Essays in Empirical Population Economics

Inaugural-Dissertation

zur Erlangung des Grades Doctor oeconomiae publicae (Dr. oec. publ.)

an der Ludwig-Maximilians-Universität München

2005

vorgelegt von

Parvati Trübswetter

Referent:	Prof. Stephan Klasen, Ph.D.
Korreferent:	Prof. John Komlos, Ph.D.
Promotionsabschlussberatung:	13. Juli 2005

Essays in Empirical Population Economics

by

Parvati Trübswetter

Submitted to the Department of Economics

in partial fulfillment of the requirements for the degree of

Doctor oeconomiae publicae (Dr. oec. publ.)

at the

Ludwig Maximilian University, Munich

2005

Thesis Supervisor:	Prof. Stephan Klasen, Ph.D.
Thesis Co-Supervisor:	Prof. John Komlos, Ph.D.
Final Committee Consultation:	July 13, 2005

Danksagung

Mein großer Dank gilt allen, die mir bis hierher beigestanden haben, exemplarisch: Stephan Klasen für die Betreuung der Dissertation und für alle halben Stellen, die er für mich auf meinem langen Weg zur Promotion gefunden hat, John Komlos für das Zweitgutachten, Stephan fürs Dasein, Carola, ohne die ich niemals diese Arbeit abgegeben hätte, Susanne für die gute Unterhaltung, Mathias und Andreas, die mich in der letzten Zeit in Ruhe gelassen haben, Iris und Thomas, dass sie immer noch an mich glauben, Cormac O Grada für die gute fachliche und menschliche Betreuung am UCD, Patrick Puhani für seine Großzügigkeit in der Endphase meiner Dissertation, Herbert für die Kooperation in unserem Migrationsprojekt, aus dem der zweite Artikel dieser Arbeit hervorging und allen anderen, die gerne genannt worden wären oder die ich nennen hätte sollen.

Contents

Main introduction

> The encounters between econometricians and data are frustrating and ultimately unsatisfactory both because econometricians want too much from the data and hence tend to be disappointed by the answers, and because the data are incomplete and imperfect.
>
> Zvi Griliches, Handbook of Econometrics

Population economics shows to be an important field in economics as much work has been done on demographic issues over the last years. Several reasons for the importance of the field can be stated. Typical subjects of population economics such as fertility, health, distribution, labour participation etc. influence outcomes of other economics disciplines such as labour economics, development economics, growth theory etc. The family, which is one of the subjects of population economics, is the first place where population outcomes are produced. With the availability of individual and household data sets, economists could investigate these indicators thoroughly and gain valuable insights into the functioning and decision making of the basic economic units - the families. The findings can be used in many related disciplines of economics as for example in development (see Deaton (1997) for an overview) or labour economics (for example Gray (1998)).

The analysis of demographic phenomena has a long tradition. Malthus (1798) started the ongoing discussion on demographics with his theory on population growth. He postulates that innate population growth is very fast and can only be stopped by shortages of subsistence goods; a theory, which was since then changed, disputed and refuted. The fertility decisions of households were exogenously given and not subject to analysis. Becker (1964, 1981) introduced a new way of looking at family behaviour with his idea that fertility of the household is endogenous and influenced by economic and non-economic factors. Decisions of families could be analysed and were based on economic reasoning.

Studies on population economics focus either on a macroeconomic topic like population growth - for example related to technical change and natural resources as in the analysis of Robinson and Srinivasan (1997) or on microeconomic topics like mortality, fertility, poverty or migration. I will focus on some of the mentioned microeconomic topics and will analyse them more thoroughly in the following papers that can be read independently of each other.

The dissertation consists of three papers that deal with excess female mortality, with migration decisions, and with nutrition and productivity. I will now present a short overview of the three fields and my contributions.

Mortality and health are the most vital subjects of population economics as life per se and health are typically seen as a positive asset without starting any economic reasoning (Sen, 1998). Mortality has been investigated with respect to diverse contexts. In the context of European or industrialised countries, mortality in the early development and industrialisation attended much attention. Fogel (1997) summarises the findings on the huge mortality decline in the Western World since 1750.

Studies on mortality in developing countries of today focus mainly on illness-related mortality and on mortality based on inequalities. Mortality based on gender inequality in access to survival related goods is called excess female mortality and is investigated thoroughly by Sen (1989) and Klasen (1994, 1998, 1999) and Klasen and Wink (2002). Most of the papers investigate the concept of missing women: missing women are the women that would be alive today additionally if no inequality in the access to resources of the household existed. The excess female mortality can not only be detected in today's developing countries like India, Pakistan or China (Das Gupta and Bhat (1997); Langford and Storey (1993); Coale and Banister (1994)), but also in the early development of Europe, for example in Germany, England or Ireland (McNay, Humphries, and Klasen (1998, 2002)). Ireland has to be seen as a special case, because excess female mortality increased from the 1850s onwards while it began to decline in other European countries. This exceptionally high excess female mortality persisted in some age groups even until the mid twentieth century.

My first paper contains an analysis of Excess female mortality in Ireland from 1871 to 1926, based on census data and vital statistics. The beginning

of my observation period lies after the great Irish famine of 1845 to 1850 with very high excess mortality, which had affected men more than women. Panel regressions on excess female mortality show that the pattern of explanatory factors is similar to the pattern found for example in England at the same time, but with a higher stock of excess female mortality. I find that women's situation improves with a higher labour market participation. But, as working conditions were detrimental to health in early industrialisation, a higher female participation in sectors such as the industrial sector ruins the gains in power over family resources. Furthermore, the high emigration rates of women and the high share of the agricultural sector hurt women.

Comparing Ireland to England and Wales, a much higher share of agricultural employment can be observed. As agricultural employment increases excess female mortality in both regions, higher Irish excess female mortality can partly be explained. Further differences in explanatory variables that help to explain the higher Irish excess female mortality are a high emigration of women out of Ireland and a different effect of poor relief on women.

Migration is a further serious indicator that can alter the economic situation of the population of a country positively or negatively. Analyses can be divided into the search for the determinants and the effects of migration (Borjas, 1994). I will focus on ideas underlying the first one, because my contribution will be on the determinants of internal migration in Germany. Greenwood (1997) emphasises the importance of local unemployment on migration decisions which is mainly relevant in the U.S., but also in East-West German migration, as will be shown in the paper. When looking at the effects of immigration, economists often face the problem that the causation between the variables is not clear; so, they are looking eagerly for exogenous and unexpected changes in the observed market because responses to exogenous shocks can be identified. These occurrences are called "natural experiments" and papers have been done for example on the effects of unexpected huge flows of migration or labour into a local labour market (Gould, Lavy, and Paserman (2004) and Card (1990)).

An empirical analysis of who decides to migrate and if migrants are selected randomly from the home population is seriously hampered by a lack of individual data in the context of international migration. Household panels in the home countries usually do not allow international emigrants to be

identified - they simply disappear from the panel. Thus, although we have rich information on immigrants in many host countries, the phenomenon of self-selection cannot be addressed systematically on the basis of standard data sources. German unification offers a unique opportunity to study the problem of the selection of migrants. There exist several micro data sets that allow East-West migrants to be identified before and after movement.

The second paper is concerned about the migration decision of East German workers to move to West Germany. Cumulative net migration from East to West Germany amounts to 1.3 million persons or 7.5 per cent of the original population in East Germany over the period from 1989 until the end of 2001. Since the inequality of earnings in East Germany has approached West German levels in the late 1990s, the standard Roy model predicts that a positive selection bias of East-West migrants should disappear. Using a switching regression model and data from the IAB-employment sample, which is a one per cent sample of individuals who are registered by the German social security system, we find that employed East-West migrants remain positively self-selected with respect to unobserved abilities. This result is consistent with the predictions of an extended Roy model, which considers moving costs that are negatively correlated with labour market abilities of individuals. Moreover, we find that wage differentials as well as differences in employment opportunities are the central forces which drive East-West migration after unification.

Poverty and hunger are outstanding problems of population economics related to developing countries. Empirical studies on the determinants of these factors are necessary to find remedies against them. Economists have shown theoretically that better nutrition increases productivity and that there is a threshold of the nutritional status below which people are not able to produce the equivalent to their subsistence wages. The theory of nutritional efficiency wages has been modeled by several authors, for example by Leibenstein (1957) and Bliss and Stern (1978). The corresponding dynamic concept of poverty traps was developed by Dasgupta (1997). Theory implies that help for the poorest of the population has to improve their nutritional status directly (for example through (re)distribution of land or food) to enable them to work.

Empirical investigations exploring the issue at hand explain the influence

of nutrition and health on income. Behrmann and Deolalikar (1989) and Strauss and Thomas (1995) have written comprehensive surveys on the existing studies on nutritional efficiency wages and they find that studies on the influence of BMI on wages indicate that there is a positive relationship while studies which investigate the influence of wages on BMI indicate that there is more or less no influence.

The last paper of my dissertation explains the reasons why individuals stay in poverty. I am looking into the empirical relationship between nutrition (Body Mass Index) and wages using the Côte d'Ivoire Living Standards Survey (CILSS) from 1985 to 88. I find that non-labour income of the individual positively influences not only the probability to be in employment, but also the amount of wages the individual will get. Non-labour income further influences wages indirectly by increasing the nutritional status of the workers. With a higher BMI, workers are able to work more productively and generate a higher wage. I conclude that a policy to help individuals out of a poverty trap should increase the non-labour income of individuals.

References

Becker, G. S. (1964): *Human Capital*. Columbia University Press, New York.

——— (1981): *A treatise on the family*. Harvard University Press, Cambridge, M.A.

Behrmann, J. R., and A. B. Deolalikar (1989): "Health and Nutrition," in *Handbook of Development Economics, Vol. 1*, ed. by H. Chenery, and T. N. Srinivasan, chap. 14. North-Holland, Amsterdam.

Bliss, C., and N. Stern (1978): "Productivity, wages and nutrition Part I: The theory," *Journal of Development Economics*, 5, 331–362.

Borjas, G. J. (1994): "The Economics of Immigration," *Journal of Economic Literature*, 32(4).

Card, D. (1990): "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review*, 43(2).

Coale, A. J., and J. Banister (1994): "Five Decades of Missing Females in China," *Demography*, 31(3), 459–479.

Das Gupta, M., and P. N. M. Bhat (1997): "Fertility Decline and Increased Manifestation of Sex Bias in India," *Population Studies*, 51(3), 307–315.

Dasgupta, P. (1997): "Nutritional status, the capacity for work, and poverty traps," *Journal of Econometrics*, 77, 5–37.

Deaton, A. (1997): *The analysis of household surveys - A microeconometric approach to development policy*. John Hopkins University Press, Baltimore and London.

Fogel, R. W. (1997): "New Findings on Secular Trends in Nutrition and Mortality: Some Implications for Population Theory," in *Handbook of Population and Family Economics, Vol. 1A*, ed. by M. R. Rosenzweig, and O. Stark, chap. 9. North-Holland, Amsterdam.

Gould, E. D., V. Lavy, and M. D. Paserman (2004): "Immigrating to Opportunity: Estimating the Effect of School Quality Using a Natural Experiment on Ethiopians in Israel," *Quarterly Journal of Economics*, 119(2).

Gray, J. S. (1998): "Divorce-Law Chnges, Household Bargaining, and Married Women's Labor Supply," *American Economic Review*, 88(3).

Greenwood, M. J. (1997): "Internal Migration in Developed Countries," in *Handbook of Population and Family Economics, Vol. 1B*, ed. by M. R. Rosenzweig, and O. Stark, chap. 12. North-Holland, Amsterdam.

Klasen, S. (1994): "'Missing Women' Reconsidered," *World Development*, 22, 1061–71.

——— (1998): "Marriage, Bargaining, and Intrahousehold Resource Allocation: Excess Female Mortality among Adults during Early German Development, 1740-1860," *The Journal of Economic History*, 58(2), 432–467.

——— (1999): "Gender Bias in Mortality in a Comparative Perspective," paper prepared for the German Cliometrics Conference, September 20-24, Toronto, Canada.

Klasen, S., and C. Wink (2002): "A Turning Point in Gender Bias in Mortality? An Update of the Number of Missing Women," *Population and Development Review*, 28(2), 285–312.

Langford, C., and P. Storey (1993): "Sex Differentials in Mortality in the Twentieth Century: Sri Lanka and India Compared," *Population and Development Review*, 19(2), 263–282.

Leibenstein, H. (1957): *Economic Backwardness and Economic Growth: Studies in the Theory of Economic Development*. John Wiley, New York.

Malthus, T. R. (1798): *An Essay on the Principle of Population and a Summary View of the Principle of Population*. Penguin, Baltimore, Originally published in 1798 and 1930, reprint from 1970.

McNay, K., J. Humphries, and S. Klasen (1998): "Death and Gender in Victoran England and Wales: Comparisons with Contemporory Developing Countries," *DAE Working Paper No. 9801*.

——— (2002): "Excess Female Mortality in Nineteenth Century England and Wales: A Regional Analysis," mimeo.

Robinson, J. A., and T. N. Srinivasan (1997): "Long-Term Consequences of Population Growth: Technological Change, Natural Resources, and the Environment," in *Handbook of Population and Family Economics, Vol. 1B*, ed. by M. R. Rosenzweig, and O. Stark, chap. 21. North-Holland, Amsterdam.

Sen, A. (1998): "Mortality as an Indicator of Economic Success or Failure," *The Economic Journal*, 108(446), 1–25.

Sen, A. K. (1989): "Women's Survival as a Development Problem," *Bulletin of the American Academy of Arts and Sciences*, 43, 14–29.

Strauss, J., and D. Thomas (1995): "Human Resources: Empirical Modeling of Household and Family Decisions," in *Handbook of Development Economics*, ed. by J. Behrmann, and T. Srinivasan, chap. 34, pp. 1883–2023. Elsevier, Amsterdam.

Gender Bias in mortality in Ireland around 1900

Abstract

Excess female mortality in Ireland rose after the famine of the late 1840's while it began to normalise in other European countries. This exceptionally high excess female mortality persisted in some age groups even until the mid twentieth century. In my paper, I show the historical background that could have led to such a special pattern, explain excess female mortality theoretically and try to find an explanation for the excess female mortality using panel regressions with county data from the Irish censuses from 1871 to 1926.

Using panel regression to explain excess female mortality, I find that the pattern of explanatory factors is similar to the pattern found for example in England at the same time. I find that women's situation improves with a higher labour market participation. But, as working conditions are detrimental to health during the observed period, a higher female participation in sectors such as the industrial sector ruins the gains in power over family resources. Furthermore, the high emigration rates of women and the high share of the agricultural sector hurt women.

Comparing Ireland to England and Wales, a much higher share of agricultural employment can be observed. As agricultural employment increases excess female mortality in both regions, higher Irish outcomes can be partly explained. Further differences in explanatory variables that help to explain the higher Irish excess female mortality are a high emigration out of Ireland and a different effect of poor relief on women.

Contents

List of Figures

List of Tables

1 Introduction

Gender inequality is pervasive in today's developing countries with severe consequences for well-being and economic development. The inequality can be related to access to resources, education, health care or any other good. Gender inequality reduces the productivity of farms and firms, because low labour market participation of women leads to inefficiencies in the allocation of factors of production. Lower labour market participation and education of women reduces overall output and hurts, finally, the economic development of countries. The most severe effect of gender inequality shows up when looking at the costs that are imposed on the people directly affected. Gender inequality means a lower health status and quality of life for them and in the end a higher mortality rate and a lower life expectation (World Bank, 2001).

But not only the economic effect of premature mortality should be taken into consideration. Living per se is valued as a positive good and thus is striven for to be as long as possible (Sen, 1998).

Sen (1989), Klasen (1994) and Klasen and Wink (2002) find that if there was no inequality between the sexes, a much higher number of women would be alive today. They calculate around 90 million "missing women", missing because of their reduced access to survival-related goods in the household.

In the line of the concept of "missing women", studies on the most affected countries of today exist, for example for India (Coale, 1991; Das Gupta and Bhat, 1997), Sri Lanka (Langford and Storey, 1993) and China (Coale and Banister, 1994) or for Northern Africa (Tabutin, 1992).

In a historical European context, Ireland stands out as a country with a particularly large disadvantage in mortality for women from late 19th to mid 20th century (Coleman, 1992; Klasen, 1999). But other countries also suffered from excess female mortality, for example Klasen (1998) shows that excess female mortality was high among adults in Germany from 1740 to 1860. Humphries (1991), McNay, Humphries, and Klasen (2002) and McNay, Humphries, and Klasen (1998) investigate the excess mortality rates of England and Wales in the 19th century, which exist not only in the adult women group, but also among adolescents.

After having suitably defined mortality disadvantage, the measure should

indicate female excess mortality arising due to gender inequities in the access to resources (health care, clean water, clear air, food).

Two main hypotheses can be extracted from earlier investigations on excess female mortality. Agricultural change hurts women more than men, because women in ante-industrial societies are mainly employed in agriculture; when agriculture becomes more mechanised, many women lose their jobs (e.g. Boserup (1970), Johansson (1991)). With upcoming industrialisation, new job opportunities arise. But besides the problem that men are preferred in industry jobs, because they are stronger, working conditions in early industrialisation are hard and safety guidelines do not exist, so the employment that helps women in financing their own life also hurts them. It is not clear whether the positive effect of new employment in the industry sector exceeds the negative effect of harmful work conditions.

The remainder of the paper is as follows. In the next chapter, I define excess female mortality, chapter 3 then presents theories on explaining excess female mortality in child and adult ages and comes out with testable hypotheses. After the description of the data and the econometric framework in Section 4, I present the mortality and explanatory variables in chapters 5, 6 and 7. Finally, chapter 8 shows the regression results from explaining excess female mortality and in chapter 9, I draw conclusions from my results.

2 Definition of Excess female mortality

The problem is that it is not clear whether the higher female mortality rates exist despite or because of equity between the sexes. One can only detect times when there were high mortality rates compared to the other time periods. Kennedy (1973) has found a high excess female mortality based on comparisons with England and the U.S.

The definition of excess female mortality is based on an assumption of how large the female relative to male mortality rates should be. Klasen (1999) argues that we cannot use a benchmark of a society with complete equity between the sexes because it does not exist. Furthermore, mortality rates can differ for reasons different from resource allocation. Today, female mortality is relatively low amongst others because of the higher cigarette and alcohol

consume of males. Because I am interested in excess female mortality that arises because of inequities in allocation, I have to normalise the mortality rates.

When we want to compare mortality rates directly, we should standardise the mortality rates to make them comparable. We can standardise the rates with different methods. McNay, Humphries, and Klasen (1998) use the diagonal regression method of Preston (1976) to standardise English mortality rates. The method calculates the expected mortality rates by regressing male on female mortality. The regression line is then taken as the expected mortality rate and a positive deviation of the observed female mortality from the line is then defined as excessive.

Another way to standardise mortality rates is to take model life tables. The life tables I use to normalise the mortality rates were developed by Coale, Demeny, and Vaughan (1983). They are based on different, mainly European, regions in the late 19th and early 20th century. The tables are based on the mortality rates and life expectancies, showing the highest mortality rates in the South and the West regions and the lowest in the North. As the West tables are based mainly on England, we use these as a benchmark for the Irish mortality rates. We can then see if Irish excess female mortality is higher than the standard in this region and to this time. The reasons to choose the second approach is that the normalisation to the surrounding region is a reasonable indicator for the speciality of Irish mortality rates.

I normalise the actual mortality rates using the following equation:

$$efm = \frac{fm_a/mm_a}{fm_n/mm_n}, \tag{1}$$

where efm = excess female mortality, fm = female mortality, mm=male mortality, the subscript a refers to the actual and n to the normal value.

3 Theories of excess female mortality

Since excess female mortality is understood as an outcome of inequalities between the sexes, theories explaining excess female mortality look into the

families as the first place to find inequalities in the allocation of survival-related goods. As I will explain both excess female mortality of children and of adults, I will divide the theoretical considerations into explanations for girls' and women's excess mortality.

The first model explaining female mortality was developed by Becker (1964, 1981). The household was by then seen as an entity and an increase in household income increased welfare of all members in the same way. As the basic model is not able to explain inequalities in the allocation of goods, new models have been developed. They are based on bargaining models and the basic idea is that the allocation of goods in the households is subject to bargaining among the family members and the relative power of the members. Bargaining models can be cooperative or non-cooperative.

Child mortality can be explained by different theories. First of all, bargaining models used for the allocation of goods among adults can be enlarged to include children so that they also have to play a role in the bargaining process. Another method is to look at children as an investment of the parents in the future, for example to assure their old age pension. The last possibility is that there are non-economic reasons to prefer one gender out of cultural tradition.

3.1 Adults: bargaining models

Distribution between men and women is normally addressed in the framework of bargaining models of marriage or models of distribution within the family in general. The first model of the household was developed by Becker and can be called the common preference model (Becker, 1964, 1981). In Becker's model, an effective altruist maximises utility of all members of the household through the allocation of resources. Female mortality depends in this model only on changes in overall changes in resource availability, because resources are always divided equally over the family, independently on who gains most.

The empirically more successful models are the bargaining or collective models (McElroy and Horney, 1981; Klasen, 1998). The important concept in bargaining models is the threat point which plays a decisive role for the intra-household resource allocation. Specialisation of labour and public goods related to the household bring gains of marriage, but the share of resources

that each household member gets is subject to a bargaining process. The strength of the partner is influenced by outside option of both partners. The outside option is determined by the earnings opportunities of both partners, the relative position in the marriage or remarriage market and other indicators that could affect the well-being of both partners after divorce. Mortality of women in this model depends not only on the overall income as in the case of Becker's model, but more on her own income. With a higher income she reaches a better bargaining position and can claim a higher share of resources. Even if the two models differ considerably, the model of Becker can be regarded as a special case of bargaining models (see Thomas (1990)).

Sen (1990) introduces the concepts of perceived interest and perceived contribution into the bargaining model. Perceived interests of the partners can influence the resource allocation in the household. As women may be more altruistic to children than men, they share their resources with the children and end up in a weaker position. Also, the contributions to the household can be perceived differently by the household members. Household work may be valued less by the husband who contributes his gained wages to the household, so that he claims a large part of the common resources. The allocation of the resources depends now not only on the breakdown position, but also on factors that influence life in the family. Mortality of women does not decrease automatically in this model when she earns more money if her perceived interests lie in the welfare of her children.

In all three models, the state can influence gender bias in mortality by providing survival related goods to individuals or households. When the endowment with survival related resources increases, the amount of resources per household member increases, even if not in the same way as proposed in the last two models.

A macroeconomic view on the change in excess female mortality could be provided by using the Kuznets relationship between development and inequality (Kuznets, 1955), which says that as countries develop, inequality first worsens and then improves again. Used for inequality between genders, the theory tells that women's economic role is weakened in early development (they are losing bargaining power) and is strengthened later on. In the huge agricultural change in Ireland that took place in the time of my analysis, women lost their income possibilities and with them their bargaining power.

In industry, the many female outworkers were in very unhealthy environments and died early from tuberculosis or other illnesses. And finally men possess more strength so they were more likely to be hired in industry than women.

The theories provide several testable hypotheses on the power of women in the household even in the case when only macro data are available. While the theoretical household models at hand try to explain what factors drive the allocation of resources between the sexes, I will explain the county-wide mortality rate. The mortality rate could be defined as a kind of relative female health status and thus as an outcome of the distribution of resources in families. The explanatory variables in the household models are proxied with aggregate county variables.

One hypothesis is that with a higher amount of spinsters of marriageable age, divorce can less be used as a threat to the husband in case of non-cooperation, because a wide remarriage market is available to the men. The special case of Ireland takes some power of this hypothesis, because divorce is especially hurting women because of the strictly catholic population in most of the counties.

A further hypothesis is related to the power in the family due to resources brought into the household. Horney and McElroy (1988) argue that earned incomes may not be a good indicator for the bargaining power of a partner, because earnings are a part of the strategy of the household to maximise utility. The better measure may be non-labour income as inheritances or property of one partner, but that cannot be used in a macro model. Thus, we use labour market participation of men and women and suggest that the higher the labour market participation of women, the higher their power in the struggle over shared resources should be and so, their mortality should be lower. Even if the non-participation of women is part of the utility maximisation of the family, women are weakened in their position because their outside option is not part of the maximisation calculus.

The hypothesis based on the Kuznets relationship reads that variables measuring agricultural change should influence excess female mortality negatively.

