , 12* (1-2), 137-152.

180) Liu, J.-T., Hammitt, J. K., & Lin, C. J. (2000). Family background and returns to schooling in Taiwan. *Economics of Education Review , 19* (1), 113-125.

181) Lorenz, W., & Wagner, J. (1990). A Note on the Returns to Human Capital in the Eighties: Evidence from Twelve Countries. *Luxembourg Income Study Working Paper No. 54* .

182) Lucas, R. E. (1993). Making a Miracle. *Econometrica , 61* (2), 251-272.

183) Lucifora, C., & Meurs, D. (2006). The Public Sector Pay Gap in France, Great Britain and Italy. *Review of Income and Wealth , 52* (1), 43-59.

184) **M**aluccio, J. (1998). Endogeneity of schooling in the wage function: Evidence from the rural Philippines. *Food Consumption and Nutrition Division.International Food Policy Research Institute.Working Paper No. 54* .

185) Mancuso, D. (1997). *Some Implications of Marriage and Assortative Mating by Schooling for the Earnings of Men.* Ph.D Dissertation Stanford University.

186) Mann, P. S., & Kapoor, B. L. (1988). Earnings differentials between public, private and joint sectors in Punjab (India). *Journal of Development Studies , 25* (1), 97-111.

187) Mare, R. D. (1991). Five Decades of Educational Assortative Mating. *American Sociological Review , 56* (1), 15-32.

188) Martins, M. F. (2001). Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal. *Journal of Applied Econometrics , 16* (1), 23-39.

189) Martins, P. S., & Pereira, P. T. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries. *Labour Economics , 11* (3), 355-371.

190) McLaughlin, D. K., & Perman, L. (1991). Returns vs. Endowments in the Earnings Attainment Process for Metropolitan and Nonmetropolitan Men and Women. *Rural Sociology , 56* (3), 339-365.

191) Meghir, C., & Palme, M. (2003). Ability, Parental Background And Education Policy: Empirical Evidence From A Social Experiment. *The Institute For Fiscal Studies. Working Paper No. 03/05* .

192) Melly, B. (2005). Public-private sector wage differentials in Germany: Evidence from quantile regression. *Empirical Economics , 30* (2), 505-520.

193) Meurs, D., & Ponthieux, S. (2000). Une mesure de la discrimination dans l'écart de salaire entre hommes et femmes. *Économie et Statistique , 337* (337-338), 135-158.

194) Miller, P., Mulvey, C., & Martin, N. (1995). What Do Twins Studies Reveal About the Economic Returns to Education? A Comparison ofAustralian and U.S. Findings. *The American Economic Review , 85* (3), 586-599.

195) Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy , 66* (4), 281-302.

196) Mincer, J. (1974). *Schooling, Experience and Earnings.* New York: National Bureau of Economic Research. Columbia University Press.

197) Montgomery, J. D. (1991). Social Networks and Labor-Market Outcomes: Toward an Economic Analysis. *The American Economic Review , 81* (5), 1408-1418.

198) Mueller, R. E. (1998). Public–private sector wage differentials in Canada: evidence from quantile regressions. *Economics Letters , 60* (2), 229–235.

199) Mwabu, G., & Schultz, T. P. (1996). Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa. *The American Economic Review , 86* (2), 335-339.

200) Mwabu, G., & Schultz, T. P. (1998). Wage Premia for Education and Location, By Gender and Race in South Africa. *Economic Growth Center, Yale University. Working Paper No. 785* .

201) **N**adaraya, E. (1964). On Estimating Regression. *Theory of Probability & Its Applications , 9* (1), 141-142.

202) Nasir, Z. M. (1998). Determinants of Personal Earnings in Pakistan: Findings from the Labour Force Survey 1993-94. *The Pakistan Development Review , 37* (3), 251–274.

203) Nasir, Z. M. (2000). Earnings Differential between Public and Private Sectors in Pakistan. *The Pakistan Development Review , 39* (2), 111–130.

204) Nasir, Z. M. (2002). Returns to Human Capital in Pakistan: A Gender Disaggregated Analysis. *The Pakistan Development Review , 41* (1), 1-28.

205) Nazli, H. (2004). The Effect of Education, Experience and Occupation on Earnings: Evidence from Pakistan. *The Lahore Journal of Economics , 9* (2), 1-30.

206) Newey, W. K. (1991). Two-step series estimation of sample selection models. *MIT Unpublished Manuscript (Later it is Published (2009) in The Econometrics Journal,Vol.12, Issue.1 Pp:S217–S229* .

