OXFORD UNIVERSITY

New Approaches to Understanding Income Differences and Current Account Imbalances

SUBMITTED TO THE DEPARTMENT OF ECONOMICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

in

Economics

By

Swarnali Ahmed

ST. ANTONY'S COLLEGE, OXFORD

September 2012

[1] Words count: 286 words per page (using page 7 as a representative) ×205 pages = 58630 words.

New Approaches to Understanding Income Differences and Current Account Imbalances

Swarnali Ahmed
Doctor of Philosophy
Michaelmas, 2012
St. Antony's College, Oxford University

ABSTRACT

This thesis employs two new approaches to explain some of the important debates in two key economic fields: labour market economics and macroeconomic studies related to current account imbalances.

Chapter 1, **Chapter 2** and **Chapter 3** begin a new strand of research by introducing the normal inverse Gaussian (NIG) distribution to describe unobserved heterogeneity in the labour market. The NIG distribution can be represented as a normal variance-mean mixture with the inverse Gaussian (IG) distribution as the mixing distribution. A 0.01% subsample of the 1980 US Census, comprising all men between 18 and 65 who are in the labour force, as well as a comparable sample from Ghana, is used to show that the NIG distribution provides a better fit of the log earnings function than the normal distribution. The prediction of right skewness of the log earnings distribution arising from the log normal skill Roy selection model is rejected in favour of left skewness. The thesis then extends the model to describe the distribution of log earnings conditioned on education. The same two datasets (US males and Ghanaian males) are used for the empirical analysis. We find that, once the unobserved heterogeneity is accounted for, the return to education is almost flat for lower levels of education in Ghana, and then

increases for education levels greater than ten years. One of the key differences between the two datasets is that skewness and unobserved heterogeneity is a function of education for Ghana but not for the US. The NIG framework is found to be a useful tool to model this heterogeneity.

Chapter 4 uses a model that allows for a rich structure of age effects similar to those predicted by the life cycle theories to argue that the demographic shifts are partly responsible for the sustained rise in the US current account deficit and the rapid increase in China's current account surplus in the last decade. However, demographics do not have an impact on the long run equilibrium or level of current accounts. Rather, they are important determinants of the short run adjustment of current accounts to their equilibrium levels. In the next twenty years, the demographic shifts are likely to push towards further current account positive adjustments in China and current account negative adjustments in the US. Developing the infrastructure, financial markets, policy tools and regulatory settings to be able to cope with the excess capital flow remains an urgent task.

Acknowledgments

I am grateful to my supervisors Bent Nielsen and Francis Teal. They have been pillars of support throughout the process. It has been an amazing learning experience to work under their guidance. I thank them for their patience and time. They have helped me understand the economic concepts, build better models, and organize my ideas in an articulate fashion. Bent Nielsen has been instrumental in shaping my economic, econometric and writing skills. Francis Teal has taught me both the foundations and the nuances of the labour market, and has very patiently helped me improve my ability to communicate through my writing. I am humbled by my association with them.

I would like to thank Dominic Wilson for introducing me to the world of current account imbalances and demographics. It is a fascinating field and I have learnt a lot from him during our research report on this topic. I had presented one of the earlier versions of the first three chapters in Centre for the Study of African Economies (CSAE) seminar. I am thankful to the participants of the seminar for their valuable insights and suggestions. I had also presented my research report (with Dominic Wilson) in a Research Brown Bag Seminar at International

Monetary Fund (IMF). I thank the participants for their comments which helped me better understand the econometric and conceptual complexities in modelling current accounts.

Contents

List of Tables

List of Figures

An Overview of the Thesis

Two Empirical Questions

This thesis studies two important issues in economics, one using micro data and the second using macro. The first involves the determinants of the distribution of incomes, the second the determination of the current account imbalances of the US and China. The contribution of the thesis is to focus on aspects of how the data may be generated in both of these cases, which previous work in these areas has largely ignored. In the case of the incomes we use a novel statistical method to analyse aspects of the conditional and unconditional distribution of incomes, which allows a direct modelling of how unobservables may impact on these distributions. In the macro modelling we show the importance of dynamics in understanding how the role of demographics may differ in determining adjustment to long run equilibrium for the current account and the determination of the equilibrium path.

The Distribution of Earnings

In estimating an earnings function the key empirical problem relates to how the unobserved differences across individuals influence the differences in income. The observed heterogeneity in the population could be due to the unobserved differences in ability, unobserved differences in social and family circumstances, upbringing, etc., leading to two individuals with the same education level and the same observed attributes earning different income. What almost all work to date in this area has done is to seek an instrument that will allow a causal inference to be made of the effect of education on earnings. In contrast in this thesis a statistical approach is adopted which provides a means of directly modelling these unobservable factors. It will be shown that such an approach can resolve some of the problems extant in the literature with respect to the Roy model and its predictions of how selection will create skewness in the distribution of incomes and insights into possible reasons why IV estimation may be difficult to use for estimating the return to education.

To understand the impact on income due to unobserved heterogeneity in the labour market, we use a distribution called the normal inverse Gaussian (NIG) distribution which allows us to directly study the impact of the unobserved differences on income. The conventional approach is to model log earnings using the normal distribution and to treat the unobserved differences as residuals. That this procedure may be problematic is apparent from the widely known empirical fact that earnings are not log normally distributed.

One of the motivations of using the NIG distribution to model log earnings is to exploit what is known as the "mixing property" of the distribution. Usually, if we mix two normal distributions, we end up with a distribution that is also normal. However, the NIG distribution is the product of mixing two different types of distribution, one of which is the normal distribution and the other the inverse Gaussian distribution.

The key idea which underlies the value of this distribution for the purposes of modelling the determinants of income is that if we are willing to make the assumption that the variance σ^2 is inverse Gaussian distributed then we can model the log of earnings, w, as the following distribution

$$(w \mid \sigma^2) \stackrel{\mathsf{D}}{=} \mathsf{N}(\mu + \beta\sigma^2, \sigma^2). \tag{1}$$

This inverse Gaussian distribution is a function of two parameters, δ and γ

$$\sigma^2 \stackrel{\mathsf{D}}{=} \mathsf{IG}(\delta, \gamma).$$

where δ is a scale parameter and $\gamma > 0$ along with the transformed parameter $\alpha^2 = \beta^2 + \gamma^2$ controls the steepness of the distribution. Full technical details are given in the chapters but here we focus on how this distribution offers us the opportunity to extend any analysis which is based on assuming the normal distribution of earnings.

By adopting this more general framework we can advance work in this area in

several ways. First, we can begin by asking if the distribution of earnings is better described by the NIG than by the normal distribution. Second, we can allow the data to tell us the degree of skewness in the distribution, offering a means to directly test the Roy selection model, which generates an aggregate distribution of earnings from two normal distributions. Third, we can identify the variance of the unobservables σ^2 and test how far their variance differs from the normal distribution. We find that the data decisively rejects the normal distribution, that the parameters imply left skewness of earnings contrary to the predictions of the Roy model and that the variance of σ^2 is right skewed.

We can then go further and ask about the distribution of earnings conditional on education (S), again both testing and relaxing the assumption of normality by means of the NIG distribution. We will focus on two conditional distributions of log earnings, w.

$$\mathsf{E}(w|S,\sigma^2) = M = \mu_0 + \mu_1 S + \beta\sigma^2(S). \tag{2}$$

$$\mathsf{E}(w|S) = \mu_0 + \mu_1 S + \frac{\beta\delta}{\gamma}. \tag{3}$$

The return to education will be μ_1 either if we condition on σ^2 or if the parameters of the IG are not functions of education which leads to our final innovation. We test if δ and γ are themselves functions of education. If they are, then the returns to education, if we do not condition on σ^2, will not be μ_1 but

$$\mu_1 + \frac{\partial}{\partial S}\left(\frac{\beta\delta}{\gamma}\right)$$

which we can derive given we have been explicit about the form of the distribution.

In summary the approach we adopt in the first three chapters will seek to develop a method of analysing the data which moves from an assumption of normality to a distribution sufficiently general to capture the degree of skewness in the distribution of earnings either unconditionally or conditioned on education. As we model the variance by the inverse Gaussian we can go further and discuss the distribution of this variable, which has an interpretation as the distribution of all the unobserved aspects of the determinants of earnings.

The rationale offered for the log normal distribution is that it will arise from any variable whose growth is the subject of independent and identically distributed disturbances. The NIG distribution relaxes this assumption as it can be interpreted as arising from disturbances which are independently but not identically distributed. We will argue that the NIG does a much better job of describing both the distribution of earnings and its distribution conditioned on education than the normal distribution. The "picture" that emerges from the data is one where the degree of heterogeneity in the determinants of earnings cannot be captured by seeking to separate out these unobservables from the observables.

Armed with this relatively new and unconventional technique of using a different distribution for modelling income, we compare how population heterogeneity across emerging markets and developed economies influences their incomes. We use the US as an example of an advanced economy and the Ghanaian labour market as an example of a developing economy. The economies differ radically in the structure of their labour markets, offering us the opportunity to assess how well

the NIG distribution can describe their labour market outcomes. In particular for the US data we confine ourselves to male earnings, thus eliminating the selection issues associated with female earning functions. In contrast our Ghana data is for wage earnings in an economy where wage employment is a highly selected outcome. We anticipate that the role of unobservables in the earnings functions for Ghana will be much more heterogeneous than for the US. The novel modelling procedure used in this part of the thesis allows for the possibility that education will influence the variance of the distribution, after we condition on education, thus ensuring a breakdown of the separation of an educational effect from the residuals which is the basis for all the IV work in this area.

Current Account Imbalances

A central issue in the determination of the current account is the role of the demographic structure. As the current account imbalance reflects differences in aggregate saving and investment in an economy, then insofar as the demographic structure of the population affects the savings rate and investment rate it will have an impact on the current account. For our second question about understanding current account imbalances we move beyond conventional approaches in three ways. First, we explore new ways of measuring the age structure. Second, we use an error correction specification of the model which allows us to distinguish between the effects of demographics on adjustment to long run equilibrium and as a determinant of equilibrium itself. Third, we set the model in a general framework

that allows a more systematic testing of the exogeneity assumptions that need to be fulfilled for valid inference from demographics to the current account.

The most famous insight into the linkage between current accounts and demographics comes from the life cycle hypothesis (LCH), which shows how the demographic structure of an economy may influence the general desire to save or invest across an entire economy. The LCH posits that savings and investment patterns are different at different points in a person's life. Early on in life, when incomes are low, people are more likely to borrow and invest (in themselves or their children). They then move through a period of 'prime savings' towards middle age to accumulate assets, and then to gradual dissavings as those assets are spent in retirement. With savings and investment schedules both likely to depend on age structure, the current account balances – the difference between savings and investment – may also shift along with demographics in particular ways. Most empirical studies have used the LCH to show how the population distribution affects the level of current accounts. However, there is also the possibility that this effect does not happen in the levels, but rather that changes in population distribution would affect the way current accounts adjust. Our model allows us to test for both the level and adjustment effects of population distribution on current accounts.

We use a model that allows for a rich structure of age effects similar to those predicted by the life cycle theories to test whether the demographic shifts are partly responsible for the sustained rise in the US current account deficit and the rapid increase in China's current account surplus in the last decade. In our model,

we feed in the demographics data as 15 age groups, 0–4, 5–9, ..., 70+. The initial regression specification can be thought of along the lines of:

$$CA_t = \beta' X_t + \alpha_1 p_{1t} + \alpha_2 p_{2t} + ... + \alpha_n p_{nt} + u_t \tag{4}$$

where CA_t is the current account expressed as a share of GDP, X_t is the vector of standard macroeconomics explanatory variables and the $p_{1t}, ..., p_{nt}$ shares of the population of n age groups, 15 in our case, across time t.

Due to the high correlation between the shares of consecutive age groups, a regression that includes all 15 age groups will encounter multicollinearity. The variation of the ordinary least squares (OLS) estimates will be large, making it difficult to identify the effect of any particular group. We therefore express the information of the entire age structure using three variables D_{1t}, D_{2t} and D_{3t} which are complicated geometric averages of the population age shares. They are constructed using a low-order polynomial to represent the fifteen population age shares. We thus transform equation 4 into:

$$CA_t = \beta' X_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + u_t \tag{5}$$

and then estimate β', γ_1, γ_2 and γ_3. In the fourth chapter, we give the motivation and derivation of this transformation in details.

In terms of econometric methodology, we use the autoregressive distributed lag (ARDL) model and choose the formulation that is often known as the error

correction model. In addition, we consider the ARDL model in the context of an economic system, the idea being that all the components of the economy should be modelled to obtain more accurate and comprehensive picture. We then conduct the necessary exogeneity tests that allow us to reduce from the multi-equation system to one linear equation with the current account as the independent variable. In the ARDL form, the change in current accounts is presented by two components: First, we have an equilibrium relationship for the change in current accounts. Second, we have an equilibrium error which accounts for the deviation of the pair of variables from the long run equilibrium. Thus, the ARDL framework gives the opportunity to test for both the long run and short effects of demographics on the current accounts. The first component captures the short term adjustment effect and the latter component captures the long run level effect. If we start off representing the change in current accounts

$$\Delta CA_t = c + \alpha CA_{t-1} + \beta' X_{t-1} + \gamma' \Delta X_t + \delta' Z_{t-1} + \varphi' \Delta Z_t + \varepsilon_t \tag{6}$$

where $Z = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix}$, we can re-write the model to represent in the long run equilibrium ECM form. The ECM formulation distinctly shows the two components of the model

$$\Delta CA_t = c + \alpha(CA_{t-1} + \frac{\beta'}{\alpha} X_{t-1} + \frac{\delta'}{\alpha} Z_{t-1}) + \gamma' \Delta X_t + \varphi' \Delta Z_t + \varepsilon_t \tag{7}$$

where the long run equilibrium, ECM_t, is given by:

$$ECM_t = CA_t + \frac{\beta'}{\alpha}X_t + \frac{\delta'}{\alpha}Z_t = 0 \tag{8}$$

The model states that the change in CA_t from the previous period consists of the change associated with movement with X_t and Z_t along the long run equilibrium path plus a part α of the deviation $CA_{t-1} + \frac{\beta'}{\alpha}X_{t-1} + \frac{\delta'}{\alpha}Z_{t-1}$ from the equilibrium. This is explained in detail in Chapter 4.

Our econometric analysis of the US and China confirms that the current account imbalances since the early 1990s in these countries can be partly explained in light of the shifting demographics of these economies. However, contrary to previous studies, our results indicate that demographics do not determine the long run equilibrium path of current accounts. Rather, demographic shifts play a role in determining the short run adjustment of the current accounts to their long run equilibrium. In other words, demographics do not have a level effect but an adjustment effect on current accounts. Though our results indicate that the role of demographics is different to what is envisaged in the literature, our findings for China are in line with the findings of existing studies. Up to the age of 35, the increase in population appears to cause the current accounts to decrease. Between ages 35 and 69, the increase in people causes the current accounts to increase. Above 69, the increase in population again tends to become a drag on the current account. However, in the US, we find evidence that the traditional age group whose incremental increase should be creating a positive adjustment

pressure on current accounts is actually creating a negative adjustment pressure on the current accounts.

Delving deeper by using a larger dataset for the US, we find that this is a persistent feature of the economy. However, we find a structural break in the response of demographics in the late 1980s, the start of the period of financial liberalization. It appears that households used their increased wealth from the stock market boom on imported goods, slightly exacerbating their response towards current account adjustments. We also run a scenario analysis using our model estimates and UN population projections. Our results show that the demographic pressure for current account negative adjustments and current account positive adjustments will continue in the US and China, respectively. This calls for developing the infrastructure, financial markets, policy tools and regulatory settings to cope with the excess capital flow.

The thesis is arranged as follows: Chapter 1 studies the issues in modelling income distribution and gives the motivation behind using a new distribution to model income. Chapter 2 introduces the proposed distribution and uses the proposed distribution to model the marginal income distribution of comparable datasets from the US and Ghana. Chapter 3 extends the analysis to model conditional (on education) distribution of income. Chapter 4 explores the role of population heterogeneity in explaining current account imbalances using datasets from the US and China.

Chapter 1

Issues in Modelling Income Distribution

Two broad issues have been prominent in research in labour economics in the past century. First, modelling the (marginal/unconditional) distribution of earnings. Second, modelling the returns to education, where the objective is to model earnings conditional on education. Log normality has been central to both research agendas. In this thesis we introduce a distribution that allows us to address the limitations of using the normal distribution.

In the former strand of research, Roy's (1951) selection model has been by far the most influential in the labour economic literature. In its two sector version, its prediction is, if selection operates, the distribution of earnings will be right skewed if the underlying distribution of skills in the two sectors is log normal. However, it has been noted that this is not what we observe in the data. Heckman and Sadlacek (1990) show that the log normal skill assumption can be seriously

questioned and is not applicable in most cases.

In the past fifteen years, an important focus in labour economics has been on addressing the second issue, that is, modelling earnings conditional on education. The coefficient on schooling in a regression of log earnings on years of schooling is often called a rate of return. This notion was first introduced as a central concept of human capital theory by Becker (1964), and later estimated and popularized by Mincer (1974). In this literature, Roy's insight that individuals will choose the best of their available options (or comparative advantage) implies that we have two problems, well summarized in Heckman *et al.* (2009). One is that the return to education may vary by individual. The second is that if some aspect of the individual is correlated with education but unobserved (ability for short), the OLS estimation of the earnings functions will give a biased coefficient on education. Traditionally, the literature has attempted to address this issue by using instrumental variables (IV), see Card (2001) for an overview. This second strand of research has paid almost no attention to the distribution of the earnings conditional on education.

Most studies, including Roy's selection model, have made inferences about the unobservables to describe the observed earnings distribution. If we are interested in how income is distributed when income is a function of unobserved ability, one option is to choose a distribution that directly allows both the mean and higher moments to be a function of the unobservables. In this thesis, we relax the assumption of log normality and introduce a more general distribution that allows us to do exactly this, that is, express mean and higher moments of log earnings

as a function of unobservables. Unlike the assumption of constant variance across individuals implied by the normal distribution, our more flexible distribution accounts for the differences in variance across individuals, and allows the mean of log earnings to be a function of these differences in individual variance. We can term these differences in variance across individuals as the unobserved heterogeneity across individuals. In other words, the unobserved heterogeneity represents the set of unobservables that drive the distribution of the earnings function.

Similarly, in addressing our second question - how is income distributed conditional on schooling - one option is to choose a distribution that is general enough to ensure that both mean and higher moments can be a function of the unobservables when income is conditioned on education. Our proposed distribution allows for the possibility that the residual (error term) of a regression of log earnings on education could be a function of education. If higher moments of the distribution are themselves functions of education, then this model provides new insights into the problems posed in using IV to find the "causal" effect of education. We will investigate the implications of the IV procedure when the dimensions of the distribution are themselves functions of education. In particular, we will see if the assumptions necessary for the IV procedure to work still hold.

Using this new approach, we use both a standard US dataset and one collated from household surveys in Ghana to investigate whether differences can be found across both high and low income labour markets:

First, we ask if the unconditional distribution of log earnings deviates from normality and whether that deviation takes the form implied by a Roy selec-

tion model. We show that both our datasets decisively reject normality and our proposed distribution provides a better fit of the log earnings function than the normal distribution. More importantly, the prediction of right skewness of the log earnings distribution arising from the log normal skill Roy selection model is strongly rejected in favour of left skewness.

Second, extending the analysis to model the log earnings distribution conditioned on education, we find that the unobserved heterogeneity present in the error from a regression of log earnings on education, using our proposed distribution, is not a function of education for the US economy. However, in Ghana, we find that the unobserved heterogeneity in the residual from the earnings function is inseparable from education, implying that the usual attempts to separate education from the unobserved heterogeneity, particularly unobserved ability, may not always give an accurate picture. Also, exploiting some of the analytical properties of our proposed distribution, we find that there is possibility of self-selection in Ghana.

1.1 The Roy Selection Model

The finding that log earnings distribution is not normal is not surprising since earlier empirical studies of income established that log earnings functions are not normally distributed (Neil and Rosen (2000)). This prompted researchers to address the fat tail of log earnings function by using other distributions, particularly Pareto and Champernovne distributions (Staele 1943, Miller 1955, Lebergott

1959, Harrison 1981). At the same time, another strand of research developed, pioneered by Roy (1951), that explained the non-normality in earnings distribution as a consequence of the selection of individuals into different sectors, based upon their observed and unobserved skills. These studies made inferences about the unobservables to explain the observed earnings data.

One of the widely accepted explanations of the long tail problem is given by Roy's selection model. This model explains occupational choice and its consequence for earnings when individuals differ in their endowments of occupation-specific skill. We will focus on the exposition of the Roy model of Heckman and Honoré (1990).

Heckman and Honoré (1990) consider a two-skill Roy model. Agents possess two skills associated with two sectors. The agents know their skills. For each sector they compute the potential log earnings for sector j as

$$w_j \stackrel{def}{=} \log W_j = \log \pi_j + \mu_j + U_j, \tag{1.1}$$

where π_j is a skill price common for all agents, μ_j is the expected skill, and U_j is the demeaned individual-specific skill level. Equation 1.1 is known to the agent. The agent then maximizes over the two potential log earnings and earns

$$w = \max(w_1, w_2). \tag{1.2}$$

In other words the agent has a choice of two sectors associated with two skills.

The agent chooses the sector in which the combination of the common skill prices, $\log \pi_j$, and the individual skill, $\mu_j + U_j$, is the highest and receives the log earning w.

Heckman and Honoré then proceed to analyse the aggregate distribution of log earnings w in population when the individual skills U_1, U_2 and parameters $\pi_1, \pi_2, \mu_1, \mu_2$ are unobserved by the econometrician. They transform the skills U_1, U_2 into two variables

$$D = U_1 - U_2, \qquad V = a_1 U_2 - a_2 U_1$$

where D is the difference between the two skills and V is a linear combination of the skills which is uncorrelated with D for appropriately chosen a_1 and a_2. The variables D, V will have zero mean while the coefficients a_1 and a_2 are chosen as:

$$\begin{aligned} a_1 &= \frac{\sigma_{11} - \sigma_{12}}{\sigma_D^2} = \frac{1 - \sigma_{12}/\sigma_{11}}{\sigma_D^2}, \\ a_2 &= 1 - a_1 = \frac{-\sigma_{22} + \sigma_{12}}{\sigma_D^2}, \end{aligned}$$

where σ_{11} is the variance of U_1, σ_{22} is the variance of U_2, σ_{12} is the covariance of U_1 and U_2, while σ_D^2 is the variance of D

$$\sigma_D^2 = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$$

In other words the sign a_1 is positive if the population regression coefficient of U_2 on U_1 is smaller than unity. The potential log earnings for sector j are then

written as

$$w_j = \log \pi_j + \mu_j + a_j D + V.$$

The actual individual earning, found by maximising over the potential sectoral earnings through $w = \max(w_1, w_2)$, see equation 1.2, can be decomposed into

$$w = V + M, \tag{1.3}$$

where V is zero mean variable while the mean of aggregate log earning w is captured by

$$M = \max(\log \pi_1 + \mu_1 + a_1 D, \log \pi_2 + \mu_2 + a_2 D). \tag{1.4}$$

In order to proceed with the analysis Heckman and Honoré make three levels of assumptions.

First, the variables D and V are uncorrelated by construction. This is not sufficient to analyze the variance of w because M is a non-linear function of D. If it is assumed that D and V are independent then M and V will be independent and hence uncorrelated. With this assumption they show that in sector $j = 1$ the skewness of the potential earnings is determined by the sign of the coefficient a_1.

Secondly, assuming further that the densities of D and V are jointly log concave their Theorem 1 shows that the distribution function of log earnings, w, is log concave. The normal distribution would be an example of log concave distribution for D and V.

Thirdly, assuming also that the skills D and V are jointly normal their Theorem 3 shows that the distribution of log earnings, w, is skewed to the right independently of the sign of the sectoral skewness which is given by the coefficients a_1 and a_2.

The key insight of the Roy model is that agents' decision between two sectors hinges on their comparative advantage in the unobserved skills D and V. The selection between sectors implies that aggregate log earnings are non-normal. The assumption that D and V are independent implies that M and V are independent in equation 1.3. Therefore the conditional variance of w given D does not depend on D, that is

$$\mathsf{Var}(w|D) = \mathsf{Var}(V + M|D) = \mathsf{Var}(V|D) = \mathsf{Var}(V)$$

so w becomes homoskedastic in D. A consequence is that when analyzing the two skill model it suffices to analyze the variable M which is a maximum over the one-dimensional variable D.

Intuitively, Roy's selection model implies that workers self-select the sector that gives them the highest expected earnings. This self-selection of agents between two sectors hinges on their comparative advantage in the unobserved skills. The economic interpretation of the model is best understood by the following

expression derived by Heckman and Honoré

$$\begin{aligned} \mathsf{E}(\log w_1 | \log w_1 &> \log w_2) \\ &= \log \pi_1 + \mu_1 + \mathsf{E}(U_1 | D > c) \\ &= \log \pi_1 + \mu_1 + a_1 \mathsf{E}(D | D > c) \end{aligned}$$

where $c = \log(\pi_1/\pi_2) + \mu_1 - \mu_2$.

This is an expression of what we can observe, namely the distribution of log earnings in sector 1. We may or may not be able to observe sector 2 depending on the context. To make inferences from the unobserved skill distribution to the observed earnings we need assumptions about the sign of a_1 and the term $\mathsf{E}(D|D > c)$. The equation shows us two potential sources of comparative advantage, the first is the high variance in sector 1 and the second is a negative covariance between the skills. If the variance in sector 1 is high, then there is a high probability that sorting will enable individuals to increase their incomes. If σ_{12} is negative $\mathsf{E}(D|D > c) \geq \mathsf{E}(D) = 0$. The only ground on which sorting will not occur is if σ_{12} is positive and sufficiently large relative to the variance in sector 1. In other words, absolute advantage has to be very high in that individuals are good in both sectors such that there is no gain from sorting. The key assumption of log normal skills allows Heckman and Honoré to make a further analysis of the distribution of the aggregate log earnings showing that this is right skewed as predicted by Roy. This is irrespective of the sign of skewness of log earnings in individual sectors. However, the empirical data do not show the right skewness predicted by Roy's

selection model Heckman and Sadlacek (1990). In this thesis, we also show that Roy's prediction is not supported by the US census data.

1.2 Selection on the Basis of Observables and Unobservables

Since Roy's classic work on self-selection, many other theories were developed to explain the skewness of earnings distribution. Some economists extended his simple model to incorporate other considerations, like the observable differences across agents or other methods of selection, while others developed completely new models based on different principles. The key expositions are studied in detail in Neal and Rosen (2000).

One dynamic extension of the selection model is the Learning, Sorting and Matching model (LSM) that explains how the size of distribution of earnings within a group evolves over time. Learning models explain earnings function based on observed differences, sorting models look at selection issues, while matching models look at the role of both observables and unobservables in explaining the earnings distribution. These models assume that both the agents and their employers are unaware of the agents' productive capacities. The productive capacity can be given the same interpretation as the unobserved skill component in the selection model described in the previous section.

The key difference between selection model and LSM is that the selection

model assumes that the agents know their own productive capacities but the employers do not, but LSM assumes that both agents and employers are unaware of the former's productive capacities. The LSM then explains the skewness in earnings distribution evolving over time as both agents and employers learn about the expected productivity of the agents from the output/performance of agents over time. Each agent's earnings grows with experience since the agent sorts away from things he does poorly as he learns about his comparative advantage in different types of employment. Under this general idea, there lies subtle differences between learning, sorting and matching models. The learning models show how earnings become more dispersed for a given group as the agents and employers gain more information about their productive capacities from their observed output. The sorting models describe how agents learn about their unobserved productivity as they generate different output in different jobs. This generates skewness in the distribution of earnings as the agents 'sort' themselves into tasks that maximize their productive capacities. On the other hand, matching models relates to events where work experience on a particular job provides no information about the agents' expected output in other jobs.

We have thus far looked at attempts to explain the marginal earnings distribution. In the next section, we discuss the theories that explain the earnings distribution conditional on the observed differences in education across individuals. We discuss the key problems in this research agenda and how the existing studies attempt to resolve the issues.

1.3 Mincer Earnings Function Explaining Conditional Distribution

The most important theory that explains the skewness in the distribution of earnings, conditional on education, is the human capital theory, formalized by Mincer (1958). The human capital theory claims that the differences in earnings is due to the differences in the level and type of training, experience and education received. Mincer's research prompted Becker (1964) to model earnings conditional on education as a central concept of human capital theory. The subsequent studies in this strand of research tried to explain earnings differences in the light of education and unobserved heterogeneity across individuals. In particular, these studies also tried to explain the level of education chosen by individuals in the light of observed differences and also unobserved differences. This unobserved differences across individuals, termed as 'ability', can be given the same interpretation as unobserved skills in the Roy selection model and the unobserved productive capacities in the LSM model. In this section, we look at the key issues related to the Mincerian approach of studying earnings function conditional on education.

1.3.1 The Mincerian Equation

Earnings functions were introduced as a part of the human capital theory by Becker (1964), and later estimated and popularized by Mincer (1974) to the extent that it is now called the Mincer model. In its simplest form the Mincer earnings

function is

$$w_i = \theta_1 + \theta_2 S_i + u_i$$

where S_i is the level of education for individual i and u_i is a, supposedly, zero mean normal variable which is independent of S_i. The latter assumption ensures that the parameters θ_1, θ_2 can be estimated by least squares regression. It must be mentioned that most of the Mincerian literature on wage equations does not assume that the residuals are normally distributed.

The most common criticism directed against the Mincerian approach questions the assumption of independence between S_i and the error term u_i. This dependence is usually associated with an unobserved ability A_i so that

$$u_i = A_i + \varepsilon_i \tag{1.5}$$

where S_i and A_i are dependent, while the error ε_i is independent of S_i, A_i. As a consequence least squares regression of log earnings w_i on education S_i is biased. This bias is usually expected to be upwards. This is because the usual assumption in the literature is that $Cov\left(S_i, A_i\right) > 0$ (Card (2001)). This implies that

$$\text{plim } \theta_{2,OLS} = \theta_2 + \frac{Cov\left(S_i, A_i\right)}{Var\left(S_i\right)} > \theta_2 \tag{1.6}$$

Hence, if we employ any strategy that attempts to correct this ability bias, the estimated θ_2 should be less than the OLS estimates. In the next section, we look at one of the most popular techniques used in literature to correct for this bias.

1.3.2 Instrumental Variables

A large body of work has addressed the problem of unobserved heterogeneity which results in an upward bias in the OLS estimates of earnings function, well summarized in Heckman, Lochner and Todd (2009). One of the most widely used methods is instrumental variable (IV) estimation by which an instrument Z_i is used. The instrument Z_i should meet two important requirements to be consistent. First, Z_i should be correlated with S_i, that is, $Cov\left(Z_i, S_i\right) \neq 0$. Second, Z_i should be uncorrelated with the unobserved A_i, that is $Cov\left(Z_i, A_i\right) = 0$. Then, the upward bias in equation 1.6 can be corrected

$$\text{plim } \theta_{2,IV} = \theta_2 + \frac{Cov\left(Z_i, A_i\right)}{Cov\left(Z_i, S_i\right)} = \theta_2 \tag{1.7}$$

since $\frac{Cov(Z_i,A_i)}{Cov(Z_i,S_i)} = 0$ as $Cov\left(Z_i, A_i\right) = 0$. Comparing equations 1.6 and 1.7, $\theta_{2,IV} < \theta_{2,OLS}$ since $\theta_{2,IV} = \theta_2$ but $\theta_{2,OLS} > \theta_2$. However, rather somewhat counter intuitively, empirical studies employing IV methods show that the IV estimates are often as big as or bigger than the corresponding OLS estimates (Card 2001).

There have been many explanations why the IV estimates tend to overestimate, rather than underestimate, the true return to education. One of the popular explanations for this puzzle, given by Griliches (1977) and later echoed by Angrist and Krueger (1991), is that the ability biases in the OLS estimates are relatively small; and the gaps between the IV and the OLS estimates mainly reflect the downward biases related to measurement errors. However, Card (2001) argues that the amount of measurement error in schooling that would have to

exist to justify that large gap between OLS and IV estimates is unreasonably large. Schooling is relatively well measured in the US, so that the measurement explanation is likely to be of second order importance. There are possibly other potential explanations for this empirical puzzle. In their review, Heckman *et al.* (2009) find that the traditional instruments are weak, resulting in indecisive empirical estimates. Hence, it is likely that the empirical puzzle persists due to the difficulty of finding strong instruments for education in the earnings function.

In section 3.5.3, we discuss why the IV method may not work in one of our datasets. The IV method works on the assumption that the instrument Z_i should be uncorrelated with the unobserved A_i. However, if education is not separable from the unobservables present in the residual of the log earnings function, it is difficult to find an instrument that would be correlated with education but separable from the component representing the unobservables. This point has also been discussed by Card (2001) and Heckman, Lochner and Todd (2009)).

1.3.3 Functional Form of Earnings Function

This strand of research involving earnings functions resulted in heated discussions about the exact functional form of the relationship between education levels and log earnings, well summarized in Belzil (2007) and Teal (2010). The assumptions of Mincerian earnings functions imply a linear relationship between education and log earnings. However, the presence of unobserved heterogeneity across individuals is likely to lead to a non-linear relationship. Heckman, Lochner and

Todd (2009) argue that the US census data does not support the linearity assumption between the log earnings and education which underlies the Mincerian model. This non-linearity leads to some potential non-trivial problems. First, a non-linear education-earnings profile would seriously question the stability of the stylised facts reported in Mincer (1974). Second, for the use of the IV estimates, the assumption that the relationship between education and log earnings is linear is crucial (Belzil and Hansen (2002) and Belzil (2007)).

The follow-up question is, if the relationship between education and earnings is not linear, is it concave or convex? Psacharopoulos (1994) and Psacharopoulos and Patrinos (2002) assert that the return to education is concave. However, this is strongly refuted by Bennell (1996a,b; 2002). Belzil (2007) and Belzil and Hansen (2002) also show that some parts of the earnings function are convex. This issue of convexity/concavity has huge policy implications since this raises important questions regarding the optimal investment in higher education, particularly in developing countries (Teal (2010)). If the earnings function is convex, then the marginal returns to education are lowest for the individuals with the least education. This implies that giving priority to investment in primary education may have little impact on incomes unless the affected individuals proceed to higher levels of education. Furthermore, the convexity of the earnings function will provide a strong incentive to pursue higher education even though the returns to lower levels of education are modest.

1.4 Concluding Remarks

Our review of literature shows that the presence of unobservables across individuals is responsible for the deviation of non-normality of the marginal distribution of log earnings and for the non-linearity of log earnings-education profile. The studies in both the fields have been conducted under the assumption of normal distribution when, as we have seen in our literature review, empirical work clearly rejects this assumption. In this thesis, we try to directly model the problem of the presence of unobserved heterogeneity across individuals by deviating from the assumption of normal distribution and using a more general distribution that has the power to extract the unobserved component directly from the observed data. For the marginal case, this gives us the advantage of a better fit than the normal distribution. We also have a new tool of checking if the skewness of log earnings distribution conforms with Roy's selection model. For the conditional (on education) case, we want to use this new approach to add to the discussion of the functional form of the earnings function. Our more general distribution allows us to derive an expression for the unobserved heterogeneity component in the earnings function. We can also add to the discussion of non-separability of education and the unobserved heterogeneity component since our proposed distribution allows for the possibility of including the unobserved component both as a separate parameter and as an interaction term with education. This helps to see the exact form in which the unobserved heterogeneity enter the earnings function. We use the proposed distribution to model the marginal distribution of

log earnings in Chapter 2 and the conditional distribution (on education) of log earnings in Chapter 3.

1.5 Bibliography

Becker, G.S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* New York: Columbia University Press.

Belzil, C. (2007): "Testing the Specification of the Mincer Wage Equation," *Centre National de Recherche Scientifique, CIRANO, CIREQ and ZIA,* Discussion Paper No. 2650.

Belzil, C. and J. Hansen (2002): "Unobserved Ability and the Return to Schooling," *Econometrica,* 70(5), 2075-2091.

Bennell, P. (1996a): "Rates of return to education: does the conventional pattern prevail in Sub-Saharan Africa?" *World Development,* 24, 183-199.

Bennell, P. (1996b): "Using and abusing rate of return: a critique of the World Bank's 1995 Education Sector Review," *International Journal of Educational Development,* 16, 235-248.

Bennell, P. (2002): "Hitting the target: doubling primary school enrollments in Sub-Saharan Africa by 2015," *World Development,* 30, 1179-1194.

Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127-1160.

HARRISON, A. (1981): "Earnings by size: a tale of two distributions," *Review of Economic Studies,* 48, 621-631.

HECKMAN, J.J. AND B.E. HONORÉ (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), 1121-1149.

HECKMAN, J.J. AND G. SADLACEK (1990): "Self-Selection and the Distribution of Hourly Wages," *Journal of Labor Economics*, 8(1), S329-S363.

HECKMAN, J.J., L.J. LOCHNER AND P.E. TODD (2009): "Earnings Function, Rate of Return and Treatment Effects: The Mincer Equation and Beyond," Chapter 7 in *Handbook of the Economics of Education,* Volume 1, eds. Erik Hanushek and F. Welch. Elsevier.

LEBERGOTT, S. (1959): "The shape of the income distribution," *American Economic Review*, 49, 328-347.

MILLER, H.P. (1955): *The Income of the American People.* New York: Wiley.

MINCER, J. (1958): "Investment in human capital and personal income distribution," *Journal of Political Economy*, 56, 281-302.

MINCER, J. (1974): *Schooling, Experience and Earnings.* New York: Columbia University Press.

NEAL, D. AND S. ROSEN (2000): "Theories of the Distribution of Earnings," in *Handbook of Income Distribution,* eds. A.B. Atkinson and F. Bourguignon. Elsevier.

PSACHAROPOULOS, G. (1994): "Returns to investment in education: a global update," *World Development,* 22, 1325-1343.

PSACHAROPOULOS, G. AND H.A. PATRINOS (2002): "Returns to investment in education: a further update," World Bank Research Working Paper 2881, The World Bank.

ROY, A.D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers,* New Series, 3(2, June 1951), 135-146.

STAELE, H. (1943): "Ability, wages and income," *Review of Economics and Statistics*, 25, 77-87.

TEAL, F. (2010): "Higher Education and Economic Development in Africa: a Review of Channels and Interactions," *Centre for the Study of African Economies, University of Oxford, UK*, CSAE WPS/2010-25.

Chapter 2

Modelling the Marginal Distribution of Log Earnings

Abstract: This chapter begins a new strand of research by introducing the normal inverse Gaussian (NIG) distribution to model the marginal distribution of income. The NIG distribution can be represented as a normal variance-mean mixture with the inverse Gaussian (IG) distribution as the mixing distribution. A 0.01% subsample of the 1980 US Census, comprising all men between 18 and 65 who are in the labour force, as well as a comparable sample from Ghana, is used to show that the NIG distribution provides a better fit of the log earnings function than the normal distribution. The prediction of right skewness of the log earnings distribution arising from the log-normal skill Roy selection model is rejected in favour of left skewness. The NIG framework is found to be a useful tool to understand how the unobserved differences across individuals affect the earnings distribution.

2.1 Introduction

One of the major issues in labour economics is modelling the marginal distribution of earnings. The key problem faced in empirical studies is that the log earnings distribution is not normally distributed. Broadly speaking, as we have seen in the previous chapter, the theories used to address the non-normality problem can be placed under one general framework. They hinge on our ability to make inferences from an unobserved heterogeneity component, representing the unobserved differences across individuals, to the observed income distribution. Roy's (1951) selection model terms this source of unobserved heterogeneity or unobserved differences across agents as unobserved skills while the Learning, Sorting and Matching model calls it unobserved productive capacity.

We take a different approach to the existing literature. We relax the assumption that the log earnings distribution is normally distributed. The probability density of the normal distribution is a function of two parameters: the mean, μ, and variance, σ^2. The mean describes the location of the distribution while the variance measures the width of the distribution. This distribution does not allow the possibility for the mean to be dependent on the variance, which can be crucial in understanding the impact of unobservables on earnings functions. In other words, the simplistic set-up of normal distribution maybe too restrictive for studying the log earnings distribution since we may require more parameters to define the dimensions of the unobserved heterogeneity across individuals.

In this thesis, we use a more general distribution that can address the limita-

tions of using the normal distribution. The normal distribution can be regarded as a special case of our proposed general distribution. Contrary to the previous models, we start with the observed data of income and use a distribution that helps us to account for the differences in unobserved variance across individuals. These differences in unobserved variance across individuals can be termed as unobserved heterogeneity across individuals. In other words, the unobserved heterogeneity represents the set of unobservables. The variance is now open to interpretation as it models the unobservables driving the distribution of earnings. Then, the advantage of our proposed distribution is that it provides a useful tool to extract the unobserved heterogeneity from the observed income. We can measure the unobservables directly and empirically test our approach. Furthermore, our proposed distribution allows us to express mean and higher moments of log earnings as a function of these unobservables. In doing so, we can determine from observed earnings how individuals' unobserved heterogeneity differs and quantify the impact of the unobservables on the earnings distribution. We would not be able to do so using the normal distribution. Unlike the Roy selection model, there is no prior about the sign of the skewness. Our proposed distribution allows the observed data determine the skewness of the earnings distribution.