3.2 Children as investment goods

For the explanation of how resources are allocated to and among children, three different theories can be applied. First, the bargaining models that are used to determine the allocation decisions among adults can be enlarged to be valid also for children. When we apply bargaining models to whole families, children will also have to bargain for their resources. As the adult family members decide on the allocation of goods in the normal case, children do have a weak standing in the bargaining process. Between parents and children, Sen's concept of perceived interest can be used to explain the allocation decisions (Sen, 1990): With a rising position of the mothers, children are better off in the allocation process because of the altruistic behaviour of mothers. If women have a tendency to prefer daughters, then a better position of women helps to lower excess female mortality in child ages.

Behrmann (1998) develops a model where children's expected income goes into the utility of the family and thus children take part in the distribution of goods with their future value. That model can be subsumed under the second strand of literature on explaining resource allocation to children: Children are seen as investment goods and thus the costs and gains of children are taken into account by parents. Having a child induces cost for raising, feeding, education etc. The gains can consist of different factors, but the background is always that children are born to transfer resources, i. e. child labour services or care for aged parents. Several reasons can exist to diminish the gains from daughters. In most regions and times it has been found that boys do have higher chances of finding paid work than girls so the maintenance of the family can be guaranteed more easily with a son. In regions where girls settle with their husband's family upon marriage and care for his family, there is little incentive to sustain daughters after this. The tradition of giving big dowries is also a hindrance to the well-being of girls.

Hill and King (1993) model the decision of parents to educate their children, again using a utility function with the expected incomes of the children as one factor. Expected income is dependent on the education, the alternative is sending the children to work. Without defining preferences or tastes of parents, the model can explain e.g. with a lower return of schooling for girls, why rational parents discriminate against girls.

When the two theories above - the bargaining model or the investment model - cannot be applied, there can still be a non-economic reason that parents discriminate against girls. The non-economic values of boys could exceed that of girls because of tradition or because (for example) a son would maintain the family name or special rites (Schultz, 1997).

4 Econometric Framework and Used Data

Theory provides a framework for testing excess female mortality. I will explain the excess female mortality rates in two age groups, 10 to 19 and 25 to 44 with the explanatories coming from theory and earlier empirical investigations. I included for example emigration because of the especially high emigration rates out of Ireland during the observed period. The model will be as follows:

$$\begin{aligned} EFM_{it} \;=\; & f(Overall\ Mortality_{it},\ Economic\ Structure_{it}, \\ & Female\ Employment_{it},\ Living\ conditions_{it}, \\ & remarriage\ market_{it},\ emigration_{it}, \\ & agricultural\ change_{it},\ education_{it},\ births_{it}) \end{aligned} \tag{2}$$

where the subscripts i and t refer to county and year. The explanatory variables are grouped into the above categories and will be defined in Section 7.

The time period of my investigation is 1871 to 1926, with observations for 1871, 1881, 1891, 1901 and 1926. These are the years when a census was conducted by the Irish government. There exists another census of 1911 which I decided not to use. As Budd and Guinnane (1991) document with the help of a linked census sample containing the censuses of 1901 and 1911, the introduction of an old age pension system in 1908 led the people to lie about their ages. To give an earlier birth date could help to receive a pension; this was possible because exact birth dates were only registered from 1864. Budd and Guinnane (1991) find in their estimation results that people from the age of 30 onward would exaggerate their ages, with women exaggerating

more than men. We decided not to use this census year, because there is also the possibility to use the 1926 census that had been undertaken shortly after the separation of North and South Ireland and is still comparable. The observed period is situated after the Irish famine of 1845 to 1850 and ends when Ireland becomes independent. Thus, it does not include the time when the female mortality reached a normal level in the mid twentieth century.

Mortality data of Ireland are published on county level in the Annual Reports on Vital Statistics (Registrar General for Ireland, several years). The data exist on a yearly basis, but we have only taken the census years, because only then explanatory variables are available.

The data set contains observations for the 26 Irish counties and for Ulster. I could not use the counties of Ulster separately because the division of the cities over the counties changed over time, so proper population numbers and numbers of deaths could not be derived. For the 26 counties, we observe 5 years, for Ulster only 4 (with 1926 missing).

5 Excess Female Mortality in Ireland

The starting point of my investigation lies after the time of the Irish Famine of 1845-1850. In famines, mortality rates of women are observed to be lower than that of men (Dreze and Sen, 1989, p. 55), and it is argued that this happens because of biological reasons. Also in Ireland, women had a lower death ratio than men during the famine (O Grada, 1993, p.180).

Calculating excess female mortality in a sensible way involves the normalisation of mortality rates with respect to the life tables. Life table mortality rates are calculated for different life expectancies of females and the respective expectancies of males. I use the male life expectancy as the normal, unbiased case and take the respective male and female mortality rates. I could have taken the female life expectancy as the unbiased case, but as we are concerned about an abnormally high female mortality, I prefer the first variant.

Life expectancy of women in Ireland is low relative to men and relative to the European standard at this time. The life expectancy both for women and men increases from the middle of the nineteenth century to the middle

of the twentieth like in other European countries.

Figure 1 shows that it is only in around 1900 that life expectancy rises at all; it shows, too, that it is only in the 1930's that women gain in life expectancy over men. This phenomenon happened much earlier in other European countries. Taking the male life expectancy as the norm, life table female life expectancy is included in the Figure and lies some five years higher than the actual female expectancy. From the 1930's onwards, life expectancy is further split into urban and rural with a 5 year higher expectancy of urban females than males. Rural life expectancies are not shown, but are higher than the urban. This is possible due to the bad housing conditions and the hard work in industry in the urban areas.

Still at the end of the seventies, Walsh and Walsh (1978) discuss the relative low female life expectancy in Ireland compared to other European countries. They argue that the normalisation of the female life expectancy went hand in hand with economic upturn so that only the economic situation caused the relatively high female mortality, while inequality between the sexes is not needed as an argument for the explanation of the life expectancy differences. But with this argument, the question remains why the Irish men are less hurt by the economic situation than women.

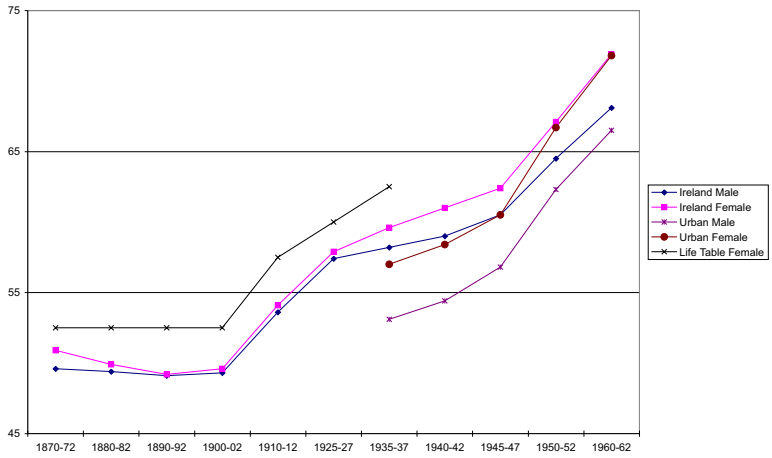

Figure 1: Life expectancy in Ireland and life tables

Having found a low life expectancy of women relative to men, I now turn to the age specific excess female mortality rates. To see if they are also high relative to the life table rates or internationally, I will compare them with

Table 1: Life table and Irish mortality

Age group	Life table West			Mortality rates Ireland			Norm
	$m(x)_f$	$m(x)_m$	$\frac{fm_n}{mm_n}$	$m(x)_f$	$m(x)_m$	$\frac{fm_a}{mm_a}$	EFM
10 to14	2.88	2.67	1.08	3.35	2.54	1.32	1.22
15 to 19	4.09	3.99	1.03	5.12	4.33	1.18	1.15
20 to 24	5.32	5.69	0.93	6.06	6.77	0.90	0.96
10 to 19	3.48	3.32	1.05	4.22	3.4	1.24	1.19
25 to 44	7.28	8.07	0.90	8.33	8.12	1.03	1.14

Life table mortality Level 14 is valid for 1871 to 1901 (1926 is Level 16)
Irish mortality rates are averages over all available years 1871 to 1926

both.

Table 1 shows the life table mortality rates and the mortality rates of Ireland and in the last row the excess female mortality in Ireland. We look at different age groups, from 10 to 14 and 15 to 19 as the age groups of adolescents, and 20 to 24 and 25 to 44 as the age groups of adults. Mortality rates of men and women are higher than the life table mortality rates; only boys aged 10 to 14 have a lower mortality than the life tables would predict. The ratios between female and male mortality are nevertheless higher for the Irish rates than for the life table, with again one exception, in the age group 20 to 24.

The highest excess female mortality exists in the age group 10 to 14, while in the age group 20 to 24 even one value indicating excess male mortality can be observed. The other values lie in between.

Figures 2 and 3 present international excess female mortality for the age groups 5 to 10 and 20 to 25, provided by Klasen and used in Klasen (1999). The excess female mortality rates stem from Ireland, Sweden, UK and Italy and refer to the nearest date to the date on the axis.

Excess female mortality in Ireland is internationally the highest in the 5 to 10 age group over all years except in 1850 which was the end of the famine. During the famine, women were relatively well off possibly because of biological reasons. In all years, Ireland shows an excess female mortality that means the ratio is bigger than one.

In the 20 to 25 age group, Ireland shows an excess female mortality below

one between 1870 and 1900. Other countries showed higher values during this period. Up to 1850 and from 1910 onwards, Ireland shows, as expected, the highest excess female mortality among the countries.

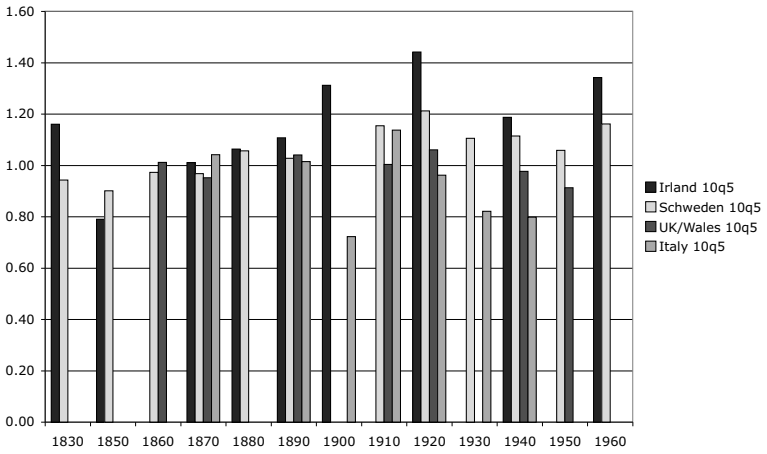

Figure 2: International normalised excess female mortality age 5 to 10

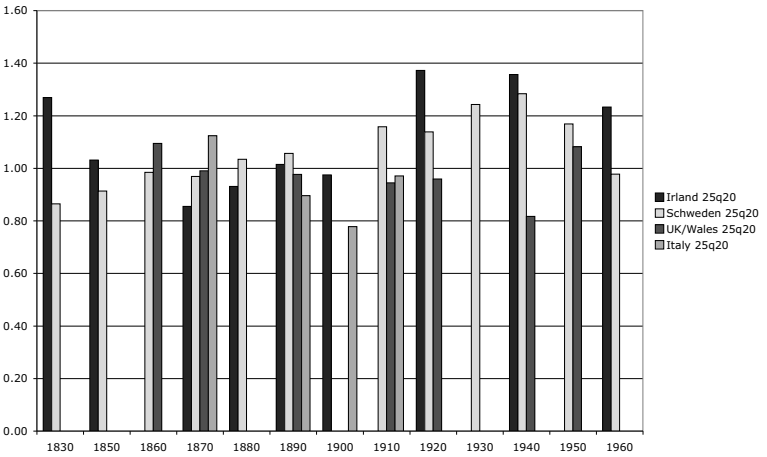

Figure 3: International normalised excess female mortality age 20 to 25

In the following, mortality rates over time and space in Ireland are presented. Figure 4 shows the age specific mortality rates of females as a percentage of males in Ireland from 1864 to 1967. The development of female relative to male mortality shows quite high differences looking at different regions, age groups and times. It is only in the 1940's that most of the age specific mortality rates decline rapidly, but some already decline from the beginning of the 19th century. Relatively high female mortality can mainly

be observed in the age groups from 5 to 24, not showing higher mortality of males until the 1950's. These values are not normalised, but as Table 1 shows, excess female mortality is expected to be high, too. Without normalisation, we nevertheless suspect disadvantages for women in certain periods when looking at the significant differences in mortality rates over time.

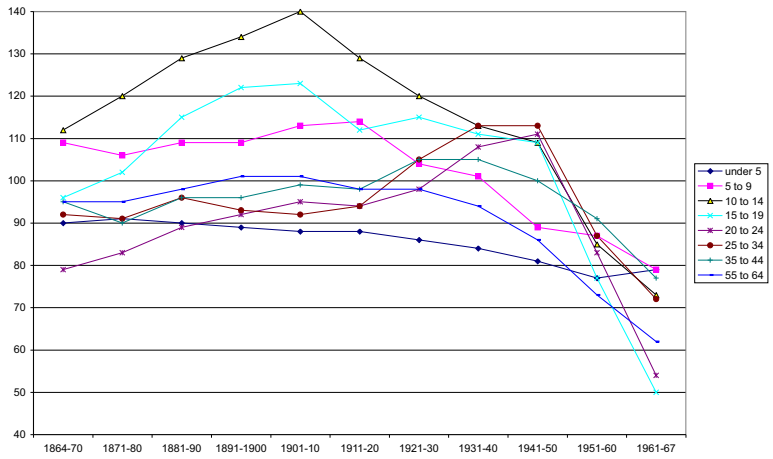

Figure 4: Mortality rates Ireland (Kennedy, 1973)

Figures 5 and 6 show the excess mortality rates for the age groups 10 to 19 and 25 to 44 on province level. The mortality rates of these age groups, but on county level, are explained in a multivariate model in the remainder of the paper. Excess female mortality in the two presented age groups is increasing in most provinces over the observed time span. The highest increase occurs in Connaught in both age groups while in Leinster the ratios do not increase (in the age group 25 to 44), or only slightly.

On county level, the following maps (Figure 7 and 8) show the development of the age groups 10 to 19 and 25 to 44 for the counties in the first year and the last two years (1871, 1901 and 1926). The two age groups stand for girls and adult women. The darker the counties are shaded, the higher the relative female mortality in the presented age groups is.

There exist two statistics to test the null hypothesis of no spatial autocorrelation, assuming that x is normally distributed. Both tests for spatial autocorrelation, Moran's I and Geary's c, show for all years and age groups that the null hypothesis of no autocorrelation cannot be rejected. The agglomeration areas of Ireland Dublin and Belfast show a relatively positive

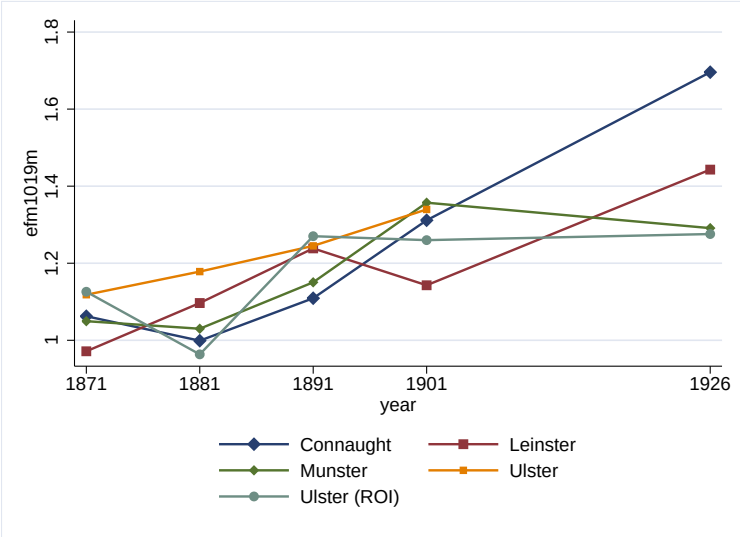

Figure 5: Excess female mortality rates age 10 to 19

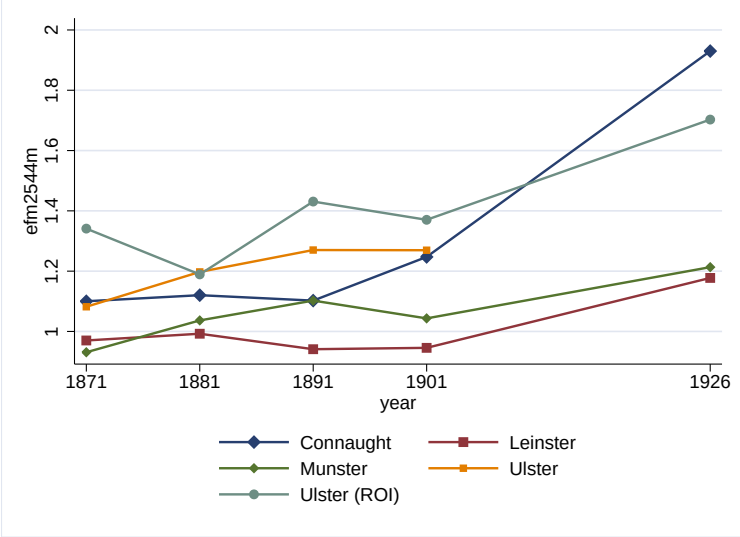

Figure 6: Excess female mortality rates age 25 to 44

picture for women, while there is no consistent picture regarding the worst areas. It seems that high excess female mortality is not directly related to the backwardness of the area. We also cannot see a pattern like a west-east division that is sometimes postulated because of the backwardness of the west (Kennedy, 1973).

One cause for the high female mortality rates could be that women had to work very hard. But hard work cannot be the only cause for the mortality differences, because we have observed the highest excess mortality rates in child ages. An extensive analysis of the causes in a multivariate model has not been carried out to my knowledge.

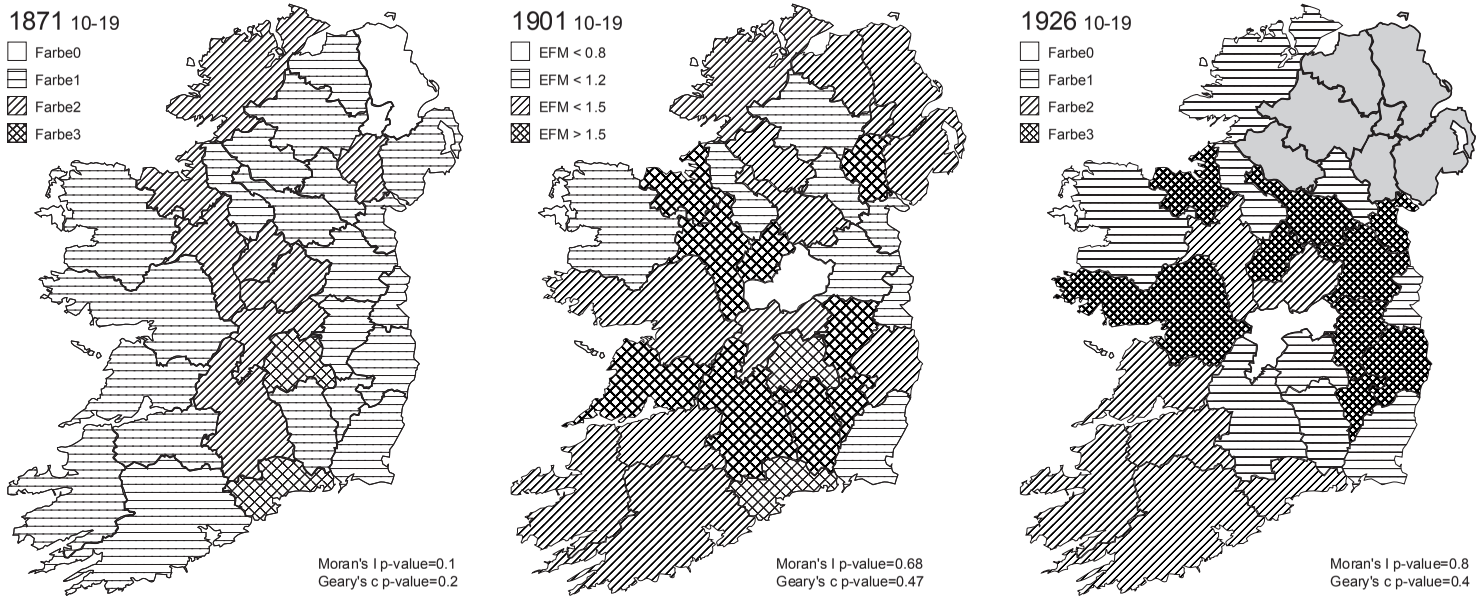

Figure 7: Relative female mortality rates in Counties age group 10 to 19, 1871, 1901 and 1926

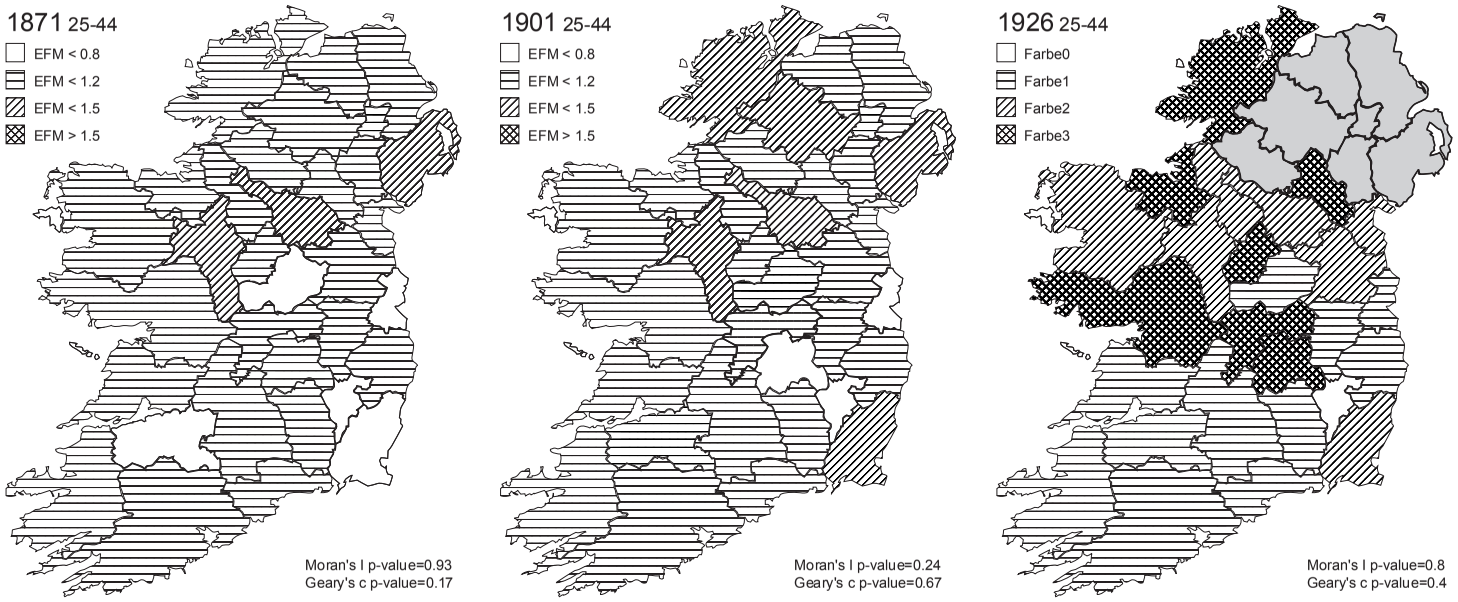

Figure 8: Relative female mortality rates in Counties age group 25 to 44, 1871, 1901 and 1926

6 Mortality from Tuberculosis

Tuberculosis was a serious problem in Ireland during the surveyed period. Even if it does not explain mortality rates, it is an interesting part of mortality that is worth looking at. I assume that excess female mortality from TB is influenced by the same factors as excess female mortality overall. This assumption is reasonable because we will see that mortality from TB and overall mortality are highly correlated.