207) Nicodemo, C. (2009). Gender Pay Gap and Quantile Regression in European Families. *IZA Discussion Papers No. 3978* .

208) Nielsen, H. S., & Rosholm, M. (2001). The public-private sector wage gap in Zambia in the 1990s: A quantile regression approach. *Empirical Economics , 26* (1), 169-182.

209) **O**lneck, M. (1977). *On the Use of Sibling Data to Estimate the Effects of Family Background, Cognitive Skills, and Schooling: Results from the Kalamazoo Borthers Study. In Kinometrics: The Determinants of Socio-economic Success Within and Between Families.* (I. P. Taubman, Éd.) Amsterdam: North Holland, Pp. 125-162.

210) Orazem, P. F., & Vodopivec, M. (1994). Winners and Loosers in Transition: Returns to Education, Experience, and Gender in Slovenia. *The World Bank Policy Research Department. Transition Economics Division.Policy Research Working Paper No.1342* .

211) Oreopoulos, P. (2003). Do Dropouts Drop Out Too Soon? International Evidence From Changes In School-Leaving Laws. *National Bureau of Economic Research (NBER). Working Paper: 10155* .

212) Ozdural, S. (1993). Intergenerational mobility: A comparative study between Turkey and the United States. *Economics Letters , 43* (2), 221–230.

213) **P**alme, M., & Wright, R. (1998). Changes in the rate of return to education in Sweden: 1968-1991. *Applied Economics , 30* (12), 1653-1663.

214) Papanicolaou, J., & Psacharopoulos, G. (1979). Socioeconomic Background, Schooling and Monetary Rewards in the United Kingdom. *Economica, New Series , 46* (184), 435-439.

215) Pastore, F., & Verashchagina, A. (2006). Private returns to human capital over transition:A case study of Belarus. *Economics of Education Review , 25* (1), 91-107.

216) Patrinos, H. A., & Sakellariou, C. N. (1992). North American Indians in the Canadian Labour Market: A Decomposition of Wage Differentials. *Economics of Education Review , 11* (3), 257-266.

217) Patrinos, H. A., Ridao-Cano, C., & Sakellariou, C. (2009). A note on schooling and wage inequality in the public and private sector. *Empirical Economics , 37* (2), 383-392.

218) Pencavel, J. (1998). Assortative Mating by Schooling and the Work Behavior of Wives and Husbands. *The American Economic Review , 88* (2), 326-329.

219) Pereira, P. T., & Martins, P. S. (2004). Returns to education and wage equations. *Applied Economics , 36* (6), 525–531.

220) Plug, E. J. (2001). Season of birth, schooling and earnings. *Journal of Economic Psychology , 22* (5), 641-660.

221) Pons, E., & Gonzalo, M. T. (2003). Returns to Schooling in Spain: How Reliable are Instrumental Variable Estimates? *LABOUR , 16* (14), 747-770.

222) Poterba, J. M., & Rueben, K. S. (1994). The distribution of public sector wage premia: new evidence using quantile regression methods. *National Bureau of Economic Research (NBER) Working Paper No. 4734* , May.

223) Powell, J. (1989). Semiparametric estimation of censored selection models. *unpublished manuscript, University of Wisconsin-Madison* .

224) Psacharopoulos, G. (1973). *Returns to Education: An International Comparison.* Amsterdam, Elsevier.

225) Psacharopoulos, G. (1977). Schooling, Experience and Earnings : The Case of an LDC. *Journal of Development Economics , 4*, 39-48.

226) Psacharopoulos, G., & Layard, R. (1979). Human Capital and Earnings: British Evidence and a Critique. *The Riview of Economic Studies , 46* (3), 485-503.

227) Psacharopoulos, G., Velez, E., & Patrinos, H. A. (1994). Education and Earnings in Paraguay. *Economic of Education Review , 13* (4), 321-327.

228) **Q**uinn, J. F. (1979). Wage Differentials among Older Workers in the Public and Private Sectors. *The Journal of Human Resources , 14* (1), 41-62.

229) Qureshi, M. G. (2012). The Gender Differences in School Enrolment and Returns to Education in Pakistan. *Pakistan Institute of Development Economics, Islamabad. Working Paper No. 84* .

230) **R**ibeiro, E. P. (1997). Conditional labor supply quantile estimates in Brazil. *Universidade Federal do Rio Grande do Sul. Discussion Paper No. 97/02* .