2.2 Non-normality in the Marginal Distribution

The normal distribution is a popular choice in modelling the distribution of log earnings. Aitchison and Brown (1957, pp.13) motivated this through the Central

Limit argument that log earnings are averages of a myriad of small independent and identically distributed factors. Neal and Rosen (2000) give an overview of stochastic process theories that lead to other distributions. Since Aitchison and Brown's seminal work in this field, other distributions have also been used extensively to model both conditional and unconditional distributions of income. Kleiber and Kotz (2003) provide a comprehensive account of the development of different distributions to understand income. In particular, recent work has employed beta-type size distributions to model income. This comprises the general form called Generalized Beta Distribution of the Second Kind containing four parameters and its subclasses, Singh-Maddala and Dagum models, containing three parameters. Beta-type size distributions also include Fisk (Log-Logistic and Lomax distributions) and Generalized Beta Distribution of the First Kind, explained in detail in Kleiber and Kotz (2003). Biewen and Jenkins (2005) provide three advantages of using the Singh-Maddala and Dagum specifications. First, these specifications provide flexibility to model heterogeneous income data. Second, estimation is easy. Third, the standard errors of all the parameters can be obtained in a straightforward manner. After discussing the characteristics of our proposed distribution (section 2.3.1), we will compare the advantages and disadvantages of our proposed distribution versus these alternatives and give the motivation for choosing our distribution over the other alternatives (section 2.3.2).

Here we start by looking at the empirical distribution of US log earnings (dataset described in section 2.4.1) and find that it deviates considerably from normality. Later we will attribute this deviation from normality to unobserved

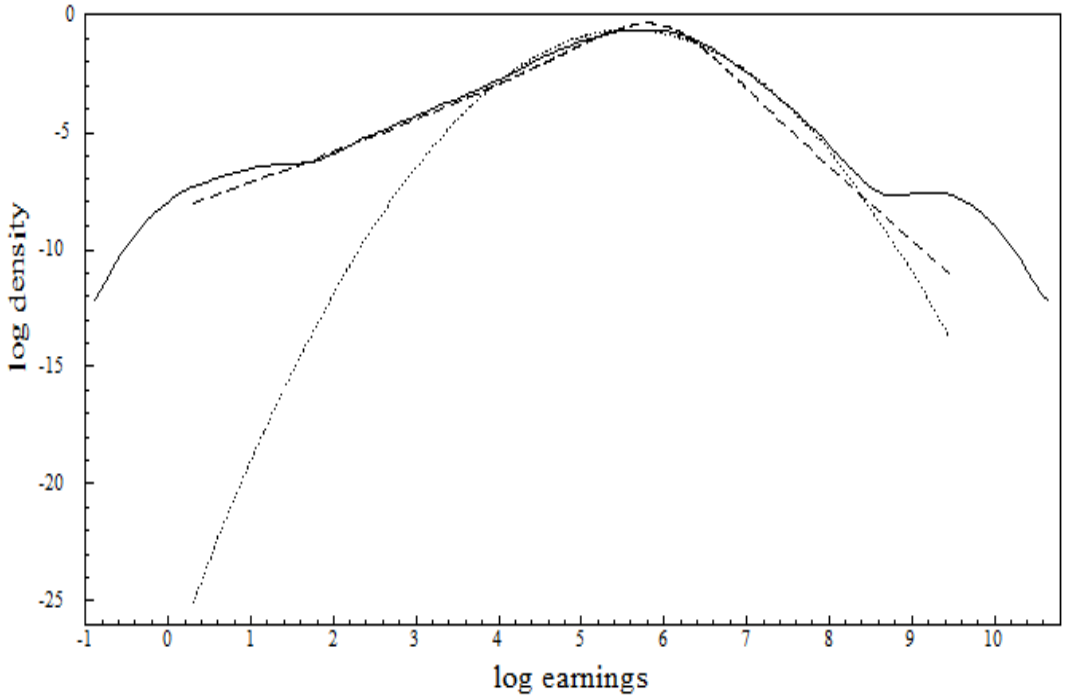

Figure 2-1: Log Histogram of the US Log Earnings. Solid line is a kernel estimate. Dotted line is a normal fit. Dashed line is a normal inverse Gaussian fit.

heterogeneity.

The non-normality is easily seen from a log histogram of the US log earnings as shown in Figure 2-1. A log histogram turns the normal density of the form

$$\mathsf{f}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

into a parabola of the form

$$\ln \mathsf{f}(x;\mu,\sigma) = -\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{(x-\mu)^2}{2\sigma^2}.$$

Deviations from the parabolic shape then indicates deviation from normality. Fig-

ure 2-1 shows that a kernel estimate of the log density deviates considerably from the parabolic shape of a normal fit. Indeed, the tails of the log distribution are close to linear. This linear shape can be captured with the normal inverse Gaussian distribution, which we will introduce later. A further observation is that the density is skewed to the left. Formal tests for normality give further, strong evidence against normality since the statistics for no skewness and no kurtosis,

$$\chi^2_{skewness} = 1121, \qquad \chi^2_{kurtosis} = 4171, \tag{2.1}$$

respectively, are both extreme when compared with a χ^2_1 distribution. The kernel estimate is a non-parametric way to estimate the probability density function of a random variable. The idea of the non-parametric approach is to avoid restrictive assumptions about the form of $\mathsf{f}(x)$ and to estimate this directly from the data. We plot the kernel estimate of the log density function using some of the in-built functions of OxMetrics software.

To show that the US log earnings deviate from normality, we found it more useful to plot the log histogram of log earnings (that is, kernel estimate of the log density), the normal inverse Gaussian distribution and the normal distribution in one chart, rather than the conventional QQ or PP plots. This is because we can present the three distributions in the same chart using our approach that gives a very good visual intuition of how the normal inverse Gaussian distribution fits the data better than the normal distribution. The QQ plot allows us to compare two distributions against each other. Here, we can compare how the

two distributions fit the observed data as we have the normal inverse Gaussian distribution, normal distribution and the kernel estimate in one chart. Hence, we can check which distribution fares better by comparing each distribution to the kernel estimate. Also, since we are using a distribution that would not be familiar to the readers, this chart shows what a normal inverse Gaussian distribution should look like. If we had graphed QQ plots, we would still need another chart to show the shape of the normal inverse Gaussian distribution. Thus, this chart gives a very good intuitive idea why the normal inverse Gaussian fits the data better. Later on, we complement this with rigorous analysis by comparing the log likelihood values for the distribution using normal inverse Gaussian distribution versus normal distribution.

2.3 Our Proposed Distribution

We propose to relax the assumption of normality in log earnings distribution and use a distribution that is more general with more free floating parameters. For this purpose, we propose to model log earnings distribution using the normal inverse Gaussian distribution that has been studied extensively, see for instance Barndorff-Nielsen (1978). We first go through the formal definition of our proposed distribution. Then we discuss the motivations for using this distribution for labour economics analysis. Finally, in section 2.3.3, we present the interpretations of the parameters of this distribution.

2.3.1 Formal Definition

Suppose log earnings w is normal inverse Gaussian (NIG) distributed, then the density can be expressed as

$$\mathsf{f}(w) = Cq(\frac{w-\mu}{\delta})^{-1} K_1 \left\{ \delta\alpha q(\frac{w-\mu}{\delta}) \right\} \exp(\beta w)$$

for $w \in \mathbb{R}$. Here $q(w) = \sqrt{1+w^2}$ and $\alpha^2 = \beta^2 + \gamma^2$ while K_1 is a modified Bessel function of the third kind, also referred to as the MacDonald function, see Section 9.6.1 of Abramowitz and Stegun (1970). The normalisation constant is

$$C = (\alpha/\pi) \exp(\delta\gamma - \beta\mu).$$

Figure 2-1 gives an idea about the general shape of the normal inverse Gaussian distribution. We discuss the fit of the distribution in detail in section 2.4.1.

The NIG distribution assumes that the distribution of the unobserved heterogeneity variable σ^2, representing the unobserved differences across individuals, is inverse Gaussian with density

$$\mathsf{f}_{\sigma^2}(x) = \frac{\delta}{\sqrt{2\pi}} e^{\delta\gamma} x^{-3/2} \exp\left\{ -\left(\delta^2 x^{-1} + \gamma^2 x\right)/2 \right\} \qquad \text{for } x > 0, \tag{2.2}$$

where $\gamma, \delta > 0$. Here $\gamma^2 > 0$ is an inverse scale parameter and $\gamma\delta$ is a shape parameter which is invariant to the scale. As a short hand notation we write

$$\sigma^2 \overset{\mathsf{D}}{=} \mathsf{IG}(\delta, \gamma). \tag{2.3}$$

It must be noted that the NIG distribution is an extension of the normal distribution with the normal distribution $\mathsf{N}(\mu, \sigma^2)$ appearing as a limiting case for $\alpha^2 \to \infty$ when $\beta = 0$ and $\delta/\alpha = \sigma^2$.

2.3.2 Motivations

One of the motivations of using the NIG distribution to model log earnings is to exploit what is known as the 'mixing property' of the distribution. Usually, if we convolute two normal distributions, we end up with a distribution that is also normal. However, the NIG distribution is the product of mixing two different types of distribution, one is normal distribution and the other is inverse Gaussian distribution. This idea is expressed in one of the key characteristics of this distribution: if w is normal inverse Gaussian distributed, then the conditional distribution of w given the unobserved heterogeneity variable σ^2 is normal

$$(w \mid \sigma^2) \overset{\mathsf{D}}{=} \mathsf{N}(\mu + \beta\sigma^2, \sigma^2). \tag{2.4}$$

If we assume that the unobserved heterogeneity variable takes the shape of IG distribution, using the NIG distribution would imply that, conditioned on this unobserved heterogeneity variable, log earnings is normally distributed. In other words, NIG distribution - the marginal distribution for log earnings - is a mixture distribution formed by mixing σ^2 and $w \mid \sigma^2$. This distributional setup is thus implicitly assuming the notion that ideally log earnings distribution should be normally distributed. One of the reasons for the deviation from normality is the

presence of unobservable differences across individuals. If we could somehow condition log earnings on the unobserved differences, the resulting distribution should be normally distributed. This idea that non-normality is due to the presences of unobservables bodes well with the earlier studies that we reviewed in the previous chapter. Our distribution allows a method of quantifying these unobserved differences.

A closer look at the expression of mean and variance of the log earnings distribution, conditioned on σ^2, in equation 2.4 shows another property of the NIG distribution useful for labour economics. The normal inverse Gaussian (NIG) distribution can be represented as a normal variance-mean mixture (Barndorff-Nielsen and Blæsild 1981). A variance-mean mixture is an expression for stating that the mean of the log earnings, conditioned on σ^2, is a function of the variance of the distribution. This property of the dependence of mean on the variance cannot be tested using normal distribution. One of the key advantages of using the NIG distribution to model log earnings is that the distribution allows the possibility that the mean and the higher moments of the distribution are a function of the variance of the distribution. This characteristics of the NIG distribution can be exploited to better understand and express the role of the unobservables in shaping the income distribution. In the next chapter, we show that this characteristics can also be exploited to disentangle the role of observables and unobservables when income is conditioned on education.

An econometric motivation for this choice is that the resulting normal inverse Gaussian distribution fits the log earnings data rather well as shown in Figure 2-

1. As mentioned in section 2.2, Aitchison and Brown (1957, p.13) motivated the use of normal distribution for log earnings distribution through the Central Limit argument that log earnings are averages of a myriad of small independent and identically distributed factors. Another motivation for using the NIG distribution is that this distribution generalises the Central Limit argument of Aitchison and Brown (1957, p.13). Halgreen (1979) showed that the normal inverse Gaussian distribution is self-decomposable which means that it arises as an average of a myriad of small independently distributed factors, see Gnedenko and Kolmogorov (1954, pp. 98). These factors are not identically distributed which addresses a critique of Aitchison and Brown's Central Limit argument. In other words, we can relax the assumption that log earnings are averages of small identically distributed factors; it is sufficient to assume they are an average of a myriad of small independently distributed factors.

The advantage of this alternative formulation is that we are able to develop a methodology where we can identify the skill composition of earnings distribution and measure the skill composition directly. We were not able to do so using the Roy selection framework. Recall that in Heckman and Honoré's (1990) model, discussed in section 1.1, V represented the linear combination of skills and M represented the mean of aggregate log earning w. The log earning function was represented with the equation $w = V + M$. Since V and M were independent in this set-up, the log earning w was homoskedastic with respect to M. In contrast to Heckman and Honoré's (1990) model in section 1.1, we introduce a distributional setup in which V and M are dependent so as to generate heteroskedasticity with

respect to M.

Suppose that the aggregate log earnings are given by $w = V + M$ as in equation 1.3. In the interpretation of Heckman and Honoré the analysis of the two sector Roy model is essentially reduced to an analysis of the univariate skills difference D. In our proposed alternative setup, the positive unobserved heterogeneity variable σ^2 represents the unobserved skill level or the ability of an individual. Both variables V and M will be functions of σ^2, the components representing unobserved heterogeneity, in the distribution of log earnings w. Conditionally on σ^2 we let M have a simple linear expression

$$M = \mu + \beta\sigma^2, \tag{2.5}$$

while the conditional distribution of V given σ^2 is assumed to be zero mean normal

$$(V|\sigma^2) \overset{\mathsf{D}}{=} \mathsf{N}(0, \sigma^2). \tag{2.6}$$

The linear expression for M therefore replaces the complicated non-linear function of the unobserved skill D in equation 1.4. A feature of this setup is that V and M are uncorrelated but dependent. The uncorrelatedness follows using the law of iterated expectations showing that V has mean zero so the correlation of V and M is given by

$$\mathsf{E}(MV) = \mathsf{E}\mathsf{E}(MV|\sigma^2) = \mathsf{E}\{M\mathsf{E}(V|\sigma^2)\} = 0. \tag{2.7}$$

Thus, rather than making inferences from the unobserved variables, we are able to give a more concrete form to the unobserved heterogeneity variable. We are assuming that the unobserved heterogeneity variable, σ^2, is inverse Gaussian distributed with two parameters δ and γ. In our empirical work, we can estimate these two parameters from the observed data.

One of the motivations of using our proposed distribution, as opposed to the potential alternative distributions used in literature as discussed in section 2.2, is that we can give the interpretation of unobserved heterogeneity or unobservables to σ^2, which is IG distributed. We cannot give such an interpretation if we model income using the other distributions. Using the other distributions, we do not have a mechanism to directly quantify the impact of unobservables on income. The NIG distribution gives us more power to understand the role of unobservables in explaining income distribution. When we use conditional (on education) income distribution, it helps us to understand the interplay between education and the unobservables. In other words, using the NIG distribution, we have the benefit of being able to interpret the estimates in terms of the role of unobservables. We cannot give such an interpretation when we use the alternative distributions. In particular, the Generalized Beta Distribution of the Second Kind is said to be a scaled mixture of generalized gamma distributions with inverse generalized gamma weights (Jones, Lomas and Rice (2011) and Kleiber and Kotz (2003)). On the other hand, the NIG distribution is the mixture of normal distribution and inverse gaussian distribution. Hence, we can make the assertion that, once the unobserved heterogeneity is accounted for, the conditional (on σ^2) distribution

is the normal distribution. This accords with literature that contends that the non-normality is due to the presence of unobservables and, once the unobserved heterogeneity is accounted for, income should be normally distributed. We cannot make similar assertions with the Generalized Beta Distribution of the Second Kind or its sub-classes, most notably Singh-Maddala and Dagum models, since they are mixtures of complicated distributions (not normal distribution). However, it must be mentioned that there are also costs in using the NIG distribution rather than the Generalized Beta Distribution of the Second Kind. One of the most notable disadvantages is that the Singh-Maddala and Dagum models, that have been commonly used to model income distribution, contain three parameters. On the other hand, the NIG distribution has four parameters. Hence, we are losing one degree of freedom and making our model more complicated. Moreover, many studies have shown that the Generalized Beta Distribution of the Second Kind model fits income distributions extremely well across different times and countries (Jenkins *et al.* (2011)). Hence, there is a possibility that we have a tradeoff between a better fit of the model using Generalized Beta Distribution of the Second Kind versus the ability to give an interpretation to the role of unobservables in shaping income distribution using the proposed NIG distribution.

2.3.3 Interpretation of the Parameters

The four parameters $\mu, \beta, \delta, \gamma$ of the normal inverse Gaussian distribution are interpreted as follows. The parameters $\mu \in \mathbb{R}$ and $\delta > 0$ are location and scale

parameters, respectively. The sign of $\beta \in \mathbb{R}$ determines the sign of skewness so that $\beta > 0$ implies that the density is skewed to the right. Finally, the parameter $\gamma > 0$ along with the transformed parameter $\alpha^2 = \beta^2 + \gamma^2$ controls the steepness of the distribution with the steepness of the distribution increasing as the value of α increases. The first three moments are:

$$\mathsf{E}(w) = \mu + \frac{\delta\beta}{\gamma}, \quad \mathsf{Var}(w) = \frac{\delta}{\gamma} + \frac{\delta\beta^2}{\gamma^3}, \quad \mathsf{Skewness}(w) = \frac{3\beta}{\sqrt{\left(\beta^2 + \gamma^2\right)(\gamma\delta)}}. \quad (2.8)$$

Recall from section 2.3 that the unobserved heterogeneity variable σ^2 is IG distributed and functions of δ and γ. Hence, the mean and higher moments of log earnings are functions of the unobservables since all the expressions include δ and γ. The first three moments of σ^2, assuming σ^2 is inverse Gaussian distributed, are

$$\mathsf{E}(\sigma^2) = \frac{\delta}{\gamma}, \quad \mathsf{Var}(\sigma^2) = \frac{\delta}{\gamma^3}, \quad \mathsf{Skewness}(\sigma^2) = \frac{3}{\sqrt{(\gamma\delta)}}. \quad (2.9)$$

2.3.4 The Generalized Hyperbolic Distributions

The normal inverse Gaussian distribution can be generalized even further by replacing the inverse Gaussian distribution with a generalized inverse Gaussian distribution resulting in a five-parameter family known as the generalized hyperbolic distributions. The name comes from the fact that the tails of the log density function have a hyperbolic shape as seen in Figure 2-1. Initially the hyperbolic distribution, another special case of generalized hyperbolic distribution, was popular in empirical studies. Barndorff-Nielsen (1978) showed that Australian personal

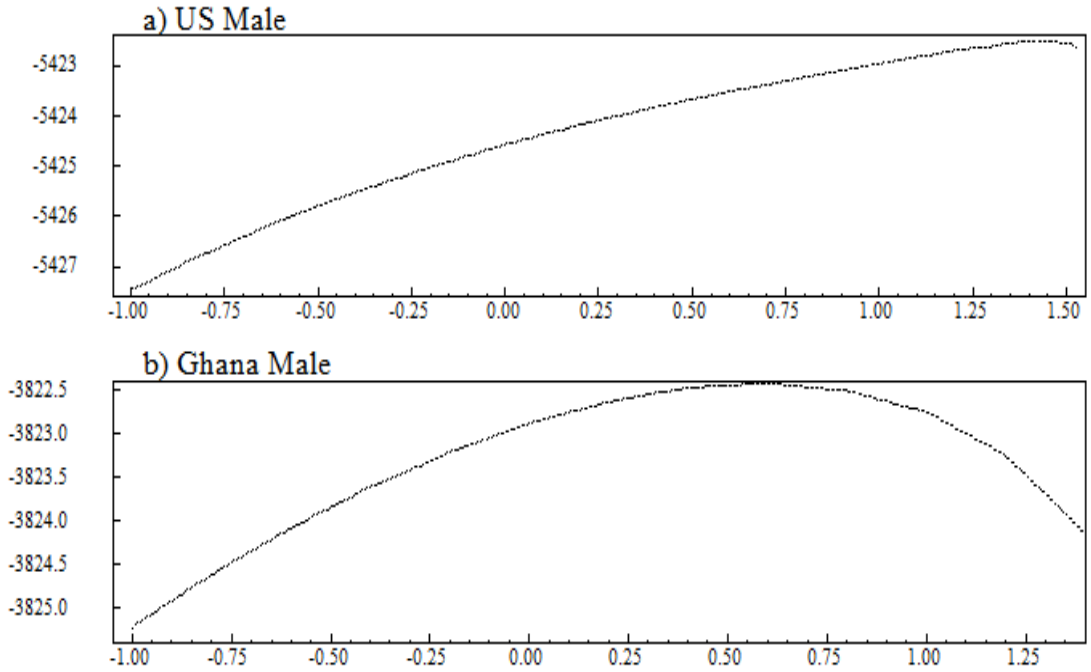

Figure 2-2: Profile Likelihood versus λ Parameter of Generalized Hyperbolic Distribution

incomes were fitted rather well by a hyperbolic distribution. Later it was found that the NIG distribution is preferable to the hyperbolic distribution for computational purposes (Barndorff-Nielsen and Stelzer (2005)). While the generalized hyperbolic distributions have been used extensively in finance, wind turbulence, and sedimentology, we do not know any other applications to the labour market.

It appears that there is not that much gain from adding the fifth parameter and the normal inverse Gaussian distribution is often preferred in empirical studies due to its flexibility in describing a wide range of shapes, useful analytical properties and tractability, see Barndorff-Nielsen and Stelzer (2005). Figure 2-2 plots the profile likelihood values against possible changes in λ where λ is the parameter

in generalized hyperbolic distribution that determines the sub-class of generalized hyperbolic distribution. We give an overview of how we estimate the parameters in section 2.4. For example, λ is equal to -0.5 for NIG and equal to 1 for hyperbolic distribution. For the US male, the log-likelihood increases as λ increases until λ is 1.4 and then falls thereafter, and for Ghana male, data described in section 2.4.2, the log likelihood value increases until λ is 0.6 and then falls. However, the difference between the log likelihood for NIG and the highest log likelihood value is negligible, 3.0 for the US male and 1.4 for Ghana male.

2.4 Empirical Studies

We will implement the normal inverse Gaussian methodology for two datasets, namely the US dataset and a similar dataset for Ghana. Our motivation for using these two datasets is that we are interested to compare the outcome of the labour market of an advanced economy with that of an emerging market. We use the US as an example of an advanced economy and the Ghanaian labour market as an example of a developing economy. The economies differ radically in the structure of their labour markets, offering us the opportunity to assess how well the normal inverse Gaussian distribution (NIG) can describe their labour market outcomes. In particular for the US data we confine ourselves to male earnings, thus eliminating the selection issues associated with female earning functions. In contrast our Ghana data is for wage earnings in an economy where wage employment is a highly selected outcome. We anticipate that the role of unobservables in the

earnings functions for Ghana will be much more heterogeneous than for the US.

In addition, the choice of Ghana allows us to look at the policy relevant questions related to the labour market of an African economy. Ghana provides a good example of an African economy where the functional form of the earnings function maybe important for policy discussions. One of the prominent policy questions related to African economies is whether the investment in education has translated into economic growth (Teal (2010)). As discussed in section 1.3.3, to answer this question, economists have tried to find if the earnings functions is convex or concave, and this issue has raised heated debate. If the earnings function is convex, then the marginal returns to education are lowest for the individuals with the least education. This implies that giving priority to investment in primary education may have little impact on incomes unless the affected individuals proceed to higher levels of education. Furthermore, the convexity of the earnings function will provide a strong incentive to pursue higher education even though the returns to lower levels of education are modest. We hope to contribute to this debate by modelling the Ghanaian labour market using the NIG distribution.

For both the normal inverse Gaussian distribution and Generalized Hyperbolic distribution, we estimate the parameters using maximum log likelihood approach. The same methodology is also used when we extend our model to the conditional case in the next chapter. For the empirical analysis, we use OxMetrics. OxMetrics has in-built functions for the probability density of the Generalized Hyperbolic distribution. This facilitates the computation of maximum log likelihood. However, since we deviate from the normal distribution, we cannot use any of the

in-built functions for the calculation of coefficient of parameters, quantiles, etc. We had to manually program each and every analysis conducted in this thesis using OxMetrics software. The programming was very lengthy and cumbersome when we extended our model to the conditional case. In particular, when we included dummy for each education level, the analysis was intricate and lengthy since we had to manually do all the coding.

2.4.1 Empirical Studies I - The US Labour Market

Data Description

The dataset considered is taken from a 0.01% subsample of the 1980 US census taken from Integrated Public Use Microdata Series at http://www.ipums.org. The 1980 census data is used because the 1990 and 2000 publicly available census extracts have been perturbed due to data protection issues. From this subsample, the selected dataset comprises all men between 18 to 65 who are in the labour force, and who gave positive values in response to questions about weeks worked, usual hours worked per week, and earnings for the year 1979. This reduces the sample to 5170 observations. For data protection reasons, earnings are top-coded at $75000 annually, and education is top-coded at 20. This means that any income above $75000 is truncated at that amount and any education level higher than 20 is top-coded at 20. There are 43 observations top-coded at $75000 and 113 observations top-coded at 20 levels of education. There are 6 observations that are top-coded at both education and income. Therefore, the total number of top-

coded observations is 150. We then look at the weekly earnings, defined as the top-coded annual income divided by the number of weeks worked.

The topcoding in the US data was not taken into account in the specification of sample likelihood contributions. Ideally, we require an adjustment factor to reflect the information loss due to censoring (Li and Wang (2003)). The estimated mean, depending on the number of observations that are top-coded, may be underestimated. Jenkins *et al.* (2011) gives an overview of other problems associated with top-coded data and measures to resolve the issue. It can be concluded from their analysis that the variance estimates may also be underestimated due to the presence of top-coded data. However, these issues maybe less problematic in our analysis since the top-coded observations comprise only 2.9% of the total observations. Hence, for our case, we can treat these potential problems as negligible.

Results

In section 2.2, we have looked at the marginal distribution of log earnings to illustrate the non-normality problem by plotting a smoothed version of the log histogram for this sample. Figure 2-1 clearly shows that, compared to the normal distribution, the normal inverse Gaussian (NIG) distribution provides a better fit, particularly in describing the tails of the distribution. Table 2.1 shows that the NIG offers a much better fit than normal distribution. The log likelihood function of NIG regression is 466 more than that of the normal regression. Since there are 5170 observations, there is a 0.1 gain in log likelihood value per observation

Type of distribution	Fit of the marginal distribution of log earnings
Normal	-5891.5
NIG	-5425.8

Table 2.1: US Male: Log Likelihood Values of the Marginal Distribution of Log Earnings

when two more parameters are added. This means that with 30 observations, the likelihood of the NIG-fit would be three more than that of the normal fit. Now, if we use the LR statistics (twice the difference between the unrestricted and restricted regressions) to check if the difference in log likelihood values is significant or not, this provides a likelihood ratio statistic of 6 which is the 95% quantile for $\chi^2(2)$. That is, 30 observations would be enough to reject the normally distributed fit in favour of the NIG distributed fit. This supports the claim that the NIG distribution fares better in fitting the log earnings data.

The parameter estimates of the distribution are

$$\underset{(298.0)}{\mu = 5.96}, \quad \underset{(12.0)}{\beta = -0.84}, \quad \underset{(16.3)}{\gamma = 1.63}, \quad \underset{(24.7)}{\delta = 0.74}. \tag{2.10}$$

The figures reported below are signed likelihood ratio statistics which are interpreted in the same way as t-statistic in regression analysis.

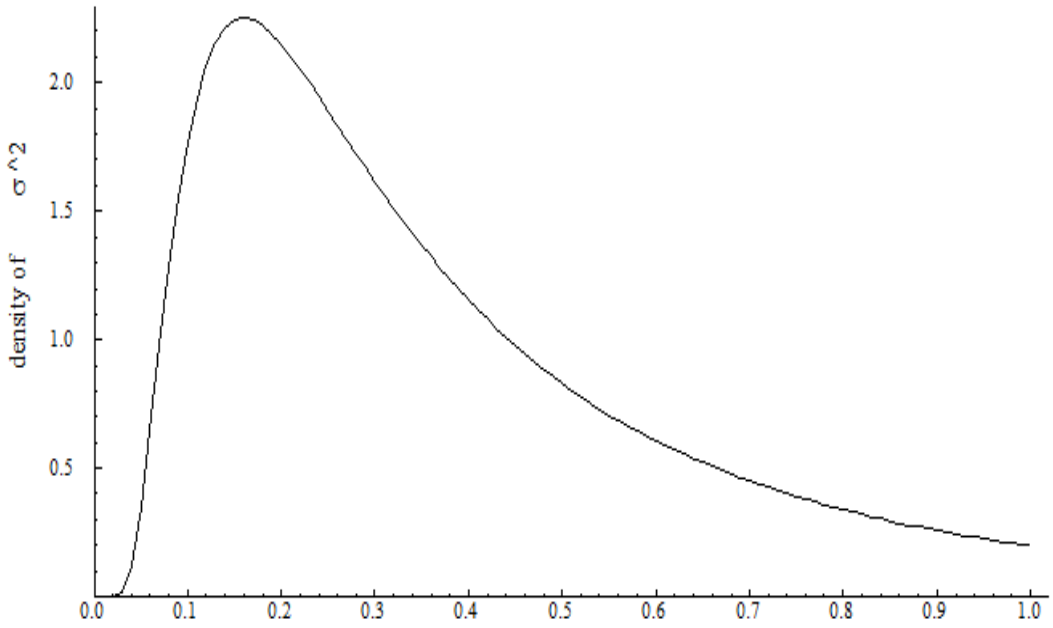

Figure 2-3: US Male: Distribution of the Unobserved Heterogeneity Variable

The mean and higher moments are

$$\mathsf{E}(w) = \mu + \frac{\delta\beta}{\gamma}, \tag{2.11}$$

$$\mathsf{Var}(w) = \frac{\delta}{\gamma} + \frac{\delta\beta^2}{\gamma^3}, \tag{2.12}$$

$$\mathsf{Skewness}(w) = \frac{3\beta}{\sqrt{\left(\beta^2+\gamma^2\right)(\gamma\delta)}}. \tag{2.13}$$

$$\mathsf{E}(w) = 5.58, \tag{2.14}$$

$$\mathsf{Var}(w) = 0.57, \tag{2.15}$$

$$\mathsf{Skewness}(w) = -1.25. \tag{2.16}$$

Contrary to the Roy's selection model prediction of right skewness, see Heck-

man and Honoré (1990, pp. 1132, Theorem 3), the US male dataset exhibits left skewness (-1.25). More importantly, our proposed distribution allows the flexibility of the skewness to be both positive and negative since the sign of the skewness is determined by the sign of β. Also, we can plot the distribution of the unobserved heterogeneity variable σ^2, as shown in Figure 2-3. The figure shows that the underlying unobservables responsible for the unobserved heterogeneity is right skewed for the US male.

2.4.2 Empirical Studies 2 - Ghana Labour Market

We conduct the same analysis on Ghana labour market - taking only male cohorts - to demonstrate that the NIG framework can be used to study log earnings distribution across varied labour markets.

Data Description

The data come from the Ghana Living Standards Survey (GLSS). The GLSS is a multi-topic household survey, designed to provide information on the living standards of Ghanaian households. The survey was conducted in four rounds (in 1987/1988, 1988/1989, 1991/1992 and 1998/1999). Each round, known as a wave, covers a nationally representative sample of households spread over a full 12-month period. The log earnings is defined the same way as that of the US dataset and is adjusted for inflation. The maximum number of years of education is twenty-four. The sample size is 2875.

The pooling of data of different rounds of the GLSS has also been done by

other studies, notably Nsowah-Nuamah, Teal and Awoonor-Williams (2010) and Owens, Sandefur and Teal (2011). The pooling of the data helps to increase the sample size so that we can get more information from the datasets and better information about the labour market. We control for year effects by controlling (having a dummy) for each wave of data. We find that, with the exception of one case, the waves are not statistically significant. The exception is the ordinary least square estimates using normal distribution for the case of education level greater than or equal to ten. We find that the third wave is statistically significant in this particular case. Also, we use real earnings, rather than the nominal values. The control of waves and the use of real earnings help to get rid of controversies related to the inflation index.

Results

Figure 2-4 shows that the NIG distribution offers a better fit than the normal distribution. Equation 2.17 clearly rejects normality. Table 2.2 compares the log likelihood values of the NIG distribution to the normal distribution. A comparison similar to that of section 2.4.1 is conducted to show that the log likelihood function of the NIG-marginal fit is 150 more than that of the normal distribution. Once again, the gain in likelihood value is 0.1 per observation and the normally distributed model is strongly rejected in favour of the NIG-distributed model.

$$\chi^2_{skew} = 56, \quad \chi^2_{kurt} = 1489, \quad \chi^2_{norm} = 1545 \tag{2.17}$$

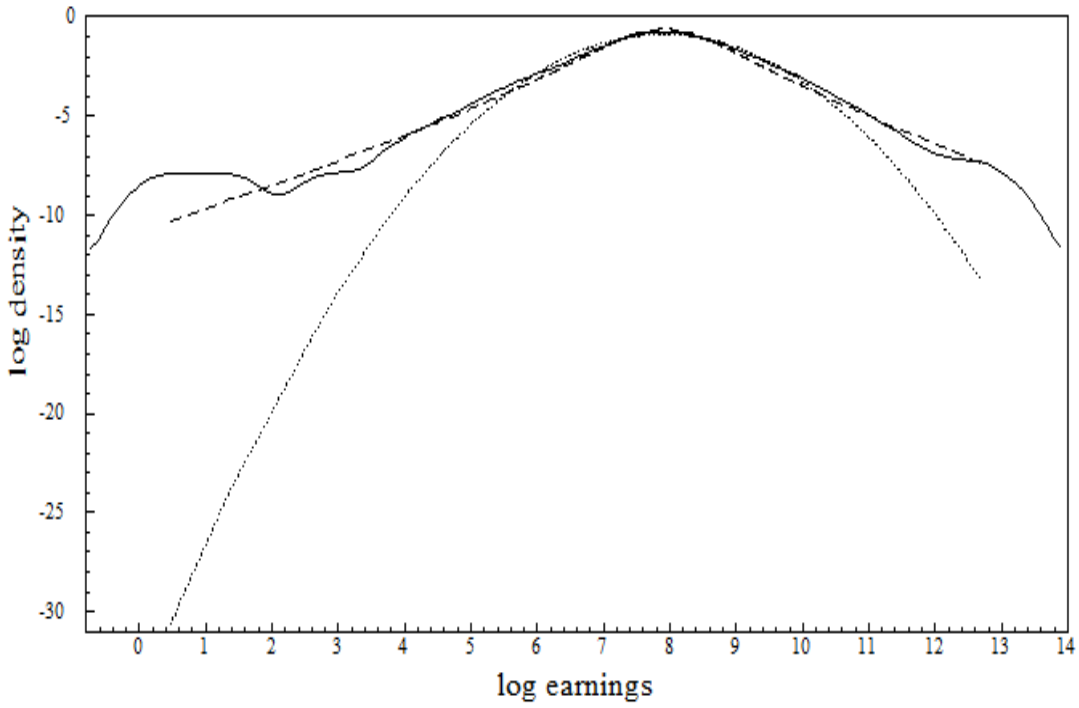

Figure 2-4: Log Histogram of Ghana Log Earnings. Solid line is a kernel estimate. Dotted line is a normal fit. Dashed line is a normal inverse Gaussian fit.

Type of distribution	Fit of the marginal distribution of log earnings
Normal	-3972.8
NIG	-3823.8

Table 2.2: Ghana Male: Log Likelihood Values of the Marginal Distribution of Log Earnings

The parameter estimates of the distribution are

$$\underset{(264.3)}{\mu = 7.93}\,, \qquad \underset{(0.8)}{\beta = -0.03}, \qquad \underset{(12.3)}{\gamma = 0.98}, \qquad \underset{(18.4)}{\delta = 0.92}. \tag{2.18}$$

The figures reported below are signed likelihood ratio statistics which are interpreted in the same way as t-statistic in regression analysis.

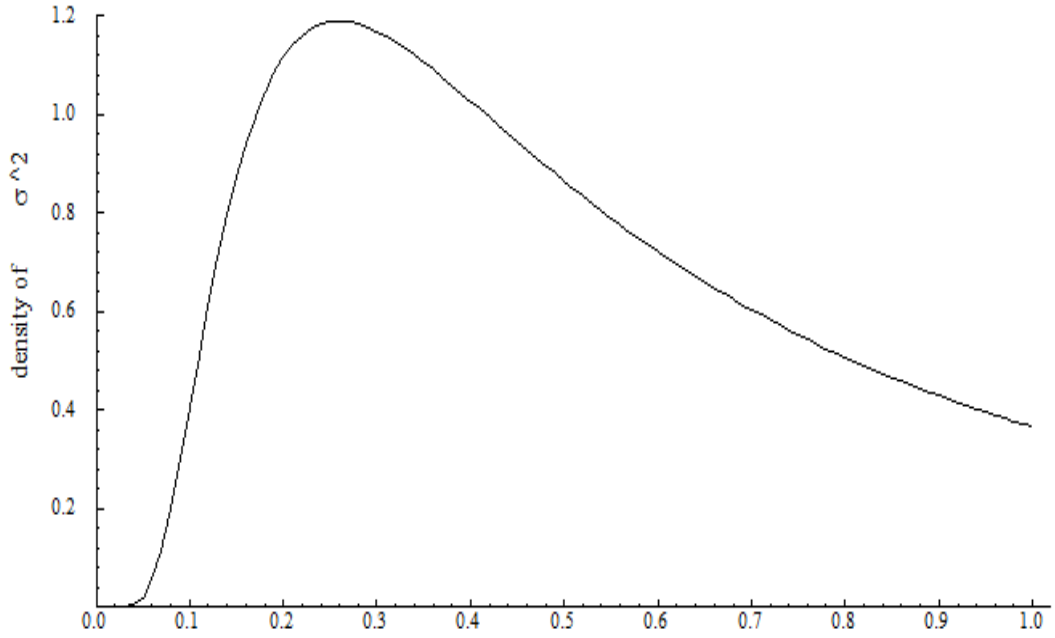

Figure 2-5: Ghana Male: Distribution of the Unobserved Heterogeneity Variable

The mean and higher moments are

$$\mathsf{E}(w) = \mu + \frac{\delta\beta}{\gamma}, \tag{2.19}$$

$$\mathsf{Var}(w) = \frac{\delta}{\gamma} + \frac{\delta\beta^2}{\gamma^3}, \tag{2.20}$$

$$\mathsf{Skewness}(w) = \frac{3\beta}{\sqrt{\left(\beta^2+\gamma^2\right)(\gamma\delta)}}. \tag{2.21}$$

$$\mathsf{E}(w) = 7.90, \tag{2.22}$$

$$\mathsf{Var}(w) = 0.93, \tag{2.23}$$

$$\mathsf{Skewness}(w) = -0.09. \tag{2.24}$$

Again, contrary to the Roy's selection model prediction of right skewness, the Ghana male dataset exhibits left skewness (-0.09). Figure 2-5 plots the distribu-

tion of the unobserved heterogeneity variable. Like the US male, the underlying heterogeneity variable is right skewed. However, the distribution is less peaked than that of the US male.

2.5 Comparing the US and Ghana Results

There are some commonalities between the results from the US and Ghana. Both decisively reject normal distribution in favour of the NIG distribution. In sharp contrast to Roy's selection model's predictions, both the datasets exhibit left skewness. As discussed in the first chapter, this is in line with other studies that find that Roy's prediction of right skewness is not borne by the observed data (Heckman and Sadlacek (1990)).

We plot the unobserved heterogeneity variable, σ^2, for the US and Ghana in Figures 2-3 and 2-5 respectively. The heterogeneity variable represents the set of unobservables. It is possible that the unobservables could predominantly represent the unobserved ability as literature is mainly concerned with the presence of the unobserved ability in log earnings distributions. However, the unobservables could also represent other features that cannot be observed but influence the earnings functions. We find that both the US and Ghana's distribution of the unobserved heterogeneity variable is skewed to the right. This is in contrast with one of the key assumptions of Roy's selection model: log skills are normally distributed.

Similar to the distribution of log earnings, the distribution of the unobserved heterogeneity variable also has a longer tail for the US and a wider distribution

for Ghana, indicating that the skewness is higher for the US but the variance is higher for Ghana. This finding raises interesting questions about the role of unobserved heterogeneity in shaping the log earnings distribution. Is the distribution of log earnings wider in Ghana since unobserved heterogeneity is more widely distributed? Or is it because of the differences in observables like education? Or because of the interaction of both education and ability? In other words, we want to determine if the distribution of the unobserved heterogeneity variable is dependent on the level of education. We attempt to get a better understanding of the heterogeneity variable in the next chapter by including education and experience in our regressions, and also looking at the interactions between education and the unobserved heterogeneity variable.

2.6 Conclusion

The aim of this chapter was to introduce a new distribution to better model the log earnings data by describing unobserved heterogeneity. Broadly speaking, the theories used to understand unobserved heterogeneity thus far can be placed under one general framework. They hinge on our ability to make inferences from an unobserved heterogeneity component to the observed income distribution. Roy's selection model terms this source of heterogeneity across agents as unobserved skills while the Learning, Sorting and Matching model calls it unobserved productive capacity.

Contrary to the previous models, we start with the observed distribution of

income and use a distribution that helps us to account for the differences in unobserved variance across individuals. Our proposed distribution allows us to address some of the limitations of using normal distribution to model earnings. It provides a useful tool to extract the unobserved heterogeneity from the observed income and allows us to express mean and higher moments of log earnings as a function of these unobservables. In doing so, we can determine from observed earnings how cohorts' unobserved heterogeneity differs across sectors and quantify the impact of the unobservables on the earnings distribution. Such analysis is not possible to conduct using the normal distribution.