The Tuberculosis epidemic was rising in the 1880's and 1890's and culminated in 1904. It had started after years of relative health in the 1860's and 1870's when the mortality from tuberculosis had even been lower in Ireland than in England, Wales and Scotland (Jones, 2001). The high mortality rates were falling very slowly in the following decades and in the 1940's they were still higher than in many European countries. The normalisation of the overall mortality rates coincides with the normalisation of tuberculosis.

Mostly youth and young adults were infected by tuberculosis (Jones, 2001). During puberty mostly females died of TB, in older age groups the death rates of TB were more equal between genders. It was more probable for poor people to have TB and urbanisation brought a higher TB rate, men being more affected.

Tuberculosis was regarded as a hereditary disease, so that the infection of children of marriageable age was often not admitted to increase the probability of marriage. Thus, an underenumeration of TB deaths in these age categories occurs. Jones (2001) reports that for example in Dublin, previous to 1879, 10% of the dead were not registered, but showed up in the returns of burials. It often occurred that deaths from TB were related to a harmless death cause as for instance bronchitis.

At that time, several explanations spread for the high mortality of tuberculosis. One was that the celtic race was specially prone to an infection with TB. This theory was contradicted by the fact that in the most celtic part of Ireland, in the West, TB was least spread. The effects of emigration were controversially discussed. One argument was that the soundest were removed from Ireland and thus mortality (overall and from Tuberculosis) was especially high in the remaining population (Jones, 2001).

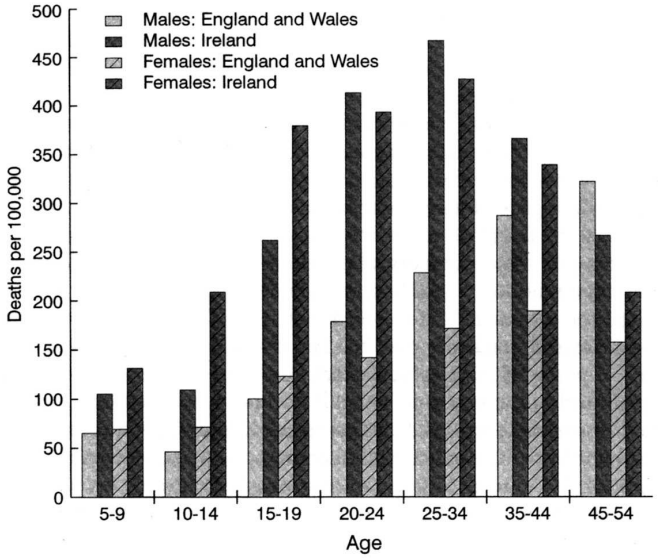

Figure 9: Death rates from tuberculosis by sex in Ireland and in England and Wales in the early twentieth century (Guinnane, 1997, Figure 4.1, p. 115)

The true reasons for the high mortality from tuberculosis and the particular susceptibility of women may have been the same as for excess female mortality: poverty, urbanisation and malnutrition. In 1871 50% of the female workers in the Linen factories of Belfast were aged between 15 and 25 years and hard work and unhealthy environment led to higher mortality rates. Many of them worked at home; the bad housing conditions in urban areas made it easy for the bacillus to spread over the women and their daughters, staying with their mothers while men and their sons worked in the fields. The women who were weakened from the bad living conditions and malnutrition contracted tuberculosis that their body otherwise may have resisted.

In the age groups 10-14 and 15-19, one can see from figure 9 that Irish female mortality from Tuberculosis exceeded mortality in other countries. In the case of 15 to 19 year old girls the percentage difference to the males was even double the mortality in England.

Figures 10 to 12 show the extent of TB mortality relative to overall mortality. Figure 10 shows that the share of TB deaths on all deaths was around 12 % for females and 11% for males in 1871. In 1926, it was still 10% for both genders, there was almost no change.

Figure 11 demonstrates the share of TB deaths on all deaths in age groups for 1871. The highest difference between males and females is in the age group 10 to 14, while the highest shares are in the ages 15 to 24 with a higher share for women than men.

Finally, the regional pattern in 1881 is shown in Figure 12. With four percentage points difference, Ulster shows the highest shares of TB deaths on all deaths (14 and 18 percent for men and women) and also the highest difference in genders. Possibly, this result is due to the high industrialisation in Belfast.

The correlation coefficients between TB mortality and overall mortality amounts to 0.98, 0.90 and 0.94 for the three graphs, so it can be argued that TB mortality is driven by the same factors than mortality from all causes.

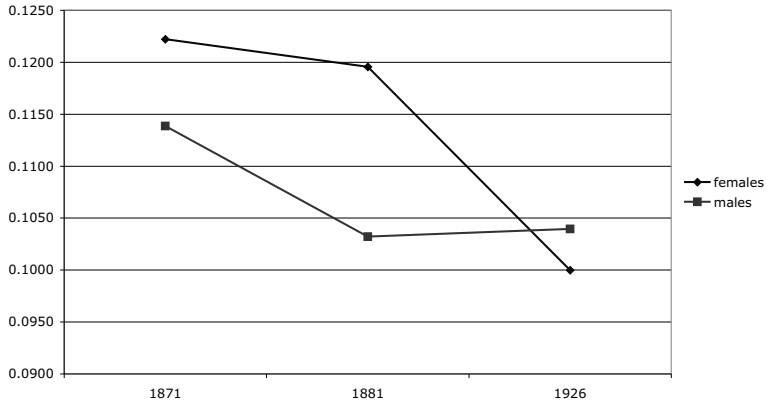

Figure 10: Death rates from TB rel. to all deaths in Ireland over time

We have seen that Tuberculosis led to a high number of deaths in teenage age which is the age of the highest excess female mortality in Ireland. Underreporting of deaths is a general problem, but if the percentage of underreporting is equal between the sexes, relative female mortality should not be biased. Furthermore, underreporting of deaths from Tuberculosis seemed to be more related to higher death numbers of other diseases like bronchitis. Jones (2001) does not suspect that there existed a difference in reporting deaths of women and men. If there was slight underreporting in deaths of any gender, it would be female mortality that is higher in reality, so that excess female mortality would also be higher.

To test if relative female mortality from Tuberculosis depends on the same

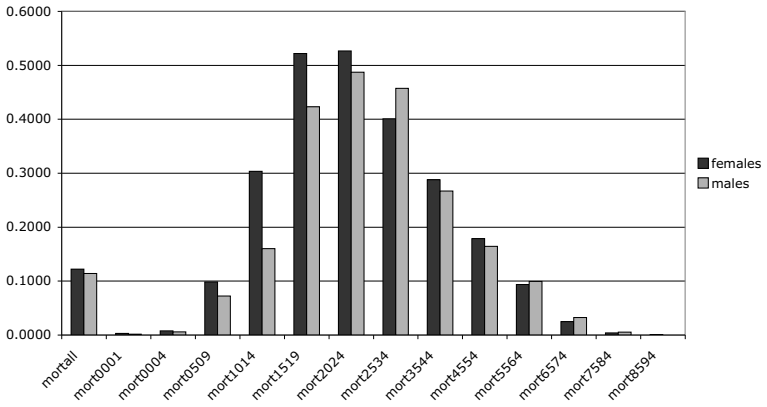

Figure 11: Death rates from TB rel. to all deaths by age groups in Ireland in 1871

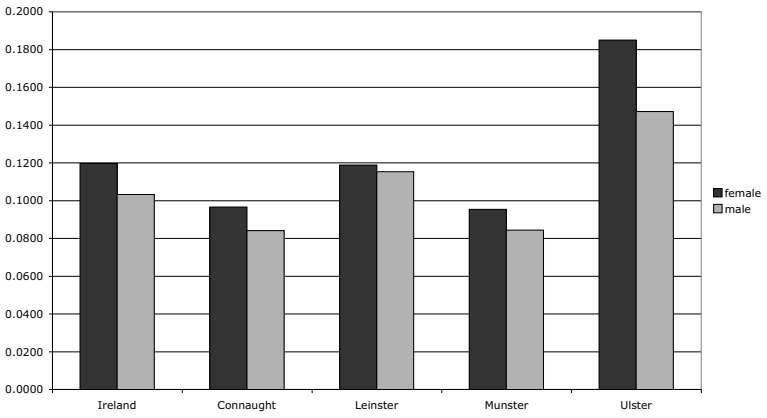

Figure 12: Death rates from TB rel. to all deaths by Province in Ireland in 1881

sources as relative female mortality as a whole, mortality from Tuberculosis could be introduced as an alternative dependent variable. But, as TB deaths are aggregated on a much higher level, I do not have enough observations to introduce it into a regression. As could be seen from the correlations between TB mortality and mortality from all causes, the results would show the same picture, only the scale would be different. I will assume in the remainder of the paper that there is no underreporting of female TB deaths relative to males.

7 Explaining relative female mortality

In the descriptive chapter it became clear that Irish relative female mortality was not only high in international comparison, but also showed differences across the regions. I will present several indicators which influence the female mortality rate in theory or in other empirical applications and which could serve as an explanation for the differences among counties. In the next chapter, I will then test which of the possible explanations show up to have a significant influence in a panel regression on excess female mortality in different age groups.

Table 4 shows an overview of the presented variables pooled over all counties and years while in the description of the variables, I aggregate over provinces and show the trends over time.

7.1 Overall mortality

Surveys about 19th century England show that women are relatively well off in terms of mortality when the general mortality is high (e.g. Woods (2000) or McNay, Humphries, and Klasen (2002)). An extreme example of a high mortality period is a famine and as mentioned before, female mortality in the Irish famine was relatively low with respect to male mortality. After the famine, overall mortality decreased, but relative female mortality was rising. When we shall later include overall mortality in the regression, we accordingly expect a negative influence on the excess female mortality. Figure 13 shows the overall mortality rates in provinces over time.

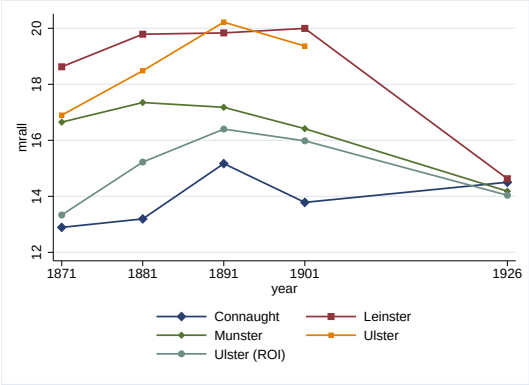

Figure 13: Overall mortality rate (based on census data; own calculation)

7.2 Economic structure

We expect that the labour market structure influences mortality of females relative to males. High participation in the labour market in general shows a well developing economy; many people earn wages and the standard of living is generally higher.

The Irish Census data split up the number of workers into five working categories. The professional class includes general or local government, defence, professional occupations (for example teachers, students, or members of religious communities). The domestic class includes all domestic offices or services. The commercial class includes commercial occupations, conveyance of men, goods and messages (for example merchants, commercial clerks and coach makers). The agricultural class includes agriculture and animals (as for example farmers, graziers, agricultural labourers and farm servants). The industrial class includes persons working and dealing in books, prints and maps, machines and implements, houses, furniture and decorations, carriages and harness, ships and boats, chemicals and compounds, tobacco and pipes, food and lodgings, etc. (for example lodging house owners, cabinetmakers, millers, greengrocers, dressmakers, tailors, shopkeepers). The last category, the indefinite and non-productive class, includes persons not producing (vagrants, schoolchildren, wives).

I aggregate the commercial and professional sectors into one sector, because neither involves physically hard work like in agriculture or industry. Leinster, with Dublin, shows the highest percentage of workers in the profes-

sional sector, while in Connaught and Ulster (ROI), the two most backward areas, the least people work in the professional sector (see figure 17).

The agricultural sector shows the opposite picture to the professional sector (see Figure 18). Now, Leinster is among the provinces with the lowest share of agricultural workers (together with Ulster which is with Dublin one of the agglomeration and industry areas) while Connaught and Ulster (ROI) are the leading agricultural provinces. This pattern persists over the whole observation period and gets even more pronounced at the end of the period.

The percentage of people not working increased until 1901 and remained stable after that. Luddy (1999) tells that the Irish commissioners were told to transfer more and more women over time into the nonproductive sector even if they worked in the domestic sector or in agriculture. The increase in the non-working population shows possibly only the changes in the method of collecting the census data, not a real change in the working behaviour.

7.3 Female employment

According to theory, participation of women in the labour market strengthens the women's position in the family, so that the mortality of girls and women relative to men and boys ought to fall with a higher employment rate of women. Women gain more power by receiving wages and are better able to get their share from the family resources. This should help them to reduce their probability of dying. Moreover, according to the theory, it should be expected that a strengthened position of the woman in the family also raises the status of the girls, as wage-gaining mothers are generally better able to support their children and especially their daughters.

Figure 20 shows the labour market participation of women relative to men. I argued before that the overall labour market participation went down because the Irish commissioners counted more and more women as non-productive even if they worked in the family business (Luddy, 1999). The full extent of this argument is displayed in Figure 20, showing that the number of women relative to men working diminishes dramatically over the observed period. I will account for the downturn in female employment with an interaction term between female employment in domestic services and the year 1871, because in this sector, the downturn is observed to be highest

among all sectors.

The first assumption, that the status of a working woman will be raised, can be expected, as it was usual in Ireland that men disposed of their wages absolutely freely and did not even tell their wives how much they earned. After having absolved the necessary investments for the farm or business (and frequenting the pub), the wife got the left-overs to provide for the family. With an own income of women, the provision for the family is expected to ameliorate, according to theory.

As many of the jobs available for women in the surveyed period were very unhealthy, the positive effect of the strengthened position in the family will not necessarily suffice to reduce the mortality risk. Accordingly, we might even observe a negative effect of female labour participation depending on the observed sector. Fahey (2003) suggests in his descriptive investigation that domestic servants will have a lower mortality, because they have better housing possibilities than others.

From 1800 to 1918 women worked mostly in agriculture, domestic service and the textile industry. Already before the famine many women worked in the textile industry, which was one of the main employers of women at the beginning of the 20th century. Industrialisation took place mostly in the North East of the country and in a reduced way in the towns of Dublin, Cork, Waterford and Limerick. For the rest, Ireland was an agricultural state. Towards the end of the 19th century chances for women in agriculture worsened considerably. The whole agricultural system shifted from tillage to dairying, there was more stock rearing and at the same time more mechanisation. All this led to a lesser work intensity of the production and a lesser demand for a female work force. The extent of the shift in agriculture is shown in the next chapter.

The second assumption, that the situation of girls ameliorates when their mothers work, cannot necessarily be applied to Irish circumstances. As there were very many out-workers in the early industrialisation in Ireland and these jobs were mostly occupied by women, the girls had to stay at home together with their mothers. The bad housing conditions may have led to a higher mortality (see chapter 6 about Tuberculosis). Moreover it was normal in Irish families that men and sons had the privilege to eat first, and that privilege would not change with the standing of the mothers, because it was seen as a

tradition (Luddy, 1999).

7.4 Living conditions

Living conditions are proxied by three variables: population density, the share of women in workhouses and the share of families living in fourth class housing. The last two indicate the poverty of the families and women, the first is an indicator for urbanity of the neighbourhood.

Figure 30 shows the population density in Irish areas, which is an indicator for backward areas. Fahey (2003) reports that women died earlier, especially in rural regions. In 1926, the census shows the mortality difference by comparing mortality in the province of Connaught with the four main cities. Female life disadvantage at birth in Connaught amounts to 1.7 years while in the main cities, women had an advantage in life expectancy of 3.1 years at birth.

The relation of women to men in workhouses is shown in Figure 28. The Irish poor law legislation of 1838 was an adaption of the poor law of England and Wales of 1834 and brought a workhouse system for the very poor to Ireland. Figure 28 shows that Ulster with 1.15 women per men in workhouses started with the lowest value, Munster started with the lowest fraction of 1.35. The fraction went down below one until 1901 in most counties; values for 1926 are not included.

The number of families living in fourth class housing can be seen in Figure 29. 4th class housing is defined as:"Houses built of mud or perishable material, and having only one room and window" (Census of Ireland, several years, p.7). 4th class housing includes also 3rd class housing (one to four rooms and windows) with more than one family, 2nd class housing (five to nine rooms and windows) with four or more families and 1st class housing (houses of a better description than the preceding) inhabited by six or more families.

Fourth class housing can be taken as an indicator for the poverty of the population, which can have negative effects especially on females, as we have seen in the theory. Accordingly, we hypothesise that in counties with particularly bad housing conditions also the relative female mortality will

Table 2: Proportion never married in an international view (in percent)

	England and Wales		France		Germany		Ireland	
Year	M	F	M	F	M	F	M	F
1841							10	12
1851	12	12					12	13
1861	10	12	11	13			15	14
1871	10	12	11	12	9	12	17	16
1881	10	12	13	13	8	11	17	17
1891	10	12	12	13	8	10	20	18
1901	11	14	10	11	8	11	24	22
1911	12	16	11	11	9	12	27	25

Source: Guinnane (1997, page 96)

be particularly high. Poverty can lead to insufficient nutrition especially of the weakest members of the household, furthermore bad housing conditions can lead to a higher female mortality from diseases that are contracted while working in bad conditions at home.

7.5 (Re-)marriage market

In Ireland, a high proportion of men and women never married. One reason for this phenomenon could be the heritage laws that split up the land among all siblings, so that it was not possible any more for a family to earn a living from the land. Also as a dowry it had not enough worth to attract a good match. Another hindrance was the very strong Catholic Church that made it almost impossible to divorce.

Table 2 shows that Ireland's number of spinsters and bachelors is higher than in other countries. Being normal until the 1860's, the percentage of people never married increases from then. In Germany, England and Wales, the rates stay approximately stable.

The Irish perspective is demonstrated in table 3 where the differences between the provinces are shown. Connaught, the province with the highest rate of emigration, lowest industry and highest agriculture, shows the lowest rate of non-married, especially for women.

Table 3: Proportion never married in the provinces of Ireland (in percent)

	Leinster		Ulster		Munster		Connaught	
Year	M	F	M	F	M	F	M	F
1841	13	14	10	14	9	11	7	8
1851	15	14	13	15	10	10	7	8
1861	19	17	16	16	12	12	10	10
1871	21	19	19	19	13	13	12	12
1881	22	20	19	20	14	13	11	9
1891	25	22	21	23	17	13	14	10
1901	28	25	24	26	20	17	19	14
1911	31	28	26	27	26	21	25	18

Source: Guinnane (1997, page 97)

In my set of data the picture is similar. In Figure 26 the number of spinsters relative to the number of bachelors is shown. The numbers remain relatively stable over the whole period with Ulster having the most spinsters per bachelor (1.2) and with Connaught only having 0.8.

7.6 Emigration

Up to the time of the Famine population growth in Ireland was very high in relation to other European countries. Exact population growth can only be calculated from 1841 onwards when the first reliable count of the population was published in the census (Guinnane, 1997). Over the period 1750 to 1845, Guinnane (1997) proposes a yearly growth rate of 1.3 percent per year. In the last years of this period, the growth rate seems to be already lower, as estimates of Mokyr and Grada (1984) propose only a growth rate of 0.5 to 0.75 percent for the 1830's. Thus, while the high growth rates were lastingly stopped by the Famine, Guinnane (1997) argues that they would have sunk to a normal measure even without the Famine, though during a longer period.

During the Famine, excess mortality ranged from one quarter of the population to almost zero depending on the region (Guinnane and O Grada, 2002). The number of deaths from the famine range from 0.5 to 1.5 million dependent on the estimate, with the most reliable estimate being one million

excess deaths which means almost 10 percent of the population (O Grada, 1993). Regarding the past population growth rates we should have expected that population numbers would have recovered quickly after the famine. Instead, population began to decrease severely. There were two main factors that accounted for this decrease: During the Famine, a huge migration movement mainly to the U.S., but also to England started, and fertility declined lastingly. In Figure 14, the amount of the population loss can be seen. To take as an example the population of age 0 to 4, the loss is almost 50 percent over 50 years. Also the number of men relative to women reverses in most age groups. For example in the age group 25 to 29, there were significantly more women in 1871, but there are as many as men in 1926.

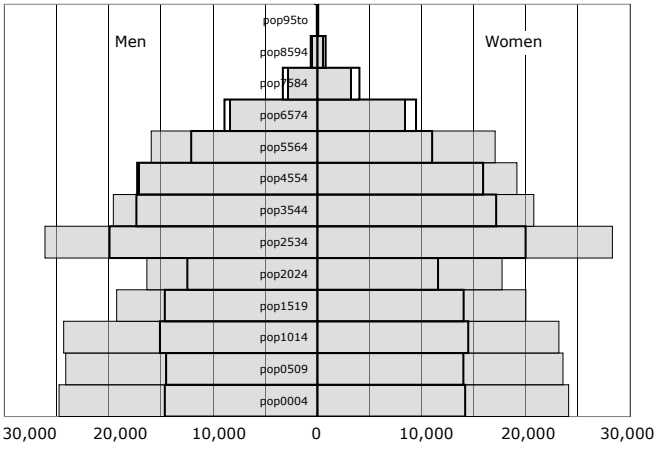

Figure 14: Population loss in Ireland mainly due to emigration

A huge wave of emigration to the USA had already started during the famine. These very high numbers make it look unlikely that the remaining population can be compared with a healthy and stable population. Yet the question arises, what the causal connection between the emigration and a negative selection of the remaining population is. It is possible that mostly the healthy and strong persons left the country, because they were still able to do so, as for instance during the famine. Thus the poor and ill ones remained in the country. On the other hand it also could be possible that the high mortality rate stimulated those, who found themselves in the risk group of dying - the poor and ill, to flee from home.

To measure the effect of emigration on excess female mortality, not only the numbers of people emigrating could play a role, but much more important

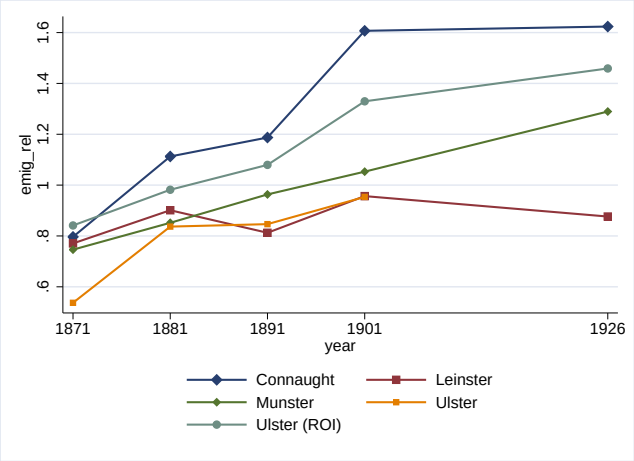

Figure 15: % of women emigrating rel. to % of men emigrating from Provinces

will be the relation between male and female emigration.

The interesting variable is the relation between emigration of women and men. In Connaught one can observe the highest emigration of women relative to men, but not the highest mortality of women in the age groups in which women are able to decide for themselves (15-24 and 25-34).

7.7 Agricultural change and rural areas

Between 1850 and 1920 agriculture in Ireland changed fundamentally. After the famine, several developments started: firstly, there was a shift from tillage to grazing. In the 1840s half of the output consisted of crops (wheat, oats and potatoes), in 1908 farmyard hen and duck contributed more to agricultural value added than all crops together (O Grada, 1993). This was made possible, among other things, by the change in the heritage practice which changed from dividing the land among all siblings to giving the whole land to one heir.

The change in agriculture is captured first by the share of small farms relative to all farms. Figure 24 shows particularly in the most agricultural district, Connaught, that the number of small farms went down. That happened possibly due to the changes in the heritage laws that put a halt to the progressive division of land.

The change of production in agriculture is proxied by the share of the county area that is cultivated with crops, fruit and horticulture. The variable is shown in Figure 25. The downturn which is expected can be confirmed, but on a very low level. The share of the area cultivated diminishes from a starting point of 5 to 8 percent by 2 percentage points, meaning a downturn of at least 25%.

The agricultural change went along with different occurrences. First of all, there was a dramatic fall in the numbers of people working on the land, so that a negative effect on land workers has to be expected. Furthermore, the regional specialisation in agriculture shifted, but it brought only little sustained growth in aggregate value of farm output. Nevertheless, the productivity growth was rather high. While output per worker remained low compared to Britain, the gap narrowed over time with a higher TFP than in Britain.

O Grada (1993) argues that changes in Irish farming patterns (induced by market conditions, embourgeoisement and technological change) reduced women's entitlement relative to men's.