231) Riboud, M. (1985). An Analysis of Women's Labor Force Participation in France: Cross-Section Estimates and Time-Series Evidence. *Journal of Labor Economics , 3* (1(Part-II)), S177-S200.

232) Riboud, M. (1974). Etude Economique de l'Accumulation du Capital Humain et de son Role Productif. *Thèse Doctorale. Université de Paris 1. Panthéon-Sorbonne*.

233) Robinson, C., & Tomes, N. (1984). Union Wage Differentials in the Public and Private Sectors: A Simultaneous Equations. *Journal of Labor Economics , 2* (1), 106-127.

234) Robinson, P. M. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. *Econometrica , 55* (4), 875-891.

235) Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica , 56* (4), 931-954.

236) Rouault, D., & Kaukewitsch, P. (1998). Les structure des salaires en France et en Allemagne en 1995 : une analyse statistique comparative des hiérarchies salariales. *Économie et Statistique , 315* (315), 3-27.

237) Rouse, C. E. (1999). Further estimates of the economic return to schooling from a new sample of twins. *Economics of Education Review , 18* (2), 149–157.

238) **S**andefur, G. D., & Scott, W. J. (1983). Minority Group Status and the Wages of Indian and Black Males. *Social Science Research , 12* (1), 44-68.

239) Sanroman, G. (2006). Returns to schooling in Uruguay. *Economics Department.Social Sciences Faculty,University of the Republic, Montevideo, Uruguay. Working Paper No. 14/06* .

240) Sargan, J. D. (1964). Wages and Prices in the United Kingdom: A Study in Econometric Methodology. *In P. E. Hart, G. Mills, and J. K. Whitaker (Eds.) Econometric Analysis for National Economic Planning, London, 1964* .

241) Schafgans, M. M. (1998). Ethnic Wage Differences in Malaysia: Parametric and Semiparametric Estimation of the Chinese-Malay Wage Gap. *Journal of Applied Econometrics , 13* (5), 481-504.

242) Schafgans, M. M. (2000). Gender wage differences in Malaysia: parametric and semiparametric estimation. *Journal of Development Economics , 63* (2), 351–378.

243) Schultz, T. P. (1988). Education investments and returns. *Handbook of Development Economics (Vol. 1, p. 543-630). Elsevier* .

244) Schultz, T. P. (1993). Investments in the Schooling and Health of Women and Men: Quantities and Returns. *The Journal of Human Resources , 28* (4), 694-734.

245) Selz, M., & Thélot, C. (2004). The Returns to Education and Experience: Trends in France over the Last Thirty-Five Years. *Population , 59* (1), 9-47.

246) Shabbir, T. (1994). Mincerian Earnings Function for Pakistan. *The Pakistan Development Review , 33* (1), 1-18.

247) Shapiro, D. M., & Stelcner, M. (1989). Canadian Public-Private Sector Earnings Differentials, 1970–1980. *Industrial Relations: A Journal of Economy and Society , 28* (1), 72–81.

248) Siddiqui, R., & Siddiqui, R. (1998). A Decomposition of Male-Female Earnings Differentials. *The Pakistan Development Review , 37* (4-II), 885–898.

249) Simonnet, V. (1996). Mobilité professionnelle et salaire : des différences entre hommes et femmes. *Économie et Statistique , 299* (299), 59-71 .

250) Simpson, W. (1985). The Impact of Unions on the Structure of Canadian Wages: An Empirical Analysis with Microdata. *The Canadian Journal of Economics , 18* (1), 164-181.

251) Skalli, A. (2007). Are successive investments in education equally worthwhile? Endogenous schooling decisions and non-linearities in the earnings–schooling relationship. *Economics of Education Review , 26* (2), 215-231.

252) Smith, S. P. (1976). Pay Differentials between Federal Government and Private Sector Workers. *Industrial and Labor Relations Review , 31* (3), 179-197.

253) Snipp, C. M., & Sandefur, G. D. (1988). Earnings of American Indians and Alaskan Natives: The Effects of Residence and Migration. *Social Forces , 66* (4), 994-1008.

254) Sofer, C. (1990). La répartition des emplois par sexe : capital humain ou discrimination. *Économie & Prévision , 92* (92-93), 77-85.

255) Staneva, A., Arabsheibani, G. R., & Murphy, P. (2010). Returns to Education in Four Transition Countries: Quantile Regression Approach. *IZA Discussion Paper No. 5210* .