We empirically test the new distributional setup using two datasets: the US Male and Ghana Male. Our empirical tests reinforces some of the useful properties of modelling income using the NIG distribution. It highlights some of the limitations of the normal distribution and how the NIG provides a more flexible option. Compared to the normal distribution, the fit of the log earnings function is better in both the datasets, the US and Ghana, when we use our proposed normal inverse Gaussian (NIG) distribution. We also find that both the datasets exhibit log earnings distribution that is skewed to the left. This is in contrast to Roy's prediction of right skewness but accords well with other empirical findings. Our empirical tests provide convincing evidence that normal inverse Gaussian distribution provides a more comprehensive way of describing the marginal log earnings function. This provides a good motivation to extend our model further to address the second issue widely discussed in labour literature, that is, modelling the returns to education, where the objective is to model earnings conditional

on education. In other words, we extend our marginal distribution framework to a conditional (on education) distribution case. This sheds lights on the use of this distribution in modelling log earnings function. In particular, we want to determine if the unobserved heterogeneity variable is dependent on the level of education, and, if so, the nature of that dependence. Also, we want to find whether the non-normality that we have witnessed in the two datasets is due to mixing individuals from all education levels or whether it is due to the presence of the unobservables.

2.7 Bibliography

Abramowitz, M. and I.A. Stegun (1970): *Handbook of Mathematical Functions*, 9th Edition, New York: Dover Publications Inc.

Aitchison, J. and J.A.C. Brown (1957): *The Lognormal Distribution, with Special Reference to its Uses in Economics*, Cambridge: Cambridge University Press.

Barndorff-Nielsen, O.E. (1978): "Hyperbolic Distributions and Distributions on Hyperbolae," *Scandinavian Journal of Statistics*, 5, 151-157.

Barndorff-Nielsen, O.E. and P. Blæsild (1981): "Hyperbolic Distributions and Ramifications: Contributions to Theory and Application," in *Statistical Distributions in Scientific Work,* Vol. 4, eds. by C. Taillie, G. Patil, and B. Baldessari, Dordrecht: Reidel.

Barndorff, O.E. and R. Stelzer (2005): "Absolute Moments of Generalized Hyperbolic Distributions and Approximate Scaling of Normal Inverse Gaussian Levy Processes," *Scandinavian Journal of Statistics,* 32(4), 617-637.

Biewen, M. and S.P. Jenkins (2005): "A framework for the decomposition of poverty differences with an application to poverty differences between countries," *Empirical Economics,* 30, 331-358.

Gnedenko, B.V. and A.N. Kolmogorov (1954): *Limit Distributions for Sums of Independent Random Variables*, Cambridge: Addison-Wesley Publishing Company, Inc.

Halgreen, C. (1979): "Self-Decomposability of the Generalized Inverse Gaussian and Hyperbolic Distributions," *Zeitschrift für Wahrscheinchkeitstheorie und verwandle Gebiete,* 47, 13-17.

Heckman, J.J. and B.E. Honoré (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), 1121-1149.

Heckman, J.J. and G. Sadlacek (1990): "Self-Selection and the Distribution of Hourly Wages," *Journal of Labor Economics*, 8(1), S329-S363.

Jenkins, S.P., R.V. Burkhauser, S. Feng and J. Larrimore (2011): "Measuring Inequality using censored data: a multiple-imputation approach to estimation and Inference," *Journal of the Royal Statistical Society A*, 174(Part 1), 63-81.

Jones, A., J. Lomas and N. Rice (2011): "Applying Beta-type Size Distributions to Healthcare Cost Regressions," *Health, Econometrics and Data Group*, WP 11/31.

Kleiber, C. and S. Kotz (2003): *Statistical Size Distributions in Economics and Actuarial Sciences,* Hoboken, New Jersey: John-Wiley & Sons, Inc.

Li, G. and Q. Wang (2003): "Empirical Likelihood Regression Analysis For Right Censored Data," *Statistica Sinica*, 13, 51-68.

Neal, D. and S. Rosen (2000): "Theories of the Distribution of Earnings," in *Handbook of Income Distribution,* eds. A.B. Atkinson and F. Bourguignon. Elsevier.

Nsowah-Nuamah, N., F. Teal and M. Awoonor-Williams (2010): "Jobs, Skills and Incomes in Ghana: How was poverty halved?" *Centre for the Study of African Economies, University of Oxford, UK*, CSAE WPS/2010-01.

Owens, T., J. Sandefur and F. Teal (2011): "Poverty Outcomes and Incomes in Ghana and Tanzania: 1987–2007. Are macro economists necessary after all?" *CSAE 25th Anniversary Conference, St Catherine's College, Oxford*, 20–22nd March 2011.

Roy, A.D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers,* New Series, 3(2, June 1951), 135-146.

TEAL, F. (2010): "Higher Education and Economic Development in Africa: a Review of Channels and Interactions," *Centre for the Study of African Economies, University of Oxford, UK*, CSAE WPS/2010-25.

Chapter 3

Modelling Earnings Function in the Presence of Unobserved Heterogeneity

Abstract: This chapter extends the methodology introduced in the previous chapter to describe the distribution of log earnings conditioned on education. The same two datasets, the US males and Ghanaian males, are used for the empirical analysis. One of the key differences between the two datasets is that skewness and unobserved heterogeneity are functions of education for Ghana but not for the US. Of particular interest is the finding that, for Ghana, once the unobserved heterogeneity is accounted for, the return to education is almost flat for lower levels of education and then increases for education levels higher than ten. The NIG framework is found to be a useful tool to model the unobserved

heterogeneity and understand the impact of the unobserved heterogeneity on the return to education.

3.1 Introduction

In the past fifteen years, an important focus in labour economics has been modelling earnings conditional on education, popularly known as determining the rate of return to education. This study faced two problems in estimating the unbiased return to education. The first is that the return to education may vary by individuals. The second is that if some aspect of the individual is correlated with education but unobserved (ability for short), the OLS estimation of the earnings functions will be biased. The OLS estimates would be biased upward since they would wrongly attribute the positive effect of ability on the earnings function to education. The literature has attempted to address this issue by using instrumental variables (IV); see Card (2001) for an overview. However, as discussed in the first chapter, puzzles in the labour field persist since the IV estimates are as big as or bigger than corresponding OLS estimates (Card 2001). That is, the IV approach tends to overestimate, rather than underestimate, the true return to education. One of the explanations for this puzzle, given by Griliches (1977) and later echoed by Angrist and Krueger (1991), is that the ability biases in the OLS estimates are relatively small; furthermore, the gaps between the IV and the OLS estimates mainly reflect the downward biases related to measurement errors. Many studies have also found other unobserved differences amongst individuals

related to upbringing and family background that could lead to inaccurate estimation of the return to education. We can term these omitted variable biases as the presence of unobserved heterogeneity – the unobserved differences across individuals that influence their earning capacity – that leads to the inaccurate estimation of the return to education when the OLS estimations are used.

Having found in the previous chapter that the normal inverse Gaussian (NIG) distribution is useful in describing the marginal distribution, we extend our model to understand the income distribution conditioned on education. We first discuss the empirical methodology of how we extend the marginal distribution to the conditional distribution case. We discuss the relevant regression framework and the statistical analyses associated with it. Once we build the model and determine the empirical results, we discuss how our findings contribute to the discussions of the existing literature related to earnings functions. In particular we try to focus on the role of the unobservables in the earnings functions. We try to contribute to the literature focusing on how the presence of unobserved heterogeneity affects the return to education. As discussed in Chapter 1, the literature has widely debated the functional form of education in earnings function and how the presence of unobserved heterogeneity influences that functional form. We use this new approach towards modelling and extracting the unobservables from the observed earnings data to determine how the functional form of education in earnings functions is affected due to the presence of unobservables.

Like the previous chapter, our more general form of the normal distribution allows us to address some of the limitations of using the normal distribution. We

can express mean and higher moments as a function of education and the unobservables. In the normal distribution, we could make only the mean a function of education. Our proposed distribution allows for the possibility that the residual (error term) of a regression of log earnings on education could be a function of education and the unobservables. Our proposed distribution helps us to account for the differences in unobserved variance across individuals. The variance is now open to interpretation as it models the unobservables driving the distribution of earnings function. These characteristics help us to take a completely different approach to studying the income distribution conditioned on education by using a general distribution that allows us to do three things. First, our more general distribution allows us to better understand the distributional form of income conditioned on education. Second, we can better express the functional form of log earnings conditioned on education by directly modelling both education and the unobservables, and their relationship with each other. The traditional approach is to leave the unobservables in the residual of the log earnings function, and worry about them only to the extent that they cause bias in estimating the impact of education on earnings. Our approach allows for the possibility that the unobserved variance across individuals is directly included and quantified in the functional form. Third, the more general distribution gives us more tools to understand and model the unobservables. In particular, we are able to test whether education and the unobserved heterogeneity are separable or not. We can also see if the unobservables are present in the mean of the distribution, its higher moments and/or in residual terms.

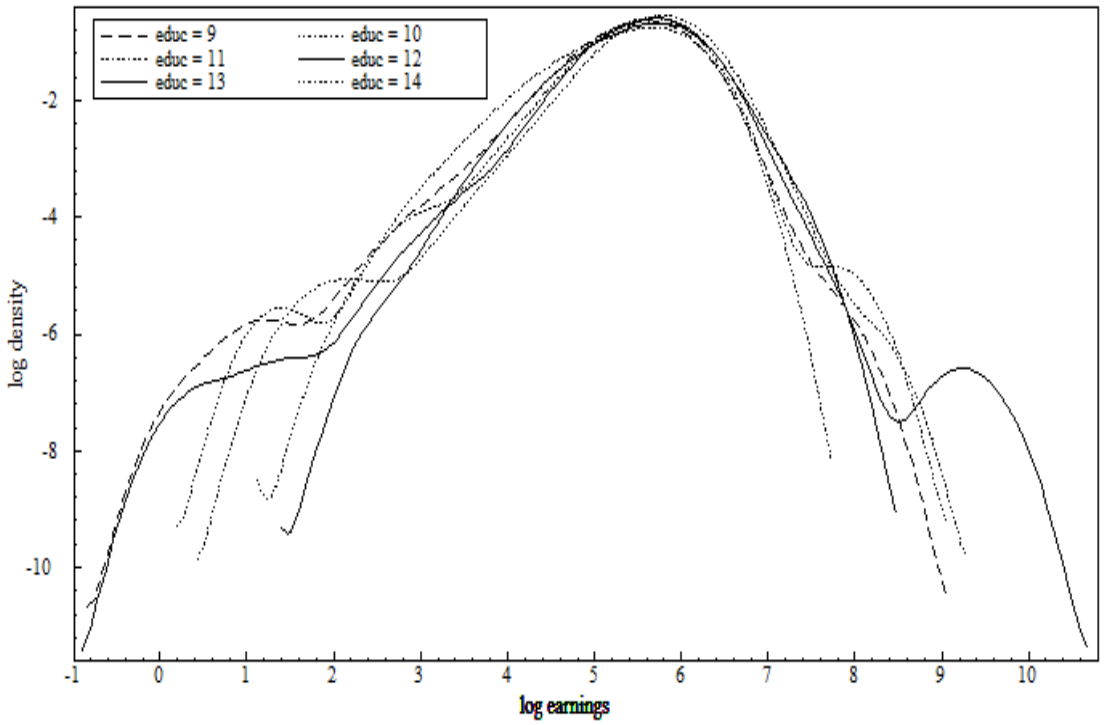

Figure 3-1: US Male: Kernel Estimates of Log Earnings (Conditional on Education)

3.2 The Presence of Unobservables in Income Distribution Conditioned on Education

In the previous chapter, we found that the marginal distribution of log earnings has a mixing feature that could be explained using our proposed distribution, the normal inverse Gaussian distribution. We now move onto the earnings functions. The question is now whether the mixing feature we saw for the marginal distribution of log earnings can be attributed to the education level or whether it has a different source. To see, Figure 3-1 shows kernel estimates of the log density of the log earnings for different levels of education, from 9 to 14 years. The log

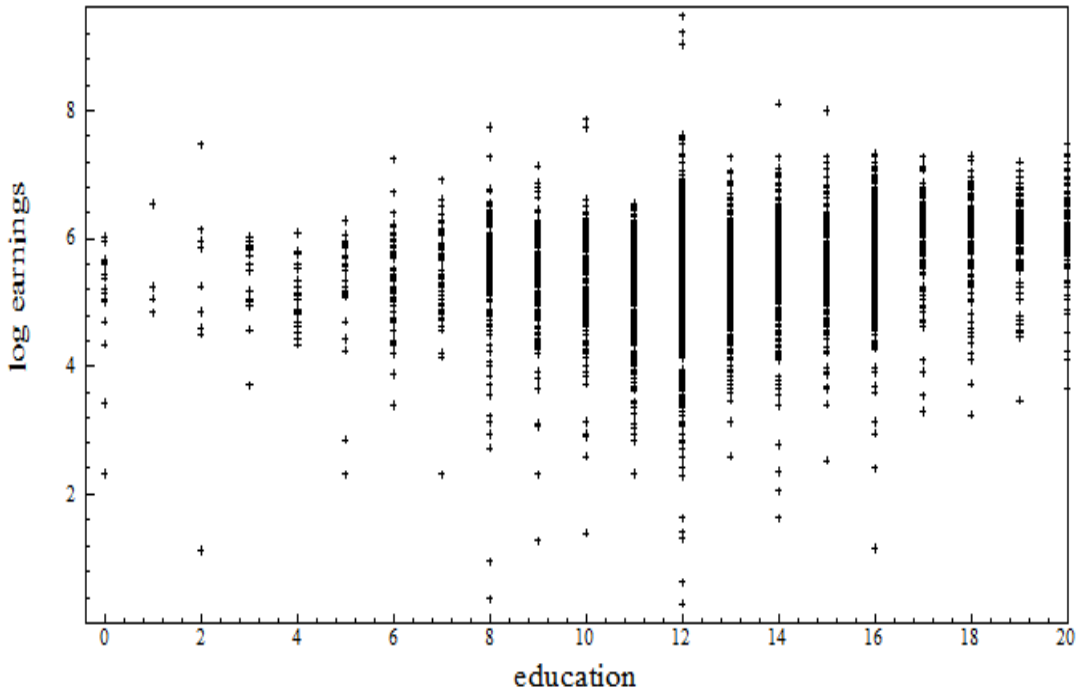

Figure 3-2: US Male: Cross Plot of Log Earnings versus Education

densities appear to have linear tails just as for the marginal distribution in the previous chapter. The mixing feature of the distribution is therefore not a feature of mixing across different levels of education, and must be attributed to some other unobserved heterogeneity.

A different issue is whether the mixing distribution depends on the level of schooling. To illustrate this Figure 3-2 shows a cross plot of log earnings against education. The shape of the conditional distributions given the education does appear to depend on the education level, so that the distribution is different for low levels of education and for higher levels of education. For higher levels of education this interaction is, however, limited. This is also seen from Figure 3-1.

In the following we look at how the unobserved heterogeneity can be adopted

in the standard earnings function.

3.3 Earnings Function and Normal Inverse Gaussian Distribution

The unobserved heterogeneity can be brought into the earnings function just as it was done for the marginal earnings distribution in the previous chapter. Recall that in section 2.3.2, we introduce a distributional setup where V, the linear combination of skills, and M, the mean of aggregate log earning w, are dependent so as to generate heteroskedasticity with respect to M. This was in contrast to Heckman and Honoré's (1990) model, discussed in section 1.1, where V and M were independent. In our proposed alternative setup, the positive unobserved heterogeneity variable σ^2 represented the unobserved skill level or ability of an individual. We showed in Chapter 2 that both variables V and M in log earnings function $w = V + M$ will be functions of σ^2. Conditioning at first on both the unobserved heterogeneity variable σ^2 and the level of education S, the equation for the conditional expectation (2.5) is modified as

$$M = \mu_0 + \mu_1 S + \beta\sigma^2 \tag{3.1}$$

whereas the equation for the conditional distribution component (2.6) is unaltered

$$(V|S, \sigma^2) \stackrel{\mathsf{D}}{=} \mathsf{N}(0, \sigma^2). \tag{3.2}$$

This implies that V is uncorrelated with S and with σ^2 through a similar argument to that for M and V in the previous chapter, see equation 2.7.

We are now in a position to compare our set-up to the Mincerian equation discussed in detail in section 1.3.1. Equation 3.1 looks very similar to the Mincerian equation. In the Mincerian equation, we had decomposed the error term, u_i into $A_i + \varepsilon_i$ (equation 1.5) where A_i was the unobserved ability and ε_i was the structural error. Comparing the Mincerian equation to our set-up, it is possible to let the heteroskedasticity variable σ^2 play of the role of ability as in equation 1.5 while the innovation V plays the role of the structural error ε. An important difference though is that although V and σ^2 are uncorrelated they are dependent with σ^2 resulting in heterogeneity. In other words, the Mincerian equation assumes that the unobserved ability A is independent from the structural error ε. However, we allow for the possibility that the unobserved ability, represented by σ^2 in our set-up, and the structural error ε, represented by the innovation V, are dependent.

From the viewpoint of the Mincer earnings functions we are interested in two different conditional expectations. First, the conditional expectation of log earnings given S, σ^2 is

$$\mathsf{E}(w_i|S, \sigma_i^2) = M = \mu_0 + \mu_1 S + \beta\sigma^2(S). \tag{3.3}$$

Secondly, the conditional expectation of log earnings given S alone is, using equa-

tion 2.8,

$$\mathsf{E}(w_i|S) = \mu_0 + \mu_1 S + \frac{\beta\delta}{\gamma}. \tag{3.4}$$

These conditional expectations result in two candidates for the return to education. First, differentiating equation 3.3 gives

$$\frac{\partial}{\partial S}\mathsf{E}(w_i|S, \sigma^2) = \mu_1. \tag{3.5}$$

If it is deemed reasonable to let the heteroskedasticity variable σ^2 play the role of ability then μ_1 is the return to education. The reason that the return to education can then be identified without involving an instrumental variable is that the ability σ^2 has been given a distributional structure. The second candidate for the return to education arises by differentiating equation 3.4. Bearing in mind that the parameters β, δ, γ could depend on S, we get

$$\frac{\partial}{\partial S}\mathsf{E}(w|S) = \mu_1 + \frac{\partial}{\partial S}\left(\frac{\beta\delta}{\gamma}\right). \tag{3.6}$$

In the case of the US data with higher levels of education, the parameters β, δ, γ do not appear to depend on S in which case this reduces to $\frac{\partial}{\partial S}\mathsf{E}(w_i|S) = \mu_1$. Then S and σ^2 are independent, which would imply that there is no ability bias and a least squares regression of log earnings on education will give a consistent estimate for the marginal return μ_1. In general the parameters β, δ, γ would depend on S. This will be seen for the Ghana data. It would also be seen for the US data with lower levels of education which are however not analysed here.

In order to understand the relationship between education and the unobserved heterogeneity variable, there are various ways of including education in the heterogeneity parameters δ and γ. This implies that, in our regression framework, we can either make δ or γ a function of education, the same way we would make μ a function of education. From an empirical viewpoint some of these variations look alike. From a more theoretical perspective it may be preferable to let the δ-parameter depend on S. This follows from the analysis of Blæsild (1981) who considered the class of two-dimensional generalised hyperbolic distributions, which is generalisation of the normal inverse Gaussian distribution. That class of distributions is closed under conditioning. Indeed, if w and S have joint bivariate generalised hyperbolic distribution then the conditional distribution of w given S could be normal inverse Gaussian where parameter μ is a linear functions of S and δ^2 is a quadratic function of S.

3.4 Empirical Studies

In this section, we model the earnings function using the normal inverse Gaussian distribution. The focus in this section is the statistical analysis. In the next section, we discuss the interpretations of our findings against the backdrop of the existing debates in the labour literature.

We will implement the normal inverse Gaussian methodology for two datasets, namely the US dataset and Ghana dataset. Both the datasets were described in the previous chapter. In addition, Table 3.1 shows how education (educ) is defined

educ	Interpretation	Category of schooling
0	Kindergarten not completed	-
1	Kindergarten completed	KINDERGARTEN
2	Grade 1 completed	ELEMENTARY SCHOOL
3	Grade 2 completed	
...	...	
9	Grade 8 completed	
10	Grade 9 completed	HIGH SCHOOL
...	...	
13	Grade 12 completed	
14	First year of undergraduate degree completed	COLLEGE
...	...	
17	Fourth year of undergraduate degree completed	
18	First year of post-graduate completed	POST-GRADUATE UNIVERSITY
19	Second year of post-graduate completed	
20	Third year of post-graduate completed	

Table 3.1: Explanation of Educ Variable in the Dataset

in the US dataset. Educ runs 0-20 in four broad categories: 0-9 primary, 10-13 high school, 14-17 undergrad, and 18-20 postgrad. Typically, an average child in the USA starts his/her grade 1 at six years of age. Since the dataset is taken from the US Census, it comprises people from various sectors and the analysis is not restricted to a few sectors of the economy only. Therefore, the cross-section analysis is useful because we do not have to worry about price index or other factors that might have an additional impact on different sectors of the economy over time.

3.4.1 Modelling Conditional Distribution - The NIG Regression Methodology

For the empirical work using the US dataset, we consider the cohorts where schooling is equal to or greater than twelve years for two reasons: 1) Our empirical work

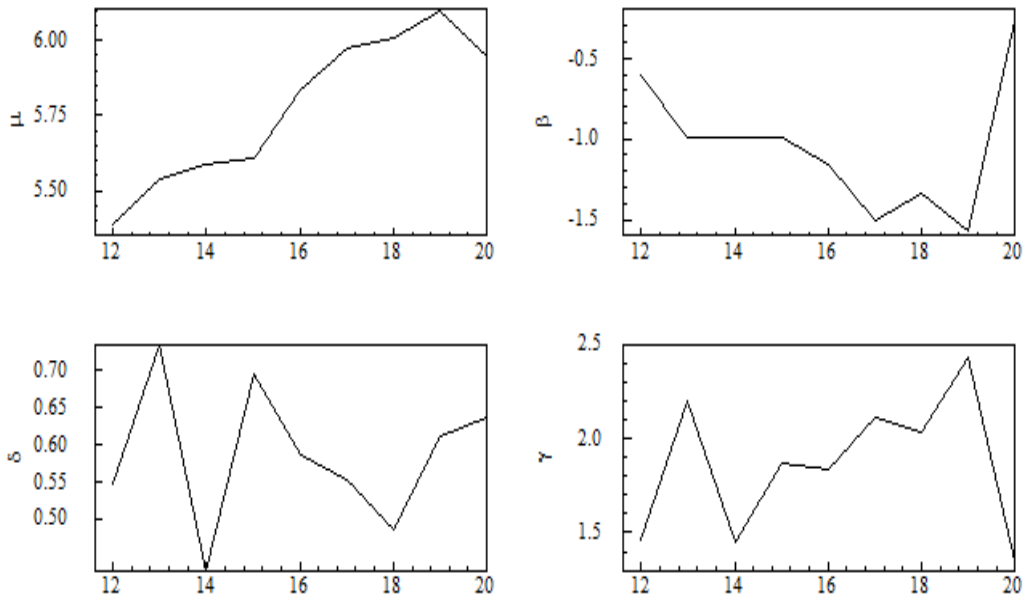

Figure 3-3: US Male - Change of the NIG Parameters with Education Level

on the entire sample supports the notion that there seems to be a structural break at educ$\geqslant$12. This result, seemingly counterintuitive at first, is not so, since in 1980 schooling was compulsory until 16 years of age in almost all the states of America (Angrist and Krueger (1991, pp. 1013)). Sixteen years of age corresponds to the completion of grade 10, which is given by eleven years of schooling ($educ$=11). Therefore, the cut-off at twelve years of schooling merely shows the self-selection of people. Those who do not want to continue schooling or are unable to do so due to financial constraints will have eleven years of schooling (until 16 years of age). 2) The number of observations for educ>11 is 3965 (77% of entire sample) and the observations are more or less evenly divided across each education level. On the other hand, for lower education levels, there are cases where the number

of observation for a particular education level is too low to conduct any rigorous analysis.

Figure 3-3 shows how the parameters $\beta, \gamma, \delta, \mu$ change with education level. It is seen that the μ-parameter develops in a more or less linear fashion, while the β, γ, δ-parameters could be piecewise constant, if not constant. There seems also a jump in the parameters for education level of 20, particularly in β and γ. If we plot β/γ against education level for educ=12 to educ=20, the function first remains almost constant, then slightly declines and increases substantially for educ=20. In our empirical analysis, we test if the jumps at educ=20 is statistically significant by using dummies for educ=20 and testing if they are statistically significant. We add the dummy for all the four parameters, each at a time and also together for β and γ. It turns out that educ=20 is not statistically significant in any of these cases.

The Saturated Model

Our concern is how wages depend on education. Initially we would like to make a very flexible functional and then find a statistically valid simpler functional form. Therefore we start by fitting an NIG distribution for each level of education, starting from twelve years of schooling. We also control for the number of years of experience in μ to make our model more robust. This is because in typical log earnings function, experience is controlled for when the impact of education on log-earnings is investigated (Mincer 1974). For both the US and Ghana dataset, experience is defined as age minus the sum of education and six (that is, experience

= age-education level-6). For each level of education, starting from educ =12 to educ =20, we have estimates of parameters $\beta, \gamma, \delta, \mu$ giving a total of $4 \times 9 = 36$ parameter values plus 1 more for the inclusion of experience in μ (control). The total number of parameters in the saturated model is thus 37. The resulting model for the log earnings is a NIG regression of the form

$$\begin{aligned} w \quad | \quad & educ = NIG(\mu, \delta, \gamma, \beta) \\ \delta \quad = \quad & \delta_0 + \sum_{j=12}^{19} \delta_j D_j, \qquad \gamma = \gamma_0 + \sum_{j=12}^{19} \gamma_j D_j, \\ \beta \quad = \quad & \beta_0 + \sum_{j=12}^{19} \beta_j D_j, \qquad \mu = \mu_0 + \sum_{j=12}^{19} \mu_j D_j + \exp er, \end{aligned} \tag{3.7}$$

where D_j is an indicator variable taking the value unity if the level of education is j and zero otherwise.

The log likelihood value of the saturated model is -3466. This is used as the benchmark when further reduction is carried out (using general-to-specific algorithm) to determine the best reduced model. We will choose a mixture of two strategies to find a statistically valid simple functional form.

The General-to-Specific Algorithm

In the first strategy we are inspired by the independent work of on the one hand Cox and Snell (1974) and the other hand Hoover and Perez (1999) and Hendry and Krolzig (2005). A simple exposition of the algorithm is given in Hendry and Nielsen (2007). In simple terms, the idea is to conduct a multiple path reduction of the general unrestricted model (GUM) and reduce the model as far as possible

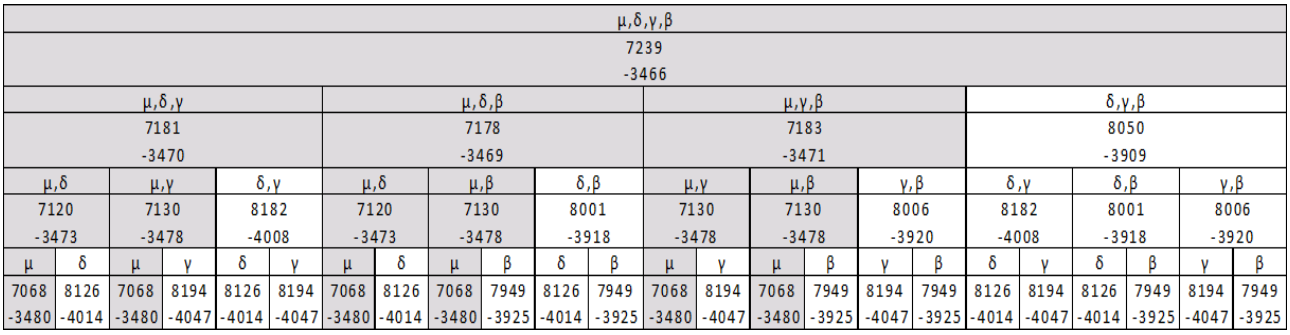

Figure 3-4: US Male - The General Unrestrictd Model Tree Showing All the Possible Paths (shaded areas denote possible reduction paths). First value shows value according to Bayesian Information Criterion, second value is log likelihood value.

by eliminating one parameter each time. GUM is "the most general, estimable, statistical model that can reasonably be postulated initially, given the present sample of data, previous empirical and theoretical research, and any institutional and measurement information available. By estimable is meant that the parameter can be uniquely determined from the evidence. The GUM is also formulated to contain the parsimonious, interpretable, and invariant econometric model at which it is hoped the modelling exercise will end." (Hendry (1995, pp. 361)).

The saturated model is given by equation 3.7. In this case, the general unrestricted model is the saturated model, that is, the model which contains one dummy for each education level for each parameter. As mentioned above, the saturated model will contain 37 variables. In each reduction, 8 variables for one parameter are removed. In other words, for each reduction, there are 8 restrictions. The only exception is when the parameter μ is removed. In that case, 9 variables are removed, 8 dummy variables for education and 1 control variable for experience. Since the first parameter to be removed could be either μ, δ, β or γ,

the first stage of reduction could be done in 4 ways. If we define each path as successful reduction from the general unrestricted model to the model containing a single parameter, there are $4! = 4*3*2 = 24$ possible paths. However, we do not have to reach the end of the path if we reach a non-rejected model beforehand. Such a non-rejected model is known as a terminal model. Therefore, we carry successive eliminations for each path until no variables can be eliminated. We use log likelihood values to determine if a terminal model is reached or further reduction is possible. The log likelihood values and the BIC criteria are discussed in Appendix A. Figure 3-4 shows the terminal nodes and the possible reduction paths in grey. Using a general-to-specific algorithm, we find that all the terminal models end in μ and a combination of (μ, δ), (μ, γ) and (μ, β).

In simple terms, the node diagram in Figure 3-4 starts from the first node (top node in the diagram) containing all the dummies for all the parameters μ, δ, β and γ. From there, we try to reduce by removing the dummies of one parameter in each reduction. If the log likelihood values permits the reduction, we colour that branch grey. Otherwise, we stop conducting further reduction and keep the branch white (where reduction is not possible). Going back to Figure 3-4, this means that we can successfully reduce from $(\mu, \delta, \beta, \gamma)$ to (μ, δ, γ) to (μ, δ) to (μ). However, once we reach (μ, δ), we cannot reduce to (δ). Hence (δ) is shaded white. In a similar fashion, as marked by the grey shades, we can reduce from $(\mu, \delta, \beta, \gamma)$ to (μ, δ, γ) to (μ, γ). From (μ, γ), as indicated by the grey and white shades respectively, we can reduce to (μ) but not to (γ). From $(\mu, \delta, \beta, \gamma)$ we cannot reduce to (δ, β, γ). Hence the remaining branches of the tree are all left white

for that segment. These node diagrams give us an intuitive indication of which parameters are varying with education level and gives an intuitive idea of what to expect in our second strategy (discussed below).

In the second strategy, we use the finding from the first strategy to determine the functional form. We start with a simple NIG regression where μ is a linear function of education and experience, and the other three parameters are constants. We then include education in linear form in γ and β and check log likelihood values to test if the addition is statistically significant. Similarly, we carry a series of regressions, involving higher order polynomials of education in γ, β, δ and μ, to determine the correct functional form. We resort to a chi-square table, see Appendix A, to select the best model.

With only μ depending on education we get likelihood and BIC taking values

$$\ell = -3480, \qquad BIC = 7068.$$

3.4.2 Results for the US

The best reduced model is summarized below. The interpretation of the results will be discussed in Section 3.5. Recall that the model has two components, the conditional model $(w \mid \sigma^2) \overset{\mathsf{D}}{=} \mathsf{N}(\mu + \beta\sigma^2, \sigma^2)$, see equation 2.4, and the mixing distribution $\sigma^2 \overset{\mathsf{D}}{=} \mathsf{IG}(\delta, \gamma)$, see equation 2.3. Initially all the parameters $\mu, \beta, \delta, \gamma$ were specified as functions of education with μ also depending on experience. It was found β, δ, γ could be taken as constants. In the best reduced model μ

reduces to a simple linear function of education and experience. The estimated parameters for the conditional model are

$$\widehat{\mu}_i = \underset{(31.8)}{4.38} + \underset{(24.6)}{0.09} S_i + \underset{(28.8)}{0.02} X_i \tag{3.8}$$

$$\widehat{\beta} = \underset{(13.5)}{-0.74} \tag{3.9}$$

while the parameters for the mixing distribution are

$$\widehat{\delta} = \underset{(27.5)}{0.55}, \qquad \widehat{\gamma} = \underset{(15.7)}{1.57} \tag{3.10}$$

The figures reported below are signed likelihood ratio statistics which are interpreted in the same way as t-statistic in regression analysis. The log likelihood values are

$$\ell = -3493, \qquad BIC = 7036, \qquad n = 3965$$

This reduced model is therefore preferred according to BIC whereas a conventional likelihood ratio test statistic would have a modest p-value of 0.001. The estimates for the conditional expectations (3.3), (3.4) are then

$$\begin{aligned} \mathsf{E}(w_i|S_i, X_i, \sigma^2) &= 4.38 + 0.09S_i + 0.02X_i - 0.74\sigma^2, \\ \mathsf{E}(w_i|S_i, X_i) &= 4.38 + 0.09S_i + 0.02X_i + \frac{(0.55)(-0.74)}{1.57} \\ &= 4.12 + 0.09S_i + 0.02X_i. \end{aligned}$$

The comparable least squares regression model is

$$\widehat{w}_i = \underset{(0.004)}{4.09} + \underset{(0.007)}{0.09}\, S_i + \underset{(0.007)}{0.02}\, X_i, \tag{3.11}$$

$$\widehat{\sigma}^2 = 0.43, \qquad \ell = -3970. \tag{3.12}$$

Since the unobserved heterogeneity parameters δ and γ do not depend on education and experience the least squares estimate for the mean is very similar to the estimate for $\mathsf{E}(w_i|S_i, X_i)$ based on the normal inverse Gaussian mixture as discussed in section 3.3.

It must be mentioned that the results seem fairly robust for the two steps used to derive the best results: the reduction from the saturated model and the determination of the functional form. In the first strategy, Figure 3-3 supports the finding that μ be a function of education. When we try to reduce by excluding μ from our regression, Figure 3-4 clearly rejects that reduction. The labour literature also supports the idea that the location parameter of log earnings will be a function of education. In the second strategy, we try different functional forms. In particular, we try to have higher polynomials for μ and also include higher polynomials for δ, γ and β. We do not find evidence of any relationship between education and the other parameters (apart from μ). Our results are robust to alternative specifications.

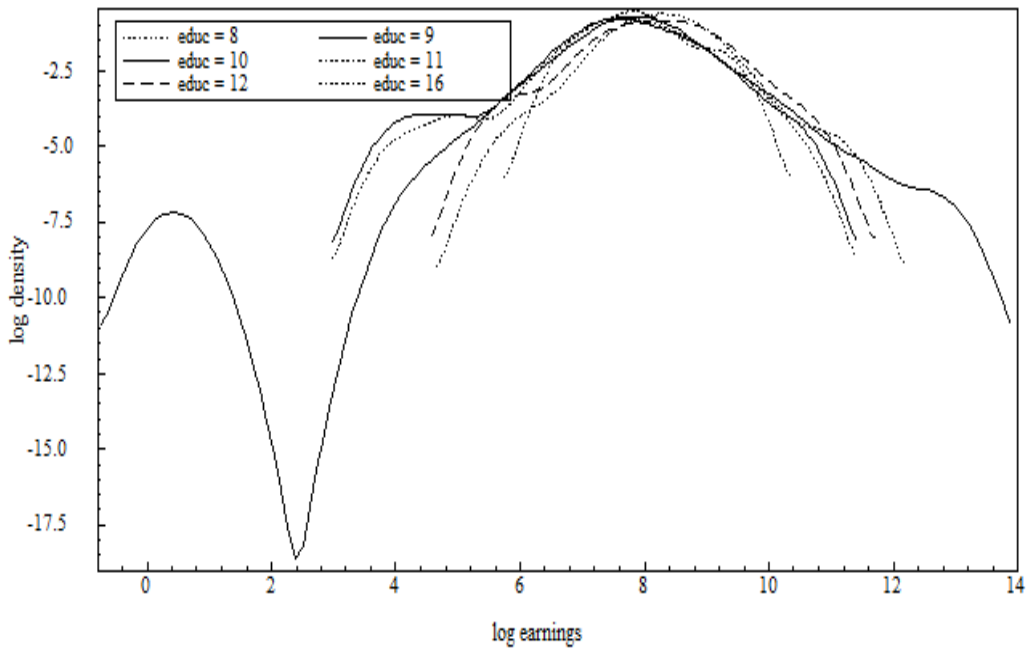

Figure 3-5: Ghana Male: Kernel Estimates of Log Earnings (Conditional on Education)

3.4.3 Results for Ghana

We conduct the same empirical analysis on the Ghana dataset. The dataset described in the previous chapter. Like the US dataset, we can conclude from Figure 3-5 that the mixing feature of the distribution is not a feature of mixing across different levels of education, and must be attributed to some other unobserved heterogeneity. Figure 3-6 shows how the parameters $\beta, \gamma, \delta, \mu$ change with each education level. We find that the parameter δ changes with the education level and seems to be a function of education.

We conduct the same two stages (general-to-specific algorithm and reduction) to deduce the best reduced model for NIG and compare it to the results derived

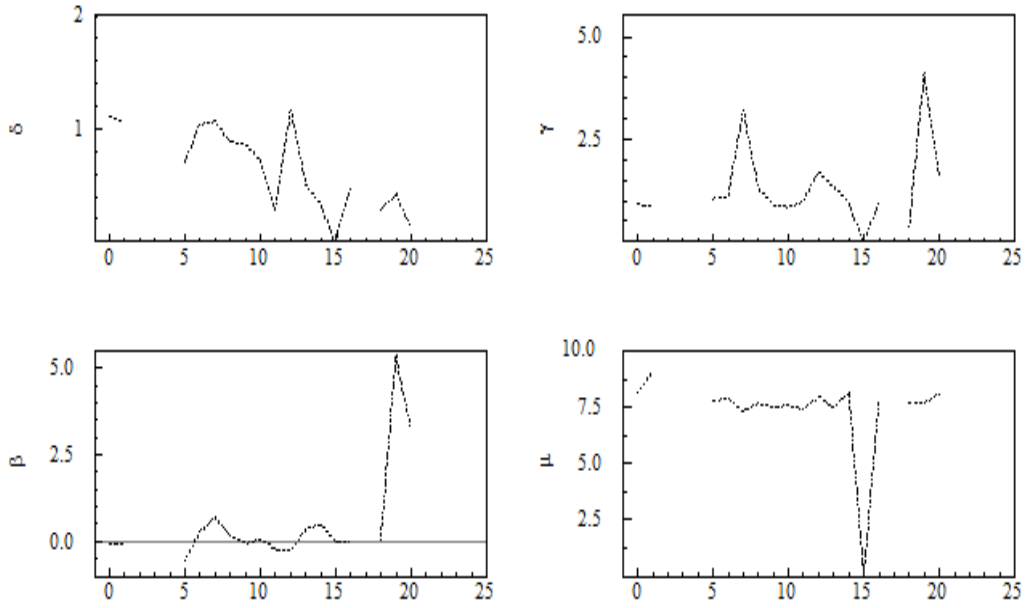

Figure 3-6: Ghana Male - Change of the NIG Parameters with Education Level from OLS (normal distribution). The saturated model has the same form as the US:

$$
\begin{aligned}
w \quad | \quad & educ = NIG(\mu, \delta, \gamma, \beta) \qquad (3.13)\\
\delta \quad = \quad & \delta_0 + \sum_{j=1}^{24} \delta_j D_j, \qquad \gamma = \gamma_0 + \sum_{j=1}^{24} \gamma_j D_j,\\
\beta \quad = \quad & \beta_0 + \sum_{j=1}^{24} \beta_j D_j, \qquad \mu = \mu_0 + \sum_{j=1}^{24} \mu_j D_j + \exp er,
\end{aligned}
$$

For Ghana, the saturated model contains 101 variables (25 variables for each parameter plus 1 variable for experience in μ). In each reduction, 24 variables for one parameter are removed. In other words, for each reduction, there are 24 restrictions. The only exception is when the parameter μ is removed. In that

μ,δ,γ,β					
7967 -3581					
μ,δ,β					
7800 -3593					
μ,δ		μ,β		δ,β	
7636 -3607		7728 -3653		7762 -3674	
μ	δ	μ	β	δ	β
7552 -3660	7795 -3786	7552 -3660	7646 -3712	7795 -3786	7646 -3712

Figure 3-7: Ghana Male - One of the Branches of the General Unrestricted Model Tree Showing the Possible Paths (shaded areas denote possible reduction paths). First value shows value according to Bayesian Information Criterion, second value is log likelihood value.