7.8 Education

Traditionally, women in Ireland had almost no rights. Girls and women had to have their meals after men and sons had eaten and they had to show consideration for them in many ways. School education was denied to more girls than boys, as women were regarded as not capable to make use of the acquired knowledge. If girls were offered school education at all, there were special curricula and books, in order to prepare them for their future life at home, that is, they learnt cooking, needlework etc. People believed that knowledge should be given to women only to a degree that should enable them to become a helpful housewife (Luddy, 1999).

Literacy and school attendance are both good indicators for the backwardness and poverty of a country. Two arguments back up this thesis: on one hand a civilisation (culture) of poverty can lower the demand for school education, on the other hand it can be that families cannot afford school education because of their poverty.

The first argument, that poverty lowers demand for education, does not apply to Ireland according to O Grada (1993). He quotes Wakefield (1812) with the statement: "I do not know of any part of Ireland so wild that its inhabitants are not anxious, nay eagerly anxious for the education of their children".

The second argument, that people cannot afford education, was weakened by the Intermediate act of 1878, which intended to ameliorate school education and preparation for university. First it was only thought for boys, until several influential women pushed it through to be valid for all pupils (Luddy, 1999). In this act it was laid down that there should exist a central institution, which organised examinations and awarded prizes to the teachers. Moreover there were grants for boys and girls equally. The Irish education act of 1892 made school attendance compulsory for all children. Only from then girls went to school almost the same as boys. Yet it was unusual for girls to continue school after primary school. Nevertheless, several schools were opened in the 1860's in Dublin and Belfast, which offered secondary education to girls. This was the most remarkable development in the school system in the Ireland of the 19th century.

In spite of all these improvements there were still discussions that the female brain was not able to perform as well as the male one. Therefore at school girls were mostly prepared for their future life as housewives: they learned cooking, household management, child care etc. There were even special schoolbooks for the education of girls (Luddy, 1999).

Apart from better education and thus higher independence, which women could reach by school attendance, also relatively marginal facts played a role for the improvement of the girls' situation attending school. At school they normally got a meal, often the only one they had during the day. Thus beside the long term perspectives the short term perspectives also improved. When the school enrolment rate is high in a low developed country, school attendance provides the girls with a healthier surrounding than is to be expected at home. A high rate of female enrolment is moreover a sign of esteem of girls, as it seems worthwhile for parents to invest in school education. Arguing with theoretical considerations, we can say that a higher literacy among adult women leads to a better bargaining position in the family, so their mortality rate should decrease.

The highest share of illiterate women relative to all women can be observed in Munster and Ulster, the lowest in Connaught (see figure 27).

7.9 Number of births

The number of births often leads to maternal mortality in reproduction ages when health care is of a low standard. As a proxy for fertility rates, I use the number of births per woman aged 20 to 44 during the year of the census. Births to women are depicted in Figure 31 and show that the lowest birth rates were in 1881 and fertility rose from then. The lowest value was almost only 0.8 children per woman in Connaught, in 1926 the rate is up at least to 1.4. Ulster always shows the highest rate, so fertility seems to be related to urbanisation.

Table 4: Summary statistics

Variable	**Mean**	**Std. Dev.**
Explanatories to 1926, 127 observations		
EFM 10 to 19	1.246	0.337
EFM 25 to 44	1.161	0.264
Overall Mortality rate	16.096	2.98
Females to males working	0.469	0.197
Share of workforce in prof. sector	0.116	0.059
Share of workforce in agriculture	0.519	0.146
Share of adults without job	0.202	0.076
Share of adult women in industry	0.102	0.063
Share of adult women in dom. sector	0.242	0.144
Share of adult women in prof. sector	0.047	0.021
Spinsters to Bachelors	0.816	0.136
Emigration women to men	1.005	0.263
No. of births per adult woman per year	0.108	0.02
Share of farms 1 to 5 acres	0.133	0.055
Share of tilled area	0.062	0.064
Population density	0.28	0.385
Explanatories only to 1901, 107 observations		
Share of illiterate women	0.75	0.12
Female to male workhouse inmates	1.13	0.147
Fam. in 4th class housing	0.066	0.069

8 Estimation Results

I want to explain the excess female mortality split up into different age groups, counties, and time spans. This will give a panel of observations, which will be explained by time and county variables. There are no time invariant county variables in my explanatory variables list so that they all can be estimated in the framework of a panel model.

The dependent variable is excess female mortality in two age groups, 10 to 19 and 25 to 44. The estimated parameters increase women's mortality relative to men's if they are positive. In case of a positive parameter one can speak about a negative influence of the explanatory variable on women.

First I estimate a random effects panel model (which turned out to be the correct specification after having carried out a Hausman test) with only time dummies as explanatory variables. If there is a considerable variation in the rates that cannot be explained by time and random effects, I will try to explain the remaining variation in a second step. The regression results are shown in Table 5. Most age categories show that excess female mortality was rising over time, the two age groups that are chosen for the next estimation step show an R^2 of around 12%. Thus, there is a considerable part of the variation in the dependent variable that can be explained by further explanatories.

Table 6 shows the regression results when we use the explanatory variables described in chapter 7 to explain excess female mortality. I only include the age groups 10 to 19 as the girls category being below motherhood age and 25 to 44 as the group of adult women. In column (1) and (3), the longer panel including 1926 is used, in column (2) and (4), I use a shorter panel only until 1901 but with more explanatory variables. After testing with the Hausman Test, the age group 10 to 19 is estimated by panel random effects, while the age group 25 to 44 is estimated by OLS. An F-Test on the fixed effects could not reject the Null Hypothesis that all fixed effects are zero and the σ_{ui} of the random effects turned out to be zero, so a classical OLS regression applies (Greene, 1997). Again, the random effects of the first age group refer to the counties and the data are county averages or sums. Overall, the regression explains between 20 and 40 percent of the variation in the data in column (1) to (4) and ends up in explaining at least another 8% of the variance of

Excess female mortality relative to regression in Table 5.

Table 5: Panel regression random effects

	(1) EFM 10 to 14	(2) EFM 15 to 19	(3) EFM 10 to 19	(4) EFM 20 to 24	(5) EFM 25 to 44
1881	-0.101	0.065	-0.008	0.076	0.036
	(0.85)	(0.59)	(0.10)	(0.87)	(0.73)
1891	0.085	0.264*	0.168*	0.164+	0.078
	(0.71)	(2.41)	(2.22)	(1.88)	(1.56)
1901	0.111	0.359**	0.242**	0.104	0.081
	(0.93)	(3.24)	(3.17)	(1.17)	(1.60)
1926	0.266*	0.363**	0.276**	0.491**	0.253**
	(2.05)	(3.03)	(3.34)	(5.17)	(4.62)
Constant	1.320**	1.065**	1.152**	0.797**	0.958**
	(14.49)	(12.56)	(19.50)	(12.81)	(22.71)
No. of obs.	127	127	127	127	127
Counties	27	27	27	27	27
R-squared	0.0524	0.1069	0.1270	0.1971	0.1257

Absolute value of t statistics in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

Now, the estimated coefficients are compared to the hypotheses and also to results of McNay, Humphries, and Klasen (2002). They have estimated excess female mortality of the age groups 10-19 and 20-44 in England and Wales in 1860.

Overall mortality has no positive effect on women, as would have been expected from the hypotheses. A possible reason is that the times are not as extreme as famine times, so that mortality of girls is not affected. Excess mortality of women rises with higher mortality rates, maybe because they form quite a big part of the overall mortality. That contradicts our hypothesis and also the findings of McNay, Humphries, and Klasen (2002) who find at least a negative relationship in the girls category.

The next variable shows the labour market participation of women relative to men with the hypothesis that a higher labour market participation should give them more power in the struggle over resources in the household.

For adult women, the hypothesis seems to hold, excess female mortality diminishes with rising labour market participation of women. The hypothesis that girls should profit from the stronger mothers cannot be accepted, the influence is not significant.

The structure of the economy is only captured with the next two variables, workers in agriculture and in the professional sector. The left-out category is the domestic sector and industry. There seems not to be an influence of the structure of the economy on women's or girls' mortality, only agricultural work influences women negatively. McNay, Humphries, and Klasen (2002) find that employment in agriculture is detrimental to women and girls, the rest of the sectors are not different from each other. The results are consistent in the finding of a negative influence of agricultural work on women. The hypothesis was that rural regions lag behind in the equality of mortality rates between genders, our findings confirm this theory. The number of people with no job does not influence the mortality of women relative to men.

The structure of women's work is one of the most important explanatory factors, because the hypotheses are the strongest. Labour market participation per se is positive for women, as we have seen. Bad work conditions in typical women's jobs could reverse the positive result on mortality. And we see that industry work is detrimental to women's health and the positive effect of higher power is reversed by unhealthy work. Domestic work shows also a tendency to be negative for women. The interaction term between the dummy for 1871 and the domestic work is not significant, even if we have seen in the descriptive chapter that there was a huge downturn in female employment from 1871 to 1881. The professional sector does not influence excess female mortality. The influence of women's work on girls' mortality is not significant at all, while we had expected a negative influence of the mothers' work on girls if girls suffer from the bad working conditions of women. McNay, Humphries, and Klasen (2002) find a positive influence of women working in agriculture on both girls and women, while women working in manufacturing are bad for girls and women. With respect to manufacturing, the effect on women is the same in England and Ireland.

The remarriage market is proxied by the percentage of spinsters on bachelors and with a lower number, women are expected to have a higher power because of their better outside option and should have a lower excess mortal-

ity. The variable does not show a significant influence. McNay, Humphries, and Klasen (2002) proxy the remarriage market with the number of women to men aged 20 to 24 and find a negative significance for both age categories meaning a positive influence on women. That contradicts the hypothesis of the remarriage variable that should enhance women's status with a lower number of competitors on the remarriage market.

The very high emigration from Ireland could lead to either a lower or a higher excess mortality of women dependent on who leaves the country, the sick to find a better living or the only strong, because the sick were too weak to leave. The regression result shows that a higher emigration of women relative to men leads to a higher excess mortality in both age groups. Thus it seems to be the sick that have to stay.

The number of births and the two agricultural indicators do not have a significant influence on excess female mortality.

Population density is good for girls, but bad for women. To rationalise the finding it could be argued that bad housing conditions and hard work in cities has a bad influence on women, while the higher availability of schools with regular meals and health care is good for girls.

The percentage of women that are illiterate and the ratio of women to men in workhouses shows a negative influence on girls, not on women; 4th class housing is not significant.

But the question still remains, why Ireland experienced such high excess female mortality. From the comparison with England and Wales, some hypotheses can be deducted. The first factor to be different between the countries seems to be the high emigration which leaves the poorest and weakest at home. A further factor might be that poor relief seems to be more efficient in England than in Ireland. While the percentage of women dependent on poor relief lowers excess female mortality in England, the share of women relative to men in workhouses in Ireland shows a tendency towards increasing excess female mortality. Ireland has a much higher share of agricultural employment than England (52 to 33%) and in both countries, the share of agricultural work increases excess female mortality. Thus, with a much higher share, Ireland's mortality pattern can be explained. The participation rate of women in industry is not significantly different between the countries

and also the height of the influence is very similar, so the structure of female labour market participation cannot be an explanation.

9 Conclusions

The aim of the paper was to go beyond a descriptive study of why Irish excess female mortality was exceptionally high at the beginning of the twentieth century. Several studies have described the phenomenon and possible explanatory factors (e.g. Fahey (2003) or Kennedy (1973)).

Many studies have estimated the factors that influence excess female mortality and have calculated the resulting number of missing women for other countries and times. For the existence of excess female mortality, inequalities in the division of resources between women and men are needed. General findings are that sectoral changes influence women's mortality if a traditional female sector shrinks or gains in efficiency by new technologies. An example of this is the agricultural sector that changed dramatically in Europe in the 19th century.

My findings on the influence of labour market participation of women are in line with the results found earlier on England and Wales (McNay, Humphries, and Klasen, 2002) and with theories of intra-household resource allocation. I find that women's situation improves with a higher labour market participation. But, as working conditions are detrimental to health during the observed period, a higher female participation in sectors such as the industrial sector ruins the gains in power over family resources. A special Irish pattern regarding labour market participation of women cannot be read from the regression results.

The emigration out of Ireland cannot be compared to any of the earlier analyses of any country, because it was exceptionally high. The hypothesis of authors working on Irish excess female mortality that the society in Ireland cannot be regarded as a sound population are confirmed by the regression results.

With a much higher share of agricultural work in Ireland than for example in England and with a negative influence of agricultural work on women, the high excess female mortality can be explained besides the high emigration.

Table 6: Panel regression 2

	(1) EFM 10 to 19	(2) EFM 10 to 19	(3) EFM 25 to 44	(4) EFM 25 to 44
Mortality rate	0.008	-0.017	0.040**	0.014
	(0.47)	(0.77)	(3.71)	(0.91)
W to m working	0.190	1.081	-1.745+	-2.086*
	(0.14)	(0.87)	(1.98)	(2.23)
Workers in prof	0.685	0.832	-1.772	-1.332
	(0.28)	(0.33)	(1.13)	(0.79)
Workers in agr	0.395	-1.177	1.615+	1.073
	(0.29)	(0.83)	(1.86)	(1.09)
Pop. with nojob	1.105	0.699	-0.373	0.376
	(0.73)	(0.49)	(0.38)	(0.34)
W in ind.	1.013	-1.089	3.130**	2.921*
	(0.59)	(0.60)	(2.81)	(2.26)
W in dom	0.550	-0.814	2.688+	2.265
	(0.25)	(0.39)	(1.87)	(1.48)
W in dom*1871	0.137	1.404	-1.334	-0.103
	(0.07)	(0.85)	(1.08)	(0.08)
W in prof	-0.716	-5.163	4.431	1.909
	(0.15)	(0.94)	(1.41)	(0.48)
Population density	-0.252+	-0.319+	0.160+	0.079
	(1.71)	(1.65)	(1.68)	(0.62)
W to m in workh.		0.635*		0.212
		(2.39)		(1.11)
4th class housing		-0.164		0.134
		(0.15)		(0.17)
Spinst. to Bach.	-0.158	-0.295	0.088	0.303
	(0.30)	(0.54)	(0.26)	(0.81)
W to m emig	-0.045	0.656**	0.217+	0.091
	(0.23)	(3.20)	(1.74)	(0.59)
Perc. of tilled area	0.141	0.848	-0.309	-0.283
	(0.23)	(0.56)	(0.79)	(0.29)
Share of small farms	0.806	0.208	0.011	0.030

Continued on next page...

... table 6 continued

	(1) EFM 10 to 19	(2) EFM 10 to 19	(3) EFM 25 to 44	(4) EFM 25 to 44
	(0.98)	(0.22)	(0.02)	(0.05)
Women illit.		0.987*		0.333
		(2.01)		(1.03)
Births to women	-0.003	0.683	0.002	0.408
	(1.11)	(0.75)	(0.77)	(0.61)
1881	0.183	0.837	-0.775+	-0.301
	(0.26)	(1.39)	(1.70)	(0.67)
1891	0.364	0.971+	-0.684+	-0.296
	(0.58)	(1.87)	(1.67)	(0.76)
1901	0.369	0.759	-0.621	-0.361
	(0.62)	(1.53)	(1.62)	(0.97)
1926	0.673		-0.085	
	(1.22)		(0.24)	
Constant	0.135	-1.017	-0.129	-0.325
	(0.08)	(0.51)	(0.12)	(0.23)
Observations	127	107	127	107
Counties	27	27	27	27
R-squared	0.29	0.35	0.42	0.21

Absolute value of t statistics in parentheses
(1) and (2): Random effects, (3) and (4) OLS
+ significant at 10%; * significant at 5%; ** significant at 1%

References

Becker, G. S. (1964): *Human Capital*. Columbia University Press, New York.

——— (1981): *A treatise on the family*. Harvard University Press, Cambridge, M.A.

Behrmann, J. R. (1998): "Intra-Household Allocations of Resources: Is there a Gender Bias?," in *Too young to die: Genes or Gender?*, ed. by U. Nations, New York. United Nations.

Boserup, E. (1970): *Women's Role in Economic Development*. George Allen And Unwin Ltd., London.

Budd, J. W., and T. Guinnane (1991): "Intentional Age-Misreporting, Age-Heaping, and the 1908 Old Age Pensions Act in Ireland," *Population Studies*, 45(3), 497–518.

Census of Ireland (several years): *General Reports*. HMSO, Dublin.

Coale, A. J. (1991): "Excess Female Mortality and the Balance of Sexes in the Population: An Estimate of the Number of "Missing Females"," *Population and Development Review*, 17(3), 517–523.

Coale, A. J., and J. Banister (1994): "Five Decades of Missing Females in China," *Demography*, 31(3), 459–479.

Coale, A. J., P. Demeny, and B. Vaughan (1983): *Regional Model Life Tables and Stable Populations*. Academic Press, New York, London, 2nd edn.

Coleman, D. A. (1992): "The Demographic Transition in Ireland in International Context," in *The development of Industrial society in Ireland*, ed. by J. H. Goldthorpe, and C. T. Whelan, vol. 79 of *Proceedings of the British Academy*, pp. 53–77. Published for the British Academy by Oxford University Press, Oxford, The third joint meeting of the royal irish academy and the british academy Oxford 1990.

Das Gupta, M., and P. N. M. Bhat (1997): "Fertility Decline and Increased Manifestation of Sex Bias in India," *Population Studies*, 51(3), 307–315.

Dreze, J., and A. Sen (1989): *Hunger and Public Action*. Clarendon Press, Oxford.

Fahey, T. (2003): "Why did women have no life expectancy advantage over than men in early twentieth century Ireland?," mimeo, The Economic and Social Research Institute, Dublin, Ireland.

Greene, W. H. (1997): *Econometric Analysis*. Prentice Hall, New Jersey, 3rd edn.

Guinnane, T. W. (1997): *The vanishing Irish Households, Migration, and the rural economy in Ireland, 1850-1914*. Princeton University Press, Princeton, New Jersey.

Guinnane, T. W., and C. O Grada (2002): "Mortality in the North Dublin Union during the Great Famine," *Economic History Review*, LV(3), 487–506.

Hill, M. A., and E. M. King (1993): "Women's Education in Developing Countries: An Overview," in *Women's education in developing countries: Barriers, benefits and policies*, ed. by E. M. King, and M. A. Hill, Proceedings of the British Academy, pp. 1–50. Published for the World Bank, The John Hopkins University Press, Baltimore and London.

Horney, M. J., and M. B. McElroy (1988): "The Household Allocation Problem: Empirical Results from a Bargaining Model," *Research in Population Economics*, 6, 15–38.

Humphries, J. (1991): "Bread and a pennyworth of treacle: Excess Female mortality in England in teh 1840s," *Cambridge Journal of Economics*, 15(4), 451–473.

Johansson, S. R. (1991): "Welfare, mortality and gender: Continuity and change in theories about male/female mortaliy differences over three centuries," *Continuity and Change*, 6, 135–177.

Jones, G. (2001): *'Captain of all these men of death' - The History of Tuberculosis in nineteenth and twentieth century Ireland*. Rodopi, Amsterdam, New York.

Kennedy, R. E. (1973): *The Irish: Emigration, Marriage and Fertility.* University of California Press, Berkeley, Los Angeles, London.

Klasen, S. (1994): "'Missing Women' Reconsidered," *World Development*, 22, 1061–71.

——— (1998): "Marriage, Bargaining, and Intrahousehold Resource Allocation: Excess Female Mortality among Adults during Early German Development, 1740-1860," *The Journal of Economic History*, 58(2), 432–467.

——— (1999): "Gender Bias in Mortality in a Comparative Perspective," paper prepared for the German Cliometrics Conference, September 20-24, Toronto, Canada.

Klasen, S., and C. Wink (2002): "A Turning Point in Gender Bias in Mortality? An Update of the Number of Missing Women," *Population and Development Review*, 28(2), 285–312.

Kuznets, S. (1955): "Economic Growth and Income Inequality," *American Economic Review*, 45(1), 1–28.

Langford, C., and P. Storey (1993): "Sex Differentials in Mortality in the Twentieth Century: Sri Lanka and India Compared," *Population and Development Review*, 19(2), 263–282.

Luddy, M. (1999): *Women in Ireland, 1800-1918 - A Documentary History.* Cork University Press.

McElroy, M. B., and M. J. Horney (1981): "Nash-Bargained Household Decisions: Toward a Generalization of the Theory Demand," *International Economic Review*, 22(2), 333–349.

McNay, K., J. Humphries, and S. Klasen (1998): "Death and Gender in Victoran England and Wales: Comparisons with Contemporory Developing Countries," *DAE Working Paper No. 9801.*

——— (2002): "Excess Female Mortality in Nineteenth Century England and Wales: A Regional Analysis," mimeo.

Mokyr, J., and C. O. Grada (1984): "New Developments in Irish Population History, 1700-1850," *Economic History Review, 2nd series*, 37(4), 473–488.

O Grada, C. (1993): *Ireland before and after the Famine*. Manchester University Press, Manchester and New York, 2nd edn.

Preston, S. H. (1976): *Mortality Patterns in National Populations*. Academic Press, New York.

Registrar General for Ireland (several years): *Annual report of the Registrar General of marriages, births and deaths in Ireland*. His Majesty's Stationery Office, London.

Schultz, P. T. (1997): "Demand for Children in Low Income Countries," in *Handbook of Population and Family Economics, Vol. 1A*, ed. by M. R. Rosenzweig, and O. Stark, chap. 8. North-Holland, Amsterdam.

Sen, A. (1990): "Gender and Cooperative Conflicts," in *Persistent Inequalities*, ed. by I. Tinker, pp. 123–149. Clarendon Press, Oxford.

——— (1998): "Mortality as an Indicator of Economic Success or Failure," *The Economic Journal*, 108(446), 1–25.

Sen, A. K. (1989): "Women's Survival as a Development Problem," *Bulletin of the American Academy of Arts and Sciences*, 43, 14–29.

Tabutin, D. (1992): "Excess Female Mortality in Northern Africa since 1965: A Description," *Population: An English Selection*, 4, 187–207.

Thomas, D. (1990): "Intra-household resource allocation: an inferential approach," *Journal of Human Resources*, 25(4), 635–664.

Wakefield, E. (1812): *An account of Ireland, Statistical and Political, II*. London.

Walsh, B. M., and D. Walsh (1978): "The male female differential in life expectancy in Ireland - a note," *Journal of the Irish Medical Association*, 71(14), 475–80.

Woods, R. (2000): *The Demography of Victorian England and Wales*. Cambridge University Press, New York.

World Bank (2001): *Engendering Development*. Oxford University Press, Oxford, New York.