256) Steunou, P.-Y. (2003). Capital Humain et Hétérogénéité : Théorie et Application. *Doctoral Thesis. U.F.R de Droit et Sciences Économiques. Université du Maine* .

257) Svejnar, J. (1999). Labor markets in the transitional Central and East European economies. *Handbook of Labor Economics (Vol. 3-B, p. 2809–2857)* .

258) Sviatova, G., Kuandykov, E., & Chulkina, M. (1988). Demographic genetic characteristics of the population of a large multiethnic city (the example of Alma-Ata). An analysis of assortative mating and migrations. *Genetika , 24* (7), 1269-1275.

259) Tahar, G., & Plassard, J.-M. (1990). Théorie du salaire d'efficience et disparités non compensatrices : évaluation à partir de l'enquête Fqp. *Économie & Prévision , 92* (92-93), 67-76.

260) Tansel, A. (1994). Wage Employment, Earnings and Returns to Schooling for Men and Women in Turkey. *Economics of Education Review , 13* (4), 305-320.

261) Taubman, P. (1976). The Determinants of Earnings: Genetics, Family, and Other Environments: A Study of White Male Twins. *The American Economic Review , 66* (5), 858-870.

262) Terrell, K. (1993). Public-private wage differentials in Haiti Do public servants earn a rent? *Journal of Development Economics , 42* (2), 293-314.

263) Theil, H. (1953). Repeated least-squares applied to a complete equation systems. *Mimeo. The Hague: Central Planning Bureau.*

264) Tomes, N. (1983). Religion and the Rate of Return on Human Capital: Evidence from Canada. *The Canadian Journal of Economics / Revue canadienne d'Economique , 16* (1), 122-138.

265) Trostel, P., Walker, I., & Woolley, P. (2002). Estimates of the economic return to schooling for 28 countries. *Labour Economics , 9* (1), 1-16.

266) **U**usitalo, R. (1999). Return to education in Finland. *Labour Economics , 6* (4), 569–580.

267) **V**ella, F. (1998). Estimating Models with Sample Selection Bias: A Survey. *The Journal of Human Resources , 33* (1), 127-169.

268) Viger, E. (2007). Les effets de la démocratisation de l'enseignement en France: Une étude empirique. *Doctoral Thesis. U.F.R de Sciences Économiques. Université Paris 1 Panthéon – Sorbonne.*

269) **W**ahba, J. (2000). Returns to Education and Regional Earnings Differentials in Egypt. *Department of Economics. University of Southampton. Discussion Papers in Economics and Econometrics* .

270) Walsh, J. R. (1935). Capital Concept Applied to Man. *The quarterly Journal of Economics , 49* (2), 255-285.

271) Warunsir, S., & Mcnown, R. (2010). The Returns to Education in Thailand: A Pseudo-Panel Approach. *World Development , 38* (11), 1616-1625.

272) Watson, G. S. (1964). Smooth Regression Analysis. *Sankhy: The Indian Journal of Statistics, Series A , 26* (4), 359-372.

273) Webbink, D., & Wassenberg, J. V. (2004). Born on the First of October: Estimating the Returns to Education Using a School Entry Rule. *Mimeo, University of Amsterdam* .

274) Weisberg, J. (1995). Returns to Education in Israel: 1974 and 1983. *Economics of Education Review , 14* (2), 145-154.

275) Weiss, Y. (1999). The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens upon Marriage and Divorce? *In Handbook of Population and Family Economics, ed. M. R. Rosenzweig and O. Stark. Amsterdam: North-Holland* .

276) White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica , 48* (4), 817-838.

277) Willis, R. (1986). *"Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions", in Orley Ashenfelter and Richard Layard, eds., the Hand Book of Labor Economics* (Vol. 1). Amsterdam: North Holland Elsevier Science Publishers.

278) **Y**asin, G., Chaudhry, I. S., & Afzal, S. (2010). The Determinants of Gender Wage Discrimination in Pakistan: Econometric Evidence from Punjab Province. *Asian Social Science , 6* (11), 239-255.

279) **Z**hang, J., Zhao, Y., Park, A., & Song, X. (2005). Economic returns to schooling in urban China, 1988 to 2001. *Journal of Comparative Economics , 33* (4), 730–752.

280) Zhang, X. (2011). The Rate of Returns to Schooling:A case study of Urban China. *International Journal of Humanities and Social Science , 1* (18), 173-180.