μ,δ		μ,γ		μ,β		δ,β		γ,β	
7636 -3607		7654 -3616		7728 -3653		7762 -3674		7723 -3654	
μ	δ	μ	γ	μ	β	δ	β	γ	β
7552 -3660	7795 -3786	7552 -3660	7805 -3791	7552 -3660	7646 -3712	7795 -3786	7646 -3712	7805 -3791	7646 -3712

Figure 3-8: Ghana Male - Testing if Reduction is Possible from Two to One Parameter (shaded areas denote possible reduction paths). First value shows value according to Bayesian Information Criterion, second value is log likelihood value.

case, 25 variables are removed, 24 dummy variables for education and 1 control variable for experience. While conducting the general-to-specific algorithm and starting from a saturated model, we find that there are some combinations of parameters where our programs do not offer any log likelihood values, indicating that those reduction paths are not supported by the data. Figure 3-7 shows one of the possible branches of the general unrestricted model. Unlike the US where we could reduce till μ in all cases, we find that in Ghana, the terminal node includes

μ and β (denoted by the last shaded region in Figure 3-7). Figure 3-8 shows the possible reduction channels when we move from two parameters to one parameter. Again, unlike the US case, we find that the terminal nodes containing μ and β. This hints that the relationship between log earnings and education may be more complicated in Ghana than the US. In particular, we find parameters other than μ depending on education. We confirm our intuition by using our second strategy discussed below.

In our second strategy, we try to find the exact functional form of the relationship between log earnings and education levels. In the case of Ghana male, we consider the entire sample since we have enough observations across each education level. We also use a cut-off at ten years of education ($educ<10$ and $educ\geq 10$). This is because secondary school in Ghana ends after nine years of education, the point at which students can choose not to attend school. Thus, self-selection starts at the end of nine years of education. In Ghana, along with controlling for experience, we need to control for waves because the household data used merges information from four waves. Although we use real earnings, controlling for waves helps to get rid of controversies related to the inflation index.

Compared to the normal distribution, the best results using NIG have a simpler functional form for cohorts with less than ten years of schooling. Also, we find that across all education levels, δ is a function of education. As explained in the later sections, this is an important finding since this implies that the unobservables cannot be separated from education. The log likelihood values are also better for NIG compared to normal distribution.

Educ $<$ 10:

The best reduced model is summarized below. The interpretation of the results will be discussed in section 3.5. Recall that the model has two components, the conditional model $(w \mid \sigma^2) \stackrel{\mathsf{D}}{=} \mathsf{N}(\mu + \beta\sigma^2, \sigma^2)$, see equation 2.4, and the mixing distribution $\sigma^2 \stackrel{\mathsf{D}}{=} \mathsf{IG}(\delta, \gamma)$, see equation 2.3. The estimated parameters for the conditional model are

$$\widehat{\mu}_i = \underset{(98.5)}{7.88} + \underset{(2.65)}{0.005} X_i \tag{3.14}$$

$$\widehat{\beta} = \underset{(0.48)}{-0.03} \tag{3.15}$$

while the parameters for the mixing distribution are

$$\widehat{\delta}_i = \underset{(7.7)}{1.15} - \underset{(2.2)}{0.04} S_i, \qquad \widehat{\gamma} = \underset{(6.6)}{0.99} \tag{3.16}$$

The figures reported below are signed likelihood ratio statistics which are interpreted in the same way as t-statistic in regression analysis. The log likelihood values are

$$\ell = -1354, \qquad n = 962$$

The estimates for the conditional expectations (3.3), (3.4) are then

$$\begin{aligned} \mathsf{E}(w_i|S_i, X_i, \sigma^2) &= 7.88 + 0.005X_i - 0.03\sigma^2, \\ \mathsf{E}(w_i|S_i, X_i) &= 7.88 + 0.005X_i + \frac{(1.15 - 0.04S_i)(-0.03)}{0.99} \end{aligned}$$

The comparable least squares regression model is

$$\begin{aligned} \widehat{w}_i &= \underset{(39.2)}{7.50} - \underset{(2.3)}{0.27} S_i + \underset{(2.1)}{0.08} S_i^2 - \underset{(2.0)}{0.005} S_i^3 && (3.17) \\ &\quad + \underset{(2.3)}{0.02} X_i - \underset{(3.4)}{0.0004} X_i^2 + \underset{(2.6)}{0.20} wave_3, && (3.18) \\ \widehat{\sigma}^2 &= 1.01, \qquad \ell = -1373. && (3.19) \end{aligned}$$

Educ $\geq$ 10

The estimated parameters for the conditional model are

$$\begin{aligned} \widehat{\mu}_i &= \underset{(0.80)}{6.66} + \underset{(15.5)}{0.08} S_i + \underset{(10.0)}{0.016} X_i && (3.20) \\ \widehat{\beta} &= \underset{(2.0)}{0.09} && (3.21) \end{aligned}$$

while the parameters for the mixing distribution are

$$\widehat{\delta}_i = \underset{(5.5)}{1.94} - \underset{(3.1)}{0.16} S_i + \underset{(2.5)}{0.004} S_i^2, \qquad \widehat{\gamma} = \underset{(1.4)}{0.93} \tag{3.22}$$

The log likelihood values are

$$\ell = -2257, \qquad n = 1913$$

The estimates for the conditional expectations (3.3), (3.4) are then

$$\begin{aligned}
\mathsf{E}(w_i|S_i, X_i, \sigma^2) &= 6.66 + 0.08S_i + 0.016X_i + 0.09\sigma^2, \\
\mathsf{E}(w_i|S_i, X_i) &= 6.66 + 0.08S_i + 0.016X_i + \frac{(1.94 - 0.16S_i + 0.004S_i^2)(0.09)}{0.93}
\end{aligned}$$

The comparable least squares regression model is

$$\begin{aligned}
\widehat{w}_i &= \underset{(0.5)}{0.95} + \underset{(3.1)}{1.24}S_i - \underset{(2.8)}{0.08}S_i^2 + \underset{(2.8)}{0.002}S_i^3 && (3.23) \\
&\quad + \underset{(7.9)}{0.05}X_i - \underset{(5.9)}{0.0007}X_i^2, && (3.24) \\
\widehat{\sigma}^2 &= 0.86, \qquad \ell = -2422. && (3.25)
\end{aligned}$$

3.5 Interpretation of the Best Model

This section investigates what we have learnt from the new model constructed using the normal inverse Gaussian (NIG) distribution. The interpretation of the model is divided into four sections: (1) distribution, (2) functional form, (3) unobserved heterogeneity, (4) putting it all together.

3.5.1 Distribution

As discussed in section 1.1, the skewness of log earnings distribution has been a subject of scrutiny since the inception of Roy's selection model which predicted that log earnings should be skewed to the right (Roy 1951). Our best model for the US male dataset shows that the log earnings distribution is actually left skewed.

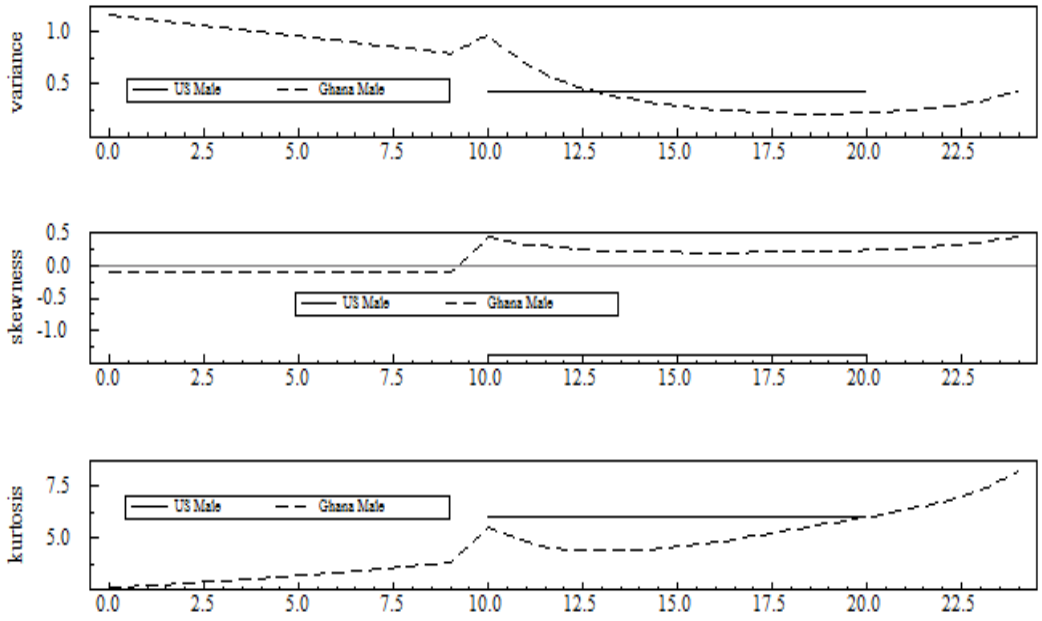

Figure 3-9: Plot of Skewness, Variance and Kurtosis versus Education

To illustrate this, Figure 3-9 plots the skewness of log earnings distribution for each education level. This is in accordance with what we found in Chapter 2 about the marginal distribution of log earnings in the US. We can conclude that the log earnings distribution in the US is skewed to the left and this skewness does not depend on the level of education. Similarly, Figure 3-9 shows that the kurtosis and variance also do not depend on the level of education. This hints that the separability between unobserved heterogeneity and education will probably be less of an issue for the US dataset since the higher moments of log earnings do not depend on education.

The more interesting case is exhibited by the best model derived from Ghana male dataset where δ was a function of education. As Figure 3-9 shows, variance,

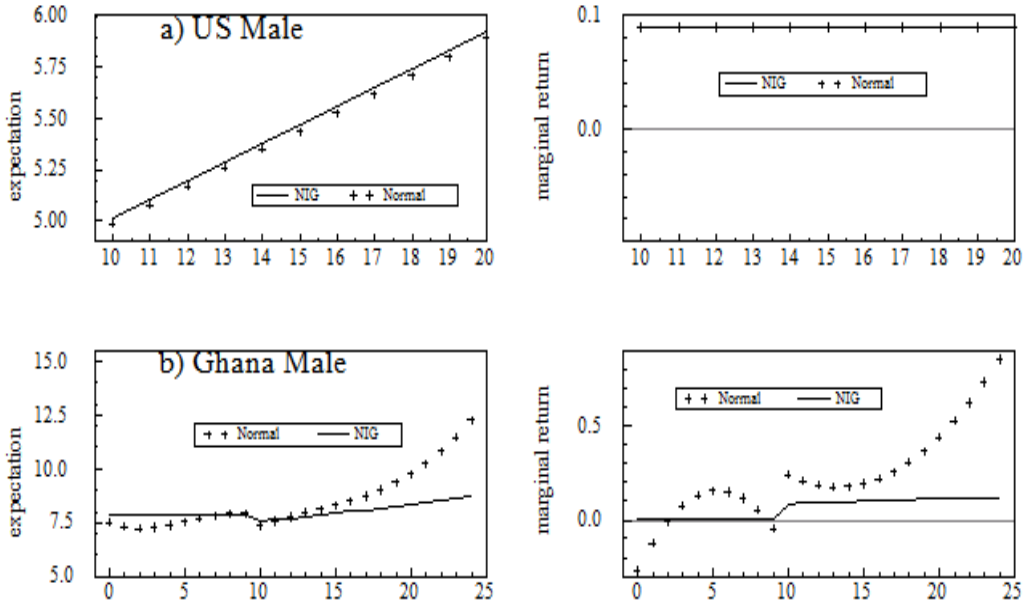

Figure 3-10: Plot of Expectation and Marginal Return versus Education

skewness and kurtosis are functions of education, a characteristic that would not be picked up if normal distribution were used instead. Here, the distribution is left skewed before ten years of education and right skewed beyond that. That is, after ten years of education, the economy follows the prediction of log normal skill Roy model. The implications of this finding will be discussed in section 3.5.4, after the discussion on unobserved heterogeneity.

3.5.2 Functional Form

Recall from section 1.3 that the key question in the labour market discussions revolves around the functional form of log earnings conditional on education. The assumptions leading to Mincerian equation would imply that log earnings is linear

in education. However, empirical studies have found both concave (Psacharopoulos (1994)) and convex (Belzil and Hansen (2002)) shapes. In the US, we find that log earnings is linear in education for education levels greater than eleven. Since, education and unobserved heterogeneity is separable for the US, we find that the results with the NIG distribution are comparable to that of the normal distribution estimates. As Figure 3-10 shows, the marginal return is also comparable to the normal OLS counterpart.

Figure 3-10 compares the expected value of log earnings and marginal return to education using NIG and normal distributions for Ghana. Using NIG methodology, the expected value of log earnings is almost flat for cohorts with less than ten years of schooling, and rises less steeply than the outcome with normal distribution for cohorts with more than ten years of education. For cohorts less than ten years of schooling, the marginal return is zero (flat) using NIG but concave using normal distribution. After ten years of education, the return is less for NIG, by around 5%. Towards very high levels of education (towards the very end of the level of education), we find slight signs of concavity in the NIG case as marginal return drops, albeit very slowly.

One of the potential explanations for the difference between our NIG and OLS results, is that, we find that the convexity of the curve decreases remarkably in the NIG case. This is because the NIG framework is able to extract the unobserved heterogeneity which is responsible for the convexity of curves. The more we can account for the unobserved heterogeneity, the earnings functions become more linear in shape as predicted by the Mincerian approach. This gives a good signal

that our approach can account for the presence of the unobservables responsible for the convexity in the earnings functions.

The most interesting insight from the Ghanaian dataset is that, even after accounting for the unobserved heterogeneity which is potentially decreasing the convexity, the earnings function is highly convex in the sense that the return to education jumps remarkably after ten years of education. This indicates that people who continue to pursue education higher than ten years receive a significant boost in their income due to their increased years of education. Most work done in this field also shows that the return to education is high for higher levels of education in Ghana (Rankin, Sandefur and Teal (2010)). Teal (2010) uses data from thirty two African countries, including Ghana, to show that the returns to education, measured both by macro production functions and by micro earning functions, are the highest for those with higher levels of education, indicating that the earnings functions are highly convex in African countries.

3.5.3 Unobserved Heterogeneity

In this section, we introduce some of the advantages of modelling returns to education using NIG regression, rather than normal distribution, and show what it implies for our results derived from the US and Ghana datasets. NIG contains four parameters that can, theoretically, be a function of education. Hence, we can model skewness and higher order moments as a function of education. This flexibility is found to be extremely useful for labour markets where the unobserved

heterogeneity is a function of education, evident from our second empirical work on Ghana dataset.

Non-separability of Heterogeneity and Education

Recall that σ^2 contains the parameters δ and γ (equation 2.3). The key difference between the USA and Ghana is that Ghana reveals non-separability between education and unobserved heterogeneity since for Ghana, δ is a linear function of education. This implies that we cannot separate the effect of unobservables from education because unobserved heterogeneity is itself a function of education. When we control for unobserved heterogeneity, we are in effect controlling for a portion of education as well. This is a critical finding because labour economists have widely attempted to separate the effect of unobserved heterogeneity from education using the Instrumental Variable Method (IV), discussed in section 1.3.2. We give below a very brief mathematical exposition to highlight the problems faced when IV approach is used in such a case. This can be seen as an extension of our analysis in section 3.3.

Following the terminology used in section 3.3, when δ is a function of S, the least squares estimator is no longer consistent, but the parameters can be estimated consistently by maximum likelihood. It is not so clear how to find an instrument that will deliver a consistent instrumental variable estimator. Suppose Z is a candidate instrument giving an estimator

$$\widehat{\mu}_{1,IV} = \frac{\sum_{i=1}^{n}(w_i - \overline{w})Z_i}{\sum_{i=1}^{n}(S_i - \overline{S})Z_i}.$$

In section 1.3.2 we discussed that there are two moment conditions required to ensure consistency. First, the instrument Z should be correlated with schooling S. Second, the instrument Z should be uncorrelated with the unobserved heterogeneity. In our case, this would imply that the instrument Z should be uncorrelated with $\beta\sigma^2 + V$. Recall that we have presented the log earnings as $w = M + V$, where V is the linear combination of skills and M is the mean of aggregate log earning w, discussed in sections 1.1 and 3.3 . The component $\beta\sigma^2$ comes from the unobserved heterogeneity present in M (equation 3.1). As discussed in section 3.3, V plays the role of structural error ε. Although V and σ^2 are uncorrelated, they are dependent with σ^2 resulting in unobserved heterogeneity. Therefore, the component V needs to be added to $\beta\sigma^2$. Thus, the moment conditions to ensure consistency are that

$$\mathsf{Cov}(S, Z) \neq 0, \qquad \mathsf{Cov}(\beta\sigma^2 + V, Z) = 0.$$

Since σ^2 has an inverse Gaussian distribution with δ depending on S it would be rather complicated to find an appropriate instrument Z. Intuitively, the fact that δ is a function of S in a normal inverse Gaussian distribution implies that if we had used normal distribution instead, education would have been present in the error term and be a function of the unobserved heterogeneity in the error term. Under such a scenario, it would be rather tricky, if not impossible, to find a suitable IV that would meet the consistency requirement. This issue of potential problems of using IV method when education is not separable from unobserved

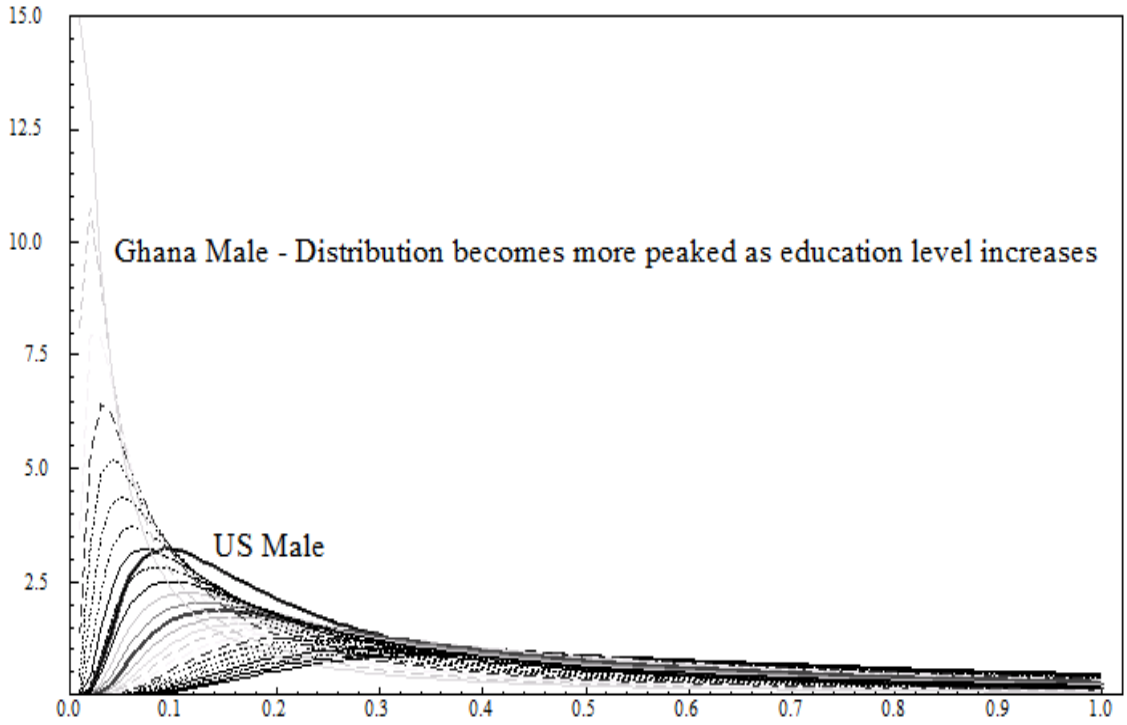

Figure 3-11: Distribution of the the Unobserved Heterogeneity Variable. Each curve represents the distribution of sigma-square for a given level of education. Using the results of Ghana, the curves show that the distribution becomes more peaked as education level increases. For the US, one single curve represents the unobserved heterogeneity distribution since the heterogeneity parameters do not depend on education.

ability has also been highlighted by Card (2001) and Heckman, Lochner and Todd (2009). The results from the NIG show that a solution to this problem requires an interaction term between education and unobserved heterogeneity on the right side of the log earnings function.

The Distribution of Unobserved Heterogeneity

Using the values of the parameters of the best model, we can carry out a lot of analysis that cannot be done with the normal distribution. Using the parameters δ and γ, we can plot the distribution of the unobserved heterogeneity. Figure 3-11

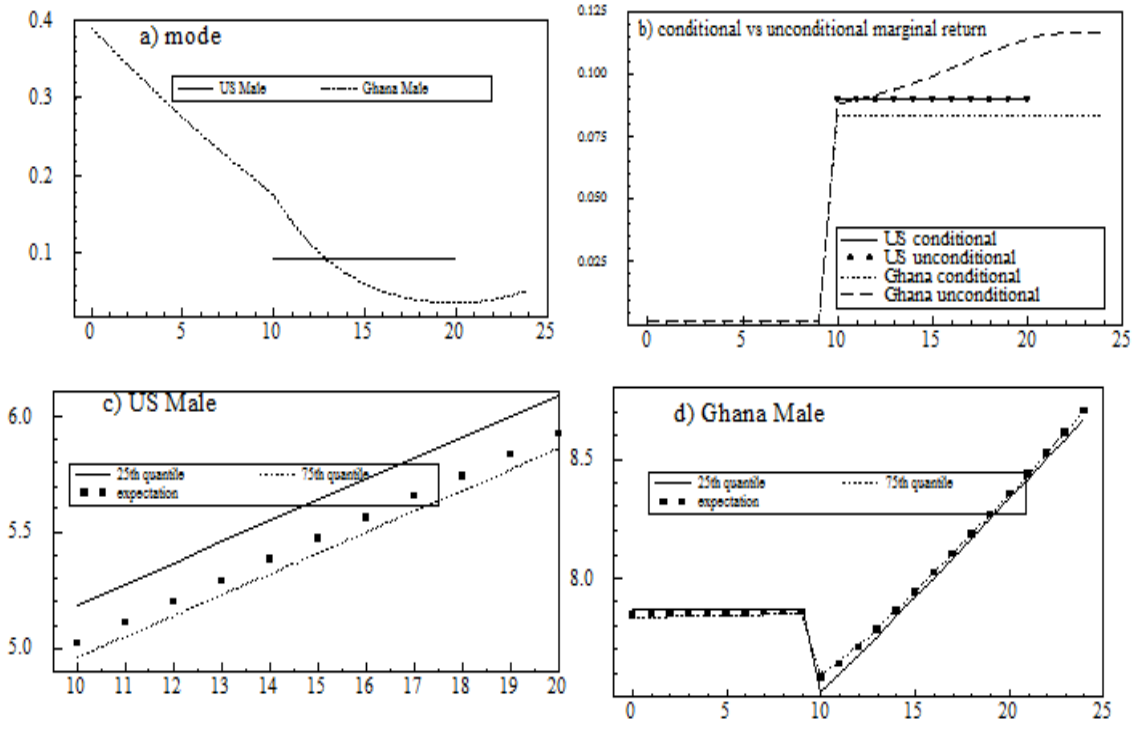

Figure 3-12: a) Mode of the Normal Inverse Gaussian distribution, b) The Unconditional and Conditional Marginal Returns (see equations 3.4 and 3.3). The Conditional Marginal Returns is Conditioned on Sigma-square; and c)/d) The Conditional Log Earnings for Different Quantile Values of Sigma-square. The x-axis of all the charts represents levels of education.

does this for the best reduced models. The advantage of the NIG is that it gives us the tools to observe how the distribution of the unobserved heterogeneity variables varies for each education level. Figure 3-11 shows that, as the education level increases, the function becomes more peaked and tends to collapse to a normal distribution. This implies that as education increases, the distribution becomes less heterogeneous. Also, as Figure 3-12 shows, the mode of the distribution for Ghana changes with education. This hints that there might a case of self-selection in Ghana labour market, in accordance with Rankin, Sandefur and Teal (2010). In their paper, Rankin, Sandefur and Teal (2010) find evidence of a pattern of

waiting by which there is a preference for a public or a large firm job, and as the probability of obtaining one declines the worker switches to the informal sector. This hints that selection is an important determinant of earnings in Ghana.

Conditional Log Earnings Function

The advantage of using NIG is that we have the tools to find the marginal return of education conditioned on σ^2, the ensemble of unobservables. That is we can find $\frac{\partial(y|\sigma^2)}{\partial educ}$ since $y \mid \sigma^2 = \mu + \beta\sigma^2$ The mode of the distribution, on average, is at σ^2= 0.09 and we derive μ and β values from the coefficients of the best reduced models. Figure 3-12 compares the unconditional marginal distribution of NIG and the conditional marginal distribution of NIG (conditioned on σ^2=0.09). The conditional marginal return is 9% for the US male, the same as the unconditional marginal return. For Ghana, the conditional marginal return is lower than the unconditional one. More importantly, we find that, once we condition for σ^2, the ensemble of unobservables, we find that the marginal return is constant. In other words, the convexity found using the normal distribution is attenuated to a greater degree when we use the NIG distribution. The small amount of convexity that is left behind is also removed when we condition the marginal return on σ^2. We can conclude that the convexity in the shape of earnings function can be attributed to the presence of the unobservables.

Quantiles of σ^2

The next step is to investigate what happens to log earnings function as σ^2 changes its values? As mentioned in the previous section, $y \mid \sigma^2 \sim N(\mu + \beta\sigma^2, \sigma^2)$. Therefore, we have the power to condition the log earnings on the latent variables because the mean of conditional log earnings is simply $\mu + \beta\sigma^2$. Using our regression analysis, we can estimate μ, β and also σ^2 since σ^2 depends on δ and γ. We can conduct an interesting simulation of finding the conditional log earnings for different quantile values of σ^2. Figure 3-12 plots the log earnings conditioned on different quantiles of σ^2. In addition, the figure shows the mean log earnings (which is $\mu + \frac{\delta\beta}{\gamma}$). For the US, we find that the presence of unobserved heterogeneity causes an upward bias on the return to education since the return indicated by 75^{th} quantile is higher than the 25^{th} quantile. In Ghana, for lower levels of education, we find that the return to education is higher for the 25^{th} quantile than the 75^{th}. However, for higher levels of education, the return is higher for the 75^{th}. In Ghana, the 25^{th} quantile and the 75^{th} quantile cross over at education level of 10 since the sign of β changes from negative to positive when education level is greater than 10 (see equations 3.15 and 3.21).

3.5.4 Putting It All Together

In this section, we discuss the implications for the labour market by connecting the different parts of our results. The key findings of our empirical analysis using the NIG distribution are:

First, we find that the returns to education increases remarkably for education levels greater than ten compared to people with education levels less than ten, indicating that the earnings function in Ghana is highly convex. Literature also finds that the return to education increases with higher levels of education in Ghana.

Second, compared to the OLS estimates, the NIG estimates are similar for the US dataset but considerably different for Ghana. This is because the unobserved heterogeneity component is separable from education in the US but not for Ghana. We find that education is linear in the earnings function in the US, in line with the Mincerian approach. This is yet another indication that, if we can account for the unobserved heterogeneity in earnings function, log earnings should be linear in education, not convex. One of the reasons for the differences in the NIG and OLS estimates for Ghana is that, once we account for the unobserved heterogeneity using the NIG distribution, the convexity found in our OLS estimates are considerably reduced and the returns to education are much lower. This is even further reduced when we condition the returns to education on σ^2, the unobserved heterogeneity variable. We can conclude that if we have a method of representing unobserved heterogeneity and education distinctly, the functional form of log earnings function will be linear, in line with the Mincerian predictions.

Third, we find evidence of self-selection in Ghana. There seems to be a pattern that, as the education level increases, the distribution of the unobserved heterogeneity component becomes more peaked and tends to collapse to a normal distribution. This implies that, as education increases, the distribution becomes

less heterogeneous. We can deduce a scenario where there are some unobserved heterogeneity components that guide who choose to have higher levels of education in Ghana.

Fourth, we find that the skewness of log earnings is left skewed for lower levels of education in Ghana but becomes more right skewed as the education level increases. In Figure 3-11, we show that, as education level increases, the distribution of heterogeneity component becomes more normal. From these two findings, we can conclude that the log earnings becomes more right skewed as the heterogeneity amongst individuals decreases. In other words, as the population becomes less heterogeneous, the skewness of log earnings are more in line with the predictions of Roy selection model.

Fifth, as we mentioned in our introduction, one of the problems in labour economics has been to estimate the return to education without bias. There are some aspects of individuals that are correlated with education but unobserved. Ability is one of the strongest candidates that fall in this category. In the presence of different abilities amongst individuals, the OLS estimates would tend to be positively biased. Any method that would attempt to correct this ability bias should therefore give results that are lower than the OLS estimates. In case of Ghana, our estimates are lower than the OLS estimates when we account for the unobserved heterogeneity. This strongly hints that our unobserved heterogeneity variable is probably predominantly capturing the unobserved ability component.

3.6 Conclusion

In this chapter, we address the second issue widely discussed in labour literature, that is, modelling the returns to education, where the objective is to model earnings conditional on education. We extend our marginal distribution framework from the previous chapter to a conditional (on education) distribution case. This sheds lights on the usefulness of using this distribution in modelling log earnings function. The NIG distribution is used in the framework of general-to-specific algorithm to model the log earnings data of two datasets, the US male and Ghana male.

We find that the NIG and OLS estimates in the US are same since education is not a function of the unobserved heterogeneity component. However, the NIG and OLS estimates are considerably different in Ghana since education is a function of unobserved heterogeneity. Consequently, one of the key differences between the two datasets is that skewness and unobserved heterogeneity is a function of education for Ghana but not for the US.

Of particular interest is the finding that, once the unobserved heterogeneity component is accounted for, log earnings function is almost linear in education. Using the NIG distribution, in Ghana, we find evidence that the return to education is almost flat for lower levels of education and increases remarkably for education levels higher than ten. After ten years of education, the return to education increases modestly. When we explicitly condition the marginal return on the unobserved heterogeneity variable, σ^2, we find that the marginal return to

education for higher levels of education is constant in Ghana, and the return to education is linear. Our results accord with the popular belief that the marginal return to education will decrease when the unobserved heterogeneity is accounted for (Card 2001).

There seems to be also sign of self-selection in Ghana labour market. As the education level increases, the distribution of unobserved heterogeneity tends to the normal distribution, an indication that self-selection filters out certain groups of people and makes the cohorts more homogeneous. In their paper, Rankin, Sandefur and Teal (2010) find evidence of a pattern of waiting by which there is a preference for a public or a large firm job, and as the probability of obtaining one declines the worker switches to the informal sector. This shows that selection is an important determinant of earnings in Ghana.

3.7 Appendix A: Choosing the Critical Values

The problem with starting from a saturated model with a huge number of variables is that the likelihood value cannot fall when variables are added to a model, so there is a tendency to overfit the model (Greene 2000). With the likelihood ratio (LR) test, there is a tendency to reject the restricted model due to the increased likelihood arising simply from the increased number of variables in the unrestricted model. Therefore, we need a mechanism of punishing for the increased number of degrees of freedom in the model. Typically, two alternatives to the LR test are used: Akaike information criterion and Schwarz criterion. The idea

of using information criterion is to adjust the equation standard error to allow for potentially over-parameterizing the model. Each criterion adds a penalty which increases with the number of regressors. Greene (2000) mentions that neither criterion has an obvious advantage over the other. However, "... [T]he Schwarz criterion, with its heavier penalty for degrees of freedom lost, will lean toward a simpler model. All else given simplicity does have some appeal." (Greene, 2000: 306). The formula for the Schwarz criterion (SC) is:

$$SC = -2 \ln L + k \ln n \tag{3.26}$$

where k is the number of free parameters, n is the sample size and L is the maximized likelihood value. One important characteristic of SC is that models based on SC are asymptotically consistent for large values of n (Burnham and Anderson 2004). This is because, using SC, the probability of rejecting true model (Type I error) goes to zero. Using SC, the cut-off value for failing to reject the restricted model is proportional to $k \ln n$ For large numbers of free parameters, the cut-off value is larger than that obtained using $\chi^2 - distribution$ for LR test.

The critical values are different for the US and Ghana as the sample size and the degrees of freedom are different. Even within the same sample, the degrees of freedom can differ depending on the reduction undertaken. For example, if we use the US sample and reduce from the saturated model containing all the dummies for all the parameters to a model excluding only the dummies for μ, the degrees of freedom would be 9 (8 dummies for education levels and 1 for experience). For

the hypothesis testing, the critical value $c_{95\%}$ (from a chi-square table with df=9 at 95% significance level) is 16.92. The critical value motivated by the $c_{BIC} \sim 9(\log 3965) = 74.57$.

3.8 Bibliography

Angrist, J.D. And A.B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics,* 106, 979-1014.

Belzil, C. and J. Hansen (2002): "Unobserved Ability and the Return to Schooling," *Econometrica,* 70(5), 2075-2091.

Blæsild, P. (1981): "The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data," *Biometrika,* 68(1), 251-263.

Burnham, K.P. And D.R. Anderson (2004): "Multimodel inference: understanding AIC and BIC in Model Selection," *Sociological Methods and Research,* 33, 261-304.

Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127-1160.

Cox, D.R. and E.J. Snell (1974): "The choice of variables in observational studies," *Applied statistics*, 23, 51-59.

GREENE, W. (2000): *Econometric Analysis*, Fourth Edition, New Jersey: Upper Saddle River.

GRILICHES, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45, 1-22.

HECKMAN, J.J. AND B.E. HONORÉ (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), 1121-1149.

HECKMAN, J.J., L.J. LOCHNER AND P.E. TODD (2009): "Earnings Function, Rate of Return and Treatment Effects: The Mincer Equation and Beyond," Chapter 7 in *Handbook of the Economics of Education,* Volume 1, eds. Erik Hanushek and F. Welch. Elsevier.

HENDRY, D.F. (1995): *Dynamic Econometrics*, Oxford: Oxford University Press.

HENDRY, D.F. AND B. NIELSEN (2007): *Econometric Modeling: A Likelihood Approach,* Princeton: Princeton University Press.

HENDRY, D.F. AND H.M. KROLZIG (2005): "The properties of automatic Gets modelling," *Economic Journal,* 115, C32-C61.

HOOVER, K.D. AND S.J. PEREZ (1999): "Data mining reconsidered: Encompassing and the general-to-specific approach to specification search," *Econometrics Journal,* 2, 167-191.

MINCER, J. (1974): *Schooling, Experience and Earnings,* New York: Columbia University Press.

PSACHAROPOULOS, G. (1994): "Returns to investment in education: a global update," *World Development,* 22, 1325-1343.

RANKIN, N., J. SANDEFUR AND F. TEAL (2010): "Learning and Earning in Africa the : Where are the Returns to Education High?" *The Centre for the Study of African Economies, Oxford University, UK,* CSAE WPS/2010-02.

ROY, A.D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers,* New Series, 3(2, June 1951), 135-146.

TEAL, F. (2010): "Higher Education and Economic Development in Africa: a Review of Channels and Interactions," *The Centre for the Study of African Economies, Oxford University, UK,* CSAE WPS/2010-25.

Chapter 4

The US and China Current Accounts Debate: A Model of Demographics

Abstract: Using a model that allows for a rich structure of age effects similar to those predicted by the life cycle theories, this paper argues that the demographic shifts are partly responsible for the sustained rise in the US current account deficit and the rapid increase in China's current account surplus in the last decade. However, demographics do not have an impact on the long run equilibrium or level of current accounts. Rather, they are important determinants of the short run adjustment of current accounts to their equilibrium levels. In the next twenty years, the demographic shifts are likely to push towards positive adjustments of current accounts in China and negative adjustments in the US. Developing the

infrastructure, financial markets, policy tools and regulatory settings to be able to cope with the excess capital flow remains an urgent task.

4.1 Introduction

The current account imbalances of the US and China have garnered a great deal of interest amongst macroeconomists in the last decade. In this chapter, we attempt to understand the role of demographics in shaping these current account imbalances. The most famous insight into the linkage between current accounts and demographics comes from the life cycle hypothesis (LCH), which shows how the demographic structure of an economy may influence the general desire to save or invest across an entire economy. Since the current account is the difference between an economy's savings and investment, one of the products of empirical work on the LCH was the observation that with savings and investment schedules both likely to depend on age structure, the current account balances may also shift along with demographics in particular ways. Most empirical studies have used the LCH to show how the population distribution affects the level of current accounts. However, there is also the possibility that this effect does not happen in the levels, but rather that changes in population distribution would affect the way current accounts adjust. We build a model that allows us to test for both the level and adjustment effects of the population distribution on current accounts. Our results indicate that demographic shifts are not determinants of the level of current accounts. Rather, they play an important role in the short run adjustment

of current accounts to their long run equilibrium.

Our approach is different from conventional studies linking current accounts to demographics in three ways.

First, our demographic specification allows us to estimate the impact of demographics on a more granular level than has been done in most studies. The common approach is to estimate the impact of the working age population on the current account. In our paper, following Higgins (1998), we divide the entire population into fifteen groups: 0–4, 5–9, 10–14,..., 70+. We use information of all the age groups by forming three demographic variables that are geometric-type averages of the age groups.

Second, Higgins (1998) has conducted the analysis in a panel framework, finding a common age group across all countries for the segments of the population that contribute towards current account surpluses and the segments of the population that contribute towards current account deficits. Most studies have also given a common age definition of prime savers for all countries (e.g. the IMF World Economic Outlook (2004) defines this group as 40–65 and Wilson and Ahmed (2010) estimate it as 35–69). However, country-specific differences could lead to varying responses of the population towards current accounts and a panel framework might fail to capture those differences. Our model finds the age groups of prime savers for the US and China separately. To preview our results, we find that the segments of the population that contribute positively and negatively towards current account adjustments are, in fact, different for the US and China.

Third, in terms of our econometric framework, we use the autoregressive dis-

tributed lag model (ARDL), which allows us to model the dynamics of the time series directly, rather than treating the autoregressive errors as a nuisance. This framework allows us to test for both the level and adjustment effects of the population distribution on current accounts. As far as we are aware, there are no studies that have combined the use of a granular demographic dataset (that can better capture the predictions of LCH) and the ARDL framework (that can incorporate both level and adjustment effects) to find the impact of demographics on current accounts.

In the last decade, global capital and current account imbalances have played a major role in a number of important macro debates. The perception of unsustainable imbalances between the US and the rest of the world, the unusual flow of capital from the emerging to the developed world and the so-called 'global savings glut' have all played a role in the recent global recession. The US and China are at the centre of this debate. The last decade has witnessed a sustained deterioration in the US current account deficit and a rapid rise in China's current account surplus. As two of the largest economies in the world with significant financial and trade linkages with other economies, these country-specific imbalances have major implications for the global imbalances. In US dollar terms, the US and China are by far the biggest contributors for the negative and positive imbalances, respectively.

Many studies have addressed this theme of global current account imbalances based on macroeconomics, finance and political factors. One of the most widely discussed studies, by Caballero, Farhi and Gourinchas (2008), attributes this issue

to the increased savings in the East Asian economies after the 1997 financial crisis and the subsequent rise in their demand for safe assets. The US, as the most financially developed economy providing the most diverse and safe set of financial instruments, was able to meet this demand. This resulted in capital flowing from the East Asian economies to the US, leading to a rapid build-up of current account surplus in the East Asian economies and a rise in the US deficit. There has also been intense political debate in the US around the notion that the rapid rise in China's current account surplus could be due to the unfair advantage of intentionally maintaining an undervalued Remnibi. Other traditional explanations revolve around the social and cultural trends of these economies: Chinese consumers do tend to have high saving rates while US consumers have one of the lowest saving rates in the world.

What has often gone unnoticed in this debate is that the past decade has also witnessed significant demographic shifts in these two economies that could be partly responsible for the rapid rise in imbalances. Our model estimates show that the change in prime saving age – the age groups in the US that contributed positively towards net savings – was negative in the US in the last two decades. This is in contrast to China, where the number of prime savers shot up. These demographic shifts are a key determinant of current account imbalances since an economy's current account is literally equal to its net saving, i.e. total savings minus total investment. Therefore, a tendency to save more across an economy due to the increase in the prime savers as a share of the total population will translate into pressure for current account positive adjustments. Similarly, a tendency to

dissave due to the decrease in the prime savers as a share of the total population will translate into pressure for current account negative adjustments.

Our empirical work in this chapter gives convincing evidence that the demographic shifts undergone in the US and China have, indeed, influenced the current accounts of their economies. Our results reinforce one of the perennial problems identified by macroeconomists: the increase in the population of age groups that usually have a positive adjustment effect on current accounts of other countries is in fact having a negative adjustment effect on current accounts in the US. This strongly hints that the prime saving age group in the US is different from the prime saving age group of other countries. Also, we find evidence that the financial liberalization in the late 1980s has somewhat exacerbated this behaviour of cohorts. This calls for structural changes in the economy that encourage positive net saving by the positive increase in people in the 35–69 age group – the age group that usually contributes positively towards current accounts in an economy. In China, we find that the demographic pressure towards positive adjustments in the current accounts is strikingly high. Going forward, our model results would imply that the demographic shifts will continue to push towards more positive adjustments in China's current accounts and negative adjustments in the US current accounts over the next twenty years. Hence, developing the infrastructure, financial markets, policy tools and regulatory settings to be able to cope with the excess capital flow remains an urgent task.

The chapter is arranged as follows. We first discuss the importance of the US and China's current account imbalances in the global picture to show the

significance of our work. We then briefly go through the evolution of the US and China current account imbalances in the past two decades. We discuss in detail how we capture the demographics in our regressions and give an outline of the demographic trends relevant for current account discussions. The economic variables that also influence current account imbalances are discussed briefly as we use them as controlled variables in our regression. We then discuss in detail how we frame our empirical analysis. This is followed by results using three datasets from the US and China. We then discuss the implications of our results in terms of explaining the current account imbalances in the past decades and also the likely influence of demographics on current accounts in the next few decades.