10 Appendix

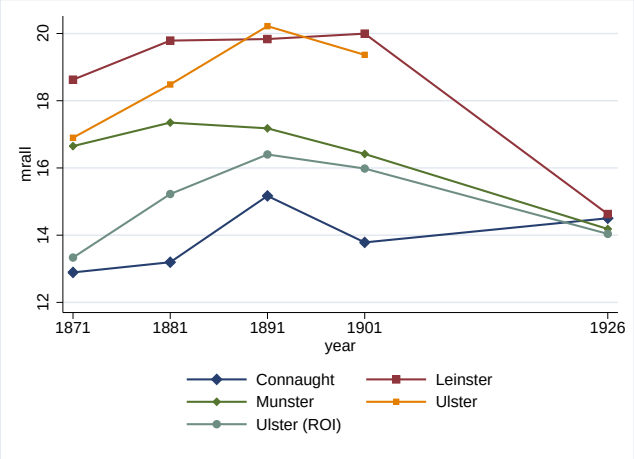

Figure 16: Overall mortality rates

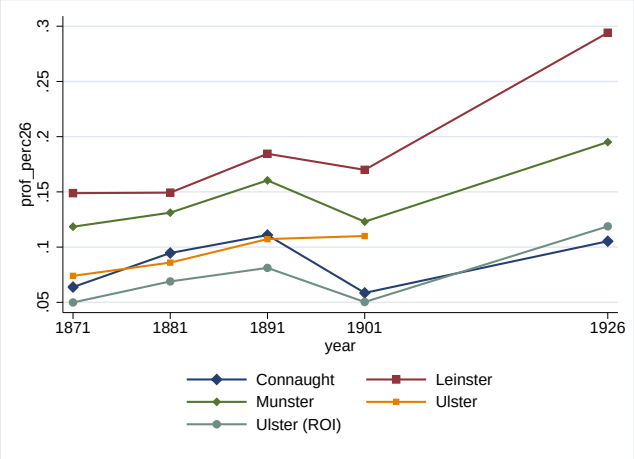

Figure 17: Percentage of Men and Women over 20 years working in the professional sector (based on census data; own calculation)

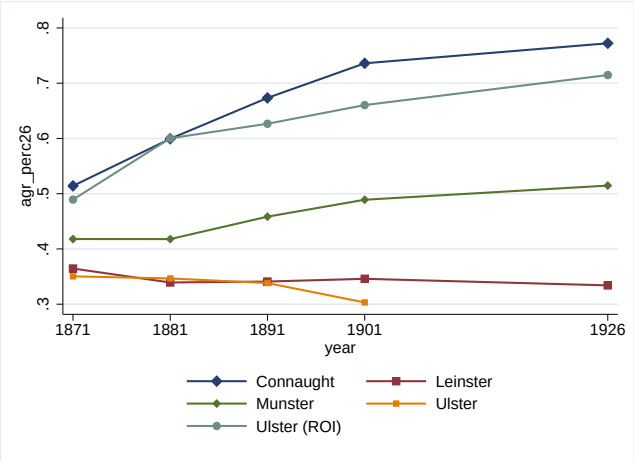

Figure 18: Percentage of Men and Women over 20 years working in agricultural sector (based on census data; own calculation)

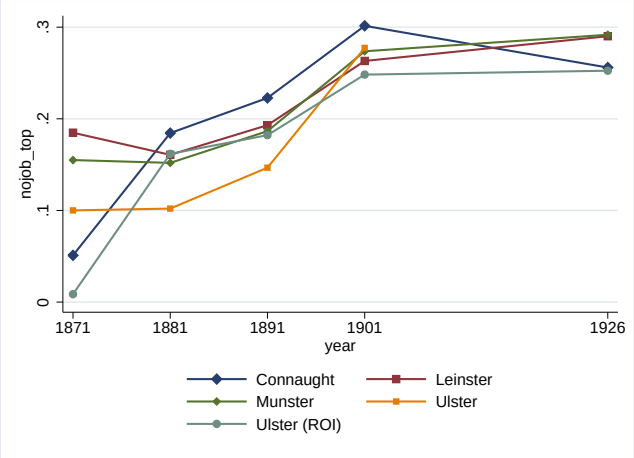

Figure 19: Percentage of Men and Women over 20 years not working rel. to population (based on census data; own calculation)

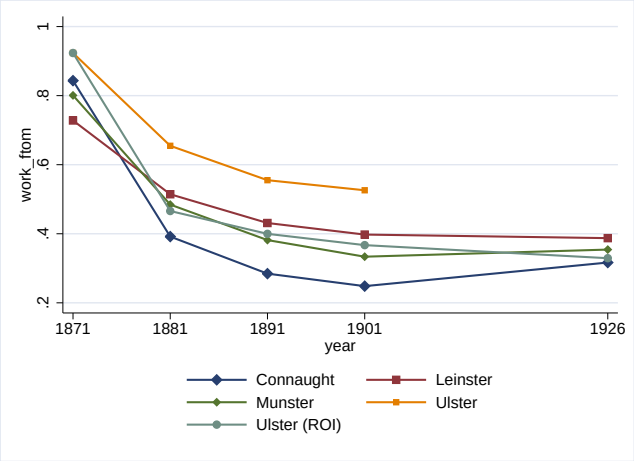

Figure 20: Women working to men (based on census data; own calculation)

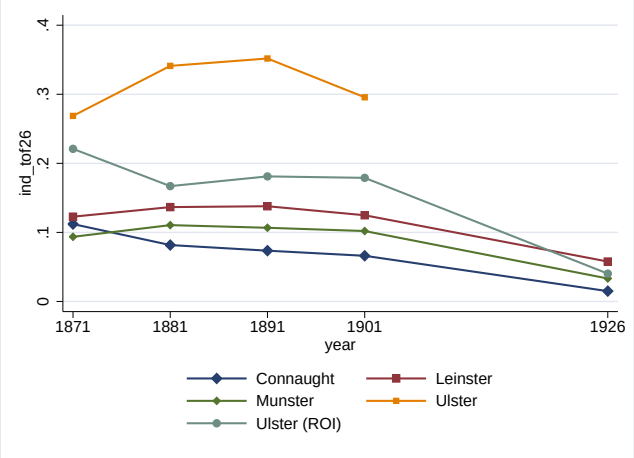

Figure 21: Share of adult women in industry (based on census data; own calculation)

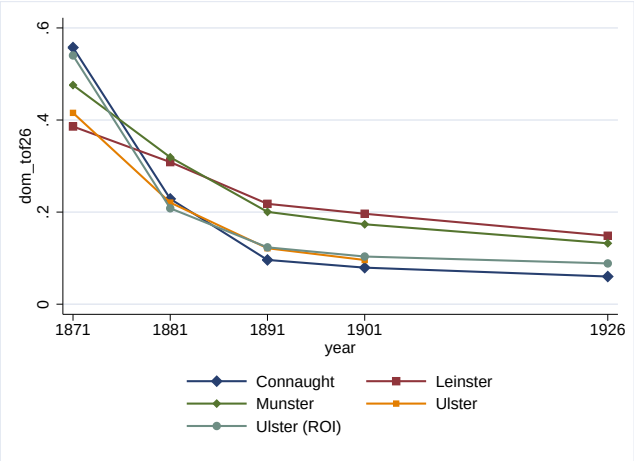

Figure 22: Share of adult women in domestic sector (based on census data; own calculation)

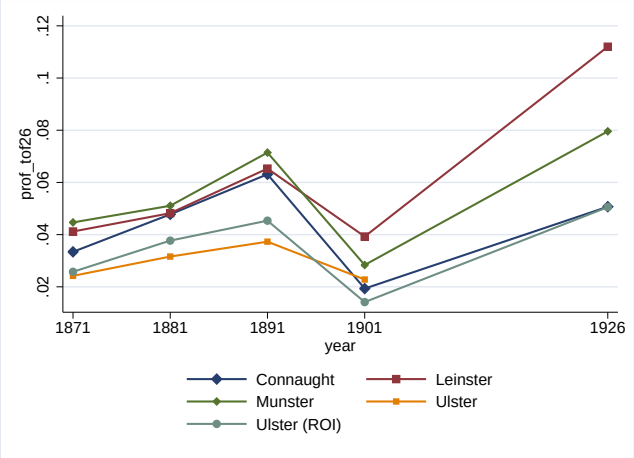

Figure 23: Share of adult women in the professional sector (based on census data; own calculation)

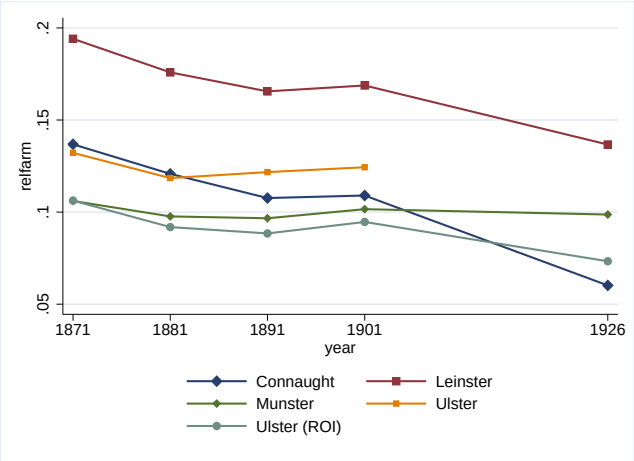

Figure 24: Small farms relative to all farms (based on census data; own calculation)

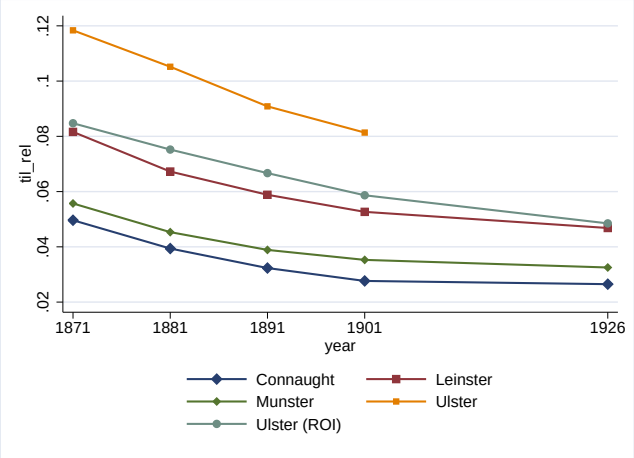

Figure 25: Share of county area cultivated with crops, fruit and horticulture (based on Irish Farm Statistics)

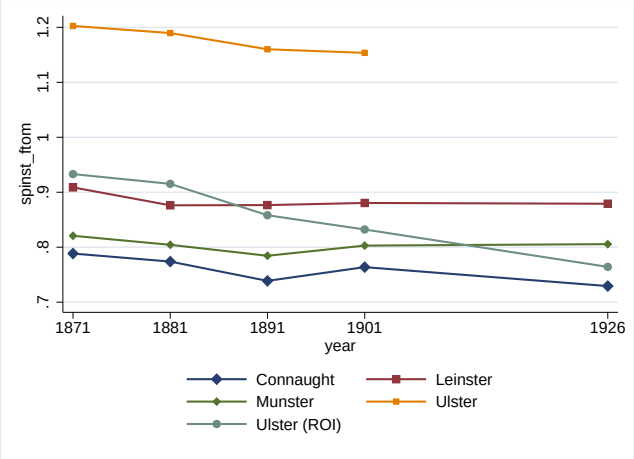

Figure 26: Spinsters rel. to bachelors (based on census data; own calculation)

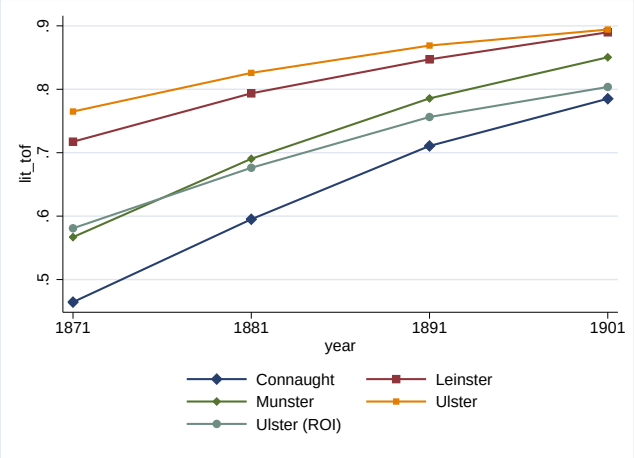

Figure 27: Share of illiterate women over five relative to all women over five (based on census data; own calculation)

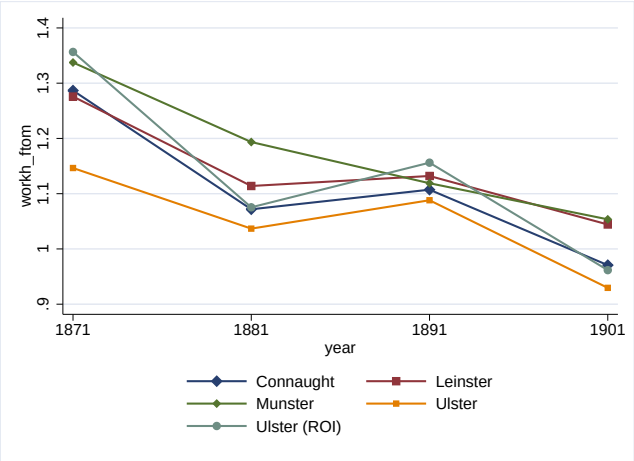

Figure 28: Women in workhouses to men in workhouses (based on census data; own calculation)

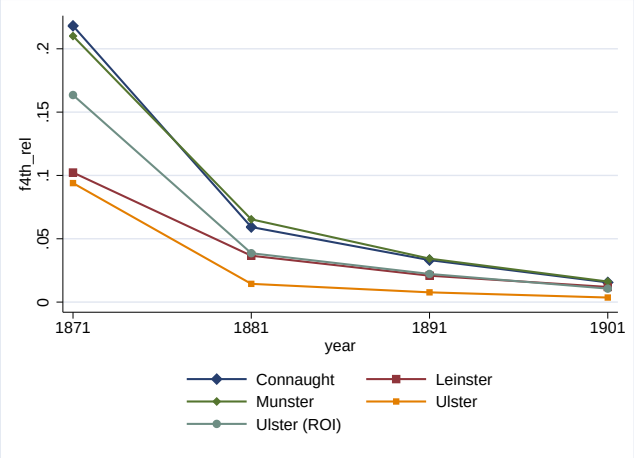

Figure 29: Share of families living in 4th class housing (based on census data; own calculation)

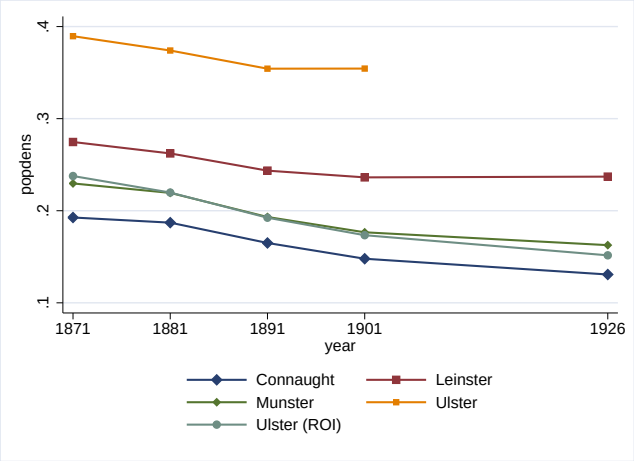

Figure 30: Population density (based on census data; own calculation)

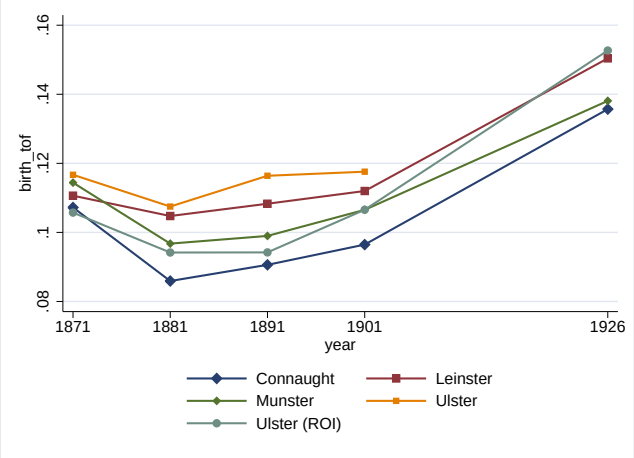

Figure 31: Number of births relative to women married women 20 to 44 (based on census data; own calculation)

Determinants of East-West Migration of Employees in Germany*

Abstract

Since the inequality of earnings in East Germany has approached West German levels in the late 1990s, the standard Roy model predicts that a positive selection bias of East-West migrants should disappear. Using a switching regression model and data from the IAB-employment sample, we find however that employed East-West migrants remain positively self-selected with respect to unobserved abilities. This result is consistent with the predictions of our extended Roy model which considers moving costs that are negatively correlated with labour market abilities of individuals. Moreover, we find that wage differentials as well as differences in employment opportunities are the central forces which drive East-West migration after unification.

*This paper has originally been coauthored by Herbert Brücker and has been called: Do the Best go West?

Contents

List of Figures

List of Tables

1 Introduction

Cumulative net migration from East to West Germany amounts to 1.3 million persons or 7.5 per cent of the original population in East Germany over the period from 1989 until the end of 2001. With that number, East Germany shows together with Albania the highest emigration rate among the countries formerly behind the iron curtain. Although net emigration rates in East Germany declined sharply after the currency union was announced, East-West migration has tended to increase again since 1996. The persistent phenomenon of East-West migration has raised increasing concerns that individuals with high abilities and qualifications migrate to the West and that this "brain drain" will contribute further to sluggish economic growth and diverging per capita income levels.

Indeed, for an understanding of the economic consequences of migration an analysis of the forces which drive the composition of the migrant population is crucial. The impact of migration on wages and employment opportunities of natives as well as the labour market performance and welfare dependency of the migrants themselves is determined by the structure of the migrant population with respect to their abilities and human capital characteristics. Thus, the self-selection of migrants has important consequences for economic growth, labour markets and the fiscal balance of the welfare state.

The self-selection of migrants has attracted increasing attention in the migration literature since the seminal paper of Borjas (1987), who applied the Roy (1951) model to the migration decision. In the Roy model, the self-selection of individuals depends essentially on the relative returns to their abilities, such that the distribution of income in the home and the host region determines the composition of migrants. As a result, migrants were positively selected at the inequality in the distribution of income is higher in the host region than in the home region – and vice versa (Borjas, 1987, pp. 551-52). Thus, the standard Roy model predicts that migrants from the East were positively selected at the beginning of the transition process, since the distribution of earnings was more equal in East Germany than in West Germany in the early 1990s. However, this positive selection bias should have disappeared over time, as the inequality of earnings in East Germany has meanwhile almost approached West German levels (see Figure 1).

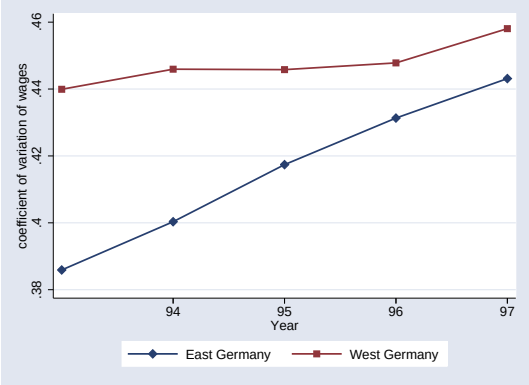

Figure 1: Wage Inequality in East and West Germany

The strong results from the standard Roy model are, however, controversial, especially because the empirical evidence provided for its predictions is not conclusive. Borjas (1987) analyses the the development of earnings of migrants in the host region, which allows no direct inference on the question of whether migrants are drawn from the upper or the lower tail of the income distribution in their home countries. Moreover, Borjas' (1987) conclusion that the "quality" of US migrants has declined over time has been questioned (Jasso and Rosenzweig, 1990). Thus, many authors argue that migrants are presumably favourably selected although the variance of earnings is higher in the home region relative to the host region.

An empirical analysis of selection bias in the context of international migration is seriously hampered by a lack of individual data. Household panels in the home countries usually do not allow international emigrants to be identified - they simply disappear from the panel. Thus, although we have rich information on immigrants in many host countries, the phenomenon of self-selection cannot be addressed systematically on the basis of standard data sources.

German unification offers a unique opportunity to study the problem of the selection of migrants. There exist several micro data sets that allow East-West migrants to be identified before and after movement. We employ the "IAB-Regionalstichprobe", which is a one per cent sample of individuals who are registered by the German social security system and includes individuals in both parts of Germany. This data set enables us to analyse whether and to what extent relative returns to abilities affect the self-sorting of East-

West migrants. The empirical framework is based on a switching regression model (Goldfeld and Quandt, 1973) with endogenous switching (Maddala and Nelson, 1975; Maddala, 1983). The switching regression model can be derived from the Roy model and was first applied in the context of migration decisions by Nakosteen and Zimmer (1984). It makes it possible not only to estimate the wage differential for (prospective) migrants consistently, but also enables us to draw inferences on the selection bias of the migrant population.

The remainder of the paper is structured as follows: firstly, we briefly review the findings of the empirical literature on East-West migration in Germany (Section 2). Secondly, we present an extended version of Roy's model which considers the correlation between moving costs and abilities of individuals in the labour market. As a result, the strong predictions of Roy's model are relaxed or even reversed (Section 3). Thirdly, we describe the data base (Section 4) and analyse the socio-economic characteristics of migrants (Section 5). Fourthly, we present the switching regression model and discuss the econometric results (Section 6). In conclusion we summarise our findings and their implications for the theory and empirics of self-selection (Section 7).

2 Empirical literature

There exists a large body of literature on East-West migration in Germany that addresses the problem of self-selection at least partially. At first glance, this literature seems to confirm the hypothesis that "those who emigrate tend to be better educated and possibly better workers in unobservable ways than stayers" (Burda and Hunt, 2001, p. 65). In her comprehensive analysis of German East-West migration, Hunt (2000) finds, based on individual data from the German Socio-Economic Panel (GSOEP), that migrants are disproportionately high skilled if controlling for age and gender. However, they are disproportionately low skilled if those control variables are not included. Hunt interprets this as evidence that, in particular, the young and high-skilled tend to move. Based on an analysis of migration intentions as revealed in the GSOEP, Burda (1993) shows that those having completed secondary schooling with an Abitur (academic secondary schooling exam) intend to move over proportionally, while those with a university or other

tertiary education degree intend to move less frequently than other education groups (Burda, 1993, p. 460). Similar results can be found in Burda, Härdle, Müller, and Werwatz (1998). In contrast, the results of Schwarze (1996) indicate that years of education are positively correlated with migration intentions as well as with actual migration. Analogously, Pischke, Staat, and Vögele (1994) show on basis of data from the "Arbeitsmarktmonitor" that East-West commuters possess higher skill levels than stayers.

Moreover, the results in the empirical literature regarding the impact of the wage differential on the propensity to migrate are ambiguous. While Schwarze (1996) finds that the wage differential affects individual migration probabilities positively, the results of Hunt (2000) and Burda, Härdle, Müller, and Werwatz (1998) indicate that no unambiguous correlation between the wage differential and the propensity to migrate exists. Burda, Härdle, Müller, and Werwatz (1998) interpret the non-linear relation between the income differential and migration probabilities in their estimates as evidence for the option value of waiting theory of migration. However, to the best of our knowledge, no paper on the determinants of East-West migration considers the self-selection of migrants. Thus, the results might be affected by selection bias.

Altogether, the results of the empirical literature regarding the question whether migrants from East Germany are positively selected are not conclusive. Furthermore, the empirical literature on the determinants of East-West migration does not consider the selection of migrants with regard to unobservable abilities in the labour market, which may bias their results.

3 Theoretical framework

The self-selection of migrants is affected by various factors since both the benefits as well as the costs of migration are not equally distributed across the population. The Roy model offers a rigourous and theoretically powerful framework to analyse the self-selection of individuals. According to the Roy model, self-selection is driven by comparative advantage of individuals. As a consequence, the distribution of income in the host and the home region determines whether individuals with higher or lower abilities tend to migrate: if the distribution of income in the host region is more equal than in the home

region, and if the correlation between the incomes of (potential) migrants in both locations is positive, migrants are chosen from the lower tail of the income distribution and vice versa (Borjas, 1987, pp. 551-52).

In its original formulation, the Roy model does not consider any switching costs. In the context of migration, pecuniary and non-pecuniary moving costs are, however, an important factor which cannot be ignored in the analysis of the migration decision. Moreover, it is reasonable to assume that abilities relevant for the labour market performance of individuals and moving costs are negatively correlated, i.e. that the same human capital characteristics which yield higher returns in the labour market allow individuals to better reduce moving costs. In this more general framework, migrants may be chosen from the upper tail of the income distribution although the distribution of income in the host region is more equal than in the home region.[1]

In what follows, we discuss first the mechanics of an extended Roy model which considers the correlation between labour market abilities and moving costs in order to derive the framework for the empirical analysis and then present the estimation model.

3.1 An extension of the Roy model

Suppose that w_1 is the wage of residents in the home region (region 1), and w_2 the wage of residents in the host region (region 2). Assume that log wages in region 1 and region 2 have a joint normal distribution, such that

$$\ln w_1 = \mu_1 + \varepsilon_1, \tag{1}$$

where μ_1 is the mean of the log wage in region 1 and ε_1 a normally distributed disturbance with zero mean and variance σ_1^2. Analogously,

$$\ln w_2 = \mu_2 + \varepsilon_2, \tag{2}$$

where ε_2 is normally distributed with zero mean and variance σ_2^2. The Roy model focuses on the impact of selection bias on the disturbances ε_1 and ε_2,

[1] A similar point has recently been made by Chiswick (2000): He demonstrates in a numerical example that the strong implications of the Roy model are relaxed if fixed moving costs are considered.

which can be interpreted as (unobservable) abilities of individuals.

The original Roy model ignores all switching costs, i.e. an individual from region 1 migrates into region 2 if $w_2 > w_1$. However, it is reasonable to assume that moving costs exist and that they are related to human capital characteristics and other abilities of individuals. Suppose that C represents the pecuniary and non-pecuniary costs of migration as a proportion of home income. Migration occurs if $\frac{w_2 - w_1}{w_1} > C$, or, approximately, if $\ln w_2 - \ln w_1 > C$. Assume that C is normally distributed with mean τ and disturbance η, i.e.

$$C = \tau + \eta, \tag{3}$$

and that $\eta \sim N(0, \sigma_\eta^2)$. The decision to migrate is then determined by the sign of the index function, I^*, which contains the wage gain from moving minus the costs of migration:

$$I^* = \mu_2 - \mu_1 - \tau + \varepsilon_2 - \varepsilon_1 - \eta, \tag{4}$$

i.e. an individual migrates if $I^* > 0$, and stays at home if $I^* \leq 0$.

Define

$$\sigma^* = \sqrt{\mathrm{Var}(\varepsilon_2 - \varepsilon_1 - \eta)}, \quad z = -\frac{\mu_2 - \mu_1 - \tau}{\sigma^*}, \quad \text{and} \quad \epsilon = \frac{\varepsilon_2 - \varepsilon_1 - \eta}{\sigma^*}.$$

Migration occurs if $\epsilon > z$. Under the normality assumptions, the migration rate m is given by

$$m = Pr(\epsilon > z) = 1 - \Phi(z), \tag{5}$$

where $\Phi()$ is the cumulative distribution function of the standard normal. Using the standard sample selection formula (Heckman, 1976, 1979), the (unobserved) wage of a migrant in the home region can be written as

$$E(\ln w_1 | I^* > 0) = \mu_1 + \sigma_{1\epsilon}\lambda(z), \tag{6}$$

and the observed wage in the host region as

$$E(\ln w_2 | I^* > 0) = \mu_2 + \sigma_{2\epsilon}\lambda(z), \tag{7}$$

where $\sigma_{1\epsilon}$ and $\sigma_{2\epsilon}$ are the covariance of ε_1 and ϵ, and the covariance of ε_2 and ϵ, respectively, and

$$\lambda(z) = \frac{\phi(z)}{1 - \Phi(z)}$$

is the inverse of Mills' ratio and $\phi()$ the density of the standard normal.

Whether migrants are better or worse off than the average person in the home and the host region depends on the sign of the second term in the equations (6) and (7). Since $\lambda(z) \geq 0$ by definition, the average migrant is better off than the average person in the home region if $\sigma_{1\epsilon} > 0$, and, analogously, better off than the average person in the host region if $\sigma_{2\epsilon} > 0$ – if we ignore the limiting case that $\lambda(z) = 0$.

An interpretation of these conditions requires that we decompose $\sigma_{1\epsilon}$ and $\sigma_{2\epsilon}$. Using the definition for the covariance, we can rewrite $\sigma_{1\epsilon}$ as

$$\sigma_{1\epsilon} = \frac{\sigma_{12} - \sigma_1^2 - \sigma_{1\eta}}{\sigma^*},$$

and $\sigma_{2\epsilon}$ as

$$\sigma_{2\epsilon} = \frac{\sigma_2^2 - \sigma_{12} - \sigma_{2\eta}}{\sigma^*}.$$

Thus, we can derive two fundamental conditions for the favourable self-selection of migrants: firstly, migrants are better off than the average person in the home population if $\sigma_{12} > \sigma_1^2 + \sigma_{1\eta}$, or if

$$\frac{\sigma_2}{\sigma_1} > \frac{1}{\rho_{12}} + \frac{\rho_{1\eta}}{\rho_{12}} \frac{\sigma_\eta}{\sigma_1}, \tag{8}$$

where ρ_{12} is the correlation coefficient between ε_1 and ε_2, and $\rho_{1\eta}$ the correlation coefficient between ε_1 and η. We assume for the further analysis that $\rho_{12} > 0$, since a negative correlation between earnings in both regions makes no sense economically. Note that the second term on the right-hand side captures the correlation between labour-market abilities and moving costs. Since we assume that labour-market abilities and moving costs are negatively correlated, i.e. that $\rho_{1\eta} < 0$, the second term is negative, and, hence, increases the probability of a favourable selection of migrants relative to the average person in the home population for a given variance of earnings in the host and the home region.

Secondly, the migrant is better off than the average person in the host region if $\sigma_2^2 > \sigma_{12} + \sigma_{2\eta}$, or if

$$\frac{\sigma_2}{\sigma_1} > \rho_{12} + \rho_{2\eta}\frac{\sigma_\eta}{\sigma_1}, \tag{9}$$

where $\rho_{2\eta}$ is the correlation coefficient between ε_2 and η. Once again, since we assume that $\rho_{2\eta} < 0$, the second term on the right-hand side increases the probability of a favourable selection of migrants relative to the average person in the host population for a given variance of earnings in the host and the home region.

3.2 Comparative Statics

Consider now the implications of the model for a change in the economic conditions underlying the (self-)selection of migrants. We can write the selection bias of migrants relative to the average person in the home population as

$$S_j = S_j\left(\omega, C, \sigma_1, \sigma_2, \eta, \rho_{12}, \rho_{1\eta}, \rho_{2\eta}\right), \quad j \in \{1, 2\}$$

where $\omega \equiv \mu_2 - \mu_1$ is the income difference between the host and the home region. The second terms in equations (6) and (7) show that the selection bias in the home region is given by

$$S_1 = \sigma_{1\epsilon}\lambda(z),$$

and in the host region by

$$S_2 = \sigma_{2\epsilon}\lambda(z).$$

We can thus write the impact of a change in any variable x on the change in S_1 and S_2 as

$$\frac{\partial S_1}{\partial x} = \frac{\partial \sigma_{1\epsilon}}{\partial x}\lambda + \frac{\partial \lambda}{\partial x}\sigma_{1\epsilon}, \tag{10}$$

and as

$$\frac{\partial S_2}{\partial x} = \frac{\partial \sigma_{2\epsilon}}{\partial x}\lambda + \frac{\partial \lambda}{\partial x}\sigma_{2\epsilon}. \tag{11}$$

The first term on the right-hand side in equations (10) and (11) captures the composition effect for a constant scale of migration, and the second term the scale effect for a given composition of the migrant population (Borjas, 1987).

We focus here on the selection bias of migrants relative to the average person in the home region. Define $k = \sigma_2\rho_{12} - \sigma_1 - \sigma_\eta\rho_{1\eta}$. k has a positive sign if $\frac{\sigma_2}{\sigma_1} > \frac{1}{\rho_{12}} + \frac{\rho_{1\eta}}{\rho_{12}}\frac{\sigma_\eta}{\sigma_1}$, i.e. if migrants are positively selected, and a negative one, if otherwise.

Consider first the impact of a change in the difference of earnings between the host and the home region. Using equation (10) it can be shown that

$$\frac{\partial S_1}{\partial \omega} = -\frac{\sigma_1}{\sigma^{*2}}\frac{\partial \lambda}{\partial z}\, k, \tag{12}$$

i.e. a change in the income differential affects the composition of migrants only via the scale effect. An increase in the difference of earnings between the host and the home region reduces the positive (negative) selection bias of the migrant population if they are positively (negatively) selected. The intuition behind this result is that a higher difference in earnings increases the share of migrants in the population, which in turn reduces the selection bias in both directions, since migrants are increasingly drawn from the mean parts of the income distribution.

Increasing the mean costs of migration has the opposite effect, i.e.

$$\frac{\partial S_1}{\partial C} = \frac{\sigma_1}{\sigma^{*2}}\frac{\partial \lambda}{\partial z}\, k, \tag{13}$$

since increasing moving costs reduces the share of migrants in the population, which in turn increases the selection bias of the migrant population.

The impact of a change in the inequality of earnings on the selection bias is ambiguous. The derivation of S_1 with respect to σ_1 yields

$$\frac{\partial S_1}{\partial \sigma_1} = \frac{2\sigma_1\, k^2 \;-\; (\sigma_1 - k)\, \sigma^{*2}}{\sigma^{*3}}\,\lambda + \frac{\sigma_1 k^2}{\sigma^{*3}}\frac{\partial \lambda}{\partial z}\, z, \tag{14}$$

where the sign of the first term – the composition effect – is positive if $2\sigma_1 \left(\sigma_1 - \rho_{12}\sigma_2 + \rho_{1\eta}\sigma_\eta\right)^2 > \left(2\sigma_1 - \rho_{12}\sigma_2 + \rho_{1\eta}\sigma_\eta\right)\sigma^{*2}$, which depends on the value of the parameters. The impact of the second term – the scale effect – depends on the sign of z. If the net difference in mean earnings (incl. moving costs) is positive (i.e. $z < 0$), the scale effect is negative, and positive if otherwise. Intuitively, with a higher inequality of earnings in the home region, the incentives to migrate for those who are better off decline, while

those at the lower tail of the income distribution still want to migrate, such that the migrant population becomes increasingly negatively selected with an increasing inequality of income in the home region.

The effect of an increasing inequality of earnings in the host region is again ambiguous. Analogously to equation (14), a derivation of S_2 with respect to σ_2 gives

$$\frac{\partial S_1}{\partial \sigma_2} = \frac{\sigma_1 \left(\rho_{12}\, \sigma^{*2} \;-\; k\, n\right)}{\sigma^{*3}}\, \lambda \;-\; k\, n\, \frac{\sigma_1}{\sigma^{*3}}\, \frac{\partial \lambda}{\partial z}\, z, \tag{15}$$

where n is defined as $\sigma_2 + \rho_{12}\sigma_1 - \rho_{23}\sigma_\eta > 0$. The composition effect has a positive sign if $\rho_{12}\sigma^{*2} > kn$, which is always the case if a negative selection bias of the migrant population exists. In the converse case, the sign of the composition effect depends on the sign of the individual parameters. The scale effect is positive if migrants are positively selected *and* the net difference in earnings between the host and the home region is positive (i.e. $z < 0$), and negative in the converse case. Thus, an increase in the inequality of earnings in the host region strengthens a given selection bias in both directions via the scale effect if the difference in net earnings is positive, while it reduces a negative selection bias via the composition effect, and is ambiguous if a positive selection bias exists.

Finally, we can assess the implications of a change in the correlation coefficients. The derivation of the change in S_1 with respect to a change in the correlation coefficients is given by

$$\frac{\partial S_1}{\partial \rho_{12}} = k\, \frac{\sigma_1^2 \sigma_2}{\sigma^*}\, \lambda \;-\; k\, \frac{\sigma_1^2 \sigma_2}{\sigma^{*3}}\, \frac{\partial \lambda}{\partial z}\, z, \tag{16}$$

$$\frac{\partial S_1}{\partial \rho_{1\eta}} = k\, \frac{\sigma_1^2 \sigma_\eta}{\sigma^*}\, \lambda \;-\; k\, \frac{\sigma_1^2 \sigma_\eta}{\sigma^{*3}}\, \frac{\partial \lambda}{\partial z}\, z, \tag{17}$$

and

$$\frac{\partial S_1}{\partial \rho_{2\eta}} = k\, \frac{\sigma_1 \sigma_2 \sigma_\eta}{\sigma^*}\, \lambda \;-\; k\, \frac{\sigma_1 \sigma_2 \sigma_\eta}{\sigma^{*3}}\, \frac{\partial \lambda}{\partial z}\, z. \tag{18}$$

In all three equations, the composition effect and the scale effect have the same sign if the net difference in mean earnings (incl. moving costs) between the host and the home is positive ($z < 0$), and the converse sign if the net difference in mean earnings is negative ($z > 0$). Thus, an increasing (positive) correlation between earnings in the home and the host region strengthens the selection bias both via the composition effect and the scale effect if the

net difference in earnings is positive. In contrast, an increasing (negative) correlation between labour market abilities and moving costs weakens the selection bias if the net difference in earnings is positive.

To sum up, the mechanics of the enhanced Roy model demonstrates that (i) a higher variance of earnings in the home region relative to the host region does not necessarily yield a negative selection bias of the migrant population, (ii) a positive selection bias is more likely to occur if we consider the correlation between labour market abilities and moving costs, (iii) increasing the difference in net earnings between the home and the host region reduces the selection bias of the migrant population, (iv) increasing the (positive) correlation between earnings in both regions strengthens the selection bias, and (v) increasing the (negative) correlation between earnings and moving costs weakens the selection bias. Increasing the inequality of earnings in the home as well in the host region has an ambiguous effect on the selection bias. The scale effect weakens the selection bias if the inequality of home earnings increases and if the net difference in earnings is positive, and strengthens the selection bias if the inequality of host earnings increases. The sign of the composition effect depends on the value of the parameters of the model.

Thus, in the case of East-West migration in Germany, different forces may have affected the selection bias of the migrant population in different directions: first, the increasing inequality of earnings in the East may have reduced the incentives to migrate for those at the upper tail of the income distribution via the scale effect, and, hence, may have reduced the positive selection bias. Second, the convergence of wages between East and West Germany may have resulted in less migration, and, hence, increased the positive selection bias. Third, reduced moving costs may have increased migration and thus reduced the positive selection bias. Altogether, the increasing inequality of earnings in East Germany does not necessarily negatively affect the selection of migrants with regard to their abilities.

3.3 Estimation

The Roy model as described above can be considered a switching regression model (Goldfeld and Quandt, 1973) with endogenous switching (Maddala and Nelson, 1975; Maddala, 1983).

Rewrite the wage equations in (1) and (2) as

$$\ln w_{1i} = X_{1i}\beta_1 + \varepsilon_{1i}, \tag{19}$$

and

$$\ln w_{2i} = X_{2i}\beta_2 + \varepsilon_{2i}, \tag{20}$$

where X_i is a vector of personal variables which is observed for each individual i. In the empirical application, we also consider regional and sectoral control variables. Suppose that the index function for the ith individual is given by

$$I_i^* = \delta(\ln w_{2i} - \ln w_{1i}) - Z_i\gamma - \eta_i, \tag{21}$$

where δ denotes a coefficient, $Z_i\gamma + \eta_i = C_i$ the migration cost, and Z_i again a vector of regional and personal variables. Note that the coefficient δ implies that migration is a log-linear function of the income difference, i.e. the functional form which we employ here for estimation differs slightly from the theoretical model. Identification of the model requires that at least one variable in Z_i is not included in the vector X_i.

It is obvious from the discussion of the Roy model that the index function cannot be estimated in structural form since the term $\ln w_{2i} - \ln w_{1i}$ is endogenous. Following Lee (1978) and Willis and Rosen (1979) the model can be estimated in three steps. In the first step we estimate a reduced form of the migration function. The reduced form of the index function I_i^* is given by

$$I_i^* = \delta(X_{2i}\beta_2 - X_{1i}\beta_1) - Z_i\gamma + \delta(\varepsilon_{2i} - \varepsilon_{1i}) - \eta_i = Z_i^*\gamma^* - \epsilon^*, \tag{22}$$

where Z_i^*, γ^* and ϵ^* are defined suitably. Define $I_i = 1$ if $I^* > 0$ and $I_i = 0$ otherwise. Based on the observations on I_i we can then use the probit Maximum Likelihood estimator in the first step to obtain a consistent estimate for the vector $\widehat{\gamma}^*$.

Wages in region 1 can be observed only for those individuals for whom $I_i = 0$, and wages in region 2 only for those individuals for whom $I_i = 1$. Estimating the wage equations therefore requires correction for this selection bias. Using the estimated vector of parameters $\widehat{\gamma}^*$, we can compute the inverse Mills' ratio for migrants and stayers as $\frac{\phi(Z_i^*\widehat{\gamma}^*)}{1-\Phi(Z_i^*\widehat{\gamma}^*)}$ and $\frac{\phi(Z_i^*\widehat{\gamma}^*)}{\Phi(Z_i^*\widehat{\gamma}^*)}$, re-

spectively. Under the normality assumptions, this allows us to correct for selection bias and to estimate in the second step the wage equations for stayers in the home region and movers to the host region by OLS:

$$\ln w_i = X_i\beta_1 - \sigma_{1\epsilon^*}\frac{\phi(Z_i^*\widehat{\gamma}^*)}{\Phi(Z_i^*\widehat{\gamma}^*)} + u_{1i}, \quad \text{for} \quad I_i = 0, \tag{23}$$

and

$$\ln w_i = X_i\beta_2 + \sigma_{2\epsilon^*}\frac{\phi(Z_i^*\widehat{\gamma}^*)}{1-\Phi(Z_i^*\widehat{\gamma}^*)} + u_{2i}, \quad \text{for} \quad I_i = 1, \tag{24}$$

which gives us consistent estimates of β_1, β_2, $\sigma_{1\epsilon^*}$, and $\sigma_{2\epsilon^*}$. Furthermore, it is possible to derive consistent estimates for σ_1^2 and σ_2^2 from the residuals of the wage equations and estimated parameters (see Maddala, 1983, pp. 225-26). In our empirical application, we estimated the reduced form probit model and each of the two Heckman selection equations in one step using a maximum likelihood estimator (Greene, 1997). The ML function uses the estimated parameters from the reduced form probit model as starting values for the estimation of the Heckman corrected wage equations.

In the final step, we again use the probit model to estimate the structural equation and to obtain a consistent estimate of δ, the coefficient of the wage differential. Substituting $\ln \widehat{w}_{1i} = X_{1i}\widehat{\beta}_1$ and $\ln \widehat{w}_{2i} = X_{2i}\widehat{\beta}_2$ for $\ln w_{1i}$ and $\ln w_{2i}$, respectively, allows us to estimate the structural probit equation. As Lee (1979) has demonstrated, the resulting estimates of γ and δ are consistent.

4 Data

We perform our empirical analysis of the self-selection of East-West migrants in Germany using individual data from the "IAB-Regionalstichprobe".[2] This data set contains a one per cent sample of all the returns of the social security files of Germany, collected by the Federal Employment Services (Bundesanstalt für Arbeit). The East German sample starts at the beginning of 1992 and the last spells are reported for 1997.

The sample covers employed persons, unemployed persons and individuals

[2]Employee sample, regional file. The IAB-Regionalstichprobe is provided by the German Institute for Employment Research (IAB) at the Federal Employment Services (Bundesanstalt für Arbeit). See Haas (2001) for a brief introduction.

who are currently taking a break from employment. Self-employed persons and those who are enrolled in educational programs are not included. Moreover, the sample is censored from above, i.e. individuals whose earnings exceed the rather high ceiling for contributions to the public pension scheme and unemployment insurance in Germany are not reported.[3] In 1995, 86.2% of the economically active population was captured by the social security files in East Germany (Bender, Haas, and Klose, 2000, p. 3).

The observations of each individual are organised as event data. Every change in the employment situation is collected with the date of its event, but also every year a control return is registered. For each individual, work history, personal characteristics, firm characteristics and regional details are collected. We choose only individuals who are employed full-time on 31 March. The employment state on 31 March of every year is used to transform the event-oriented data into a panel of yearly observations.

It is not trivial to identify East Germans in the data set. For the purpose of our analysis, we define East Germans as follows: (i) they have their first spell in 1992 or later, since East Germans were not included in the IAB sample before; (ii) they are registered in an East German pension scheme[4] if they are employed, and (iii) they work at a company in East Germany if they are employed. These definitions imply that we do not include those East Germans in our analysis who have migrated to West Germany before 1992.

On the basis of these definitions, we distinguish two groups of individuals: *stayers* are all individuals that have been registered as employees or unemployed in East Germany for the whole time span from 1992 to 1997 and all individuals who are registered in East Germany, but will later move to West Germany during the observed time period; *movers* are individuals who are in a spell in the West for the first time, i.e. those who moved during the previous year. Since our regressions are based on a cross-sectional analysis, we do not consider the spells of East Germans who have been residing in West Germany for more than two years. By definition, our sample contains observations for stayers over the whole period from 1992 to 1997, for movers

[3]The ceiling was 5,300 DM in 1992 and 7,100 DM in 1997, while the mean incomes in our sample amount to 2,695 and 3,097 for the two years.

[4]Landesversicherungsanstalt (LVA) Ost, Bundesversicherungsanstalt für Angestellte (BfA) Ost, or Knappschaft.

Table 1: Number of observations

year	movers	stayers	perc. of movers
1994	428	33038	1.30
1995	380	33677	1.13
1996	389	32418	1.20
1997	364	30878	1.18

only from 1993 to 1997.

Table 1 shows that the number of individuals in our sample is slightly declining. Some of the individuals may vanish from the sample due to death, international migration, leaving the labour force or becoming unemployed. Moreover, those who reside in West Germany for at least the second year vanish from the sample. Numerically, the highest number of movers is achieved in 1994. In relative numbers, the share of movers diminishes only slightly from 1.3% in 1994 to 1.2% in 1997.

5 Characteristics of East-West migrants

The following graphs provide some initial insights into the socio-economic characteristics of East-West migrants. Summary statistics for all the variables used in our regressions for the individual years can be found in Table 11.

Figure 2 displays the development of the mean log of wages per day in DM, divided into three groups: stayers and movers as defined above, and a third category, called *prospective movers*. The last category contains all individuals who are still in East Germany, but will move during the observation period. We include this third group here because the difference between prospective movers and stayers is striking: wages of prospective movers start well below those of stayers, although they receive much the same wage increase as stayers. Note that the lower wage level may reflect the difference in the age of stayers and migrants (see below). However, movers already working in West Germany receive wages above those of stayers in East Germany.

The descriptive statistics of the education variables seem to indicate that East-West migrants are slightly less skilled than the average person in East

Figure 2: Mean wage

Germany. We distinguish three groups by their highest education degree: those who possess no vocational education degree, those who possess a vocational education degree, and those who possess a degree from a university or a university of applied sciences.[5] Figure 3 displays the share of individuals without vocational education degree in the groups of movers and stayers. The share of unskilled is, at around seven per cent, relatively low in the group of stayers, and shows a very strong decline in the group of movers (from around 15 to below 10 per cent).

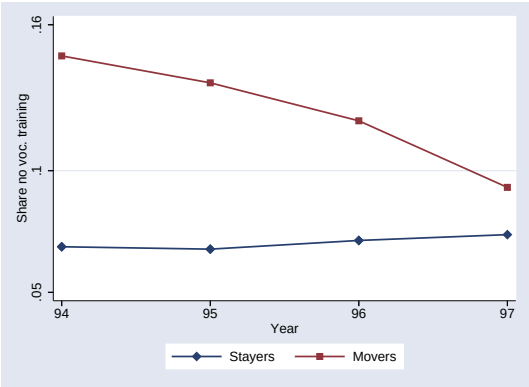

Figure 3: Mean of no vocational training degree

Figure 4 shows the individuals with completed vocational training, which is the most common education group in East Germany. The share of individuals with completed vocational training is persistently high at around 74

[5] A fourth group, unknown or missing education, is used in the regressions, but we do not display it here.

per cent among the group of stayers, while the share in the group of movers is rising from 60 to 70 per cent, with a spike in 1996.

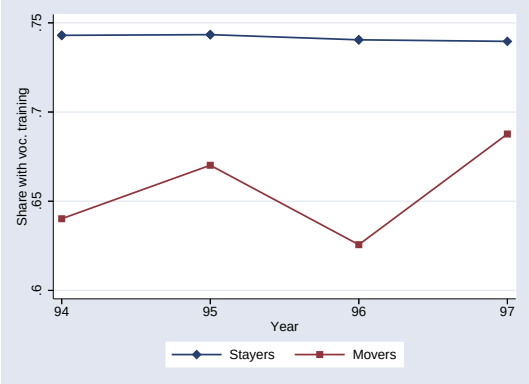

Figure 4: Mean of vocational training degree

Finally, we observe that the share of individuals with an academic education degree is lower in the group of movers than in the group of stayers. Note that, at around 11 per cent, the share of individuals with an academic degree is relatively low in the East German population. While the share of individuals with an academic degree is stable among the group of stayers, we observe that it is increasing from 1995 onward (Figure 5). It is moreover worth noting that our sample covers only individuals who have already participated in the East German labour market, i.e. students and those with an academic degree who take up their first job in West Germany are not considered here. Moreover, individuals with a wage above the ceiling of the social security records are not covered. These exclusions can heavily bias the results with regard to high-skilled jobs.

To summarise, we observe that employed migrants are less skilled than the sample average.