4.2 The Importance of the US and China in Global Imbalances

In the past decade, one of the most widely discussed macroeconomic trends has been the rapid rise in the US current account deficit and China's current account surplus (Figure 4-1). In a sample comprising the G20 countries and individual European Union countries, China had the second highest current account surplus when expressed as a share of GDP in 2007, the last year before the recent financial crisis, while the US was the seventeenth highest deficit economy. However, due to the economic size and the importance of these economies in terms of financial and trade linkages with other countries, these imbalances map into significant

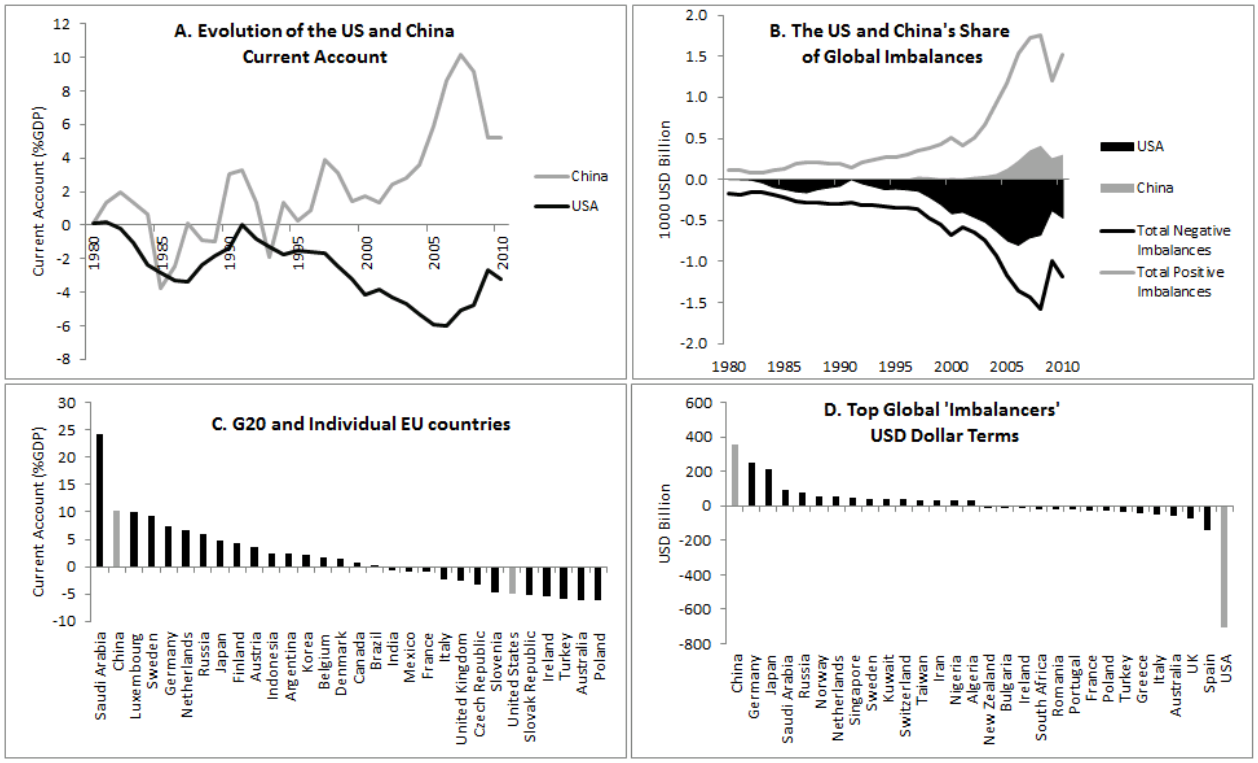

Figure 4-1: The US and China Key for Global Imbalances. Figures C and D represent current account imbalances for 2007. Source: IMF WEO (September 2011).

amounts in absolute terms (Figure 4-1D). When expressed in US dollars, the US has consistently run the largest current account deficit since the early 1980s. Japan was the main counterpart to US deficits during the 1990s, and China's surpluses became the largest in the world in absolute terms in 2006.

There is now a wide body of work that addresses this issue from the perspective of the rapid rise of surplus in emerging markets, mainly emerging Asia, and the concomitant rise in the deficit in the developed world. The key idea in this body of research is that the underdeveloped local financial markets in the emerging markets are part of the reason why savings are larger than they 'should' be, explaining why current account surpluses are also surprisingly large. Caballero, Farhi and Gourinchas (2008), for instance, rationalized the sustained rise in US current account deficits, the decline in real rates and the rise in the US assets in global portfolios as a rational response to differing capacities to generate financial assets from real investments. Under this theory, the emerging market crises at the end of the 1990s and their subsequent rapid growth prompted them to search for sound and liquid financial instruments, which the US markets were positioned to provide but their own markets were not. Of course, the US 'exorbitant privilege' from its position as a reserve currency may also help in this regard. Others that have found supporting evidence for such a dynamic include Gourinchas and Rey (2005) and Cooper (2007). More recently, Goyal *et al.* (2011) show that the divergence in the financial depth of advanced economies versus emerging markets has also contributed towards rapid global imbalances.

Our research complements this body of work by looking into one of the key

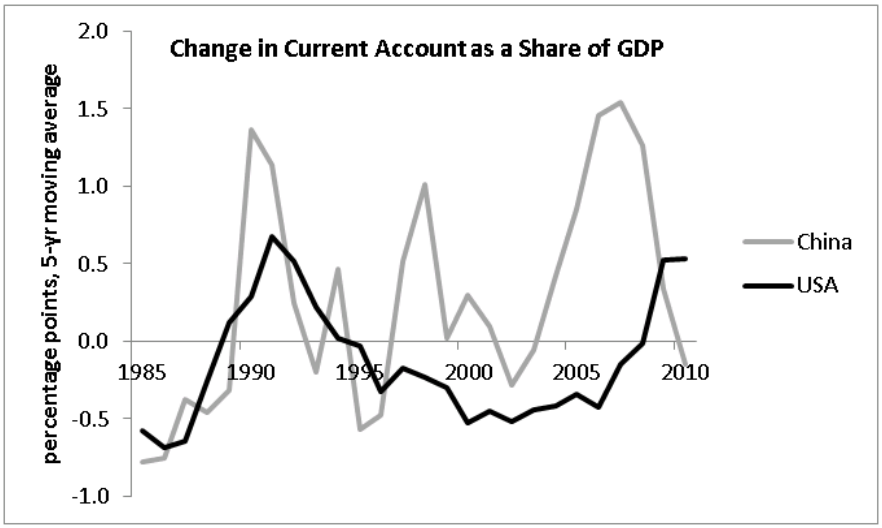

Figure 4-2: The Change in Current Accounts (as a Share of GDP). Source: IMF WEO (September 2011).

structural factors responsible for the current account positions: the demographic shifts of an economy. We try to show that one of the factors responsible for the current account positions of the US and China is the dramatic demographic shifts that these economies have undergone in the past few decades. We show that the demographic shifts in the US have supported negative adjustments in the current account whereas the demographic shifts in China have supported positive adjustments in the current account.

4.3 The Evolution of the US and China Current Accounts

Figure 4-1A shows the evolution of the US and China current accounts since 1980. The US current account has deteriorated steadily while China's has increased since early 1990s. Figure 4-2 shows that the pace of deterioration intensified for the US

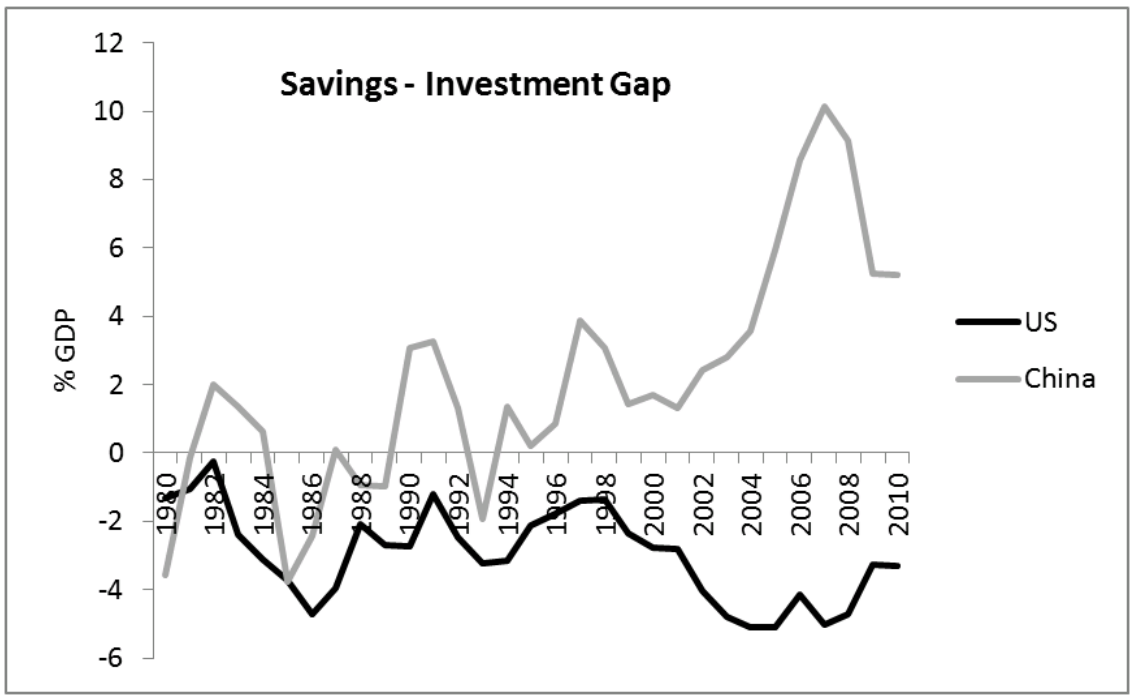

Figure 4-3: Gross National Savings Minus Investment. Source: IMF.

since mid-1990s and continued to worsen till the inception of the financial crisis in 2008. For China, the pace of current account surplus accumulation increased in mid-1990s, then slowed a bit in the late 1990s. However, the pace of increase in current account surplus became strong again since early 2000. Going back to Figure 4-1A, following the financial crisis, both showed signs of current account correction in 2008 and 2009. The current account is essentially the difference between the savings and investment balances in an economy. Figure 4-3 shows that China's savings-investment gap continued to increase while that of the US deteriorated sharply. Figure 4-4 shows that, for both these countries, the rise in the savings–investment gap in the past two decades has been predominantly driven by the changes in savings balances rather than the changes in investment balances: in the US, savings as a share of GDP has fallen more than investment,

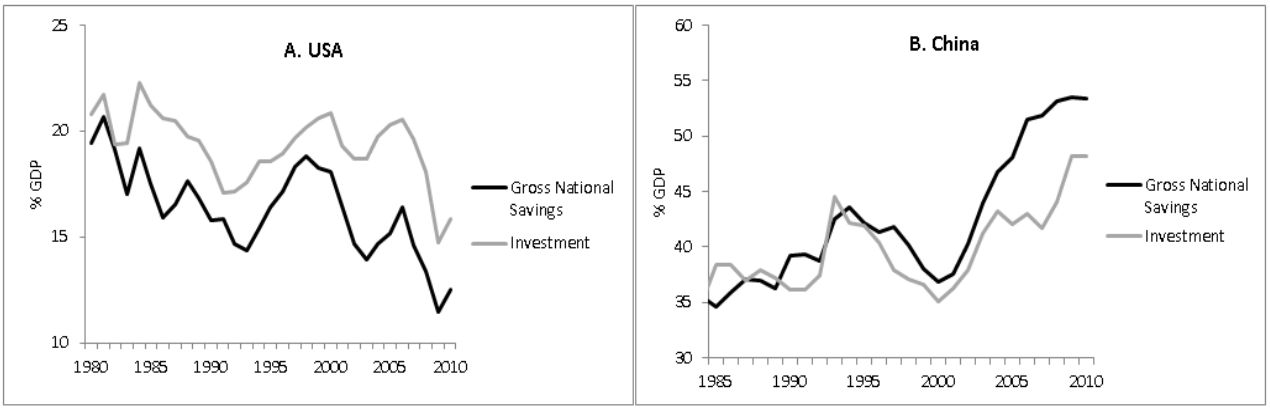

Figure 4-4: The Savings-Investment Gap Increased in Both Countries Due to the Change in Savings. Source: IMF.

leading to a drag on the current account. On the other hand, in China, the rise in savings has outpaced the rise in investment, leading to a surplus in the current account.

Blanchard, Giavazzi and Sa (2005) have attributed the large US current account to two factors. First, an increase in the US demand for foreign goods, partly because of relatively higher US growth, partly because of shifts in demand away from the US goods towards foreign goods. Second, an increase in foreign demand for US assets, starting with the high private demand for US equities in the second half of the 1990s, shifting to foreign private and then central bank demands for US bonds in the 2000s. Blanchard and Milesi-Ferretti (2009) divide the evolution of the US current account in three phases. In the first phase, 1996-2000, the current account deteriorated from -1.6% of GDP in 1996 to -4.2% in 2000. The widening of the deficit reflected a sharp rise in US investment in response to a buoyant US economic growth, which exceeded the increase in the US saving due to fiscal consolidation. FDI and portfolio equity flows linked to the productivity boom and

dot-com bubble accounted for 40% of US current account deficit and were larger than the current account deficit itself. In the second phase, 2000-2004, following the unwinding of the dot-com bubble and recession in the advanced economies, the US current account deficit continued to decline, albeit at a much slower pace. In this phase, the deterioration in the current account was mainly due to the sharp fall in the US savings more than offsetting the decline in the investment relative to the early period. In the third phase, from 2005 until the onset of the financial crisis, the US current account deficit remained high: the weakening of the dollar prompted an adjustment in real trade flows but this was offset by deteriorating terms of trade following the rise in the oil prices.

Blanchard and Milesi-Ferretti (2009) provide three explanations for China's high saving rate and high current account surplus. First, cultural factors in China encourage high savings and low consumption. Second, the underprovision of social insurance to households and poor governance of firms encourage savings with precautionary motives. Third, China's high current account surplus reflects an intentional undervaluation of the exchange rate, together with an appropriately high savings rate to avoid overheating of the economy.

Yongding (2007) provides a comprehensive overview of the forces behind China's current account surplus that also touches on the issues raised by Blanchard and Milesi-Ferretti (2009). Before the late 1970s, China was an autarky economy with negligible foreign trade and no capital inflow except Japanese government aids. The key elements for China's open policy were FDI attraction for technology and export promotion. Usually developing countries use foreign saving to obtain an

investment rate higher than what the domestic saving rate would support. Consequently, they run capital account surplus and current account deficit. However, China was an exception to this rule. Decades of the country's own history with poverty and volatility in income and Latin American debt crisis in the early 1980s significantly influenced the government's policies and also households' saving motives. The government deliberately prevented the inflow of capital to be translated into current account deficit by encouraging Chinese enterprises to use their comparative advantage to run trade surpluses and earn foreign exchange reserves as much as possible. Yongding (2007) contends that, due to the policies of the government and the households' inherent trend of maintaining high savings, the savings-investment gap of China can be somewhat regarded as a structural pattern that is unlikely to dissipate anytime soon.

While most of the studies have highlighted the financial, political and cyclical factors responsible for the piling up of huge imbalances by the US and China, we attempt to explain the imbalances in the light of one of the structural characteristics of these two economies. The difference in savings and investment pattern could also be due to the demographic structure of the economies. Put simply, if the proportion of prime savers fall in an economy, there would be pressure for negative adjustment in the current accounts, and vice verse. In a country where the younger population is growing at a rapid pace, the adjustment on current accounts should be negative as the growth of investment demand surpasses the growth of savings. In this paper, we investigate the role of demographic shifts in the current account imbalances debate by disentangling two important effects of

demographics. First, the current account imbalances could be due to the different demographic compositions of the two economies. Second, the current account imbalances could be due to the differing savings-investment behavior of the same age group in different economies. In addition, our model allows to check if the demographics have a level or adjustment effect on current accounts.

4.4 The Impact of Demographics on the Current Account

Economists have long understood that demographic trends influence savings and investment behaviour. The most famous insight into that linkage comes from the life cycle hypothesis (LCH) of savings, originally proposed by Fisher (1930), Harrod (1948), and Modigliani and Brumberg (1954). LCH posits that savings patterns are different at different points in a person's life. Early on in life, when incomes are low, people are more likely to borrow and invest (in themselves or their children). They then move through a period of 'prime savings' towards middle age to accumulate assets, and then to gradual dissavings as those assets are spent in retirement. Consequently, the saving patterns of individuals over their lifetime are hump-shaped, with individuals saving more during their working years and dissaving after retirement.

The LCH suggests that the population distribution should affect the current accounts through their connection in determining the savings and investment be-

haviour. Literally speaking, the LCH suggests that the population distribution should affect the level of current accounts. There is also a possibility that this effect suggested by the LCH does not happen in levels, rather the changes in population distribution would affect the way current accounts adjust. In our empirical work in section 4.8, we set-up a model that allows to test both the level and adjustment effects of population distribution on the current accounts, and find that the evidence points towards the population distribution predominantly having an adjustment effect.

In this section, we briefly describe the intuition behind why demographic structure of the economy should matter for the current account imbalances under the light of the life cycle hypothesis theories. Then we give a brief overview of one of the models, postulated by Higgins and Williamson (1996), that shows the underlying links between savings, investment, current accounts and demographics. We then briefly discuss what their framework implies for the linkage between demographics and current accounts.

4.4.1 The Intuition

The life cycle hypothesis (LCH) provides some clues as to how the demographic structure of an economy – that is, how many young people there are, how many of 'prime saving' age, how many elderly, etc. – may influence the general desire to save or invest across an entire economy. By definition, the current account – the flipside of it, the capital flows that finance it – is the difference between an

economy's savings and investment. One of the products of empirical work on the LCH was the observation that with savings and investment schedules both likely to depend on the age structure, the current account balance may also shift along with demographics in particular ways. Investment demand is closely related to the share of young people in an economy, through its connection with labour force growth, whereas savings supply should be related to the share of mature adults, through its connection with retirement needs (Higgins and Williamson (1996)).

When economies are closed to trade and capital flows, savings and investment must be equal and interest rates move to bring them into balance. This is true for the world as a whole, which is a closed unit. However, in practice, the vast majority of the world's economies are relatively open and current account imbalances and capital flows are sizeable. Therefore, those economies that have younger populations or have fewer 'prime savers' on average are likely to borrow from those who have more. And, as a result, it is also plausible that relative demographic profiles can have important effects in determining savings and investment outcomes across countries and hence the current accounts and capital flows between them. This implies that, for a financially open economy, a shift in the population age distribution towards younger cohorts should produce current account deficits as the increase in investment demand outweighs the fall in savings. Similarly, as the age distribution shifts towards the older working population, there would be a current account shift to a surplus as the rise in savings becomes more pronounced. There is now a host of evidence that strongly suggests such occurrences, including Higgins (1998), Lane and Milesi-Ferretti (2001) and Wilson and Ahmed (2010).

Going through the same logic used in explaining the link between population distribution and level of current accounts, we can say that when the proportion of prime savers fall in an economy, there is a pressure for the current accounts to adjust downwards. In other words, the decrease in prime savers should be characterized by a pressure on current accounts to go decrease or fall towards deficit side. On the flipside, if the proportion of prime savers is increasing in an economy, there should be an upward adjustment pressure on the current accounts. In other words, the incremental increase in prime savers in an economy should create an incremental positive pressure on the current accounts to increase. If the level of current account is negative, the increase in prime savers should decrease the deficit. In the level of current accounts is positive, the increase in prime savers should create more surplus.

4.4.2 The Model

Using predominantly the life cycle hypothesis and other existing literature, Higgins and Williamson (1996) builds a simple model to show the connection between current accounts and demographics. They contend that a dynamic model is required, one capable of accommodating the demographic effects on both savings supply and investment demand. We outline below the main underpinnings of the model, based on Higgins and Williamson (1996). Before going through the model in detail, it may be useful to give more concrete definitions for savings and investment. Hence, first we give the data definitions.

Data Definitions

Savings can be measured in different ways and that has implications for the final results showing the impact of population age structure on savings. Bosworth, Bryant and Burtless (2004) give a detailed survey on the various ways of measuring savings and their likely impact. Broadly speaking, there are two alternative measures for savings (Bosworth, Bryant and Burtless (2004)). First, saving can be viewed as income minus consumer expenditures. This concept of savings is used in the US National Income and Product Accounts (NIPA). This simple definition is also used by OECD. Bosworth, Bryant and Burtless (2004) also discuss different extensions of this definition that includes capital gains and losses, revaluation of assets, inflation adjustments and other considerations. For the purpose of macroeconomic analysis conducted in this chapter, this seems the more relevant definition to use. Second, the savings data can also be retrieved from household surveys for microeconomic analysis.

For investment, most studies, example Helliwell (2004), use gross capital formation. World bank defines gross capital formation as the sum of gross fixed capital formation, changes in inventories and acquisitions less disposals of valuables. Gross fixed capital formation consists of: 1) acquisitions less disposals of new or second hand tangible fixed assets in the form of machinery and equipment, dwellings, other buildings and structures, cultivated assets (trees and livestock that are used repeatedly, or continuously over long periods of time to produce goods such as rubber, fruit, milk, wool, etc.); 2) major improvements to existing

fixed or natural assets, including land; 3) acquisitions less disposals of intangible fixed assets (e.g. computer software). Change in inventories is acquisitions less disposals of stocks held by producers. Acquisitions less disposals of valuables is precious metals or stones, expensive jewels, works of art, etc. held as investments.

Higgins and Williamson's Model

Higgins and Williamson (1996) start off with a neoclassical growth model, inhabited by an overlapping generations population. The model allows for demographic effects on both savings supply and investment demand, and through those links, the model allows for demographics effects on current accounts. The model's demographic structure allows for three periods of life – youth, the 'prime of life' and old age. The adult population at time t consists of $N_{1,t}$ prime-age adults in the labor force, $N_{2,t}$ retired elderly, and $N_{0,t}$ dependent young. The population of dependent young is determined by $N_{0,t} = n_t N_{1,t}$, that is each prime age adult bears n_t surviving offspring. The authors then assume that the prime age adults care about their dependent offspring, but that the elderly adults have no bequest motive. Prime age adults are endowed with one unit of time, which is inelastically supplied to the labor force. Labor income is divided among current consumption, child support and savings for old age. The lifetime budget constraint of a representative prime age adults at time t can then be written as

$$W_t = C_{1,t} + \frac{C_{2,t+1}}{1 + r_{t+1}} + n_t C_{0,t}$$

where $C_{1,t}$ and $C_{2,t+1}$ refer to consumption during the prime years and retirement, respectively. $C_{0,t}$ is consumption per dependent offspring and r_{t+1} is the interest rate on assets acquired at t and held until $t+1$.

Preferences are described by an additively separable utility function of the form

$$V_t = \frac{(C_{1,t})^{1-\theta}}{1-\theta} + (1+\rho)^{-1}\frac{(C_{2,t+1})^{1-\theta}}{1-\theta} + (n_t)^{1-\epsilon}\gamma^{\frac{1}{\theta}}\frac{(C_{0,t})^{1-\theta}}{1-\theta}$$

where $\frac{1}{\theta}$ is the intertemporal elasticity of substitution. In other words, $\frac{1}{\theta}$ describes the substitutability of consumption across periods. For $\frac{1}{\theta} > 1$, higher interest rates lead to increased savings. The pure rate of time preference is given by ρ. The parameter $0 \leq \epsilon \leq 1$ decides if the weight of the young generation in the parental utility function is less than proportional to the number of children. The parameter $\gamma \leq 1$ allows for the possibility that a child can attain a given level of utility while consuming less than an adult.

The model then uses a neoclassical production function to express output, $Y_t = F(K_t, L_t)$, where $L_t = A_t N_{1,t}$ represents the aggregate labour supply measured in efficiency unites. Technological progress is considered exogenous and occurs at the rate $g-1$, so that $A_{t+1} = gA_t$. The authors use lower-case letters to represent variables defined per unit of effective labor input to write $y_t = f(k_t)$ and rely on the standard assumption that the size of the capital stock in period t is determined by the savings and investment choices made at $t-1$, so that $K_{t+1} = K_t + I_t$.

The model further assumes that the capital intensity of production depends on the extent to which the economy is linked to the international capital market

and takes the cases of perfect and zero capital mobility as bounds. For example, for the case of perfect capital mobility, the model assumes that prime age adults lend in the international capital market when their savings are greater than the value of the capital stock required for the next period. This would lead to current account surplus. Similarly, prime age adults borrow in the opposite case, leading to a current account deficit. If the economy is closed to capital flows, domestic savings supply and investment demand must be equal, with the marginal product of capital equated to the marginal rate of substitution in consumption between the two periods of adult life.

Using the aforementioned assumptions as the basic set-up of the model, the authors derive the following steady-state values for savings as a share of GDP for an open economy

$$s(n^*) = \bar{s}(n^*)\frac{w(k^*)}{f(k^*)}\left(1 - \frac{1}{n^* g}\right)$$

where $\bar{s}(n^*)$ represents saving as a share of labor income. This expression captures two distinct but opposing channels through which population growth affects the savings rate. First, higher population growth increases savings by increasing the population of prime age adults relative to the elderly who tend to dissave. This effect is captured by the term $\left(1 - \frac{1}{n^* g}\right)$. Second, higher population growth decreases savings by increasing the youth-dependency burden. This is captured by the term $\bar{s}(n^*)$.

Similarly, the steady-state investment rate is given by the following expression

$$i(n^*) = (n^* g - 1)\frac{k^*}{f(k^*)}$$

so that $i'(n^*) = g\frac{k^*}{f(k^*)}$. This expression indicates that faster population growth always leads to an increase in the investment rate.

4.4.3 Model Implications

Combining the two expressions for steady-state of savings and investment rates, we can deduce that, broadly speaking, for a country with rapid economic growth, higher steady-state fertility brings a lower savings rate but a higher investment rate, leading to a tendency toward current account deficits. On the flip side, for a country with slow economic growth, savings and investment both increases. The authors show that the latter effect dominates, so that higher fertility always reduces the current account balance.

The authors further conduct few simulations using this model to derive two strong conclusions: First, simple patterns of demographic change may result in complex and often counterintuitive savings and investment dynamics as these are dependent on the population growth, the share of prime-age population, the youth dependency ratio and other features that might often act as opposing forces. For example, the authors observe in their simulation that savings rises between $t + 1$ and $t + 2$ even as the youth dependency ratio falls and the share of prime age rises. This pattern is possible because the share of the elderly is falling quickly.

This observation strongly highlights the need for using the entire information contained in the population age distribution, rather than focusing on the working age population and/or the youth and elderly dependency ratios. Second, the demographic 'center of gravity' for investment demand is likely to be earlier in the age distribution than that for savings supply, since investment demand is closely linked to the share of young people in the population via its connection with labor force growth, and savings supply is linked to the mature adult share of the population via its connection with retirement needs. Consequently, as mentioned above, a shift in the population age distribution towards younger ages leads to a tendency towards current account deficits and foreign capital dependency, and vice versa.

4.5 Demographic Trends Relevant for Current Account Positions

Most studies use working age population to show the impact of demographics on current accounts (Debelle and Faruqee (1996)). As shown in Higgins (1998) and Wilson and Ahmed (2010), what matters more in the context of global capital flows is the related notion of those in the 'prime saving' age group. While definitions vary as to when that age occurs, Wilson and Ahmed (2010) estimate that positive net saving is mostly explained by the portion of the population between the ages of 35 and 69. The IMF defines it as somewhat narrower band, from 40

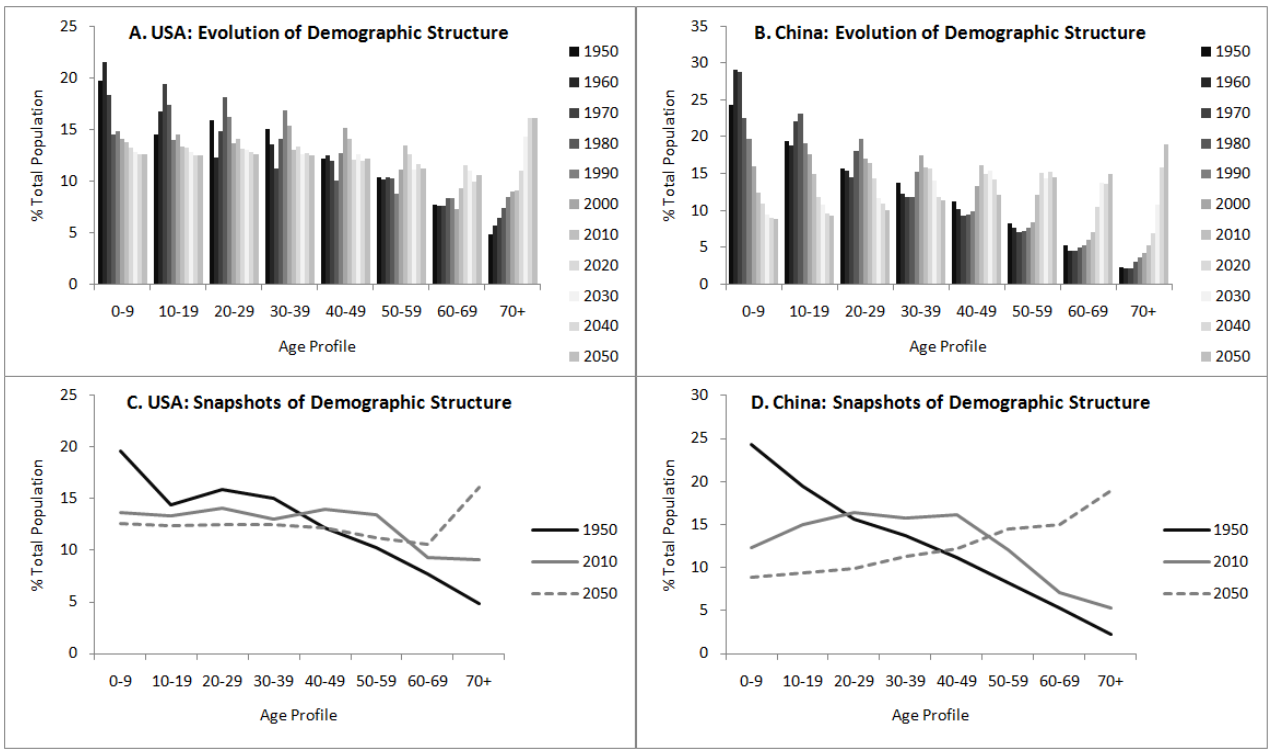

Figure 4-5: Evolution of The Age Structure. Source: UN.

to 65. Although working age and prime saving age population are clearly related, the shifts in the 'prime age' population are different from the working age population, in terms of timing and impact. The working age population group is better known and important in defining fiscal burdens and other significant policy issues whereas the prime saving age group is important for policy discussions related to capital flows and current accounts.

Following Higgins (1998) and Wilson and Ahmed (2010), we use the entire information on the age structure of the economy to show the impact of demographics on current accounts. The demographics data is represented in 5 year interval age groups. In other words, there are 15 age groups, 0–4, 5–9, ..., 70+. However, both these previous studies are conducted using a panel framework and

give a single definition of the 'prime saving' age group for more than thirty countries, consisting of both emerging markets and advanced economies. While this provides a good average approximation, the prime saving age group is likely to vary in a set of diverse countries due to factors like country-specific household saving patterns and government policies. Hence, unlike these studies, we estimate the impact of different age groups on the current account separately for the US and China.

Before estimating the impact of each age group on the current account, we present the historical evolution of demographic trends and the UN projections until 2050 (Figure 4-5). The age distribution of both countries has gradually shifted, with a higher proportion of the population moving to the relatively older age shares. In the US, the highest share of the population, 11%, resided in the age group 0–4 in the 1950s. In the 1960s, the highest share of the population, 10%, resided in the age group 5–9. This pattern continued until the 2000s, where the highest share of the population, 9%, was above 70. China has followed the same general trend but is demographically younger than the US. In the 1950s and 1960s, China's highest share of population was in the age group 0–4, at 16% and 15% respectively, higher than the US in the 1950s. In the 2000s, the highest share of the population, 9%, was in the age group 35–39.

Going forward, the age structure will continue to shift towards an ageing population for both countries. In the US, the pace of further demographic shifts will be moderate, which is evident from Figures 4-5A and 4-5C. The pace of ageing will be faster in China than the US. The proportion of the population above 70

is projected to increase from 10% in 2010 to 16% in 2050, while the share of the population in the age group 0–4 is likely to decline from 7% to 6%. In China, the share of population above 70 will increase from 5% to 19% over the same period while the share of population in the 0–4 age group will decline from 6% to 4%.

4.6 Capturing the Demographic Profile

As aforementioned, the usual practice is to represent the demographic variables in current account regressions by using the working age population or the dependency ratio. But this approach fails to exploit the rich datasets that we have for the entire age structure of economies. One of the problems of using multiple age groups is the problem of multicollinearity - the variance of the ordinary least squares estimates will be large, making it difficult to identify the effect of any particular group. Fair and Dominguez (1991) extended the analysis of Almon (1965) to create an innovative way of representing the demographic variables that could contain the information of the entire age structure in a parsimonious form without causing econometric problems. Later, this approach was popularized by Higgins (1998). In this section, we explain in detail this new method of representing the information of the entire age structure in the form of three geometric averages. We then try to get a better sense of these demographic variables by showing that this approach is essentially the same as orthogonalizing the demographic variables and then re-scaling them. We use the normalized demographic variables to get a better intuition of how the changes in the age structure of economies are reflected in the

three demographic variables.

4.6.1 Constructing the Three Demographic Variables

The demographic variables in our model are captured using three variables D_{1t}, D_{2t} and D_{3t}, which are complicated geometric averages of the population age shares. They are constructed using a low-order polynomial to represent 15 population age shares with each age share containing 5 years: 0–4, 5–9, ..., 70+ (Higgins (1998)). This allows for a richer structure of age effects similar to those predicted by the life cycle model.

In this section, we show the derivation of D_{1t}, D_{2t} and D_{3t}. The initial regression specification that can be thought of along the lines of

$$CA_t = \beta' X_t + \alpha_1 p_{1t} + \alpha_2 p_{2t} + ... + \alpha_n p_{nt} + u_t \tag{4.1}$$

where CA_t is the current account expressed as a share of GDP, X_t is the vector of explanatory variables and the $p_{1t}, ..., p_{nt}$ shares of the population of n age groups, 15 in our case, across time t.

Our aim is to transform equation 4.1 into

$$CA_t = \beta' X_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + u_t \tag{4.2}$$

and then estimate β', γ_1, γ_2 and γ_3. In the rest of this section, we give the motivation and derivation of this transformation.

Due to the high correlation between the shares of consecutive age groups, a regression that includes all 15 age groups, like equation 4.1, will encounter multicollinearity. The variation of the ordinary least squares (OLS) estimates will be large, making it difficult to identify the effect of any particular group. Traditionally, to avoid this problem, the demographic structure has been measured by the working age population (Debelle and Faruqee (1996)). This single variable representation makes the regression simpler but explicitly assumes that the population is homogeneous – the marginal reaction of individuals to saving and investment decisions is thus treated as the same regardless of their age. Stoker (1986) shows that this condition does not hold and a better approach is to use models that allow for age-based population heterogeneity. This prompted Fair and Dominguez (1991) to propose a technique for capturing the information contained in the entire age distribution while maintaining a parsimonious parametrization. This technique is similar to Almon's (1965) polynomial-distributed lag technique. In our case, this technique involves, as mentioned in the rest of the section, constraining the coefficients of the p_{nt} variables ($n = 1, 2, ..., 15$). On the other hand, in Almon's case, the coefficients of the lagged values of some variables were constrained using the same technique.

Intuitively, n is the number of age groups that we have chosen or the number of demographic variables that are used in the initial regression. Since we use the UN data and the UN data is divided into 21 age groups, 0–4, 5–9, ..., 95–99, 100+, the highest possible value for n, in our case, is 21. If we used a much granular dataset, the highest value of n would be the number of possible age groups the

dataset could be divided into. Following Higgins (1998), we have chosen to group the population above 70+ in one group. We have labelled the 15 age groups as $n = 1, 2, ..., 15$. As discussed in Almon (1965), it makes no difference where in the interval [1,15] these labels lie. The labels do not even have to be integers. However, it is most convenient to choose them in the standard form. Following Higgins (1998), we thus opt for a simple labelling of groups as $n = 1, 2, ..., 15$.

Following Higgins (1998), the first step of constructing variables that are able to retain all information and yet have a parsimonious specification is to constrain the coefficients of the population shares to lie on a third degree polynomial

$$\begin{aligned} \alpha_1 &= \gamma_0 + \gamma_1 1 + \gamma_2 1^2 + \gamma_3 1^3 \\ \alpha_2 &= \gamma_0 + \gamma_1 2 + \gamma_2 2^2 + \gamma_3 2^3 \\ &\vdots \\ \alpha_n &= \gamma_0 + \gamma_1 n + \gamma_2 n^2 + \gamma_3 n^3 \end{aligned} \tag{4.3}$$

where γ_0, γ_1, γ_2 and γ_3 are the coefficients for the constant, linear, quadratic and cubic terms respectively. The restriction that the population shares lie on a third order polynomial implies that the relationship between current account and the population age shares change smoothly, in line with the predictions of the life cycle model.

Since the sum of $p_{1t}, ..., p_{nt}$ across n is 1 and there is a constant in equation

4.3, a restriction on the α coefficients must be imposed for estimation. The age group coefficients are restricted to sum to zero, (see Fair and Dominguez (1991) and Higgins (1998))

$$\alpha_1 + \alpha_2 + ... + \ \alpha_n = 0 \tag{4.4}$$

Equating equations 4.3 and 4.4

$$\begin{aligned} & \alpha_1 + \alpha_2 + ...\alpha_n \qquad (4.5) \\ = \ & n\gamma_0 + \gamma_1(1 + 2 + ... + n) + \gamma_2(1^2 + 2^2 + ... + n^2) + \gamma_3(1^3 + 2^3 + ... + n^3) \end{aligned}$$

Once we estimate γ_1, γ_2 and γ_3, and impose the condition 4.4

$$\begin{aligned} \gamma_0 &= -\frac{[\gamma_1(1 + 2 + ... + n) + \gamma_2(1^2 + 2^2 + ... + n^2) + \gamma_3(1^3 + 2^3 + ... + n^3)]}{n} \\ &= -\frac{(\gamma_1 \sum_{j=1}^{n} j + \gamma_2 \sum_{j=1}^{n} j^2 + \gamma_3 \sum_{j=1}^{n} j^3)}{n} \end{aligned} \tag{4.6}$$

Substituting $\alpha_1 ... \alpha_n$ in equation 4.1 and grouping all the four parameters $\gamma_0, \gamma_1, \gamma_2$ and γ_3, we have

$$\begin{aligned} CA_t &= \beta' X_t + p_{1t}(\gamma_0 + \gamma_1 1 + \gamma_2 1^2 + \gamma_3 1^3) + p_{2t}(\gamma_0 + \gamma_1 2 + \gamma_2 2^2 + \gamma_3 2^3) \qquad (4.7) \\ &\quad + ... + p_{nt}(\gamma_0 + \gamma_1 n + \gamma_2 n^2 + \gamma_3 n^3) + u_t \\ &= \beta' X_t + \gamma_0 \sum_{j=1}^{n} p_{jt} + \gamma_1 \sum_{j=1}^{n} j p_{jt} + \gamma_2 \sum_{j=1}^{n} j^2 p_{jt} + \gamma_3 \sum_{j=1}^{n} j^3 p_{jt} + u_t \end{aligned}$$

With γ_0 from equation 4.6

$$CA_t = \beta' X_t + \gamma_1(\sum_{j=1}^{n} jp_{jt} - \frac{1}{n}\sum_{j=1}^{n} j) + \gamma_2(\sum_{j=1}^{n} j^2 p_{jt} - \frac{1}{n}\sum_{j=1}^{n} j^2) + \gamma_3(\sum_{j=1}^{n} j^3 p_{jt} - \frac{1}{n}\sum_{j=1}^{n} j^3) + u_t \tag{4.8}$$

From here, we derive the form of the three demographic variables, which allows us to basically capture the same information contained in $\alpha_1....\alpha_n$, but in a parsimonious form

$$\begin{aligned} D_{1t} &= \sum_{j=1}^{n} jp_{jt} - \frac{1}{n}\sum_{j=1}^{n} j, \\ D_{2t} &= \sum_{j=1}^{n} j^2 p_{jt} - \frac{1}{n}\sum_{j=1}^{n} j^2, \\ D_{3t} &= \sum_{j=1}^{n} j^3 p_{jt} - \frac{1}{n}\sum_{j=1}^{n} j^3. \end{aligned} \tag{4.9}$$

The regression is then run to find the parameters γ_1, γ_2 and γ_3 from

$$CA_t = \beta' X_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + u_t$$

The estimated coefficients of the new variables D_{1t}, D_{2t} and D_{3t} have no structural interpretation. However, the implicit age distribution coefficients can be easily recovered using equations 4.6 and 4.3. In simple terms, this reconfiguration imposes the restriction that the impact of age on current accounts across the different age groups follows a cubic 'shape'.