The difference in lagged unemployment rates of movers and stayers is relatively stable with a difference of 10 to 15 percentage points, i.e. unemployment seems to influence the migration decision strongly. Note that the lagged unemployment rate is underreported in our sample, since we excluded those who are unemployed in the present year. Thus, long-term unemployed are not covered by our sample.

Over the whole time span, the mean age of movers is, at round 34 years,

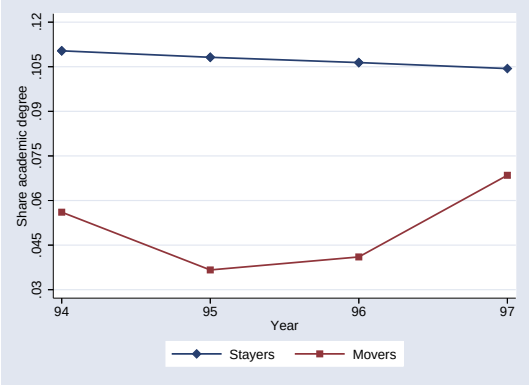

Figure 5: Mean of academic degree

persistently five years below the mean age of stayers (39 years). The age increase is less than proportional, because young people are allowed to enter the sample after 1992 if they start working in East Germany. Altogether, the descriptive statistics confirm the hypothesis of the human capital theories of migration that young people have a higher propensity to move.

The other personal characteristics show the following pattern (not displayed here): at the sample mean, the share of males among movers varies between 62 and 71 per cent, while the share of males among the stayers is constant at around 57 per cent. Moreover, there is a persistent difference in the marital status of migrants and stayers: while around 57 per cent of the stayers are married, only around 40 per cent of the movers are. Once again, these results confirm a standard hypothesis from the human capital theory of migration, i.e. that family ties affect the costs of migration.

6 Regression results

As has been outlined above, the estimation of the switching regression model consists of three steps: firstly, we estimate a reduced form probit model in order to obtain a consistent estimate of the individual probability to move. The results from this estimate are used as starting values for a maximum-likelihood estimation of a Heckit selection model of the wage equations for East Germans in West and East Germany. In the final step, consistent estimates of the parameters in the wage equations are used for estimating a

structural probit model which includes the estimated potential wage differential for each individual.

The explanatory variables are derived from the human capital theories of migration (Sjaastad, 1961): Beyond the (expected) differential in wages it is assumed that personal characteristics such as age, family ties and education affect the costs and returns to migration, and, hence, the decision to migrate. Moreover, following the traditional approach of Harris and Todaro (1970), we assume that employment opportunities affect expected earnings.

The wage regressions have the traditional Mincer form, i.e. log wages are explained by human capital variables such as education, age brackets, gender, and by occupational status. We do not include marital status in the wage regressions, since this variable is one of those used to identify the model. This is possible because family status turned out not to be significant in determining gross wages and should also not be correlated with the error term of the wage regression.

Furthermore, branch dummies are included. In order to account for regional and branch differences in economic prospects, we also include dummies for the East German Federal States (Bundesländer) and branches. We do not include the West German Bundesländer because we can not construct a counterfactual for stayers, had they moved. The counterfactual Bundesland of movers, had they stayed, is their Bundesland of origin.

Thus, in the wage regressions for East Germany, dummies for the home region (Bundesland), dummies for the home industry branch and the lagged unemployment rate are included. In the decision equation, we include the lags of most variables in order to account for the fact that the decision to migrate was made the year before, which implies that expectations have been formed on basis of the explanatory variables of the past year.

6.1 Results from the Heckit regressions

We estimate four cross-sections for the years 1994 - 1997. The results of the Heckman selection ('Heckit') estimates are reported in Tables 3 to 6. The coefficients of λ as reported in the tables are the covariances $\sigma_{1\epsilon^*}$ and $\sigma_{2\epsilon^*}$, as defined in the wage regressions for stayers and movers (Equations 6 and 7).

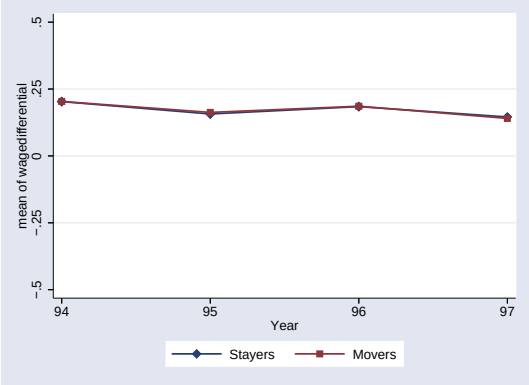

Figure 6: Mean of the wage differential with linear prediction $x_j b$

Note that $\sigma_{1\epsilon^*} = \rho_{1\epsilon^*}\sigma_1$ and that $\sigma_{2\epsilon^*} = \rho_{2\epsilon^*}\sigma_2$, where $\rho_{1\epsilon^*}$ and $\rho_{2\epsilon^*}$ are the correlation coefficients between the disturbances of the probit and the wage equations. The values of these coefficients are also reported in Tables 3 to 6.

The signs of the estimated coefficients for $\rho_{1\epsilon^*}$ and $\rho_{2\epsilon^*}$ determine whether the unobserved abilities of individuals are positively or negatively correlated with the wage levels. We find in all four cross-sections that the sign of the correlation coefficient is negative and significant for stayers, and positive and significant for movers. Thus, this can be interpreted as strong evidence that East-West migrants tend to be *positively* selected with respect to their unobserved abilities. Although the positive selection bias varies somewhat in the different cross-sections, we do not observe that the positive selection bias disappears over time. Thus, the increasing inequality of earnings in Eastern Germany do not seem to have affected the selection bias in our regressions.

The wage differential is calculated on the consistent estimate of wages which are not biased by a selection effect (see Figure 6). If we take into account the selection effect, the wage differential is much higher (see Figure 8). Note that this wage differential cannot be included in the probit regression directly because of the huge difference among stayers and movers which would produce a perfect prediction on who moves and who stays.

The education variables show the expected signs and have a stable effect on wages in East Germany over time. The wage premium for individuals with completed training over individuals without completed training amounts to 13 to 15 per cent. Individuals with a university degree get a wage premium

Figure 7: Selection bias for movers in West Germany and stayers in East Germany

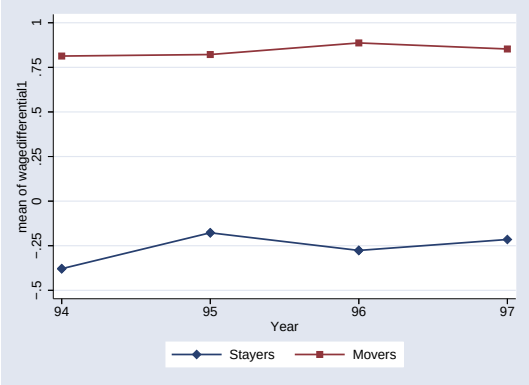

Figure 8: Mean of the wage differential calculated taking into account the sample selection

of 37 to 42 per cent. The wage gap between no education and university education amounts, then, to about 40 DM per day. Unknown education also gets a wage premium over no education, but the premium is decreasing over time (from 13.7 per cent to 8 per cent). In West Germany, unknown education is not rewarded differently than no training. The wage premium for university education is similar to East Germany with a higher volatility of between 33 to 61 per cent. Also the premium of training is similar to East Germany, but the trends of the curves are less smooth, possibly due to the smaller sample size in West Germany.

Workers with completed vocational training receive four to seven per cent higher wages than workers without completed training in East Germany, and white collar workers and foremen receive the highest wages (between 30 and 40 per cent more than workers without vocational training). In West Germany, we observe that only clerks and foremen earn significantly more than the other two groups. Skilled workers get significantly more than unskilled only in the years 1996 and 1997. The insignificant difference between skilled and unskilled workers might be interpreted as evidence for a devaluation of work-specific human capital.

With regard to the age effect on wages, there is again a difference between movers and stayers. Stayers get a shrinking wage increase while moving through the age brackets. This is consistent with the normal findings in wage regressions where the coefficients for age are positive and the coefficients for age squared are negative. From the age bracket 41-50 or latest from the age bracket 51-64 onwards, the wage even decreases. For movers, the picture is less smooth. The standard result of decreasing returns to age is also observable, but there exists a dip in all the observed years with individuals in their thirties getting lower wages than younger movers. Movers also get a less steep wage increase than stayers, which can be explained by the missing tenure of movers.

Men get consistently higher wages than women over time in East and West Germany.

6.2 Results from the probit estimates

The results of the *reduced form* probit regressions are reported in Tables 7 and 8, and of the *structural form* probit regressions in Tables 9 and 10. We show only the results for the probability to move. The results for the probability to stay can be derived from the results shown by just switching the sign of the coefficients. In the first step (reduced form probit), one can observe the overall effect of the explanatory variables. The effect is composed of the direct effect and the indirect effect which goes through the wage differential. In the second step (the structural probit), only the exogenous variables are included. The variables which are expected to affect wages are excluded because of collinearity between these variables and the wage differential.

The probit regressions explain the probability of an individual to work in West Germany in year t instead of staying in East Germany and working there. The explanatory variables refer to year $t-1$, i.e. they measure the status of a migrant or a stayer in East Germany the year before the observation. We included once again only individuals who are employed in year t. All regressions include a large number of individuals, of whom only very few decide to migrate (see Table 1).

In the reduced form probit equation we observe some surprising results. Regarding the education variables, we do not find a clear picture like those found in other studies. None of our education variables turn out to be significantly different from zero over time. Only two of the parameters are significant at the ten percent level, but they do not bring any insights regarding the effects of education on the decision to migrate over time. We conclude, therefore, that the education level does not affect the migration decision of workers directly.

Regarding the occupational status, skilled workers tend to move less than unskilled workers with significantly negative parameters from 1994 to 1996. Also clerks and foremen are potentially less likely to move than workers with no vocational training with only one significantly negative parameter in 1996.

The lagged unemployment rate exerts a very important impact on the migration decision. We expected the effect of unemployment to be large because we already observed in the descriptive analysis a huge difference in the lagged unemployment rates of movers and stayers (see Section 5).

Among the personal characteristics, the marital status variable has the expected significant negative effect on migration in almost all cross sections (except 1995). The age brackets show a clear picture during the whole time period. The cohorts with an age between 21 and 25 are most likely to move. The oldest age group in the sample, i.e. the cohorts with an age from 51 to 64 years, have the highest propensity to stay in East Germany. All other age groups behave like the reference category aged 17 to 21.

The other personal and family variables also show the expected results. Males are significantly more likely to move, while married individuals are less likely to move. The children dummy is not significant, perhaps due to the fact that children are underreported in the social security records.[6] Sector and regional dummies do not show a clear pattern over time. Saxony is the only exception which shows a significant attraction preventing workers from moving to West Germany from 1995 onward. Mecklenburg-Western Pomerania shows a significant positive effect in the structural probit only in 1994.

The wage differential has the expected positive sign in all four cross-sections in the *structural form* probit estimates. The estimated effect of the wage differential has to be interpreted as the minimum effect of the wage differential, because we used the linear predictions not taking into account the selection effect. If we consider the selection effect, the impact of the wage differential would be much higher (see Graph 6).

The second important explanation for the decision to move is lagged unemployment in the *structural form* probit estimates. This replicates the results from the reduced form probit. It has a positive and highly significant impact on the probability to move. Marital status now has significantly negative effects in all years, while the dummy for children is still not significant. The results for the home regions are also similar to the reduced form. Thus, the economic variables in the structural probit regressions do have the expected signs and are significant.

[6]The main reason for underreporting is that only one parent receives welfare benefits from children (like Kindergeld, Kinderfreibetrag), such that the other parent appears as childless in the social security records.

Table 2: Employment and unemployment among stayers and movers

	Stayers			Movers		
year	employed	unempl.	percent	employed	unempl.	percent
1994	33038	4574	13.8	428	51	11.9
1995	33677	4638	13.8	380	44	11.6
1996	32418	6502	20.1	389	82	21.1
1997	30878	6602	21.4	364	88	24.2

6.3 Comparison of Probit estimates with and without unemployed

We have excluded the unemployed from our regressions, which may bias the results if the unemployment rates in the group of movers and stayers and their socio-economic characteristics differ. Table 2 displays the number of employed and unemployed in the groups of movers and stayers. The figures show very similar unemployment rates in both groups. The difference in the characteristics are not displayed here. However, including the unemployed in the probit regressions does not change the results much: the parameters have a similar size and we observe the same pattern of significance.[7] Thus, we can conclude that the exclusion of the unemployed has not affected our results.

7 Conclusion

East Germany inherited from central planning greater equality in the distribution of earnings, but the inequality in earnings has almost approached West German levels in less than ten years since the beginning of transition. The standard Roy model predicts that migrants are favourably selected if the variance of earnings is higher in the host region than in the home region and if the correlation between earnings in both regions is sufficiently strong. Thus, we can derive from the Roy model the hypothesis that migrants from East Germany should have been favourably selected in the beginning, but that the selection bias should have disappeared over time.

[7]The results are available from the authors upon request.

Our findings do not confirm this hypothesis. We analysed the self-selection of East-West migrants in Germany on basis of a switching regression model. The results from our selection regressions provide strong evidence that migrants from East Germany are positively selected with regard to unobserved abilities. Thus, migrants tend to earn more than their staying counterparts if we control for observable human capital characteristics. East-West migrants do, however, remain positively selected over time although the inequality in the distribution in East Germany has substantially increased. This result clearly contradicts the predictions of the standard Roy model. However, if we relax the assumptions of the standard model and consider moving costs which are negatively correlated with abilities relevant for the labour market performance of individuals, then a persistent positive selection bias is in line with the theoretical expectations resulting from an extended Roy model. Nevertheless, our results should be interpreted with caution, since our sample covers only five years.

In contrast to studies based on other data sources, we do not find that individuals with higher education degrees tend to have a higher propensity to migrate. Conversely, the descriptive statistics in our data set show that skill levels of the migrants are below those of the average person who remains in East Germany. Moreover, we do not find in our reduced form probit regressions any significant impact of education on the propensity to migrate. This holds true for the whole time period covered by our sample. However, these results can be traced back at least partly to the fact that an important group of high-skilled migrants, students and individuals with an academic degree who started their working career in West Germany, is not included in our sample.

Interestingly enough, we find a *negative* correlation between occupational status and migration probabilities in our reduced form probit regressions. We interpret this as an indication that occupational status is at least partly lost by moving from the East to the West, i.e. that migrants are only partly able to transfer their work experience and firm-specific human capital to the West.

The wage differential shows a strong positive impact on the propensity to migrate in our structural probit estimates. Moreover, an unemployment event in the period preceding migration turns out to be highly significant.

These findings are in contrast to parts of the empirical literature based on individual data sets which do not control for selection bias of migrants. Thus, the results from our switching regression estimates reconcile some of the paradoxical findings in the empirical literature on the impact of wages and other key economic variables on East-West migration in Germany.

Finally, the results from our probit regressions confirm some standard findings from the human capital theories of migration: the propensity to migrate declines with age, married individuals tend to migrate less than unmarried, and males tend to migrate more than females.

References

Bender, S., A. Haas, and C. Klose (2000): "IAB Employment Subsample 1975-1995 - Opportunities for Analysis Provided by the Anonymised Subsample," *IZA Discussion Paper, No. 117*, IZA Bonn.

Borjas, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 77(4), 531–553.

Burda, M. C. (1993): "The Determinants of East-West German Migration: Some First Results," *European Economic Review*, 37, 452 – 461.

Burda, M. C., W. Härdle, M. Müller, and A. Werwatz (1998): "Semiparametric Analysis of German East-West-Migration Intentions: Facts and Theory," *Journal of Applied Econometrics*, 13, 525–41.

Burda, M. C., and J. Hunt (2001): "From German Reunification to Economic Integration: Productivity and the Labor Market in Eastern Germany," *Brooking Papers of Economic Activity*, 2:2001, 1–91.

Chiswick, B. R. (2000): "Are Immigrants Favorably Self-Selected? An Economic Analysis," *IZA Discussion Paper*, 131.

Goldfeld, S., and R. Quandt (1973): "The Estimation of Structural Shifts by Switching Regressions," *Annals of Economic and Social Measurement*, 2, 475 – 485.

Greene, W. H. (1997): *Econometric Analysis*. Prentice Hall International, London, third edn.

Haas, A. (2001): "Die IAB-Regionalstichprobe 1975-1997," *ZA-Informationen*, (48), 128–141.

Harris, J. R., and M. Todaro (1970): "Migration unemployment and development: a two sector analysis," *American Economic Review*, 60, 126 – 142.

Heckman, J. J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Model," *Annals of Economic and Social Measurement*, 15, 475–492.

——— (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153 – 161.

Hunt, J. (2000): "Why do People Still Live In East Germany?," *DIW Discussion Paper, No. 201*, DIW Berlin.

Jasso, G., and M. R. Rosenzweig (1990): "Self-selection and the Earnings of Immigrants: Comment," *American Economic Review*, 80(1), 298–304.

Lee, F. (1978): "Unionism and Wage Rates: A Simoultaneous Equation Model with Qualitative and Limited Dependent Variables," *International Economic Review*, 19, 415–433.

——— (1979): "Identification and Estimation of Binary Choice Models with Limited (Censored) Dependent Variables," *Econometrica*, 47, 977 – 996.

Maddala, G. S. (1983): *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press, Cambridge.

Maddala, G. S., and F. D. Nelson (1975): "Switching regression models with exogenous and endogenous switching," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 423–426.

Nakosteen, R. A., and M. Zimmer (1984): "Migration and Income: The Question of Self-Selection," *Southern Economic Journal*, 46(3), 840–851.

Pischke, J. S., M. Staat, and S. Vögele (1994): "Arbeitslosigkeit, Löhne und Weiterbildung: Warum pendeln Ostdeutsche in den Westen," in *Arbeitsmarktdynamik und Unternehmensentwicklung in Ostdeutschland*, ed. by H. König, and V. Steiner. Nomos.

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

Schwarze, J. (1996): "Beeinflusst das Lohngefälle zwischen Ost- und Westdeutschland das Migrationsverhalten der Ostdeuschen?," *Allgemeines Statistisches Archiv*, 80, 50 – 68.

Sjaastad, L. (1961): "The Costs and Returns of Human Migration," *Journal of Political Economy*, 70, 80 – 93.

Willis, R., and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, S7 – S36.

Table 3: Heckit East German Wages 1994-95 with Selection: Stayer

Variable	**Coeff.94**	**(Std.Err.)**	**Coeff.95**	**(Std.Err.)**
Dependent variable: log(wage)				
With training	0.149**	(0.008)	0.130**	(0.008)
University or UAS	0.395**	(0.010)	0.400**	(0.010)
Unknown education	0.137**	(0.009)	0.102**	(0.010)
Skilled workers	0.052**	(0.006)	0.058**	(0.006)
Clerks and foremen	0.354**	(0.006)	0.371**	(0.006)
Age 21-25	0.494**	(0.012)	0.512**	(0.013)
Age 26-30	0.597**	(0.012)	0.618**	(0.013)
Age 31-35	0.638**	(0.012)	0.659**	(0.012)
Age 36-40	0.655**	(0.012)	0.685**	(0.012)
Age 41-45	0.660**	(0.012)	0.688**	(0.012)
Age 46-50	0.662**	(0.012)	0.684**	(0.013)
Age 51-64	0.665**	(0.011)	0.679**	(0.012)
Sex	0.226**	(0.004)	0.240**	(0.004)
Brandenburg	0.043**	(0.006)	0.055**	(0.006)
Meck.-W. Pom	0.032**	(0.006)	0.034**	(0.006)
Saxony	-0.001	(0.005)	0.008	(0.005)
Saxony-Anhalt	0.017**	(0.006)	0.019**	(0.006)
Agric., Cons. Goods	-0.055**	(0.005)	-0.042**	(0.005)
Goods Production	-0.005	(0.005)	0.033**	(0.005)
Constr., transport	-0.093**	(0.005)	0.079**	(0.005)
Intercept	3.443**	(0.011)	3.453**	(0.012)
$\rho_{1\epsilon^*}$	-0.845**	(0.012)	-0.877**	(0.010)
σ_1	0.301**	(0.001)	0.311**	(0.001)
λ	-0.255**	(0.004)	-0.273**	(0.004)
N	33466		34057	
N cens.	428		380	
Log-likelihood	-8465.171		-9478.105	
$\chi^2_{(20)}$	25372.17		25473.09	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Reference category: No training, Unskilled worker, Female, Not married, Thuringia, Building and construction

All variables except ln(wage) refer to the year before

Table 4: Heckit East German Wages 1996-97 with Selection: Stayer

Variable	**Coeff.96**	**(Std.Err.)**	**Coeff.97**	**(Std.Err.)**
Dependent variable: log(wage)				
With training	0.151**	(0.009)	0.143**	(0.009)
University or UAS	0.412**	(0.011)	0.415**	(0.011)
Unknown education	0.104**	(0.010)	0.082**	(0.010)
Skilled workers	0.054**	(0.006)	0.069**	(0.006)
Clerks and foremen	0.378**	(0.006)	0.392**	(0.007)
Age 21-25	0.519**	(0.014)	0.516**	(0.014)
Age 26-30	0.663**	(0.013)	0.671**	(0.013)
Age 31-35	0.707**	(0.013)	0.713**	(0.013)
Age 36-40	0.733**	(0.013)	0.739**	(0.013)
Age 41-45	0.747**	(0.013)	0.768**	(0.013)
Age 46-50	0.722**	(0.013)	0.743**	(0.013)
Age 51-64	0.718**	(0.013)	0.733**	(0.013)
Sex	0.250**	(0.004)	0.250**	(0.004)
Brandenburg	0.048**	(0.006)	0.045**	(0.006)
Meck.-W. Pom.	0.036**	(0.007)	0.030**	(0.007)
Saxony	0.006	(0.005)	0.001	(0.006)
Saxony-Anhalt	0.019**	(0.006)	0.019**	(0.006)
Agric., Cons. Goods	-0.035**	(0.006)	-0.036**	(0.006)
Goods Production	0.050**	(0.006)	0.062**	(0.006)
Constr., transport	0.076**	(0.005)	0.072**	(0.005)
Intercept	3.408**	(0.012)	3.405**	(0.012)
$\rho_{1\epsilon^*}$	-0.918**	(0.007)	-0.919**	(0.008)
σ_1	0.319**	(0.001)	0.322**	(0.001)
λ	-0.293**	(0.003)	-0.296**	(0.003)
N	32807		31242	
N cens.	389		364	
Log-likelihood	-9893.626		-9739.047	
$\chi^2_{(18)}$	26672.68		26377.80	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Reference category: No training, Unskilled worker, Female, Not married, Thuringia, Building and Construction

All variables except ln(wage) refer to the year before

Table 5: Heckit West German Wages 1994-95 with Selection: Mover

Variable	**Coeff.94**	**(Std.Err.)**	**Coeff.95**	**(Std.Err.)**
Dependent variable: log(wage)				
With training	0.198**	(0.055)	0.098†	(0.060)
University or UAS	0.613**	(0.094)	0.334**	(0.113)
Unknown education	0.094	(0.061)	0.055	(0.066)
Skilled workers	-0.019	(0.050)	0.074	(0.049)
Clerks and foremen	0.200**	(0.051)	0.354**	(0.053)
Age 21-25	0.196*	(0.082)	0.196†	(0.088)
Age 26-30	0.153†	(0.083)	0.158*	(0.091)
Age 31-35	0.107	(0.085)	0.331**	(0.089)
Age 36-40	0.193*	(0.091)	0.315**	(0.091)
Age 41-45	0.099	(0.088)	0.220*	(0.095)
Age 46-50	0.247**	(0.096)	0.335**	(0.103)
Age 51-64	0.240*	(0.097)	0.183†	(0.103)
Sex	0.260**	(0.046)	0.243**	(0.044)
Agric., Cons. Goods	-0.033	(0.057)	0.049	(0.060)
Goods Production	0.095†	(0.052)	0.264**	(0.066)
Constr., transport	0.114*	(0.048)	0.159**	(0.043)
Intercept	3.548**	(0.171)	3.694**	(0.185)
$\rho_{1\epsilon^*}$	0.529**	(0.230)	0.313†	(0.182)
σ_1	0.387**	(0.058)	0.340**	(0.022)
λ	0.205**	(0.118)	0.106	(0.068)
N	34198		35020	
N cens.	33779		34654	
Log-likelihood	-2263.904		-1999.873	
$\chi^2_{(16)}$	165.00		170.80	