D_{1t}, D_{2t} and D_{3t} are essentially complicated geometric averages of the population age shares. As the population ages, all the three demographic variables

are expected to increase. This is because, an ageing population means that the population share of higher age groups is increasing while that of the lower age groups is falling. Looking at equation 4.9, population ageing would imply that the population share, p_{jt}, of higher age groups, j, will increase leading to a higher value of $\sum_{j=1}^{n} jp_{jt}$ in the computation of D_{1t}. As the $\frac{1}{n}\sum_{j=1}^{n} j$ portion of D_{1t} is a constant, higher $\sum_{j=1}^{n} jp_{jt}$ would lead to a higher value of D_{1t}. Using the same logic, D_{2t} and D_{3t} should also increase as population ages since $\sum_{j=1}^{n} j^2 p_{jt}$ and $\sum_{j=1}^{n} j^3 p_{jt}$ would increase with $\frac{1}{n}\sum_{j=1}^{n} j^2$ and $\frac{1}{n}\sum_{j=1}^{n} j^3$ remaining constant. The increase in D_{3t} will be higher than D_{2t} since D_{3t} contains a cubic term, j^3, while D_{2t} contains a quadratic term, j^2. Hence, for any given j, the cubic term would grow faster than the quadratic term. Using the same logic, D_{2t} will grow faster than D_{1t}. Thus, an economy with ageing population should have all the demographic variables increasing with the pace of increase in D_{3t} the strongest and the pace of increase in D_{1t} the weakest. The case where D_{1t}, D_{2t} and D_{3t} are all equal to zero represents a situation when the population is equally distributed amongst all the age shares. In other words, p_{1t}, p_{2t}, ..., p_{nt} are all equal.

The coefficients γ_1, γ_2 and γ_3 determine the sensitivity of current account imbalances to D_{1t}, D_{2t} and D_{3t}, respectively. The advantage of this formulation is that, once γ_1, γ_2 and γ_3 are estimated, we should be able to retrieve the age coefficients α_1, α_2,.....α_n using equations 4.6 and 4.3. α_1, α_2,.....α_n represent the sensitivity of current account imbalances to the n- age shares (in our case, $n = 15$). We can then plot a graph of the the coefficients of the age shares (representing the sensitivity of each age share to current account imbalances) versus the age shares.

When γ_1, γ_2 and γ_3 are not zero, we can assume that the impact of current account across the age profile should take a cubic shape, in accordance with the life cycle theory predictions. When $\gamma_3 = 0$, the resulting shape would be quadratic. If both γ_2 and γ_3 are zero, the resulting shape would be linear. Thus, γ_1, γ_2 and γ_3 represent the linear, quadratic and cubic portions of the curve, respectively. The signs of γ_1, γ_2 and γ_3 will determine the shape of the curve.

4.6.2 The Difference Forms for the Three Demographic Variables

We can determine the difference operators for the three demographic variables in two ways. Using equation 4.9

$$\begin{aligned} \Delta D_{1t} &= D_{1t} - D_{1t-1} && (4.10) \\ &= \sum_{j=1}^{n} j p_{jt} - \frac{1}{n}\sum_{j=1}^{n} j - \sum_{j=1}^{n} j p_{jt-1} + \frac{1}{n}\sum_{j=1}^{n} j && (4.11) \\ &= \sum_{j=1}^{n} j p_{jt} - \sum_{j=1}^{n} j p_{jt-1} \end{aligned}$$

Similarly, using equation 4.9

$$\begin{aligned} \Delta D_{2t} &= \sum_{j=1}^{n} j^2 p_{jt} - \sum_{j=1}^{n} j^2 p_{jt-1}, \\ \Delta D_{3t} &= \sum_{j=1}^{n} j^3 p_{jt} - \sum_{j=1}^{n} j^3 p_{jt-1}. \end{aligned}$$

Alternatively, we can derive the same expressions for the the difference forms of the three demographic variables if we start from equation 4.1 but add the difference terms and follow the same methodology as we had done for the finding the three demographic variables in level terms. In other words, start with

$$\Delta CA_t = \beta_d' \Delta X_t + \alpha_{d1}\Delta p_{1t} + \alpha_{d2}\Delta p_{2t} + ... + \alpha_{dn}\Delta p_{nt} + u_t \tag{4.12}$$

Here $\Delta p_{1t} = p_{1t} - p_{1t-1}$, $\Delta p_{2t} = p_{2t} - p_{2t-1}$,....., $\Delta p_{nt} = p_{nt} - p_{nt-1}$. This means that, if $p_{1t}, ..., p_{nt}$ refer to the share of the population of n age groups, $\Delta p_{1t}, ..., \Delta p_{nt}$ give the change in the share of the population of the age groups since the last period. Note that we are using d in the subscript to show that these terms are the difference (or delta) equivalent terms for the same expressions in the level form. We can now arrange the α_{d1}, α_{d2}, ...,α_{dn} parameters the same way we had done α_1, α_2, ...,α_n parameters in equation 4.3 and continue following the same steps. The difference will come in equation 4.7. In the case of level, the term $\gamma_0 \sum_{j=1}^{n} p_{jt} = \gamma_0$ since $\sum_{j=1}^{n} p_{jt} = 1$. In the case of difference operators, the term $\gamma_{d0} \sum_{j=1}^{n} \Delta p_{jt}$ will be zero since $\sum_{j=1}^{n} \Delta p_{jt} = 0$. Now, in equation 4.8, we express the three demographic variables in the current accounts equation as $\gamma_1(\sum_{j=1}^{n} jp_{jt} - \frac{1}{n}\sum_{j=1}^{n} j)$, $\gamma_2(\sum_{j=1}^{n} j^2 p_{jt} - \frac{1}{n}\sum_{j=1}^{n} j^2)$ and $\gamma_3(\sum_{j=1}^{n} j^3 p_{jt} - \frac{1}{n}\sum_{j=1}^{n} j^3)$. For the difference operators, since $\sum_{j=1}^{n} \Delta p_{jt} = 0$, we would only have $\gamma_{d1} \sum_{j=1}^{n} j\Delta p_{jt}$, $\gamma_{2d} \sum_{j=1}^{n} j^2 \Delta p_{jt}$ and $\gamma_{d3} \sum_{j=1}^{n} j^3 \Delta p_{jt}$. Thus, the three difference operators for the demographic variables would be $\Delta D_{1t} = \sum_{j=1}^{n} j\Delta p_{jt}$, $\Delta D_{2t} = \sum_{j=1}^{n} j^2 \Delta p_{jt}$ and $\Delta D_{3t} = \sum_{j=1}^{n} j^3 \Delta p_{jt}$. We can retrieve the information of the change in current

accounts versus the age in population distribution Δp_{1t}, Δp_{2t}, ..., Δp_{nt} the same way we had done for the level case. In other words, we use model estimates of the coefficients γ_{d1}, γ_{d2} and γ_{d3} to determine γ_{d0} using equation 4.6. Then we use the values of γ_{d0}, γ_{d1}, γ_{d2} and γ_{d3} to determine the coefficients of current account changes α_{d1}, α_{d2}, ..., α_{dn} versus the population changes Δp_{1t}, Δp_{2t}, ..., Δp_{nt} using equation 4.3.

4.6.3 Alternative Expression for the Three Demographic Variables

In this section, we try to capture the demographic profile described in section 4.6.1 in a different manner that allows us to better understand the underpinnings behind the three demographic variables D_{1t}, D_{2t} and D_{3t}. In particular, it helps us to understand how the three demographic variables move in response to the movements in the underlying population data. We must clarify that this is merely a simple exercise to obtain a better intuitive grasp of the three demographic variables D_{1t}, D_{2t} and D_{3t} and what the changes of the three demographic variables imply about the movements in the underlying population data. We do not use these alternative formulation in our econometric analysis reported later. This is because, as discussed in this chapter, we find that the econometric results from the alternative expressions are the same as the econometric results using D_{1t}, D_{2t} and D_{3t}. There is no change in the interpretation of the results. The only difference is that the coefficients on the demographic variables are re-scaled. Nevertheless,

we found that expressing the demographic variables using this alternative formulation helps us to better understand the underpinnings of the demographic variables, and later on, understand the results of our econometric analysis better. Hence, we choose to present this alternative formulation to the readers.

Using the same notations as section 4.6.1, we express the three demographic variables in the following manner

$$\begin{aligned}
\widetilde{D_{1t}} &= \frac{\sum_{j=1}^{n} j(p_{jt} - \overline{p})}{\sum_{j=1}^{n} j}, \\
\widetilde{D_{2t}} &= \frac{\sum_{j=1}^{n} j^2(p_{jt} - \overline{p})}{\sum_{j=1}^{n} j^2}, \\
\widetilde{D_{3t}} &= \frac{\sum_{j=1}^{n} j^3(p_{jt} - \overline{p})}{\sum_{j=1}^{n} j^3}.
\end{aligned}$$

where $\overline{p}$ is the average of the population age shares. In terms of regression methodology, this yields the same results as the previous demographic parametrization. The only difference is that the coefficients of the demographic variables are changed in order to adjust for the new scaline. When we do econometric analysis in the later sections, we conduct the same analysis and end up with the same results with the same t-statistics, log likelihood values and R^2 using this alternative set-up. The coefficients of the controlled variables are also same. The only difference is that the coefficients of the demographic variables have changed. This indicates that orthogonalizing the demographic variables in this manner captures the same essence of the parsimonious parametrizations used by Higgins (1998). This is merely re-scaling the demographic data. Hence, relating to equation 4.2,

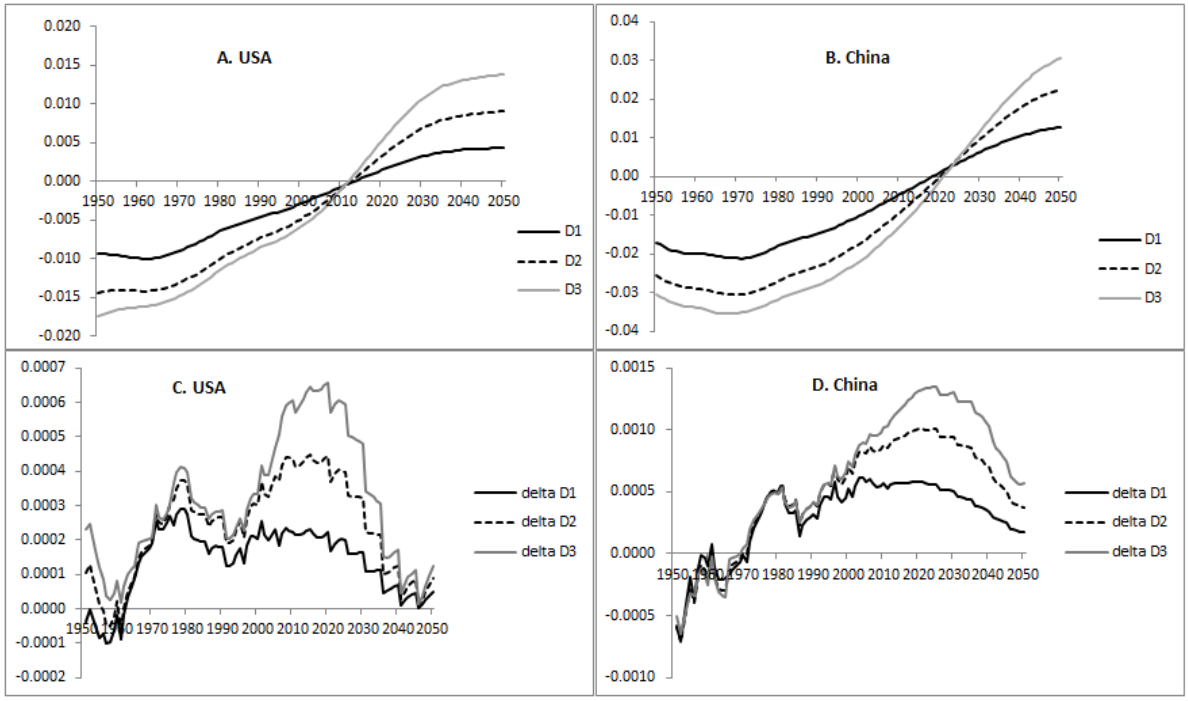

Figure 4-6: The Evolution of the Demographic Variables (using Alternative Expression). Source: Both history and projections from UN Population Statistics

we can write

$$\gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} = \widetilde{\gamma}_1 \widetilde{D}_{1t} + \widetilde{\gamma}_2 \widetilde{D}_{2t} + \widetilde{\gamma}_3 \widetilde{D}_{3t} \tag{4.13}$$

The re-scaling implies that

$$\begin{aligned} \widetilde{\gamma}_1 &= a\gamma_1 \\ \widetilde{\gamma}_2 &= b\gamma_2 \\ \widetilde{\gamma}_3 &= c\gamma_3 \end{aligned}$$

where a, b and c are scalar constants.

Figures 4-6A and B show the evolution of the three demographic variables with

time and give an intuitive idea of how the demographic variables change with time as the population shares change. Recall that Figure 4-5 shows that the share of old population have increased over time for both the countries. Figures 4-6A and B show that this maps into a higher value for all the three variables for both the countries, in particular D_3. As aforementioned, since D_3 represents the cubic term, the rise in population in older age-group is associated with a higher change in the cubic term D_3 than the other two variables. That is, as the population age shares show a more ageing population, all the three demographic variables increase with time.

Figures 4-6A and B also help us to compare the relative magnitude of the three demographic variables. When population structure is tilted towards a younger population, the first demographic variable, D_1, is higher than the other two demographic variables. As the population structure tilts towards elder segment of the population, all the demographic variables increase with the relative difference between the demographic variables decreasing. When the elderly segment of the population significantly increases, the third demographic variable, D_3, becomes higher than the first and second demographic variables, D_1 and D_2. As we have seen in section 4.6.1, the US has a more mature demographic structure than China. Hence, by 2010, the third demographic variable, D_3, in the US is almost close to the first demographic variable, D_1. In 2010, on the other hand, China still has comparatively a younger population, evident from the fact that the first demographic variable, D_1, is still quite a bit higher than the third demographic variable, D_3.

Furthermore, in section 4.6.1, we saw that, going forward, the UN projections suggest that the demographic structure of both the countries will gradually change towards more elderly population, with the pace of ageing faster in China than in the US. Figure 4-6A and B also maps that information in the three demographic variables. As aforementioned, as the population structure changes with more elderly population, the third demographic variable, D_3, becomes higher than the other two demographic variables, $D_1$and D_2. Similarly, the second demographic variable, D_2, becomes higher than the first demographic variable, D_1. This is evident for both countries. In the US, D_2 and D_3 becomes higher than D_1 earlier than in China as the US is the more mature economy (demographically speaking). By 2020, D_3 is already higher than D_1 in the US, but in China, D_3 is still lower than D_1.

Figures 4-6C and D show the change in the three demographic variables over time using the alternative expression. Both the figures show that the change in demographic variables shot up since late 1980s or early 1990s for both countries, indicating that both countries were undergoing considerable demographic shifts around that time. The change in the demographic variables continue to rise until around mid 2020 for both countries. Comparing both the countries, the pace of the increase in the 1990s seem to be higher for the US than China. In line with our observations in the evolution of the three demographic variables, the pace of increase seems to be stronger for the third demographic variable D_3 than the second demographic variable D_2 and the first demographic variable D_1.

4.7 Other Key Determinants of Current Account Balances

While we focus on the impact of demographic shifts on current account imbalances, it should be emphasized that demographic shifts are only one of the factors at play. Economists have long identified a menu of important structural, cyclical and financial factors that are likely to influence current account imbalances. Based on a host of studies, in particular Debelle and Faruqee (1996), we identify and include in our regressions some of the other important determinants of current account imbalances.

Income per capita growth, INC_t: This variable represents the stage of development of the economy and accounts for the capital and production costs of a country. The growth per capita captures the notion that high-growth countries are more likely to attract more capital and run current account deficits than lower-growth countries.

Output gap, GAP_t: The domestic output gap is defined as the difference between real and potential output, as a share of potential output. This variable captures the stage of business cycle effects on the current account and has been found to be important in industrial countries, see Debelle and Faruqee (1996). If the government wants to boost real output above the potential level, this can only be done by reducing savings leading to a deterioration of the current account.

Relative price of investment goods, RPI_t: The relative price of investment goods is included to control for its possible effects on savings supply or

investment goods, something that has been found to be important in some cross-country work on investment and savings. The increase in the real price of investment is likely to boost investment leading to a deterioration of the current account.

Government budget surplus, as a share of GDP, $FISCAL_t$: The stance of fiscal policy can have long term implications on the current account. In the absence of Ricardian equivalence in a life cycle framework, tax policies will have implications for national savings. In particular, changes in public savings and debt, that is the timing of the taxes, will not be fully offset by changes in private saving, leading to changes in the current account balance. A fiscal surplus, that is a contractionary fiscal policy, is likely to improve current account balance.

Changes in terms of trade, TOT_t: The trends in terms of trade may affect current account via the trade balance and their implications for the level of national income. The improvement in terms of trade is likely to positively affect the current account.

Real effective exchange rate, $EXRATE_t$: Real exchange rate changes affect the current account because they reflect the changes in the prices of domestic goods and services relative to foreign. Theories would predict that the rise in real effective exchange rate would increase the demand of imports and decrease the demand of exports as imports become cheaper for the local people and exports more expensive for the foreigners. However, this notion assumes that the volume of exports and imports change immediately with the change in price. In practice, the volume of goods is sticky in the short run which means that an appreciation would

initially lead to a current account surplus as the volume of imports and exports are same but the price of imports decrease and the price of imports increase. This is known as the J-curve effect, the empirical observation that a depreciation of the currency initially increases current account deficit as the volume of goods remain relatively sticky and leads to current account surplus only with time when the volume of goods change with the change in price.

Long term real interest rate, $RINT_t$: Longer term variation in real interest rates might affect the current account through their implications for interest payments (receipts) on net external debt (assets). The long tong term real interest rates also change the cost of capital, and through that channel, influences real investment levels. An increase in long term real interest rate is likely to decrease the level of investment, and thus, improve current account balances.

Oil import growth, OIL_t: Oil revenue matters for oil-producing countries by significantly improving current account balances, example Norway, Saudi Arabia. For our purpose, both China and the US are net oil importers, and the oil import growth is a potential candidate for causing a drag on the current account balances.

Capital openness, $OPEN_t$: Financial liberalization and changes in capital mobility is likely to have a long term implications on the current account imbalances. Countries that maintain a relatively closed capital account through barriers and controls are likely to have a smaller current account position.

Inflation rate, INF_t: The rise in inflation is likely to cause a drag on a country as it loses its trade competitiveness due to the increase in price levels of

exported goods.

4.8 Econometric Methodology

Our approach to modelling current account balances as a function of demographics is closely related to that of Higgins (1998), as well as having a lot in common with a frequently cited study by Lane and Milesi-Ferretti (2001). In terms of choosing the regression methodology, we opt for an autoregressive distributed lag model (ARDL). One of the advantages of the ARDL framework is that the methodology allows us to test for both the level and adjustment effect of demographics on current accounts. We choose the formulation that is often called the error correction model. We provide the motivation and the description of methodology in detail in the first subsection. We then show how the ARDL framework, under a set of assumptions, is the one dimensional case of the multidimensional model that we would have created if we had started with modelling with the view point of an economic system. This idea was introduced by Hendry (1995) to express the concept that, to understand how a system functions, its components and framework, require modelling. In other words, we need to start from a multidimensional model so that we can capture the links between the different components of the system. Using this idea, we present a stylised model that attempts to explain the links between current accounts, its controlled variables and demographics. We then present the exogeneity tests that would allow interpretation of the single equation ARDL model from a system view point.

Before proceeding further, we provide the rational for using country specific time series, rather than the panel framework using multiple countries. In the context of the analysis of the role of demographics on current accounts, country specific regressions have tremendous advantages in understanding the country specific factors that shape the evolution of current accounts. As our results show, the control variables are not necessarily the same for China and the US. More importantly, the response of current accounts towards demographic behavior is also sufficiently different to warrant country specific analysis. The response of different age groups maybe shaped by country specific policies and cultural factors that may not be captured in a panel framework. The savings and investment behavior of individuals across countries may also differ, and that may result in different impact of demographics across countries. Against such a background, the strong assumptions required to conduct panel analysis may not hold. However, we must also mention the caveats of using a time-series based ARDL model compared to alternative approaches such as Higgin's country panel. One of the main disadvantages of our approach, compared to using country panel, is that our sample size would be smaller than a panel framework. We have less degrees of freedom than the alternative approach, particularly when we add more control variables.

While the small sample size may seem to be a problem, we are assured on two grounds. First, to preview our results, when we compare our results of the short versus the long US dataset, the model implications are same. This is a good indication that in our case the potential problems of a small sample size maybe

negligible and it does not change the conclusions of the results. The comparison of the short and the long dataset gives a good robustness check. Also, the sample size is not an issue for the US since we also have a longer dataset spanning five decades. Second, the results in China are in line with the predictions of the life cycle theory. Also, as we discuss during the interpretation of our results, the findings in China are in line with what we should expect from a country that has one of the highest personal savings rates in the world. Overall, we find that the advantages of ARDL based time series model clearly outweigh the potential problems of having a smaller sample.

4.8.1 The Autoregressive Distributed Lag Model

We choose the regression methodology that is widely known as the autoregressive distributed lag model (ARDL), see Hendry and Nielsen (2007, pp. 213). We choose the formulation that is often called the error correction model. The ARDL framework has some intuitive and technical advantages. This specification is used because most of the time series used in our regressions, in particular the current account and the demographic variables, show a tendency to increase or decrease over time without an indication of reverting to the mean. The presence of unit roots in each of these series was tested empirically using a Dickey–Fuller test (Dickey and Fuller, 1979). Some of the series (for example the current account balances expressed as a share of GDP), follow a 'stochastic trend', and are said to be I(1) – integrated of order one – processes; their first difference is a stationary, or

I(0). Under this scenario, the OLS estimates will be biased and inconsistent and the residuals will have serial correlation and heteroskedasticity. The appropriate way to manipulate such non-stationary series is to use differencing and other transformations, such as seasonal adjustment, to reduce them to stationarity. The ARDL framework allows us to model the dynamics directly, rather than treating the autoregressive errors as a nuisance.

Using ARDL methodology, our general model can be expressed as

$$\begin{aligned}\Delta CA_t &= c + \gamma'\Delta X_t + \varphi_1\Delta D_{1t} + \varphi_2\Delta D_{2t} + \varphi_3\Delta D_{3t} \\ &+ \sum_{j=1}^{k-1}(\delta_j\Delta CA_{t-j} + \zeta_j'\Delta X_{t-j} + \eta_j\Delta D_{1t-j} + \iota_j\Delta D_{2t-j} + \kappa_j\Delta D_{3t-j}) \\ &+ \alpha(CA_{t-1} + \frac{\beta'}{\alpha}X_{t-1} + \frac{\delta_1}{\alpha}D_{1t-1} + \frac{\delta_2}{\alpha}D_{2t-1} + \frac{\delta_3}{\alpha}D_{3t-1}) + \varepsilon_t\end{aligned} \tag{4.14}$$

where Δ represents the first difference operator and k is the sample size. For example, $\Delta CA_t = CA_t - CA_{t-1}$. CA_t is the current account imbalances expressed as a share of GDP. X_t are the other control variables used in the regression. D_{1t}, D_{2t} and D_{3t} are the demographic variables, described in detail in section 4.6.1.

When conducting our regression analysis, we find that the lagged values of more than one year is not significant. If we therefore get rid of the lagged values that are more than one years old and also present $Z = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix}$, we can re-write

the ARDL model in equation 4.14 in a much simpler and clearer form

$$\Delta CA_t = c + \alpha CA_{t-1} + \beta' X_{t-1} + \gamma' \Delta X_t + \delta' Z_{t-1} + \varphi' \Delta Z_t + \varepsilon_t \tag{4.15}$$

We can re-write the model in the long run equilibrium ECM form

$$\Delta CA_t = c + \alpha(CA_{t-1} + \frac{\beta'}{\alpha} X_{t-1} + \frac{\delta'}{\alpha} Z_{t-1}) + \gamma' \Delta X_t + \varphi' \Delta Z_t + \varepsilon_t \tag{4.16}$$

where the long run equilibrium, ECM_t, is given by

$$ECM_t = CA_t + \frac{\beta'}{\alpha} X_t + \frac{\delta'}{\alpha} Z_t = 0 \tag{4.17}$$

This form of the model is in the error correction form, see Greene (2003). In this form, we have an equilibrium relationship, $\Delta CA_t = c + \gamma' \Delta X_t + \varphi' \Delta Z_t + \varepsilon_t$ and the equilibrium error, $\alpha(CA_{t-1} + \frac{\beta'}{\alpha} X_{t-1} + \frac{\delta'}{\alpha} Z_{t-1})$ or αECM_{t-1}, which account for the deviation of the pair of variables from the equilibrium. The model states that the change in CA_t from the previous period consists of the change associated with movement with X_t and Z_t along the long run equilibrium path plus a part α of the deviation $CA_{t-1} + \frac{\beta'}{\alpha} X_{t-1} + \frac{\delta'}{\alpha} Z_{t-1}$ from the equilibrium.

Equation 4.15 can therefore be also expressed as

$$\Delta CA_t = c + \alpha ECM_{t-1} + \gamma' \Delta X_t + \varphi' \Delta Z_t + \varepsilon_t \tag{4.18}$$

We also include a trend component in our regression.

4.8.2 The Viewpoint of An Economic System

Under a set of assumptions, discussed later in this section, the ARDL framework can be regarded as a special case of a framework where we model current accounts and all the components of current accounts in a multidimensional set-up. Thinking about our model in this light helps to go beyond the commonly determined roles of demographics on current account imbalances in traditional literature and check for the possibility of other alternative ways in which demographics could shape current account imbalances. This idea of starting with a general framework and modelling all the components of the economy, called "The Economy is a System", was introduced by Hendry (1995, pp. 318). Hendry argues in favour of conducting general-to-specific modelling from the joint density comprising all the relevant economic variables. Hendry and Doornik (1994) discuss ten interrelated reasons for commencing econometric analyses of economic time series from the joint density. One of the most obvious reasons for joint modelling is that, to understand how a system functions, its components and framework both require modelling. The econometric analysis should provide room for checking if all variables are endogenously determined by the economic mechanism and these interact in many ways.

Using this idea and methodology of expressing economic systems from Hendry (1995), our econometric analysis would start off from a stylised model formulating an economic system comprising the current account imbalances CA_t, the demographic variables D_{1t}, D_{2t} and D_{3t} expressed in the system as Z_t, and the control

variables X_t.

$$\Delta \begin{bmatrix} CA_t \\ X_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \begin{bmatrix} \pi_1 & \pi_2' & \pi_3' \end{bmatrix} \begin{bmatrix} CA_{t-1} \\ X_{t-1} \\ Z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{CA,t} \\ \varepsilon_{X,t} \\ \varepsilon_{Z,t} \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} [\pi_1 CA_{t-1} + \pi_2' X_{t-1} + \pi_3' Z_{t-1}] + \begin{bmatrix} \varepsilon_{CA,t} \\ \varepsilon_{X,t} \\ \varepsilon_{Z,t} \end{bmatrix}$$

where α_1, α_2, α_3, π_1, π_2', and π_3' are scalar coefficients and $\varepsilon_{CA,t}$, $\varepsilon_{X,t}$ and $\varepsilon_{Z,t}$ are error terms. It must be mentioned that in our stylised model, we ignore the lagged short term dynamics. Solving the system would result in the following multiple single-equations

$$\begin{aligned} \Delta CA_t &= \alpha_1(\pi_1 CA_{t-1} + \pi_2' X_{t-1} + \pi_3' Z_{t-1}) + \varepsilon_{CA,t} \\ \Delta X_t &= \alpha_2(\pi_1 CA_{t-1} + \pi_2' X_{t-1} + \pi_3' Z_{t-1}) + \varepsilon_{X,t} \\ \Delta Z_t &= \alpha_3(\pi_1 CA_{t-1} + \pi_2' X_{t-1} + \pi_3' Z_{t-1}) + \varepsilon_{Z,t} \end{aligned} \tag{4.19}$$

Equation 4.19 can be further transformed to

$$\begin{aligned} \Delta CA_t &= \alpha_1(\pi_1 CA_{t-1} + \pi_2' X_{t-1} + \pi_3' Z_{t-1}) + \omega_{12}' \Delta X_t + \omega_{13}' \Delta Z_t + \varepsilon_{CA.X,Z,t} \\ \Delta X_t &= \alpha_2(\pi_1 CA_{t-1} + \pi_2' X_{t-1} + \pi_3' Z_{t-1}) + \varepsilon_{X,t} \\ \Delta Z_t &= \alpha_3(\pi_1 CA_{t-1} + \pi_2' X_{t-1} + \pi_3' Z_{t-1}) + \varepsilon_{Z,t} \end{aligned}$$

where $\omega'_{12} = \frac{cov(\varepsilon_{CA},\, \varepsilon_x)}{vaz(\varepsilon_x)}$, $\omega'_{13} = \frac{cov(\varepsilon_{CA},\, \varepsilon_Z)}{vaz(\varepsilon_Z)}$ and $\varepsilon_{CA.X,Z,t} = \varepsilon_{CA} - \omega'_{12}\varepsilon_X - \omega'_{13}\varepsilon_Z$.

This multiple-equation system can be transformed into one-equation system or separated into one-equation system in the presence of weak exogeneity. This happens when $\alpha_2 = \alpha_3 = 0$.

$$\begin{aligned}
\Delta CA_t &= \alpha_1(\pi_1 CA_{t-1} + \pi'_2 X_{t-1} + \pi'_3 Z_{t-1}) + \omega'_{12}\Delta X_t + \omega'_{13}\Delta Z_t + \varepsilon_{CA.X,Z,t} \qquad (4.20)\\
\Delta X_t &= \varepsilon_{X,t}\\
\Delta Z_t &= \varepsilon_{Z,t}
\end{aligned}$$

Now we can estimate $ECM_{t-1} = CA_{t-1} + \pi'_2 X_{t-1} + \pi'_3 Z_{t-1}$ efficiently by OLS (Johansen (1992)). Short of analysing the full system, we test $\alpha_2 = \alpha_3 = 0$ by testing $\nu_2 = \nu_3 = 0$ in the auxiliary regressions (Harbo, Johansen, Nielsen and Rahbek (1998))

$$\begin{aligned}
\Delta X_t &= \nu_2 ECM_{t-1} + \mu_{X,t}\\
\Delta Z_t &= \nu_3 ECM_{t-1} + \mu_{Z,t}
\end{aligned}$$

Alternatively we can also test $\nu_2 = \nu_3 = 0$ in

$$\Delta X_t = \nu_2 ECM_{t-1} + \nu_4 \Delta ECM_{t-1} + \nu'_5 \Delta X_{t-1} + \mu_{X,t} \qquad (4.21)$$

$$\Delta Z_t = \nu_3 ECM_{t-1} + \nu_6 \Delta ECM_{t-1} + \nu'_7 \Delta Z_{t-1} + \mu_{Z,t} \qquad (4.22)$$

In our case, it would be interesting to go one step further and condition X on Z

so that the equivalent system to equation 4.20 would be

$$\begin{aligned}
\Delta CA_t &= \alpha_1(\pi_1 CA_{t-1} + \pi_2' X_{t-1} + \pi_3' Z_{t-1}) && (4.23)\\
&\quad + \omega_{12}' \Delta X_t + \omega_{13}' \Delta Z_t + \varepsilon_{CA.X,Z,t} && (4.24)\\
\Delta X_t &= \omega_{23}' \Delta Z_t + \varepsilon_{X.Z,t}\\
\Delta Z_t &= \varepsilon_{Z,t}
\end{aligned}$$

where $\omega_{23}' = \frac{cov(\varepsilon_X,\, \varepsilon_Z)}{vaz(\varepsilon_Z)}$ and $\varepsilon_{X.Z,t} = \varepsilon_X - \omega_{23}' \varepsilon_Z$. Now $\varepsilon_{X.Z,t}$ and $\varepsilon_{Z,t}$ are independent. If $\omega_{23}' = 0$, then we get

$$\begin{aligned}
\Delta CA_t &= \alpha_1(\pi_1 CA_{t-1} + \pi_2' X_{t-1} + \pi_3' Z_{t-1}) && (4.25)\\
&\quad + \omega_{12}' \Delta X_t + \omega_{13}' \Delta Z_t + \varepsilon_{CA.X,Z,t} && (4.26)\\
\Delta X_t &= \varepsilon_{X.Z,t}\\
\Delta Z_t &= \varepsilon_{Z,t}
\end{aligned}$$

We get a stronger interpretation here since we can separate demographic and economic effects. For this, we have to test $\psi_1 = \psi_2' = 0$ in the auxiliary regression

$$\Delta X_t = \psi_1 ECM_{t-1} + \psi_2' \Delta Z_t + \mu_{X.Z,t} \qquad (4.27)$$

To summarize our econometric framework, we employ the autoregressive distributed lag model (ARDL) for our regression analysis. We present the idea of a system, the more general framework. However, we do not conduct regression

analysis on multi-equation systems. Rather, we conduct the necessary tests that allows us to look at the single equation from the multi-dimensional system. Using this methodology, in the next section, we conduct the empirical analysis to study the impact of demographics on current accounts.

4.9 Empirical Studies

In this section, we conduct econometric analysis to determine the impact of demographics on the current account of the US and China. We start off the section with data description. We have three datasets: two datasets for the US and China, including 30 years of annual data. We use another longer time series for the US, starting from early 1960. All the econometric analysis in this chapter has been conducted using Eviews. At first, we follow Higgins' (1998) framework in terms of including the explanatory variables but use our econometric methodology. For these particular set of regressions, we include demographics, income per capita growth and real price of investment goods only as explanatory variables for describing current accounts. We find that the demographic variables do not turn out to be significant and the model does a poor job of explaining current accounts. We, therefore, extend our analysis by including the other control variables, described in section 4.7. We start off with a regression that includes all the control variables and a trend component, and remove the variables that are not statistically significant. Using this methodology, we find that the demographic variables turn out to be significant. We find the best results for both countries

and conduct the necessary tests to establish the robustness of our results. We also conduct the necessary exogeneity tests to confirm that we can reduce from the multi-equation system to a single equation with the current accounts as the dependent variable.

We present our results in the following manner: we first present the results for all the three time series using Higgins framework. When we include the other control variables and present our best results, we first present and discuss the results of the short series for the US and China. We then include the results for the longer time series for the US. We should clarify that, while conducting the analysis using Higgins framework, we find that the demographic variables are not statistically significant when we run country-specific regressions. Hence, we extend the model to include other economic variables that literature has found to be statistically significant. Our empirical strategy is thus to use the results from Higgins framework only to the extent that it shows why panel framework may not always be applicable for country-specific cases. Beyond that, our discussions of the role of demographics on current accounts revolve around the best results derived using the additional predictors. Therefore, beyond this section, we use only the best results presented in Tables 4.5 (US short dataset), 4.6 (China dataset) and 4.13 (US long dataset) to frame our discussion about the role of demographics on current accounts. Hence, our final empirical strategy is to use Higgins (1998) formulation of representing the entire age structure in the form of three demographic variables. In addition, we use the autoregressive distributed lag model (ARDL) framework to conduct our empirical analysis and use additional predictors of cur-

rent accounts in our regressions. In the subsequent sections, we use the results derived from this empirical strategy (the combination of Higgins' representation of demographic variables, ARDL framework and the addition of more predictors) when we discuss in detail about the role of demographics in shaping the evolution of current accounts. In the subsequent sections, there is no weight given to the results using Higgins' framework of using only income per capita growth and relative price of investment goods as controlled variables.

4.9.1 Data Description

We express current account as a share of GDP and include the ten other explanatory variables discussed in section 4.7. For all the regressions, the sources of the data for current account and the controlled explanatory variables are the IMF World Economic Outlook (April 2011), Penn World Table and World Bank Development Indicators. The output gap is constructed using the HP filter (in Eviews). The population statistics (both history and projections) are taken from the UN statistical database. As mentioned in section 4.6.1, the population statistics are grouped in 15 age shares: 0–4, 5–9, ..., 70+. Using the UN data, we construct the three demographic variables as mentioned in section 4.6.1. We include 30 years of annual data, from 1980 to 2010. In China, we find choppy and noisy data for the early 1980s and no data for some data series. Hence, we use 1985 to 2010. It must also be mentioned that all the econometric analysis in this chapter has been done using Eviews.

One of the potential criticisms of our datasets could be that our sample size is small. We try to address this by running another set of regressions using a longer time series for the US, dating back from early 1960. Due to data limitations, we are unable to do the same for China. The demographic variables has the same data source as the shorter series. For the longer times series for the US, we resort mainly to the US Bureau of Economic Analysis (BEA) and Bank for International Settlements (BIS) as additional sources. The current account expressed as a share of GDP and income per capita growth are taken from BEA. Like the shorter time series, the output gap is constructed using the HP filter. The underlying data of real GDP comes from BEA. The real price of investment goods is taken from World Development Indicators. The long term real interest rate is taken from World Bank. The longer series for real effective exchange rate is taken from BIS.

4.9.2 Higgins Framework

We start our regression analysis by initially including only two control variables, income per capita growth and real price of investment goods, used by Higgins (1998). In terms of regression methodology, we use our autoregressive distributed lag (ARDL) framework to run regressions for each country separately, see section 4.8. The results for the US and China are given in Tables 4.1 and 4.2. For both US and China, the demographic or the lagged demographic variables are neither individually nor jointly significant. The only exception to this rule is ΔD_{1t} of China. In terms of the control variables, income per capita growth, both lagged

Dependent Variable: ΔCA_t				
Sample (adjusted): 1982 2009				
Included observations: 28 after adjustments				
Variable	Coefficient	Std. Error	t-statistics	Prob.
c	-130.62	133.96	-0.98	0.35
ΔD_{1t}	-399.48	518.79	-0.77	0.45
ΔD_{2t}	55.30	79.09	0.70	0.50
ΔD_{3t}	-2.14	3.27	-0.65	0.52
ΔINC_t	-0.27	0.11	-2.44	0.03
ΔRPI_t	0.05	0.16	0.30	0.77
CA_{t-1}	-0.30	0.26	-1.15	0.27
D_{1t-1}	238.66	408.36	0.58	0.57
D_{2t-1}	-51.21	63.18	-0.81	0.43
D_{3t-1}	2.33	2.54	0.92	0.37
INC_{t-1}	-0.31	0.13	-2.38	0.03
RPI_{t-1}	0.04	0.23	0.15	0.88
$@TREND_t$	1.93	1.79	1.08	0.30
R-squares	0.74		Log likelihood	-12.57
Adj. R-squares	0.54		F-statistic	3.64
Wald test for joint significance of $\Delta D_{1t}, \Delta D_{2t}, \Delta D_{3t}$: P(F-stat)=0.79				
Wald test for joint significance of $D_{1t-1}, D_{2t-1}, D_{3t-1}$: P(F-stat)=0.11				

Table 4.1: Results for the US using Income per Capita Growth and Relative Price of Investment Goods as Control Variables

Dependent Variable: ΔCA_t				
Sample (adjusted): 1985 2009				
Included observations: 25 after adjustments				
Variable	Coefficient	Std. Error	t-statistics	Prob.
c	346.03	269.67	1.28	0.22
ΔD_{1t}	2542.79	1149.89	2.21	0.05
ΔD_{2t}	-349.02	185.32	-1.88	0.08
ΔD_{3t}	9.54	9.28	1.03	0.32
ΔINC_t	-0.33	0.22	-1.47	0.17
ΔRPI_t	-0.73	0.27	-2.71	0.02
CA_{t-1}	-0.32	0.25	-1.28	0.22
D_{1t-1}	-86.15	853.63	-0.10	0.92
D_{2t-1}	7.97	185.23	0.04	0.97
D_{3t-1}	0.36	9.50	0.04	0.97
INC_{t-1}	-0.25	0.20	-1.24	0.24
RPI_{t-1}	-0.65	0.37	-1.77	0.10
$@TREND_t$	1.94	5.62	-0.35	0.74
R-squares	0.74		Log likelihood	-38.27
Adj. R-squares	0.48		F-statistic	2.81
Wald test for joint significance of $\Delta D_{1t}, \Delta D_{2t}, \Delta D_{3t}$: P(F-stat)=0.22				
Wald test for joint significance of $D_{1t-1}, D_{2t-1}, D_{3t-1}$: P(F-stat)=0.19				

Table 4.2: Results for China using Income per Capita Growth and Relative Price of Investment Goods as Control Variables

and contemporary values, in the US are significant and have the expected sign. The relative price of investment goods does not have the correct sign and is not significant for the US. In China, both income per capita growth and the relative price of investment goods is significant. However, only the contemporary value of relative price of investment goods is statistically significant. The trend component is also not statistically significant. The removal of the trend component does not make any difference in the regressions, in terms of the statistical significance of the dependent variables. We can conclude that this set-up fails to extract the impact of demographics on current account imbalances when individual country regressions are conducted, as opposed to panel framework. This is because there are other forces at play that have an impact on the current account, and without taking those control variables into account, our regressions suffer from omitted variable bias.

The fact that Higgins' framework does not give significant results for demographic variables when individual country regressions are run is also substantiated when we use the longer time series for the US. Tables 4.3 and 4.4 give the results. Table 4.3 shows that income per capita growth and real price of investment goods have the expected signs. The demographic variables are mostly not statistically significant on a 5% level. The joint tests for demographic variables show that they are not jointly significant for ΔD_{1t}, ΔD_{2t} and ΔD_{3t}. The demographic variables are just about significant on a 10% level for D_{2t-1} and D_{3t-1}. We present another version of the results where we remove the variables that are not statistically significant. Table 4.4 presents the results. Again, the income per capita growth and

Dependent Variable: ΔCA_t				
Sample (adjusted): 1962 2009				
Included observations: 48 after adjustments				
Variable	Coefficient	Std. Error	t-statistics	Prob.
c	37.17	17.74	2.10	0.04
ΔD_{1t}	197.59	274.20	0.72	0.48
ΔD_{2t}	-28.24	42.96	-0.66	0.52
ΔD_{3t}	1.12	1.69	0.66	0.51
ΔINC_t	-0.11	0.05	-2.21	0.03
ΔRPI_t	-0.02	0.07	-0.35	0.73
CA_{t-1}	-0.15	0.11	-1.29	0.20
D_{1t-1}	28.15	20.08	1.40	0.17
D_{2t-1}	-4.28	2.21	-1.93	0.06
D_{3t-1}	0.27	0.11	2.35	0.02
INC_{t-1}	-0.13	0.07	-2.01	0.05
RPI_{t-1}	-0.13	0.06	-2.14	0.04
$@TREND$	-0.47	0.26	-1.81	0.08
R-squares	0.50		Log likelihood	-31.49
Adj. R-squares	0.33		F-statistic	2.91
Wald test for joint significance of $\Delta D_{1t}, \Delta D_{2t}, \Delta D_{3t}$: P(F-stat)=0.74				
Wald test for joint significance of $D_{1t-1}, D_{2t-1}, D_{3t-1}$: P(F-stat)=0.09				

Table 4.3: USA: Results using a Longer Times Series. This regression includes only income per capita growth and real price of investment as controlled variables

Dependent Variable: ΔCA_t				
Sample (adjusted): 1962 2009				
Included observations: 48 after adjustments				
Variable	Coefficient	Std. Error	t-statistics	Prob.
c	21.89	8.28	2.64	0.01
ΔINC_t	-0.12	0.04	-2.80	0.01
CA_{t-1}	-0.17	0.10	-1.78	0.08
D_{1t-1}	14.16	7.01	2.02	0.05
D_{2t-1}	-4.08	1.64	-2.49	0.02
D_{3t-1}	0.28	0.11	2.58	0.01
INC_{t-1}	-0.15	0.06	-2.55	0.01
RPI_{t-1}	-0.10	0.04	-2.48	0.02
$@TREND$	-0.26	0.11	-2.44	0.02
R-squares	0.48		Log likelihood	-32.36
Adj. R-squares	0.37		F-statistic	4.51
Wald test for joint significance of $D_{1t-1}, D_{2t-1}, D_{3t-1}$: P(F-stat)=0.08				

Table 4.4: USA: Results using a Longer Times Series. This regression includes only income per capita growth and real price of investment as controlled variables and includes only significant variables

real price of investment goods have the expected signs and are statistically significant. D_{1t-1}, D_{2t-1} and D_{3t-1} are individually significant but the joint test for D_{1t-1}, D_{2t-1} and D_{3t-1} is not statistically significant on a 5% level (but significant on a 10% level). The R^2 of the regression is also very low. Overall, the regression results do not provide convincing argument for the importance of demographics in determining current account imbalances.

Having found that we cannot extract the impact of demographics on current account using the two variables, income per capita growth and real price of investment goods (like Higgins (1998)), we run regressions including other control variables that literature has found to be useful in explaining current account imbalances, see section 4.7. The results are presented and discussed in the next section.

4.9.3 Results Including Other Control Variables

In this section, we use ARDL framework and run regressions to explain current account using the three demographic variables and the control variables mentioned in section 4.7. We start off with a general model that includes all the control variables but remove them if they are found to be statistically insignificant. We also include a trend component in our regression and remove it from the regression if it is found statistically insignificant. We first present the results using the shorter time series. This includes both China and the US. We then present the results for the US with the longer time series. For all our best results, we carry out the usual

residual diagnostics test conducted when performing such analysis and make sure that our results are free from of any such errors.

Results using Short Data Series

We start with a general specification of ARDL that includes all the ten control variables and the demographic variables. In addition, for the US, we include time dummies and another variable that captures the proportion of household financial assets that are in the form of corporate equities (denoted by $EQUITY_t$ in regression result presentations). The data for this variable has been taken from the US Flow of Funds. The motivation for using this variable is the idea that the deterioration in the US current account could be due to the financial market liberalization in the US since early 1980s (Obstfeld (2005), OECD (2011), Obstfeld and Rogoff (2005)). The idea here is that the rapid stock market liberalization since mid-1980s resulted in increase in household wealth which households have used to import goods (rather than in domestic goods). This has resulted in the deterioration in the current account deficit. While the capital openness measures the openness of all the sectors of the economy, this variable put specific attention on the specific impact of the increase in households' equity ownership. It must be mentioned that in our regressions, we had also included other wealth-related variables but they did not turn out to be significant. We used both households' assets in housing as a share of total assets and households' total mortgage as a share of total liabilities. These variables were intended to capture the impact of households' housing wealth and liabilities on the current account. We also

Dependent Variable: ΔCA_t				
Sample (adjusted): 1981 2010				
Included observations: 30 after adjustments				
Variable	Coefficient	Std. Error	t-statistics	Prob.
c	15.49	2.88	5.37	0.00
ΔD_{1t}	1094.60	223.81	4.89	0.00
ΔD_{2t}	-165.47	35.06	-4.72	0.00
ΔD_{3t}	6.22	1.36	4.57	0.00
CA_{t-1}	-0.63	0.09	-7.44	0.00
$RINT_{t-1}$	0.22	0.05	4.29	0.00
$EXRATE_{t-1}$	-0.04	0.01	-4.25	0.00
RPI_{t-1}	-0.06	0.02	-2.52	0.02
$EQUITY^*_{t-1}$	-0.18	0.02	-7.94	0.00
$DUMMY^{**}$	-0.91	0.46	-1.98	0.06
R-squares	0.90		Log likelihood	2.16
Adj. R-squares	0.86		F-statistic	21.08
Wald test for joint significance of $\Delta D_{1t}, \Delta D_{2t}, \Delta D_{3t}$: P(F-stat)=0.00				

Table 4.5: Best Results for the US. *EQUITY measures the proportion of household financial assets in equity stocks. **Dummy is a dummy variable for the period 1980-87.

included other variables like households' total deposits, credit market instruments, consumer credit, municipal securities, bank loans, other loans and advances, and commercial mortgages. All these variables were collated from the US Flow of Funds. Apart from $EQUITY_t$, all other variables turned out to be statistically insignificant.

We find that not all the ten control variables turn out to be significant for the US and China. Debelle and Faruqee (1996) also find that not all the traditional determinants of the current account are significant for each country. The results of the impact of demographics on current account imbalances are summarized in Tables 4.5 and 4.6. All the variables are statistically significant. For the three

Dependent Variable: ΔCA_t				
Sample (adjusted): 1985 2009				
Included observations: 25 after adjustments				
Variable	Coefficient	Std. Error	t-statistics	Prob.
c	6.90	2.25	3.07	0.01
ΔD_1	-1128.88	371.96	-3.03	0.01
ΔD_2	201.67	76.96	2.62	0.02
ΔD_3	-8.80	3.84	-2.29	0.04
$\Delta FISCAL$	1.28	0.17	7.31	0.00
ΔRPI	-0.31	0.05	-5.76	0.00
CA_{t-1}	-0.77	0.09	-8.86	0.00
$FISCAL_{t-1}$	2.56	0.29	8.77	0.00
INC_{t-1}	-0.33	0.06	-5.11	0.00
$EXRATE_{t-1}$	-0.05	0.006	-8.51	0.00
R-squares	0.95		Log likelihood	-16.81
Adj. R-squares	0.92		F-statistic	33.66
Wald test for joint significance of $\Delta D_1, \Delta D_2, \Delta D_3$: P(F-stat)=0.00				

Table 4.6: Best Results for China

demographic variables, we also perform two tests, the Wald test and the log-likelihood test, to show that they are jointly significant for both countries. The coefficients of the demographic variables and their implications are discussed in the next section (section 4.10.1).

In terms of the control variables, as the results in the tables show, all the control variables have the expected signs, as discussed in section 4.7. Income per capita growth and real effective exchange rates seem to be important for both the countries. For the US, long term real interest rates is important. On the other hand, in China, fiscal stance and real price of investment goods seem to be important. In the US, we also find that households' share of equity assets, as expected, play a negative impact on the current accounts. Also, we find that for

Exogeneity Tests							
Sample (adjusted): 1982 2009							
Included observations: 28 after adjustments							
Variable	Coefficient	t-statistics	Prob.		Variable	Coefficient	t-statistics	Prob.
Dependent Variable: $\Delta EQUITY_t$					Dependent Variable: ΔRPI_t			
c	4.38	1.22	0.23		c	-1.77	-0.79	0.44
ECM_{t-1}	-0.52	-1.29	0.21		ECM_{t-1}	-0.27	-1.07	0.29
ΔD_{1t}	649.20	0.70	0.49		ΔD_{1t}	379.40	0.68	0.51
ΔD_{2t}	-125.51	-0.91	0.37		ΔD_{2t}	-56.93	-0.68	0.50
ΔD_{3t}	4.64	0.87	0.39		ΔD_{3t}	2.22	0.68	0.50
Dependent Variable: $\Delta RINT_t$					Dependent Variable: $\Delta EXRATE_t$			
c	1.10	0.49	0.63		c	5.06	0.50	0.62
ECM_{t-1}	0.09	0.36	0.72		ECM_{t-1}	-0.04	-0.04	0.97
ΔD_{1t}	-0.81	-0.00	1.00		ΔD_{1t}	4763.01	1.89	0.07
ΔD_{2t}	-5.99	-0.07	0.94		ΔD_{2t}	-652.42	-1.73	0.10
ΔD_{3t}	0.35	0.11	0.91		ΔD_{3t}	25.48	1.74	0.09

Table 4.7: Exogeneity Test Results for the US. Note: ECM refers to the error correction term

the US, the time dummy capturing the period 1980-87 is significant.

Exogeneity Tests We perform two types of exogeneity tests, discussed in section 4.8.2. For the controlled variables, we perform the exogeneity test presented in equation 4.27. This test not only allows us to reduce from the multi-equation system to the single equation presentation, but also makes sure that the demographic and economic effects are separate. For the demographic variables, we perform the exogeneity test presented in equation 4.22. Both these tests (Harbo, Johansen, Nielsen and Rahbek (1998)) check that the single equation presenting the cointegrating relation is efficient so that our estimation of the ECM_t is efficient (Johansen (1992)) . The results of the exogeneity tests on the economic variables for the US and China are given in Tables 4.7 and 4.8. We find that none

Exogeneity Tests							
Sample (adjusted): 1982 2010							
Included observations: 29 after adjustments							
Variable	Coefficient	t-statistics	Prob.		Variable	Coefficient	t-statistics	Prob
Dependent Variable: $\Delta FISCAL_t$					Dependent Variable: ΔRPI_t			
c	-1.81	-0.84	0.41		c	10.22	1.23	0.23
ECM_{t-1}	0.12	1.74	0.10		ECM_{t-1}	0.34	1.24	0.23
ΔD_{1t}	135.34	0.38	0.71		ΔD_{1t}	2397.88	1.74	0.10
ΔD_{2t}	-10.32	-0.15	0.89		ΔD_{2t}	-499.96	-1.83	0.08
ΔD_{3t}	0.10	0.03	0.97		ΔD_{3t}	25.20	1.89	0.07
Dependent Variable: ΔINC_t					Variable: $\Delta EXRATE_t$			
c	-3.29	-0.42	0.68		c	-15.04	-0.33	0.75
ECM_{t-1}	0.04	0.16	0.87		ECM_{t-1}	0.13	0.09	0.93
ΔD_{1t}	-96.44	-0.07	0.94		ΔD_{1t}	3914.61	0.51	0.61
ΔD_{2t}	34.39	0.13	0.90		ΔD_{2t}	-667.53	-0.44	0.66
ΔD_{3t}	-1.94	-0.16	0.88		ΔD_{3t}	32.33	0.44	0.67

Table 4.8: Exogeneity Test Results for China. Note: ECM refers to the error correction term

Exogeneity Tests			
Variable	Coefficient	t-statistics	Prob.
Dependent Variable: ΔD_{1t}			
c	0.01	1.74	0.09
ΔD_{1t-1}	0.65	4.32	0.00
ECM_{t-1}	-0.00	-0.82	0.42
ΔECM_{t-1}	-0.00	-0.85	0.40
Dependent Variable: ΔD_{2t}			
c	0.00	0.02	0.98
ΔD_{2t-1}	0.91	8.45	0.00
ECM_{t-1}	-0.00	-1.01	0.32
ΔECM_{t-1}	-0.00	-0.74	0.47
Dependent Variable: ΔD_{3t}			
c	-0.40	-1.20	0.24
ΔD_{3t-1}	1.01	13.56	0.00
ECM_{t-1}	-0.06	-1.08	0.29
ΔECM_{t-1}	-0.06	-0.89	0.38

Table 4.9: US Short Data Series. Note: ECM refers to the error correction term

Exogeneity Tests				
Variable	Coefficient	Std. Error	t-statistics	Prob.
Dependent Variable: ΔD_{1t}				
c	0.004	0.008	0.48	0.64
ΔD_{1t-1}	0.79	0.11	6.91	0.00
ECM_{t-1}	0.0007	0.0007	1.01	0.32
ΔECM_{t-1}	0.0008	0.0006	1.37	0.19
Dependent Variable: ΔD_{2t}				
c	0.01	0.08	0.16	0.87
ΔD_{2t-1}	0.89	0.08	10.84	0.00
ECM_{t-1}	0.008	0.01	1.04	0.31
ΔECM_{t-1}	0.009	0.01	1.47	0.16
Dependent Variable: ΔD_{3t}				
c	-0.06	0.88	-0.07	0.95
ΔD_{3t-1}	0.94	0.06	14.64	0.00
ECM_{t-1}	0.08	0.08	1.01	0.32
ΔECM_{t-1}	0.10	0.07	1.53	0.14

Table 4.10: China. Note: ECM refers to the error correction term

of the controlled variables are dependent on the error correction term, ECM_t. In particular, the real effective exchange rate is not dependent on the error correction term. They are also not dependent on the demographic variables, indicating that the effects of the economic variables and the demographic variables are separate.

The results on the exogeneity tests on the demographic variables for the US and China are presented in Tables 4.9 and 4.10. For our exogeneity tests on the demographic variables, we find that the demographic variables are not dependent on the ECM_t. Though Tables 4.9 and 4.10 show that the lagged dependent variables are significant, this is not a problem for exogeneity. Both the tests justify the reduction of the multi-equation system into a single equation with current accounts as the independent variable and the economic and demographic variables as the dependent variables.

Dependent Variable: ΔCA_t				
Sample (adjusted): 1981 2010				
Included observations: 30 after adjustments				
Variable	Coefficient	Std. Error	t-statistics	Prob.
c	23.63	6.90	3.43	0.00
ΔD_{1t}	958.12	244.04	3.93	0.00
ΔD_{2t}	-147.44	37.19	-3.96	0.00
ΔD_{3t}	5.76	1.39	4.16	0.00
CA_{t-1}	-0.57	0.10	-5.96	0.00
$RINT_{t-1}$	0.22	0.05	4.50	0.00
$EXRATE_{t-1}$	-0.04	0.01	-4.21	0.00
RPI_{t-1}	-0.12	0.05	-2.31	0.03
$EQUITY_{t-1}$	-0.16	0.03	-6.09	0.00
$DUMMY_t$	-0.82	0.46	-1.79	0.09
$@TREND$	-0.09	0.07	-1.30	0.21
R-squares	0.91		Log likelihood	3.43
Adj. R-squares	0.87		F-statistic	19.78
Wald test for joint significance of $\Delta D_{1t}, \Delta D_{2t}, \Delta D_{3t}$: P(F-stat)=0.00				

Table 4.11: USA: Trend Component is Insignificant

Results Including Trend Components It must be mentioned that in the process of finding the best model, we initially do include trend components in our regressions to capture any long term movement that is not explained by the demographic variables and the control variables. As Tables 4.11 and 4.12 show, the trend components turn out to be statistically insignificant for both US and China. After including the trend components, the demographic variables are still statistically significant, both individually and jointly, for the US and China.

Error Correction Component Figure 4-7 also shows the error correction term for the US and China, including both the trend and without trend. For both the countries, the trend and without trend components are almost same in shape but

Dependent Variable: ΔCA_t				
Sample (adjusted): 1985 2009				
Included observations: 25 after adjustments				
Variable	Coefficient	Std. Error	t-statistics	Prob.
c	14.41	6.98	2.06	0.06
ΔD_1	-1402.60	440.19	-3.19	0.01
ΔD_2	244.79	85.15	2.87	0.01
ΔD_3	-10.15	3.98	-2.55	0.02
$\Delta FISCAL$	1.17	0.20	5.92	0.00
ΔRPI	-0.33	0.06	-5.87	0.00
CA_{t-1}	-0.76	0.09	-8.78	0.00
$FISCAL_{t-1}$	2.54	0.29	8.77	0.00
INC_{t-1}	-0.33	0.06	-5.20	0.00
$EXRATE_{t-1}$	-0.06	0.01	-5.72	0.00
$@TREND$	-0.25	0.22	-1.14	0.28
R-squares	0.96		Log likelihood	-15.71
Adj. R-squares	0.93		F-statistic	31.01
Wald test for joint significance of $\Delta D_1, \Delta D_2, \Delta D_3$: P(F-stat)=0.01				

Table 4.12: China: Trend Component is Insignificant

the scaling differs. In China, the error correction term seems to have a slight trending property. However, that is not eliminated when the trending component is added.

Results using Long Data Series

We conduct the same analysis as before but using the longer data series for the US. There are many motivations for using a longer data series. It helps to better understand how the role of demographics has evolved over the years. The presence of dummy in the best results for the shorter time series hints that there is a possibility of structural break when the longer time series is used. We play with various structural breaks for the US using the longer data series and find that we

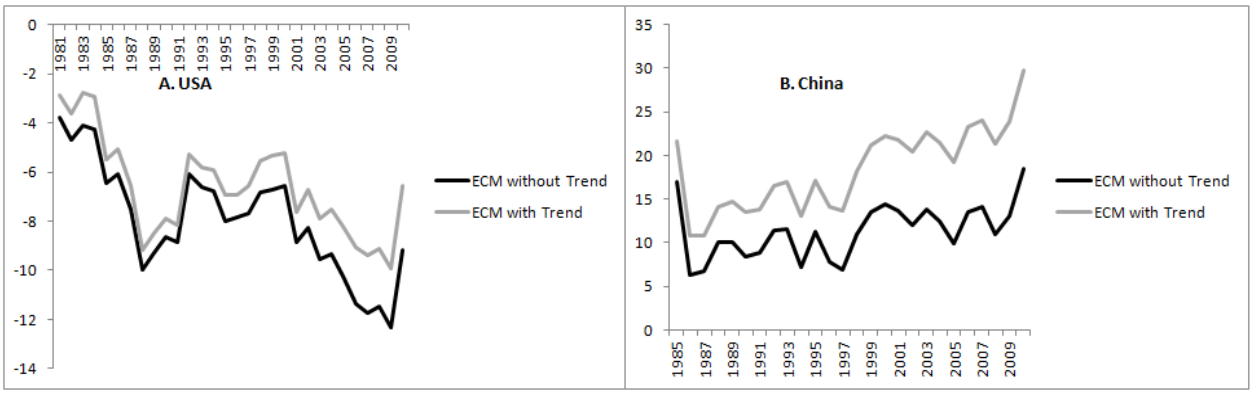

Figure 4-7: The Error Correction Component for the Short Data Series (with and without trend)

have the best result when we impose structural break at 1986. In section 4.10.1, we will attempt to explain the economic significance of the structural break at 1986 when we discuss the age coefficients. For the time being, we just present the best results using the longer data series. Table 4.13 shows the best results using the demographic variables and all other control variables. Like the shorter time series version, the demographic variables are individually statistically significant. However, we see that now we have two different responses of demographics towards current accounts, one before 1986 and the other after 1986. This is evident from the presence of the interactive terms with the dummy variables that are used to show structural breaks in the economy. Surely, something has happened in the economy during that time that results in the change in behaviour of cohorts towards current account deficits. In the next section, we will attempt to understand that.

In terms of controlled variables, we find that output gap is statistically significant. This is in line with Debelle and Faruqee (1996) who claim that output

Dependent Variable: ΔCA_t				
Sample (adjusted): 1965 2010				
Included observations: 46 after adjustments				
Variable	Coefficient	Std. Error	t-statistics	Prob.
c	8.98	1.42	6.31	0.00
ΔD_{1t}	938.90	203.75	4.61	0.00
ΔD_{2t}	-139.48	29.09	-4.79	0.00
ΔD_{3t}	5.21	1.11	4.68	0.00
$\Delta D_{1t} * DUM86$	-557.04	188.68	-2.95	0.01
$\Delta D_{2t} * DUM86$	82.49	26.93	3.06	0.00
$\Delta D_{3t} * DUM86$	-3.56	1.17	-3.04	0.00
ΔGAP_t	-0.22	0.05	-4.06	0.00
$\Delta GAP_t * DUM86$	0.19	0.06	3.08	0.00
CA_{t-1}	-0.57	0.10	-5.92	0.00
$EXRATE_{t-1}$	-0.05	0.01	-6.06	0.00
$EQUITY_{t-1}$	-0.06	0.02	-3.56	0.00
R-squares	0.79		Log likelihood	-11.13
Adj. R-squares	0.72		F-statistic	11.65
Wald test for joint significance of:				
$\Delta D_{1t}, \Delta D_{2t}, \Delta D_{3t}$: P(F-stat) = 0.00				
$\Delta D_{1t} * DUM86, \Delta D_{2t} * DUM86, \Delta D_{3t} * DUM86$: P(F-stat) = 0.03				
$\Delta D_{1t}, \Delta D_{2t}, \Delta D_{3t}, \Delta D_{1t} * DUM86, \Delta D_{2t} * DUM86, \Delta D_{3t} * DUM86$: P(F-stat) = 0.00				

Table 4.13: USA: Best Results using a Longer Times Series. DUM86=1 before 1986 and 0 since 1986.

Exogeneity Tests							
Sample (adjusted): 1965 2010							
Included observations: 46 after adjustments							
Variable	Coefficient	t-statistics	Prob.	Variable	Coefficient	t-statistics	Prob.
Dependent Variable: ΔGAP_t				Dependent Variable: $\Delta EXRATE_t$			
c	1.15	0.86	0.40	c	0.24	0.06	0.95
ECM_{t-1}	0.61	1.85	0.07	ECM_{t-1}	1.65	1.67	0.10
ΔD_{1t}	-462.33	-1.72	0.09	ΔD_{1t}	-1536.90	-1.91	0.06
ΔD_{2t}	69.91	1.63	0.11	ΔD_{2t}	244.67	1.91	0.06
ΔD_{3t}	-2.43	-1.50	0.14	ΔD_{3t}	-8.86	-1.83	0.07
Dependent Variable: $\Delta EQUITY_t$							
c	1.44	0.76	0.45				
ECM_{t-1}	-0.28	-0.60	0.55				
ΔD_{1t}	-24.98	-0.07	0.95				
ΔD_{2t}	-10.42	-0.17	0.87				
ΔD_{3t}	0.22	0.09	0.92				

Table 4.14: Exogeneity Test Results for the US Long Data Series. Note: ECM refers to the error correction term

gap should be a determinant of current accounts for advanced economies. In addition, we find that the impact of output gap on current accounts has intensified since 1986. Also, we find that the real effective exchange rate and the share of household financial assets using corporate equities are important determinants of the current accounts. It must be mentioned that, like the shorter data series, we also included variables like households' assets in housing, total mortgage, total deposits, credit market instruments, consumer credit, municipal securities, bank loans, other loans and advances, and commercial mortgages. Apart from $EQUITY_t$, all other variables turned out to be statistically insignificant.

Exogeneity Tests Table 4.14 gives the exogeneity tests and shows that there is no inefficiency in the system due to the inclusion of the controlled variables.

Exogeneity Tests			
Variable	Coefficient	t-statistics	Prob.
Dependent Variable: ΔD_{1t}			
c	0.01	3.12	0.00
ΔD_{1t-1}	0.77	9.45	0.00
ECM_{t-1}	0.00	0.88	0.38
ΔECM_{t-1}	-0.00	-1.91	0.06
Dependent Variable: ΔD_{2t}			
c	0.04	1.92	0.06
ΔD_{2t-1}	0.88	10.76	0.00
ECM_{t-1}	-0.00	-0.51	0.61
ΔECM_{t-1}	-0.01	-1.32	0.20
Dependent Variable: ΔD_{3t}			
c	0.16	0.65	0.52
ΔD_{3t-1}	0.98	14.15	0.00
ECM_{t-1}	-0.01	-0.33	0.74
ΔECM_{t-1}	-0.09	-0.98	0.33

Table 4.15: US Long Data Series. Note: ECM refers to the error correction term

There is also no inefficiency in the estimation of the error correction component. The controlled variables are not dependent on the error correction term after accounting for the lagged values of the controlled variables. Table 4.15 gives the exogeneity test for the demographic variables. Again, we find that the demographic variables are not dependent on the error correction component ECM_{t-1}. As we mentioned when discussing the exogeneity results for the shorter samples, there is no problem for exogeneity tests if the lagged dependent variables are significant.

Error Correction Component Figure 4-8 shows the error correction term, both with and without trend. The trend and without trend versions are almost similar, with glimpses of trending behaviour for both.

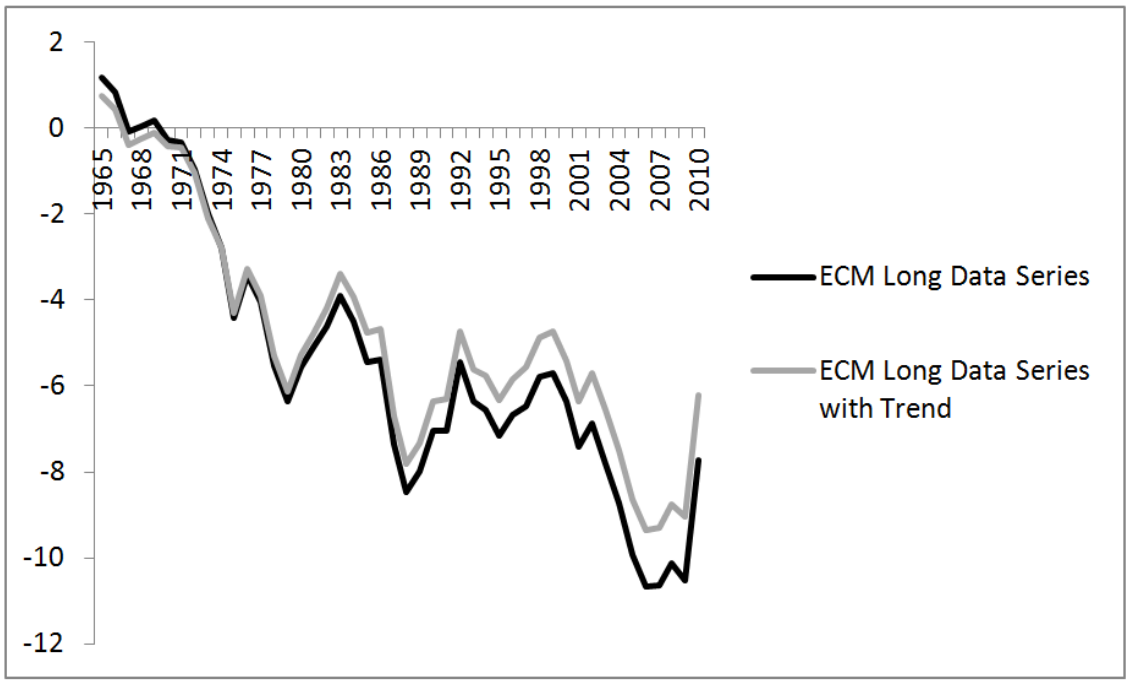

Figure 4-8: The Error Correction Component of Long Data Series (USA)

4.9.4 Interpretation of Empirical Models

In all our results, we find that demographics have a short run impact on current accounts. This implies that our model states that the change in CA_t from the previous period consists of the change associated with movement in demographics along the long-run equilibrium path. However, demographics do not determine the long run equilibrium of the current accounts. Like the previous sections, if we present $Z = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix}$, our results indicate that the models are of the form given in equation 4.15 where $\delta' = 0$, so

$$\Delta CA_t = \alpha CA_{t-1} + \beta' X_{t-1} + \gamma' \Delta X_t + \varphi' \Delta Z_t + \varepsilon_t \tag{4.28}$$

We can re-write the model in the ECM form (the same expression as equation 4.16 but re-written with $\delta' = 0$)

$$\Delta CA_t = \alpha(CA_{t-1} + \frac{\beta'}{\alpha}X_{t-1}) + \gamma'\Delta X_t + \varphi'\Delta Z_t + \varepsilon_t$$

where $ECM_{t-1} = \alpha(CA_{t-1} + \frac{\beta'}{\alpha}X_{t-1})$. Our model then indicates that the ECM_t does not depend on demographics. This means that the long run current accounts is determined by the economic variables $X_t's$ only. However, the adjustment to the long run depends on ΔZ_t. To understand what this implies, assume that $ECM_{t-1} = 0$ which indicates that the current accounts is in its long run equilibrium and also assume that $\Delta X_t = 0$. Going back to equation 4.28, a change ΔZ_t will change ΔCA_t by $\varphi'\Delta Z_t$. Then, in the next period, $ECM_t \neq 0$ because CA has changed even though $X's$ have not changed. Next, ΔCA_{t+1} will adjust to the disequilibrium though αECM_t. Thus, if $\varphi' > 0$ and $\alpha < 0$, then a change in $\Delta Z_t > 0$ will make $\Delta CA_t > 0$ and $\Delta CA_{t+1} < 0$. If we had used equation 4.23, this would have been more complicated because ΔX_t would have moved by ω'_{23} in response to ΔZ_t. However, we have already performed the required exogeneity tests to show that we could reduce from equation 4.23 to equation 4.25.

Since the level effect of demographics is insignificant, our empirical studies show that the story is not about demographics affecting the long run equilibrium of current accounts. Rather, we find that the changes in demographics affect the adjustment of current accounts to a long run equilibrium. Our empirical work thus accords with Higgins (1998) and other studies to the extent that it shows that

demographics shifts are important for current account positions. However, our empirical work gives a different interpretation of why demographics affect current accounts. Higgins (1998) and other studies start from the life cycle hypothesis to show that demographics matter for the long run equilibrium of the economy. In other words, demographics have a level effect on the current account imbalances. Our empirical work, for both USA and China, show that demographics have no long run impact on current account positions. Demographics shifts, however, still matter since the changes in demographics affect how the current accounts adjust to their long run equilibrium. The changes in demographics matter for the changes in current accounts. Though demographics do not have a level effect, they have an important adjustment effect. In the next sections, we delve deeper on the nature of this adjustment effect. We first see how each age group contributes towards this adjustment effect. Then, we see the impact of overall demographic changes on the changes in current account positions.

4.10 The Impact of Each Age Group on Current Accounts

In this section, we look at the impact of each age group on current accounts. In the next section, we look into the overall impact of demographics on current accounts. We first focus on the age coefficients that informs us of the response of each age group towards current account deficits and surpluses. We first present

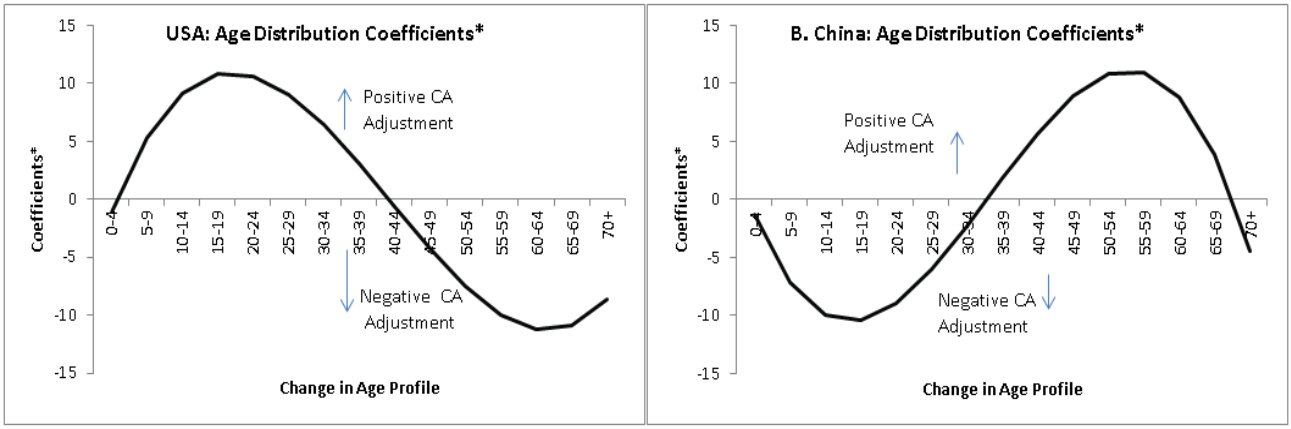

Figure 4-9: The Relationship between the Change in Current Accounts versus the Change in Demographics using the Short Data Series. Note that the age coefficients for the US and China are taken from the best results presented in Tables 4.5 and 4.6 respectively. *The age distribution coefficients represent the change in current account associated with a 0.01 change in the corresponding population age shares (expressed as ratios).

the results from the short data series of the US and China, followed by the results from the longer times series of the US.

4.10.1 The Age Coefficients

The formulation we have used allows us to map the impact of changes in population at particular age groups on the changes in current account positions. We represent the population distribution in the form of three demographic variables and our regression analysis determines the coefficients, representing the impact on current accounts from these three demographic variables. From here, we retrieve the coefficients for the entire age structure. The methodology of doing this has been discussed in detail in sections 4.6.1 and 4.6.2. Briefly speaking, the three coefficients of our regression analysis are γ_{d1}, γ_{d2} and γ_{d3} in section 4.6.2. From here, we can retrieve γ_{d0} using 4.6. Once we have γ_{d0}, γ_{d1}, γ_{d2} and γ_{d3}, we can

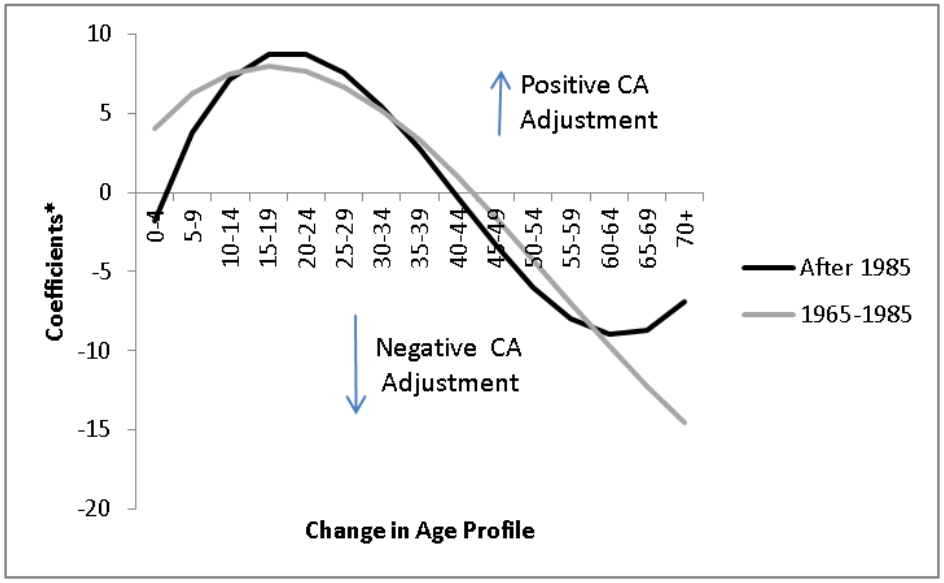

Figure 4-10: The Relationship between the Change in Current Accounts versus the Change in Demographics using the Long Data Series for the US. Note that the age coefficients for the US are taken from the best results presented in Table 4.13. *The age distribution coefficients represent the change in current accounts associated with a 0.01 change in the corresponding population age shares (expressed as ratios).

determine the change in current accounts for the change in every age group using equation 4.3.

For China, Figure 4-9B shows that up to the age of 35, the positive change in population appears to be a drag on the current account position – in other words, for this age group, the rate of investment is more than the rate of saving, on average. Between ages 35 and 69, the change in population seems to give a positive adjustment push on the current accounts. Here, the rate of savings is more than the rate of investment. The people in the age group 35–69 are traditionally known as the prime savers and having more of them in the population would tend to improve the current account position. Above 69, the population changes again tend to become a drag on the current account, creating pressure for negative

adjustments.

For the US, our estimate suggests that the change in population in age groups 5–39 create a positive current account adjustment pressure. This age group starts and ends earlier than in our estimates in China. Figure 4-9A shows that the change in current account profile versus the change in age group for the US. Figure 4-10 compares the age coefficients versus the change in age profile using the longer time series with structural break at 1986. The change in current account coefficient versus change in age profile of the series since 1985 matches the shorter time series: both showing that the positive change in the age group from 5–39 contributes positively towards current account adjustments and changes in population in groups 0–4 and 40+ contribute negatively towards current account adjustments. The series before 1985 also has similar message with just two differences: the change in age groups 0–4 and 40–44 contributes positively for the series before 1986 but contributes negatively for the series since 1986. This suggests that the change in current account coefficient versus the change in age profile has, broadly speaking, remained same over the years with the differences that the change in age groups 0–4 and 40–44 used to contribute positively towards current account adjustments before 1986 but has been contributing negatively adjustments since then. The coefficients are more or less of the same magnitude. In line with the findings of the shorter data series, it seems that the change in age groups 40–69 – the change in the population of age groups that contributes positively towards current account adjustments in China – are in fact contributing negatively towards current account adjustments in the US.

The age coefficients versus the change in age profile charts provide two key insights on the role of demographics in current account discussions.

First, contrary to existing studies, we find an adjustment effect on current accounts, rather than a level effect. This means that we cannot determine the age group of prime savers. However, if the pressure for saving is coming from the prime savers, the increase in population in that age groups should create a positive adjustment pressure on the current accounts. Using that logic, we can say that the prime savers refer to those age groups where the incremental increase in people results in an incremental increase in current accounts. Our model estimates show that the change in age groups that contribute positively towards current account adjustments is different for the US and China. This is a very good indication that the prime savers in the US and China are different, suggesting that the panel approach may not be the best methodology since it hides country-specific factors that may significantly change results. In terms of policy decisions related to capital flows, this indicates that policy makers should not automatically assume that the age group 35–69 contributes to a current account surplus without accounting for country-specific factors.

Second, though we find an adjustment effect of demographics on current accounts positions and literature finds a level effect, our age groups of prime savers in China is in line with Wilson and Ahmed (2010) and Higgins (1998), who estimate the prime saving age group to be 35–69 in a panel framework. The shape of the evolution of the change in current account profile versus change in population in each age group for China is given in Figure 4-9B. For the US, the same infor-

mation is presented in Figure 4-9A. The shape of the evolution of the change in current account with the change in the age profile is different from China's and the panel estimates since the US has historically run one of the lowest saving rates in the world (Obstfeld and Rogoff (2009), Blanchard and Milesi-Ferretti (2009)). Figure 4-9A suggests that the low saving rate could occur because the traditional prime savers in the US are not saving enough. On the flipside, Chinese consumers are characterized by one of the highest saving rates in the world (Dew and Martin (2011)) and Figure 4-9B suggests that the pressure for the change in population in prime saving age group to push for positive current account adjustments is significant in China.

4.10.2 Explaining the US Current Account versus Age Profile

In the US we find that the pressure for positive adjustment in current accounts comes from the change in the population in age groups 5–39. This is earlier than the findings in China and other studies (Wilson and Ahmed (2010) and Higgins (1998)). This seems also against the predictions of the life cycle theories. If individuals following the savings and investment behaviour predicted by life cycle theories, the pressure for net saving should be higher for the age group 35–69. Consequently, the incremental addition of population to the age group 35–69 should create a positive adjustment pressure on the current accounts. In this section, we attempt to understand why the US profile is different based on

our results and what existing literature has to say on the issue. The problem with using the existing literature is that the existing studies are mostly concerned with explaining the level of current accounts, rather than the adjustment of current accounts. However, visiting these studies are still useful since we get important clues about the adjustment process by understanding the anomalous features about the determinants of the level of current accounts in the US.

The fact that the long and shorter data series yielded similar shapes of age profiles hints that it is possibly a persisting feature of the economy. Though not looking at granular demographic datasets, there is a host of studies that support the notion that the pressure to consumer more, save less and invest more is a feature of the US economy. For example, Mann (1999) contends that the puzzle of the continued widening US trade deficit, even when the world grows faster than the US, can be attributed to a greater appetite for imports by the US consumers and business than foreigners' appetites for the US exports. This is known as Houthakker-Magee effect and this has been a feature of the US trade for the whole of postwar period. This phenomenon goes against economic principles. This feature of the US economy is also supported by Blanchard, Giavazzi and Sa (2005).

In a similar vein, Holman (2001) attributes the surge in current account deficit in the 1990s to two factors. First, there was a surge in investment spending caused by accelerating US productivity (output per hour worked) compared to workers of other countries. From 1980-1994, the US productivity in the manufacturing sector had grown, on average, by 2.7 percent every year. On the other hand, the same

indicator grew by, on average, 4.8 percent every year from 1995-1999. In particular, in 1999, the manufacturing productivity grew by 6.4 percent. The surge in productivity led to an investment boom. This is because, holding everything else constant, the rise in productivity causes increase in investment spending causing total domestic spending to exceed total domestic production. In order to fill up the excess spending over production, countries usually imports more goods and services than it exports. Second, in line with the growth in productivity, the US stock market went through a boom period. This increased consumer spending due to the increase in household wealth. The US consumers satisfied a large part of their increased demand in goods and services with imported goods, leading to a drag in current account balances. This argument is also supported by Obstfeld (2005) who thinks that greater financial liberalization allows the US to run much bigger deficits for much bigger periods.

Our econometric results seem to be in tune with the findings of existing literature. The US economy has historically run one of the lowest savings rates and the spending on consumption has been high. Both the characteristics resulted in the increase in the population in the traditionally prime saving group to cause the current accounts to deteriorate rather than improve. Our results suggests that, amongst the traditional prime savers, it is only the increase in population in the age group 35–39 that causes a positive adjustment effect on the current accounts. In other words, the increase in the age group 35–39 increases the current account of the US. The change in the remaining age groups that fall in the traditional prime saving group actually causes a negative adjustment effect on the current

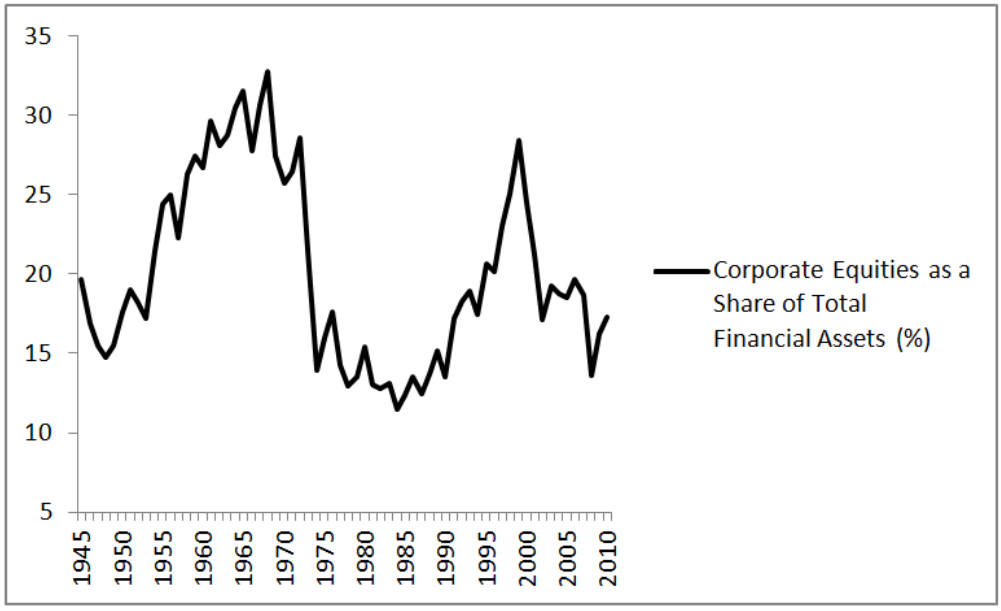

Figure 4-11: The Share of Equities in Households' Total Financial Assets in the US

account. In that light, it seems that the incremental pressure for positive current account adjustment is coming from the age groups 5–39, strongly hinting that the prime saving age group in the US is not 35–69, rather close to 5–39.

Our econometric analysis provides some convincing evidence in support of Holman (2001)'s second point. Using the longer dataseries, we find a structural break in the response of demographic variables in 1986. Figure 4-11 shows that the structural break is around the same time as the rise in corporate equities in household's total financial asset composition. The shape of the graph is similar if we plot households' corporate equities as a share of total assets. One can make an argument that the households used the boom period in the US stock market to increase their wealth, which they subsequently used to satisfy their demand for imported goods. However, looking at the age coefficient profile, the change before 1986 and afterwards is not too substantial. As mentioned before, the only

substantial difference is that the change in age groups 40–44 also started pushing for a negative adjustment effect on the current accounts since 1986. We can thus argue that the fact that the change in age groups that should traditionally pose positive adjustment effect on the current account actually created a negative adjustment effect on the US current account seems to be a persistent feature of the US economy. The asset market boom of the late 1980s had slightly exacerbated the situation. As mentioned when we presented the econometric results, we did also check the possibility if households' wealth from the housing market played a role in changing the response of demographics towards current accounts. We also looked at other variables like households' deposits, consumer credit, etc. We found no evidence that these variables made any impact in changing the impact of demographics on current accounts.

4.11 The Overall Impact of Demographics on Current Account Imbalances

In the previous section, we showed how different age groups contribute towards current account positive/negative adjustments. In this section, we discuss how this adds up to the bigger picture – that is – the overall impact of demographics on current account imbalances as implied by our model results. We see how the different age groups have evolved in the US and China in the past three decades and what that implies about the role of demographics in shaping current

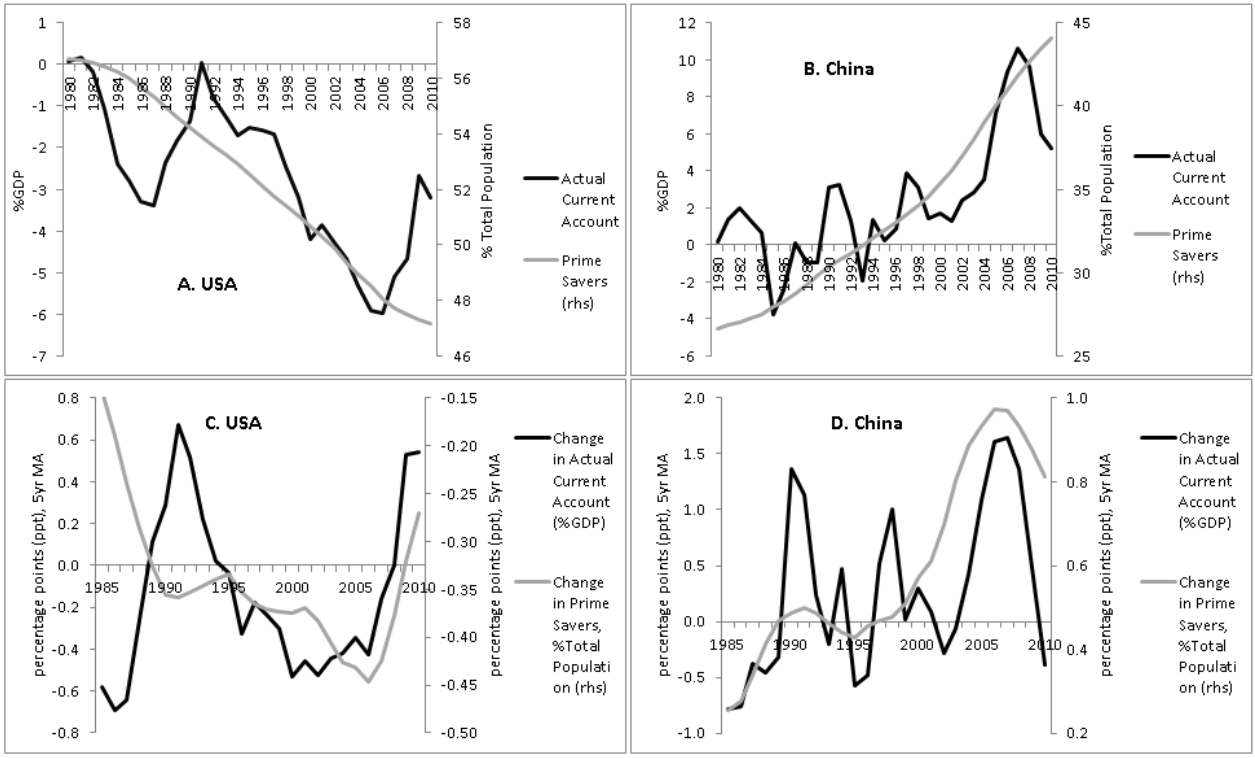

Figure 4-12: The Current Accounts and the Prime Savers, in Levels and Differences. The prime savers include the age groups derived from model estimates. Source: IMF WEO (April 2011), UN.

account trends. Finally, we try to run a scenario analysis to deduce how future demographic trends might influence the current account trends.

We first frame our discussion using the results from the shorter series of China and the US. Once we establish the key similarities and differences between the US and China, we have a separate section (section 4.11.4) discussing the key similarities and differences between the results using the longer and shorter times series for the US.

4.11.1 The Current Account and Demographics Link

Figure 4-12 plots four charts showing the evolution of actual current accounts and demographics, both in level and adjustment terms. For the prime savers, in each country, we use the age groups whose incremental increase poses a positive adjustment effect on the current account. As shown previously, our model estimates show that the group is 5–39 for the US and 35–69 for China. In the US, the share of population that contributes positively towards the current account peaked in 1979 and expected to continue to fall till 2050, albeit at a slower pace since late 2000s. The share of that portion of the population dropped from 57% in 1979 to 47% in 2010. While there are many factors at play here, this decline in the share of population that usually contributes positively towards current account surplus in the US is also partly responsible for the deficit, as shown in Figure 4-12A. In China, the opposite occurred. The share of population in the prime saving age group troughed in 1975 and has been increasing since then. In particular, the pace of the increase in the prime saving age group has increased remarkably since the early 1990s and continues to grow at an increased pace. As Figure 4-12B shows, this rapid rise in the age group that contributes positively towards the current account surplus is partly responsible for the rise in that surplus since 1994.

For our purpose, the more important link is that between the current accounts and demographics in adjustment terms since our model shows that the adjustment effect of demographics on current accounts is significant. Figures 4-12A and 4-12B plot these links for the US and China, respectively. In the US, we find that the

decrease in the prime saving age population, particularly since mid-1990s, have also caused the current accounts to decrease significantly. The adjustment effect of demographics on current accounts seems even stronger for China, where the rise in the prime savers since mid-1980s has caused the change in current accounts to be positive. Like the US, this trend is even more vivid since mid-1990s. These charts show that, while there were many forces at play, the change in demographics was also one of the reasons that shaped how the current accounts changed in the US and China, particularly since mid-1990s. In China, the change in demographics was supportive for positive changes in the current accounts. However, in the US, the change in demographics was one of the factors that results in the negative changes in current accounts, which prompted the current accounts to fall in further deficits.

4.11.2 The Impact of Demographics Implied by our Model

Our model suggests that the short term impact of demographics is substantial, particularly in China. Figure 4-13 compares the overall impact of demographics on the current account imbalances for the US and China over the past three decades. The overall impact of demographics has been calculated as the sumproduct of the coefficients of the three demographic variables (from Tables 4.5 (US short dataset) and 4.6 (China dataset)) and the corresponding values of the three demographic variables for the periods specified in the figure. As we move in different periods in the chart, the coefficients for each country remain same but the demographic

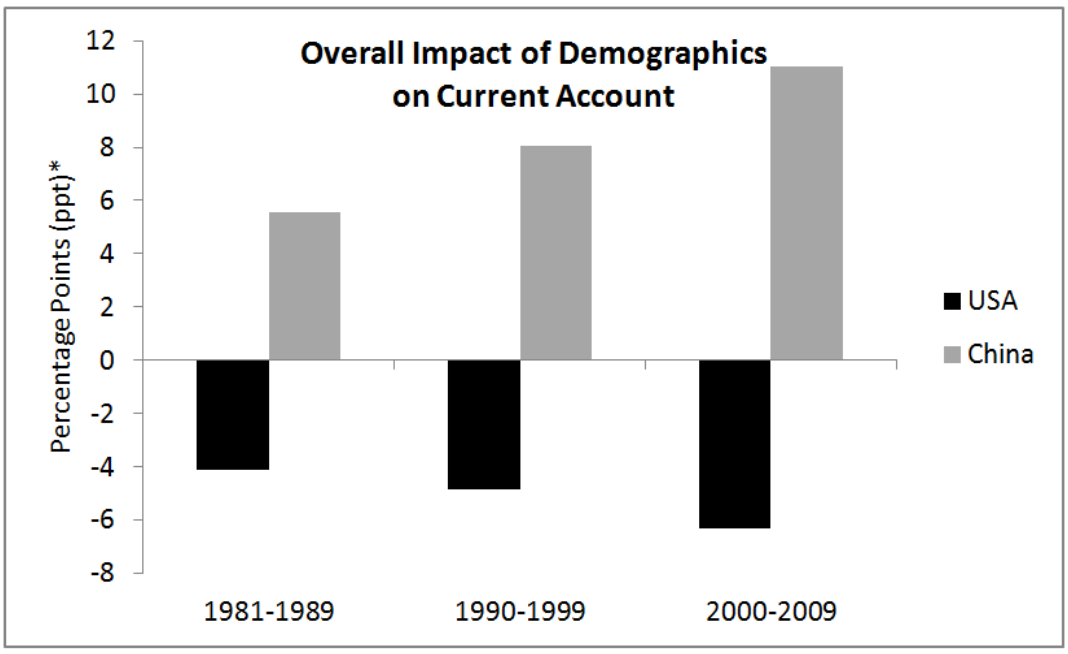

Figure 4-13: The Impact of the Change in Demographics Stronger in China. *The change in current account (expressed as a share of GDP) due to the change in demographics, on average per year.

variables change to reflect the corresponding time period. The impact of demographics has contributed negatively towards current account in the US but positively in China. In the US, for the period 1981–1989, the change in demographics decreased the current account (expressed as a share of GDP) on average every year by 4.1 percentage points. The equivalent number for the period 1990–1999 is 4.8 percentage points and for the period 2000–2009, it is 6.3 percentage points. In China, for the period 1981–1989, the change in demographics increased the current account (expressed as a share of GDP) on average every year by 5.5 percentage points. The equivalent number for the period 1990–1999 is 8.1 percentage points and for the period 2000–2009, it is 11.0 percentage points.

The impact of demographics in China seems to be stronger that the US. Our results indicate that the demographic shifts in China have been favourable in

terms of significantly increasing (from a low base) the share of the population that contributes positively towards the current account adjustments. The impact of demographics on the current account seems striking for China but should be seen in the context of a country whose personal saving rate is one of the highest in the world. This is mirrored by a correspondingly low proportion of spending on domestic demands (Dew and Martin (2011)). Furthermore, China's household consumption of GDP is lower than that of other Asian countries during similar stages of development. It fell from 51% of GDP in 1985 to 35% of GDP in 2009.

Comparing our results to the existing literature, Higgins' (1998) general conclusion is that the impact of demographics on the current account for the US is less powerful, in line with our findings. For the period 1985–89, Higgins (1998) estimates that the demographic shifts have changed the current accounts, as a share of GDP, by 0.4 percentage points. Higgins does not have estimates for China. However, his estimates for developing countries are also substantially stronger than the US. For example, demographic shifts are responsible for a 5.4 percentage point increase in the current account, expressed as a share of GDP, in Korea for the periods 1960–64 and 1985–89. For Kenya, demographic changes deteriorated the current account by 7.6 percentage points over the period 1985–1989. For Bangladesh, demographics decreased the current account by 6.4 percentage points over the period 1980–84. Higgins also generally contends that the estimated effect of demographics on current accounts exceeded more than 6% of GDP over the period 1968–1998, and predicts that the impact maybe even higher going forward, given impending demographics pressure.

While this is not the focus of the thesis, one could possible assess the magnitude of the demographic effects compared to the other effects by conducting the same sort of analysis done in this section. In other words, we could compute the sumproduct of the coefficients (of our best results represented in Tables 4.5 and 4.6) and the corresponding values of the controlled variables for different time periods. Then we could compare that value to the results presented in Figure 4-13 to derive a sense of the magnitude of the demographic effects compared to the other effects.

4.11.3 Demographics Likely to Continue to Impose Same Adjustment Pressures

How are demographics likely to influence current accounts going forward? To assess this, we run a scenario analysis using UN population projections. We use the model to look at the underlying shifts in current accounts up to 2050. These are not forecasts but rather a framework for describing a potential scenario using the existing information we have about demographics. There are of course a host of developments that cannot be captured by this kind of exercise.

The first stylized fact relevant to this exercise is that, in the US, the pace of the decline in the prime saving age population has peaked in early 2000. In China, the pace of the increase in the prime savers has also peaked in early 2000. Going forward, in the US, the prime savers are expected to continue to decline but at a slower pace. In China, almost the opposite is expected to occur. The

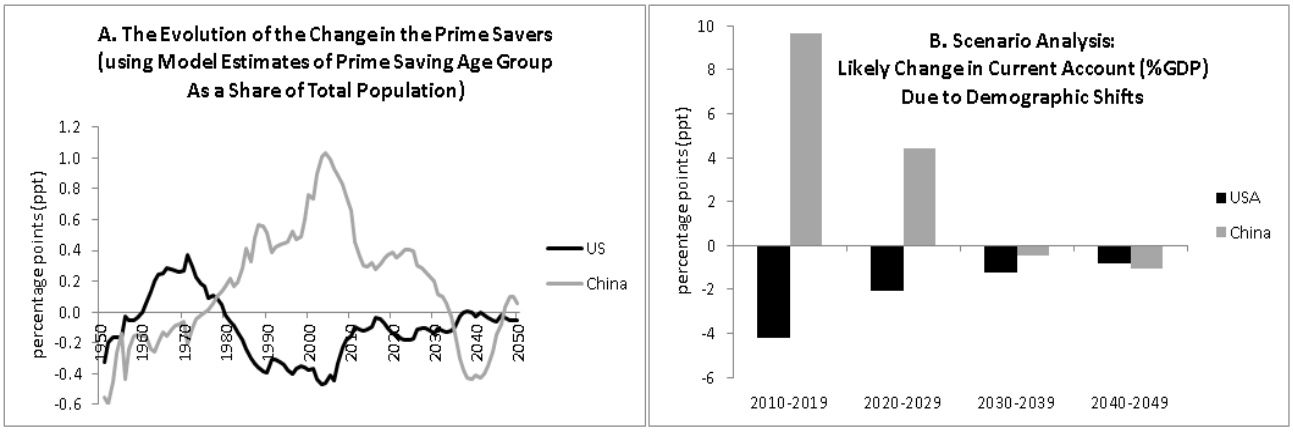

Figure 4-14: The Likely Impact of the Change in Demographics on Current Account Adjustments, on average per year

increase in prime savers is expected to remain positive until mid-2030s. However, the pace of the increase in prime savers is expected to fall.

This demographic shifts imply that the change in the prime saving age population will continue to exert downward adjustment pressure on the US current account. However, this pressure is likely to ease with time, as shown in Figure 4-14B. Figure 4-14B has been computed applying the same logic as 4-13, using the sumproduct of the coefficients of the three demographic variables (from Tables 4.5 (US short dataset) and 4.6 (China dataset)) and the corresponding values of the three demographic variables for the periods specified in the figure. The values of the three demographic variables have been computed using the UN population projections for the periods specified in the chart. On the other hand, Figure 4-14A suggests the change in prime saving population will continue to be positive in China, implying that the demographic pressure will point towards further positive current account adjustments. However, the pace of the increase in the current account due to demographics is likely to slow in China in the next five years due

to the decline in the pace of increase in the number of prime savers. Hence, our model suggests that the demographic pressure on the current account surplus for China will still be very strong (on average the change in current account due to demographics is likely to be 7 percentage points for the next twenty years) but less than the last decade. From early 2030s, however, the demographic pressure is likely to cause a drag in the current account in China as the population start ageing at a rapid pace.

It should be underlined that the numbers in the scenario analysis should be taken with a pinch of salt, not least because the analysis assumes that the structural factors in the economy remains the same for an extended period of time. Our analysis assumes that Chinese households' savings behaviour is likely to remain the same, which seems a strong assumption. Moreover, the controlled variables might provide offsetting forces that would dampen the upward pressure from demographics. Nevertheless, while we cannot commit to the numbers involved, the underlying message of this scenario analysis seems real: in the next twenty years, the demographic pressure in China will continue to point towards a significant current account surplus.

The second caveat to be made in relation to this exercise is that our model suggests that the traditional prime savers, of the age group 35–69, in the US are not saving enough. In fact, the incremental addition to the age groups 40–59 are contributing negatively towards the current account adjustments. Traditionally, this group should have contributed positively. The structural reforms after the financial crisis may likely change that. The latest data already shows evidence

that the savings rate of US households has increased. If this trend continues, the US current account is likely to get an upward boost due to demographics as more members of the traditional age group start saving. Also, the age coefficient for this age group is likely to be higher than our estimates suggest for the existing prime savers in the US. The UN Population projections suggest that the 35–69 age group peaked in 2011 and is likely to fall, albeit gradually, going forward. However, for the next few years, the positive level effect of more people of the correct age groups saving is likely to be stronger than the negative effect from the slow decline in the population in the age group 35–69.

Our findings in the scenario analysis is in line with Wilson and Ahmed (2010) and Higgins (1998). Both these reports predict that the greying population in the advanced economies is likely to cause a drag in current accounts. Higgins (1998) predicts that this would happen since the investment in the developed countries are likely to fall more sharply than the savings. Wilson and Ahmed (2010) also find that demographic pressure is likely to lead to further current account surplus for developing countries, including China.

4.11.4 The Long versus Short Time Series in the US

Figure 4-15 shows the impact on the change in current accounts due to the change in demographics for the results using the US long dataset. Figure 4-15 has been computed applying the same logic as 4-13, using the sumproduct of the coefficients of the three demographic variables (from Table 4.13 (US long dataset)) and the

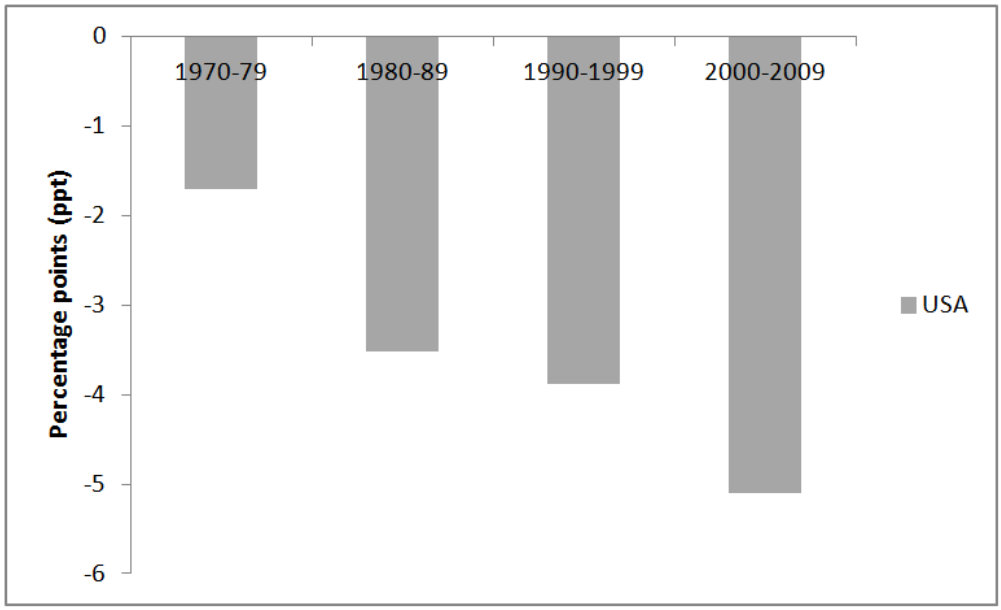

Figure 4-15: The Change in Current Account due to the Change in Demographics, on average per year, using the Long Data Series

corresponding values of the three demographic variables for the periods specified in the figure. Similar to the results with the shorter sample, we find that the impact of demographics has contributed towards a negative adjustment in current account in the past four decades. On average per year, the change in demographics have resulted in 1.7 percentage points deterioration in current accounts for the period 1970–1979, 3.5 ppt deterioration for the period 1980–1989, 3.8 ppt deterioration in current accounts for the period 1990–1999, and 5.1 ppt deterioration in current account for the period 2000-2009.

Figure 4-16 compares the model implications for current accounts, for both history and projections. This figure has also been computed applying the same logic as 4-13, using the sumproduct of the coefficients of the three demographic variables (from Tables 4.13 (US long dataset) and 4.5 (US short dataset)) and the corresponding values of the three demographic variables for the periods specified

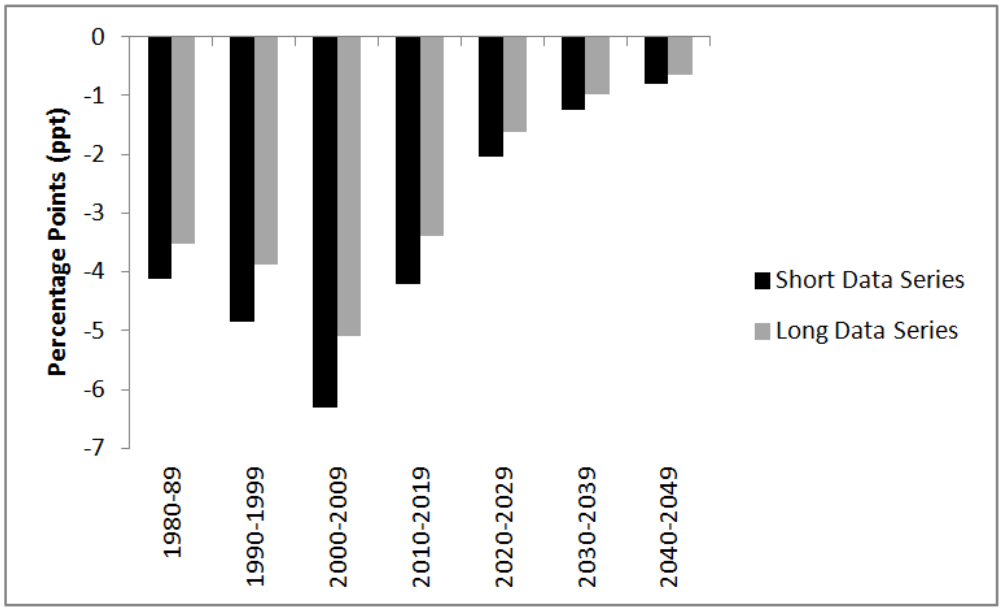

Figure 4-16: The Change in Current Accounts (%GDP) due to the Change in Demographics, on average per year. Demographic projections from the UN taken (after 2010) for this exercise.

in this figure. The estimates are somewhat different for the two regressions (with short and long data series), and with these sort of exercises, it is extremely difficult to pinpoint the accurate numbers. However, in terms of direction, both the models provide same results. Starting from early 1980s, the change in demographics in the US have continued to gradually contribute towards negative adjustments in the US current account with the decade 2000–2009 being the worst. Once again, this sort of exercise reminds the reader that we cannot estimate the exact numbers but can get a sense of direction from linking the demographic projections with the model underpinnings.

4.12 Policy Implications and Concluding Remarks

Our aim in this paper was to explain the US and China current account debate using a model of demographics in order to present one of the structural explanations of the current account imbalances. Following Higgins (1998), we used a model that allows for a rich structure of age effects similar to those predicted by the life cycle theories. However, instead of employing panel framework used by Higgins (1998) and other studies, we estimated the impact of demographics on current accounts separately for China and the US. In terms of regression methodology, we used the autoregressive distributed lag model (ARDL) that helps to directly model the dynamics, instead of taking the autoregressive errors as a nuisance. Our econometric framework, richer than the ones used by earlier studies, allows us to answer more questions about the exact way how demographics impact current accounts. In particular, we can check if the impact of demographics on current accounts is long run or short run. In addition, we consider the ARDL model in the context of an economic system, the idea that all the components of the economy should be modelled to get the more accurate and comprehensive picture. We then conducted the necessary exogeneity tests that allows us to reduce from the multi-equation system to one linear equation with the current account as the independent variable.

Using this new approach, we conduct econometric analysis on three datasets: 1) current accounts of China dating from early 1980s, 2) current accounts of the US dating from early 1980s, and 3) a longer series of the US dating from early 1960s.

Our results confirm that the imbalances in the US and China can be explained partly in light of the shifting demographics of these economies. However, our results do not conform with earlier studies based, most notably Higgins (1998), which imply that demographics have a long run level effect on the current account imbalances. Our richer econometric framework shows that demographics have a significant short run adjustment effect. In other words, demographics are a key determinant of how the current account positions adjust back to their long run equilibrium, but they are not determinants of the long run equilibrium.

We also find that the responses of age groups towards current accounts for the US and China are sufficiently different to justify estimating them separately, rather than in a panel framework. Our estimates show that the change in population in the traditional age group that should be contributing positively towards current account adjustments is not saving enough in the US. The results hold even when a longer time series is used, starting from 1960. For China, we find that, though our reason for why demographics matter is different from the existing literature, the responses of the different age groups are in line with the findings of earlier studies.

In the US, we conclude that the fact that the traditional age group of the US is not saving enough is a persistent feature of the economy. This was exacerbated slightly since late 1980s due to the financial liberalization that led to a stock market boom. The increased wealth from stock market prompted households to import more foreign goods leading to a drag on the current account. Our results from the US calls for structural changes within the economy that encourage saving

by people in the 35–69 age group. The recent improvement in the personal saving rate after the financial crisis should be viewed positively against that backdrop. As the US struggles to reduce the current account deficit, the demographic pressure should mitigate the problem to some extent due to two factors. First, the share of the population in the prime saving age group, using our model estimates, is likely to continue to decrease but at a slower pace in the next decade. Hence, demographics would continue to have a negative adjustment effect on current accounts, but the negative adjustment effect is likely to be smaller than the past. Second, following the financial crisis the US current account is likely to get an upward boost as more members of the traditional 'prime age group', 35–69, start saving. Although that group peaked in 2011, the positive level effect of more people of the correct age groups saving is likely to be stronger than the negative effect of the decline in population in that age group.

In China, we find that, up to the age of 35, the change in population appears to have a negative adjustment effect on the current accounts. Between ages 35 and 69, the increase in population seems to have a significant positive adjustment effect on current accounts, as the rate of savings surpasses the rate of investment. Above 69, the change in population again tends to create a negative adjustment effect on the current accounts. Furthermore, we find that the demographic pressure is striking in China. One of the reasons for this striking impact of demographics is that the personal saving rate in China is one of the highest in the world. This is mirrored by a correspondingly low proportion of spending on domestic demands (Dew and Martin (2011)). China's household consumption of GDP is lower than

that of other Asian countries during similar stages of development. It fell from 51% of GDP in 1985 to 35% of GDP in 2009.

Our scenario analysis suggests that the change in demographics will continue to create a negative adjustment effect on the US current accounts and a positive adjustment effect on China's current accounts. As shown in Wilson and Ahmed (2010), this is representative of a general trend that significant net flows of capital across borders – from emerging markets to advanced economies – are likely to be a feature of the next decade. Hence, a global perspective is needed to understand what have often been examined as local issues. In particular, looking at the story of savings, investment and capital flows from the perspective of the large developed markets may end up being misleading in a world where the global picture is increasingly shifting. In the final reckoning, a great deal of policy coordination will be required to address the imbalances arising from strong saving in the emerging markets and dissaving in the advanced economies. Developing the infrastructure, policy tools and regulatory settings to be able to cope with the international capital flows remains an urgent task. From the perspective of the developed countries like the US, policy makers need to urgently ensure that the financial markets of the developed economies are prepared to receive the influx of capital from emerging markets like China. From the perspective of developing countries like China, policies need to focus on two aspects. First, the successful transmission of the excess savings in the developed world in the form of choosing the proper financial instruments. Here, policy coordination between the emerging markets and developed economies is strongly recommended to assure

that both sides benefit from this capital flow. Second, China and other emerging markets need to also deepening their local capital markets to reduce some of the excessive pressure to export savings. Further adjustments in domestic demand and currency are also required in China. China should also encourage increase in domestic demand to reduce the pressure of excessive saving and increase the living standard of its citizens.

From the perspective of policies involving demographics, our study has few interesting implications. There has been a lot of debate in the advanced economies about the prospect of increasing the retirement age to reduce dependency rates and alleviate fiscal pressure. Our results for the US imply that increasing the retirement age may not help if the economies continue to have low savings due to low personal savings rate of the consumers. A better policy would be to encourage the consumers of the appropriate age groups to save more. If the consumers do not save enough, the increase in retirement age may not translate into more savings. In China, the results from our study suggest that one of the policy implications could be to increase birth rate moderately so that there is more investment demand that could use the excessive savings. In that regard, the policy makers may want to re-visit China's one-child policy and other demographic laws to assure that they are still relevant and modify these policies to some extent. The policy makers should also encourage the consumers to reduce saving and increase domestic demand so that there is less pressure of excessive current account imbalances within the economy. The social safety networks should be improved to better protect the consumers so that they are encouraged to reduce precautionary

savings. The reduction in volatility in economic growth and inflation may also encourage the consumers to reduce precautionary savings. Overall, this study suggests that global policy coordination is required in both financial markets and social dynamics (like future demographic planning, pension system, social reforms, retirement age) to mitigate the problem of excessive capital flows from the emerging markets like China to the advanced economies like the US.

4.13 Bibliography

Almon, S. (1965): "The Distributed Lag Between Capital Appropriations and Expenditures," *Econometrica,* 33(1), 178-196.

Blanchard, O., F. Giavazzi And F. Sa (2005): "The US Current Account and the Dollar," *Department of Economics, Massachusetts Institute of Technology,* Working Paper Series 05-02.

Blanchard, O. and G. Milesi-Ferretti (2009): "Global Imbalances: In Midstream?" *IMF Staff Position Note,* SPN/09/29.

Bosworth, B., R. Bryant and G. Burtless (2004): "The Impact of Aging on Financial Markets and the Economy: A Survey," mimeo, Brookings Institution.

Caballero, R., J. Farhi And P. Gourinchas (2008): "An Equilibrium Model of Global Imbalances and Low Interest Rates," *American Economics Review,* 98(1), 358-393.

Cooper, R. (2007): "Living with Global Imbalances," *Brookings Papers on Economic Activity,* 2, 91-107.

Debelle, G. And H. Faruqee (1996): "What Determines the Current Account? A Cross-sectional and Panel Approach," *IMF Working Paper*, WP/96/58.

Dew, E. And J. Martin (2011): "China's Changing Growth Pattern," *Bank of England*, Quarterly Bulletin, Q1 2011.

Dickey, D. And W. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

Fair, R. And K. Dominguez (1991): "Effects of the Changing U.S. Age Distribution on Macroeconomic Equations," *American Economic Review*, 81(5), 1276-1294.

Fisher, I. (1930): *The Theory of Interest*, New York: Macmillan.

Gourinchas, P. And H. Rey (2005): "From World Banker to World Venture Capitalist: US External Adjustment and The Exorbitant Privilege," NBER Working Paper No. W11563.

Goyal, R., C. Marsh, R. Narayanan, S. Wang And S. Ahmed (2011): "Financial Deepening and International Monetary Stability," *IMF Staff Discussion Note*, SDN/ 11/16.

Greene, W. (2003): *Econometric Analysis*, Fifth Edition, New Jersey: Upper Saddle River.

Harbo, I., S. Johansen, B. Nielsen And A. Rahbek (1998): "Aysmptotic Inference on Cointegrating Rank in Partial Systems," *Journal of Business & Economic Statistics*, 16(4), 388-399.

Harrod, R. (1948): *Towards a Dynamic Economics*, London: Macmillan.

Helliwell, J. (2004): "Demographic Changes and International Factor Mobility," Federal Reserve Bank of Kansas City Symposium paper.

Hendry, D.F. (1995): *Dynamic Econometrics*, Oxford: Oxford University Press.

Hendry, D.F. and B. Nielsen (2007): *Econometric Modeling: A Likelihood Approach*, Princeton: Princeton University Press.

Hendry, D.F. and J.A. Doornik (1994): *PcFiml 8: An Interactive Program for Modelling Econometric Systems*. London: International Thomson Publishing.

Higgins, M. (1998): "Demography, National Savings and International Capital Flows," *International Economic Review*, 39(2), 343-369.

Higgins, M. And J. Williamson (1996): "Asian Demography and Foreign Capital Dependence," *NBER Working Paper No. 5580*.

Holman, J. A. (2001): "Is the Large US Current Account Sustainable?" Federal Reserve Bank of Kansas City.

Johansen, S. (1992): "Testing Weak Exogeneity and the Order of Cointegration in UK Money Demand Data," *Journal of Policy Modelling,* 14(3), 313-334.

Lane, P. And G. Milesi-Ferretti (2001): "Long-Term Capital Movements," *NBER Working Paper No. 8366.*

Mann, C.L. (1999): "On the Causes of the US Current Account Deficit," Peterson Institute for International Economics, *Briefing for the Trade Deficit Review Commission (19 August 1999).*

Modigliani, F. And R. Brumberg (1954): *Utility Analysis and the Consumption Function: An Interpretation of Cross-section Data.* In: Kurihara, K.K. (ed.): Post-Keynesian Economics. New Brunswick, NJ: Rutgers University Press.

Obstfeld, M. (2005): "America's Deficit, the World's Problem," *Keynote Speech Prepared for the Twelfth International Conference of the Institute for Monetary and Economic Studies*, Bank of Japan, Tokyo, Japan, May 30-31, 2005.

Obstfeld, M. And K. Rogoff (2009): "Global Imbalances and the Financial Crisis: Products of Common Causes," *Paper Prepared for the Federal*

Reserve Bank of San Francisco Asia Economic Policy Conference, Santa Barbara, CA, October 18-20, 2009.

STOKER, T. (1986): "Simple Tests of Distributional Effects on Macroeconomic Equations," *Journal of Political Economy*, 94 (August), 763-795.

WILSON, D. AND S. AHMED (2010): "Current Account and Demographics: The Road Ahead," *Goldman Sachs Research*, Global Economics Paper 202.

YONGDING, Y. (2007): "Global Imbalances and China," *The Australian Economic Review*, 40(1), 3-23.

Concluding Remarks

The aim of this thesis was to use two new approaches to explain some of the important debates in two key economic fields: labour markets economics and macroeconomic studies related to current account imbalances. In the first three chapters, we introduce a new distribution, called the normal inverse Gaussian (NIG) distribution, with the aim to better model the log earnings data by describing unobserved heterogeneity. We use this distributional set-up to discuss two of the widely discussed problems in labour economics: modelling the marginal distribution of earnings and estimating the return to education, that is, modelling the conditional (on education) distribution of earnings. We show in the first chapter that the previous studies try to make inferences from an unobserved heterogeneity component to the observed income distribution. Contrary to the previous models, we start with the observed distribution of income and use a distribution that helps us to account for the differences in unobserved variance.

We empirically test the new distribution using two datasets: the US males and Ghanaian males. Our empirical tests reinforce some of the useful properties of modelling income using the NIG distribution. For the marginal distribution case,

we find the fit of the log earnings function is better in both datasets. In contrast to Roy's prediction of right skewness and in accordance with other empirical findings, we find that both datasets exhibit a log earnings distribution that is skewed to the left.

For the conditional (on education) case, we find that the NIG distribution is useful in modelling the return to education in Ghana since education and unobserved heterogeneity cannot be separable. Under such circumstances, the NIG distribution is found to contain useful tools to describe the labour market. However, in the US, education and unobserved heterogeneity can be separated. Here, both NIG distribution and normal distribution yield the same results. Of particular interest is the finding that, once heterogeneity is accounted for, the return to education is almost flat for lower levels of education and then increases remarkably after ten years of education, indicating that the earnings function is highly convex. After ten years of education, the return to education increases modestly with subsequent years of education. Many studies have also shown that the return to education in Ghana and other African countries increases with higher levels of education. Our findings strongly suggest that the NIG distribution would be a useful tool to model the labour market of developing countries like Ghana where normal distribution may not be able to pick up the intricate interactions between education and the unobservables. The NIG distribution can contribute towards policy discussions about how resources should be devoted to get positive returns to schooling for poor countries.

In our fourth chapter, we use a model of demographics that allows for a rich

structure of age effects, similar to those predicted by the life cycle theories, to explain the sustained rise in the US current account deficit and China's current account surplus. Our results confirm that the imbalances in the US and China can be explained partly in light of the shifting demographics of these economies. However, contrary to the previous studies, we find that demographics have an important short run adjustment effect on current accounts, not a long run level effect. Though our results indicate that the role of demographics is different than what the interpretations of previous studies demonstrate, our results in China are in line with the findings of previous studies. However, in the US, our results differ and indicate that the traditional age group that should be saving is not saving enough in the US. Delving deeper by using a longer dataset, we conclude that this is a persistent feature of the US economy that became slightly worse since the financial liberalization of the late 1980s.

Our work in the fourth chapter raises a few issues relevant for research on current accounts for other countries as well. Our results show that the panel framework should probably not be used to determine the relationship between demographics and current accounts since there are country specific factors that may not be picked up in a panel framework. Also, most studies use the working age population or dependency ratios to show the link between demographics and current accounts. Again, this may not be the correct approach since the working age population is more relevant for discussions related to fiscal policies. For current accounts, the more pertinent age group is the prime saving age group and this age group may differ from country to country. Finally, our results are

representative of a general trend mentioned by other studies: significant net flows of capital across borders – from emerging markets to advanced economies – are likely to be a feature of the next decade. A great deal of policy coordination is required to address the imbalances arising from strong saving in the emerging markets and dissaving in the advanced economies.