Significance levels : † : 10% * : 5% ** : 1%

Reference category: No training, Unskilled worker, Female, Not married, Thuringia, Building and Construction

All variables except ln(wage) refer to the year before

Table 6: Heckit West German Wages 1996-97 Selection: Mover

Variable	**Coeff.96**	**(Std.Err.)**	**Coeff.97**	**(Std.Err.)**
Dependent variable: log(wage)				
With training	0.171**	(0.053)	0.037	(0.068)
University or UAS	0.352**	(0.094)	0.430**	(0.098)
Unknown education	0.083	(0.056)	-0.010	(0.072)
Skilled workers	0.076†	(0.044)	0.181**	(0.052)
Clerks and foremen	0.337**	(0.043)	0.466**	(0.054)
Age 21-25	0.263**	(0.089)	0.185†	(0.101)
Age 26-30	0.291**	(0.089)	0.257**	(0.100)
Age 31-35	0.351**	(0.089)	0.274**	(0.100)
Age 36-40	0.265**	(0.090)	0.254*	(0.103)
Age 41-45	0.268**	(0.095)	0.224*	(0.108)
Age 46-50	0.275**	(0.098)	0.266*	(0.109)
Age 51-64	0.312**	(0.095)	0.299**	(0.111)
Sex	0.285**	(0.040)	0.254**	(0.044)
Agric., Cons. Goods	0.047	(0.059)	-0.081	(0.067)
Goods Production	0.117*	(0.053)	0.161**	(0.058)
Constr., transport	0.121**	(0.039)	0.175**	(0.049)
Intercept	3.525**	(0.174)	3.646**	(0.205)
$\rho_{1\epsilon^*}$	0.477**	(0.143)	0.359†	(0.191)
σ_1	0.325**	(0.027)	0.332**	(0.026)
λ	0.155**	(0.058)	0.119†	(0.071)
N	33512		31886	
N cens.	33127		31526	
Log-likelihood	-2034.372		-1939.869	
$\chi^2_{(16)}$	204.42		268.25	

Significance levels : † : 10% * : 5% ** : 1%

Reference category: No training, Unskilled worker, Female, Not married, Thuringia, Building and Construction

All variables except ln(wage) refer to the year before

Table 7: Probit first step 1994-1995

Variable	**Coeff.94**	**(Std.Err.)**	**Coeff.95**	**(Std.Err.)**
With training	-0.002	(0.080)	-0.128†	(0.084)
University or UAS	-0.068	(0.110)	-0.156	(0.140)
Unknown education	0.008	(0.095)	-0.144	(0.095)
Skilled workers	-0.159*	(0.063)	-0.152*	(0.068)
Clerks and foremen	-0.104	(0.067)	-0.085	(0.072)
Unemployed	0.640**	(0.078)	0.587**	(0.064)
Age 21-25	0.216*	(0.108)	0.355**	(0.016)
Age 26-30	0.013	(0.114)	0.046	(0.016)
Age 31-35	-0.141	(0.115)	-0.014	(0.016)
Age 36-40	-0.239*	(0.119)	-0.071	(0.016)
Age 41-45	-0.175	(0.118)	-0.097	(0.016)
Age 46-50	-0.211†	(0.126)	-0.188	(0.016)
Age 51-64	-0.408**	(0.121)	-0.377**	(0.016)
Sex	0.207**	(0.048)	0.191**	(0.055)
Marital status	-0.129**	(0.045)	-0.057	(0.051)
Children	0.020	(0.094)	0.076	(0.051)
Agric., Cons.Goods	-0.120†	(0.063)	-0.048	(0.122)
Goods Production	-0.008	(0.058)	0.022	(0.078)
Construct., Trans.	-0.064	(0.055)	0.016	(0.094)
Brandenburg	0.112†	(0.065)	-0.024	(0.071)
Meck.-W. Pom.	0.101	(0.071)	0.005	(0.077)
Saxony	-0.080	(0.062)	-0.193**	(0.065)
Saxony-Anhalt	0.057	(0.065)	-0.093	(0.071)
Intercept	-2.111**	(0.106)	-2.137**	(0.266)
N	34207		35034	
Log-likelihood	-2160.126		-1944.478	
$\chi^2_{(23)}$	305.073		260.177	

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Reference category: No training, Unskilled worker, Female, Not married, Thuringia, Building and Construction

All variables except ln(wage) refer to the year before

Table 8: Probit first step 1996-1997

Variable	**Coeff.96**	**(Std.Err.)**	**Coeff.97**	**(Std.Err.)**
With training	0.005	(0.086)	-0.032	(0.088)
University or UAS	-0.007	(0.119)	0.078	(0.117)
Unknown education	0.176†	(0.098)	0.113	(0.100)
Skilled workers	-0.134*	(0.063)	-0.097	(0.068)
Clerks and foremen	-0.127†	(0.068)	-0.073	(0.072)
Unemployed	0.476**	(0.074)	0.455**	(0.070)
Age 21-25	0.359**	(0.128)	0.395**	(0.127)
Age 26-30	0.246†	(0.135)	0.207	(0.137)
Age 31-35	0.169	(0.135)	-0.002	(0.139)
Age 36-40	0.082	(0.138)	-0.052	(0.141)
Age 41-45	-0.084	(0.143)	-0.168	(0.144)
Age 46-50	-0.023	(0.146)	-0.128	(0.146)
Age 51-64	-0.157	(0.141)	-0.298*	(0.145)
Sex	0.193**	(0.050)	0.137**	(0.050)
Marital status	-0.179**	(0.047)	-0.091†	(0.050)
Children	0.125	(0.089)	0.081	(0.088)
Agric., Cons. Goods	-0.141*	(0.070)	-0.199**	(0.074)
Goods Production	-0.183**	(0.071)	-0.069	(0.066)
Constr., Transp.	0.056	(0.054)	-0.082	(0.058)
Brandenburg	-0.195**	(0.065)	-0.033	(0.066)
Meck.-W. Pom.	-0.195**	(0.073)	-0.033	(0.072)
Saxony	-0.261**	(0.057)	-0.205**	(0.061)
Saxony-Anhalt	-0.130*	(0.062)	-0.139*	(0.068)
Intercept	-2.169**	(0.119)	-2.142**	(0.118)
N	33516		31890	
Log-likelihood	-1982.069		-1865.945	
$\chi^2_{(23)}$	276.253		248.205	

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Reference category: No training, Unskilled worker, Female, Not married, Thuringia, Building and Construction

All variables except ln(wage) refer to the year before

Table 9: Probability to move 1994-1995, structural results

Variable	**Coeff.94**	**(Std.Err.)**	**Coeff.95**	**(Std.Err.)**
Unemployed	0.658**	(0.074)	0.589**	(0.068)
Wage differential	0.553**	(0.186)	0.327*	(0.151)
Marital status	-0.248**	(0.044)	-0.242**	(0.043)
Children	0.004	(0.044)	0.068	(0.083)
Brandenburg	0.141*	(0.072)	-0.034	(0.066)
Meck.-W. Pom.	0.119†	(0.078)	0.024	(0.070)
Saxony	-0.081	(0.067)	-0.210**	(0.060)
Saxony-Anhalt	0.060	(0.072)	-0.094	(0.065)
Intercept	-2.017**	(0.133)	-2.129**	(0.052)
N	33457		34043	
Log-likelihood	-2163.287		-1938.589	
$\chi^2_{(8)}$	176.718		168.843	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Reference category: Employed, Not married, no children, Thuringia

All variables except the wage differential refer to the year before

Table 10: Probability to move 1996-1997, structural results

Variable	**Coeff.96**	**(Std.Err.)**	**Coeff.97**	**(Std.Err.)**
Unemployed	0.471**	(0.072)	0.482**	(0.067)
Wage differential	0.406*	(0.163)	0.144	(0.162)
Marital status	-0.299**	(0.043)	-0.281**	(0.044)
Children	0.140	(0.086)	0.068	(0.085)
Brandenburg	-0.154*	(0.065)	-0.017	(0.066)
Meck.-W. Pom.	-0.157*	(0.071)	-0.008	(0.071)
Saxony	-0.259**	(0.057)	-0.210**	(0.061)
Saxony-Anhalt	-0.133*	(0.061)	-0.126†	(0.067)
Intercept	-1.930**	(0.054)	-2.083**	(0.055)
N	32803		31238	
Log-likelihood	-2012.15		-1894.15	
$\chi^2_{(8)}$	163.838		141.101	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Reference category: Employed, Not married, no children, Thuringia

All variables except the wage differential refer to the year before

Table 11: Summary 94-97

Variable	**Mean94**	**Mean95**	**Mean96**	**Mean97**
log(wage)	4.565	4.608	4.627	4.637
	(0.397)	(0.408)	(0.425)	(0.434)
Age 17-20	0.029	0.028	0.030	0.031
Age 21-25	0.079	0.080	0.075	0.071
Age 26-30	0.123	0.113	0.105	0.102
Age 31-35	0.158	0.156	0.154	0.148
Age 36-40	0.157	0.155	0.154	0.156
Age 41-45	0.154	0.159	0.157	0.155
Age 46-50	0.101	0.099	0.113	0.129
Age 51-64	0.198	0.211	0.211	0.207
Male	0.580	0.576	0.565	0.563
Mover Dummy	0.013	0.011	0.012	0.011
Unemployed	0.042	0.067	0.061	0.080
Marital status	0.558	0.565	0.571	0.574
Children	0.047	0.059	0.053	0.059
Without training	0.080	0.083	0.082	0.086
With training	0.743	0.733	0.733	0.729
UAS or University degree	0.110	0.107	0.106	0.104
Unknown education	0.076	0.077	0.079	0.081
Trainees and unskilled	0.152	0.163	0.164	0.168
Skilled workers	0.382	0.375	0.369	0.365
Clerks and foremen	0.465	0.462	0.467	0.467
Brandenburg	0.181	0.179	0.177	0.176
Mecklenburg-Western Pomerania	0.123	0.123	0.124	0.125
Saxony	0.337	0.338	0.339	0.344
Saxony-Anhalt	0.195	0.194	0.193	0.193
Thuringia	0.164	0.166	0.167	0.163
Agriculture, Consumption Goods	0.155	0.145	0.138	0.137
Goods Production	0.161	0.149	0.146	0.146
Construction, Transport	0.210	0.225	0.226	0.225
Trade, Services	0.473	0.481	0.489	0.492
Number of observations	34207	35034	33516	31889
No. of obs. lnwage	33457	34043	32803	31238

All variables except ln(wage) refer to the year before

Nutritional efficiency wages in Côte d'Ivoire

Abstract

It is generally accepted that people have to eat in order to be able to do their work. On the other hand, the wage or production output enables them to subsist. This circle has been investigated from different points of view, as different policy implications arise from the possible relations between nutritional status and ability to work. Economists have shown theoretically that better nutrition increases productivity and that there is a threshold of the nutritional status below which people are not able to produce the equivalent to their subsistence wages. This implies that help for the poorest of the population has to improve their nutritional status directly (for example through (re)distribution of land or food) to enable them to work. I am looking into the empirical relationship between nutrition (Body Mass Index) and wages using the Côte d'Ivoire Living Standards Survey (CILSS) from 1985 to 88. I find that the probability to earn wages depends as predicted by theory on the Body Mass Index as a short-run indicator of nutrition and on body height as a long-run indicator. The probability to work is not dependent on the nutritional status, but on non-labour income, which theory also interprets as a productivity enhancing factor.

Contents

List of Figures

List of Tables

1 Introduction

The first of the international development goals that have been defined at the millennium summit is to "eradicate extreme poverty and hunger". All United Nations Member states have pledged to meet the goals that should be reached by 2015 and comprise changes in poverty, education, gender inequality etc. The first goal to reduce extreme poverty is divided into the goal to halve the proportion of people whose income is less than $1 a day from 1990 to 2015 and into the goal to halve the proportion of people starving (for details see United Nations (2000)).

The international organisations show in their report that especially in Sub-Saharan Africa, the path of development out of poverty is by far not enough to reach the goal to halve poverty from 1990 to 2015 (IMF, OECD, United Nations, and World Bank, 2000). Africa has by far the worst starting position and the progress out of poverty is almost not existent. The proportion of people living on less than 1$ a day went down from 48 to 46% in Sub-saharan Africa, while in the second worst region, South Asia, the drop was from 44 to 40%. All other regions do have much lower poverty numbers.

Only when the causes for poverty and hunger in Sub-saharan Africa are identified, policy measures can be designed to successfully fight against them. One theoretical explanation for the persistence of poverty is formulated in the efficiency wage theory first introduced by Leibenstein (1957). Efficiency wages are wages above the market clearing wage and induce involuntary unemployment. While in developed countries efficiency wages are paid to prevent workers from shirking (when information is imperfect), efficiency wages in developing countries are necessary for the workers to maintain their productivity. Workers in developing countries are expected to be so undernourished from the beginning that firms maximising profit arrive at paying a wage level above the equilibrium.

The influence of the nutritional status on productivity and thus on the ability to gain wages was explained theoretically by several models, but the empirical evidence is nevertheless less conclusive. For the empirical test of the efficiency wage theory, a developing country like Côte d'Ivoire can be used. Even if Côte d'Ivoire has to be seen as a high-wage country among the members of the West-African currency union and one of the most industrialised countries among the sub-Saharan countries, poverty remains very high. For an overview of the economic indicators of Côte d'Ivoire over the time span under consideration see Table 7 in the Appendix.

In Table 1, two measures of Foster, Greer, and Thorbecke (1984) are displayed to show the extent of poverty in the regions of Côte d'Ivoire for

the years 1985 to 1988. The measures are calculated with the formula

$$P_\alpha = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{z - x_i}{z} \right]^\alpha$$

where n is the population size, z is the poverty line, x_i is the expenditure of individual i. With rising α, poverty gaps are higher penalised. With α being zero, the measure corresponds to the head count ratio, which is the share of individuals living below the poverty line. The $\alpha = 2$ measure is used most frequently, because it takes the distribution among poor into account (Deaton, 1997). As the poverty line, 128,600 CFAF per year is chosen. This line is defined so that 30% of the population fell below that line at 1985 Abidjan Prices (World Bank, 1997).

The headcount measure P_0 in Table 1 shows a slight increase in poverty from 1985 to 88 for Côte d'Ivoire with widely varying numbers among the five regions. While in Abidjan, the biggest agglomeration in the country, the poverty rate is relatively low, but rising over time, poverty in the two least developed regions East Forest and Savannah, is persistently high over time. The measure P_2 shows that poverty in Côte d'Ivoire went down in the three rural areas from 85 to 86, but afterwards it increased, too. The table shows that if we take into account the average per capita expenditure of the poor and the distribution of expenditures among the poor (use P_2 instead of P_0), the poverty increase is moderated.

Table 1: Poverty measures by region and year

	1985		1986		1987		1988	
	P_0	P_2	P_0	P_2	P_0	P_2	P_0	P_2
Côte d'Ivoire	0.300	0.045	0.299	0.032	0.348	0.043	0.459	0.063
Abidjan	0.034	0.004	0.166	0.012	0.074	0.009	0.139	0.006
Other Cities	0.236	0.037	0.223	0.024	0.224	0.019	0.410	0.040
East Forest	0.479	0.069	0.395	0.045	0.435	0.041	0.494	0.062
West Forest	0.178	0.013	0.200	0.013	0.376	0.043	0.553	0.064
Savannah	0.502	0.088	0.481	0.058	0.578	0.093	0.652	0.131

Poverty Line: 128,600 CFAF/year, Source: Grootaert (1993)

Table 2 shows the dynamics of poverty and the probabilities of individuals to move out of or into poverty. The advantage of measuring poverty changes on the household level and not on the economy-wide or regional level is that changes in poverty of the same individuals are measured. Thus, measuring poverty on the individual level gives a more precise picture on poverty changes. Grootaert and Kanbur (1993a) do not only use the poverty line of 128,600 CFAF/year that defines 30% of the population of 1985 as poor, but a second poverty line of 75,000 CFAF/year to indicate very poor

individuals. With the latter poverty line, 10% of the population of 1985 are declared as very poor. In Table 2, the probability to move from one poverty class to another is displayed, with classes being defined as very poor, poor and non-poor. In the first line, one can see that poverty in the panel is a little less than the higher poverty percentage of the two corresponding cross-sections in the first line of Table 1. In detail, the values do not tell a positive story about the years in consideration. The probability to move into a worse poverty class is much higher then the probability to move upwards. In the second two panels, even the descent into the very poor category is much more probable than the ascent out of the very poor category.

Table 2: Probability to be poor

Poverty in Panels	85/86	86/87	87/88
Prob. to be poor in panel	28.1	26.7	40.1
Prob. to stay in poverty class	74.4	73	69.1
Prob. to improve poverty class	12.6	8.5	6.3
Prob. to worsen poverty class	12.7	18.5	24.5
Prob. to improve from very poor	5.7	1.2	2.1
Prob. to worsen to very poor	2.2	5.1	9.2

Source: Grootaert and Kanbur (1993a), own calculations

The development goals seek to induce an amelioration in poverty and hunger; theory tells that a better nutrition leads to a higher productivity meaning higher income. If the theory on nutritional efficiency wages holds, a policy on generating income would not help the very poor to flee unemployment and find gainful employment. The theoretically most efficient policy should ameliorate the nutritional status of the unemployed so that they would be able to work productively. The goal of better nutrition for the poor could for example be reached by providing them with land. The generated nutrition effect from the cultivation of the land would then enable the poor to work more productively in a job and flee the vicious circle of unemployment and undernutrition.

In this paper I try to investigate the phenomenon of nutritional efficiency wages and nutritional poverty traps. For the investigation I use a rotating panel data set of Côte d'Ivoire over the years 1985 to 1988. The data set is one among very few panel data sets that exist for developing countries and it has a wide variety of questions relevant to my investigation. Furthermore, the data set is on a country that has a huge number of poor people and is thus suited for our purpose. After presenting theory and the data in more detail, I will estimate the influence of nutrition on employment and wages.

2 Theory

The basis to all the theories on nutritional efficiency wages is the energy balance equation. As poor people cannot afford to buy enough food, they are undernourished and their work capacity is affected. Energy balance is only ensured when energy input (w) is equal to energy use. Energy use consists of the resting metabolism (r), the work performance (q) and an increase or decrease of stocks as the balancing factor (b). This relation of the flow measures can be written as:

$$w = r + q + b, \tag{1}$$

The effort h that an individual can put into his work is $q + b$, but if b is smaller than zero, the individual will use his reserves and will not be able to maintain the work effort over a longer time. That can be seen in the stock equation.

The nutrition stock m in $(t + 1)$ as for example measured by the BMI is only changed by the balancing factor, so the stock equation can be written as:

$$m_{t+1} = m_t + b_t. \tag{2}$$

If an individual puts too much effort into his work, the nutritional stock will decrease over time, so that he will starve. Nutritional efficiency wages can be derived from these equations. They result from the capacity curve which illustrates the relation between the income w that is - according to assumptions - spent only on nutrition and the resulting work capacity (or output) h:

$$w = h(w - r), \tag{3}$$

At very low income levels, nutrition is used only for sustaining the resting metabolism r and not for increasing the work capacity. h and w do not any more show a linear relationship as in equation 1, because h depends now on w and r: With rising income work capacity rises more than proportionately; at very high incomes, capacity rises under-proportionately or even decreases. The curve is depicted in Figure 1. The nutritional efficiency wage w^* is defined as the wage level that has to be paid in order to receive an adequate work capacity and thus to be able to produce the desired output. The equilibrium wage will be explained in detail in the following chapter.

The labour supply function can be derived by combining the capacity curve with different wages. The labour demand function is not continuous, but has a gap because of the special form of the capacity curve. If the labour demand function crosses the supply curve in this gap, the equilibrium wage is hard to determine. The equilibrium only exists, if unemployment is possible. In this situation, supply is greater than demand, but the wage cannot adjust downwards to dispose of unemployment. At a lower wage level, it is not possible to offer the same work capacity. Wages are thus determined by the nutritional situation, not by supply and demand (for further details see Ray (1998, Ch. 13)).

Putting efficiency wages in a dynamic setting, the dependency of the work capacity and wages can lead to a vicious circle. A relatively high labour supply leads to low wages that are in a range where they influence the work capacity. An undernourished individual will not be able to offer work on this wage level because of his low productivity. Once an individual is unemployed, the probability to find a job decreases, because he lives on his reserves and becomes more and more undernourished. In the next period, he would need even higher wages. The situation leads to the existence of poverty traps and is theoretically described by Dasgupta (1997). I will show the details of the model in section 2.2.

2.1 A Static Model

Mirrlees (1975) and Stiglitz (1976) developed an efficiency wage model that can be applied to developing countries. The main assumption for the model in order to be valid is that the workers are reliant on their wages in order to insure their nutrition. They also have to use the bulk of their income for food. The idea is that individuals need a certain amount of food in order to maintain their resting metabolism as it has been described by equation 1. A wage that cannot deliver this amount of food cannot be accepted by any worker. With increasing wages the productivity of a worker will increase more than proportionately up to a point when the increase will only be less than proportionate.

To model the relation between nutrition and productivity, the authors introduce efficiency hours in contrast to clock hours; whereas the first measures the productivity of the workers, the last the actual hours. Workers being more productive produce a higher amount of efficiency hours during the same amount of clock hours. The amount of efficiency hours (h) produced per clock hour depends on the consumption of the worker: h(c).

The number of hours worked per day is denoted by l and the whole wage

w is consumed ($w = c$), so that the number of efficiency hours worked per day is $l \cdot h(w)$. I can then calculate the wage per efficiency hour from the daily wage w as w/h(w). This corresponds to the costs of the employer for one efficiency hour of work. The output is dependent on the number of efficiency hours and can be denoted as $y = f[l \cdot h(w)]$.

To derive the efficiency wage which corresponds to the equilibrium wage w*, one can consider three variants of the model. I will follow the argumentation of Bliss and Stern (1978) in the following model description where not otherwise specified. Bliss and Stern (1978) model the relation between productivity, wages and nutrition. They assume that there exists a monopsony on the labour demand side. In the first variant, only landless workers supply work, in the second variant, landless workers and workers with land coexist, but they supply labour in two different labour markets, and finally in the third variant, both kinds of workers exist and compete in the same market. In all these variants, a monopsony on the labour demand side is supposed and will be abolished only in a further step. With a competitive demand side, the outcome of the model changes.

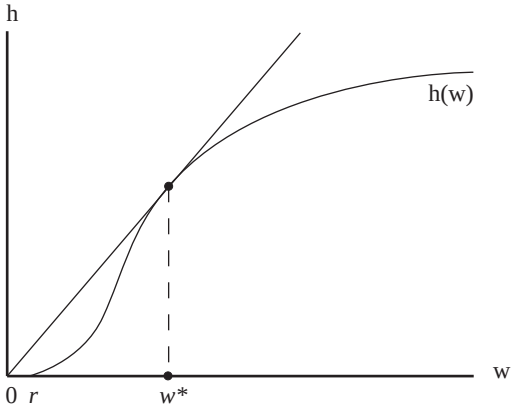

Figure 1: Wage capacity curve

Nutritional efficiency wages result from the work capacity curve which shows the relation between income being spent only on nutrition and output of a worker. The wage-productivity curve shown in Figure 1 starts at a positive wage assuming that a worker needs a certain level of consumption to be able to provide labour. The curve is then rising at an increasing rate, which means that at low levels of income, there is a huge effect of better nutrition on efficiency. From a certain level onwards there are decreasing returns to scale. The equilibrium of the model is at the point where the marginal productivity equals the average productivity, the equilibrium wage is thus w^*.

Up to this point, Bliss and Stern (1978) used the model of Mirrlees (1975) and Stiglitz (1976), but they have since then developed three different vari-

ants of the model.

2.1.1 First variant of the model: only landless workers

In the first variant of the model only landless workers supply labour in the labour market. Workers use wages w (at least to a high and fixed share) for consumption of food. The (only) employer minimises his costs, given that he produces at least the output $\overline{y}$.

The minimisation problem faced by the employer can be written as:

$$\min_{w,l} w \cdot l \tag{4}$$

subject to

$f[l \cdot h(w)] \geq \overline{y}$

where $[l \cdot h(w)]$ is the number of efficiency hours worked per day. Thus, the employer minimises his wage bill subject to producing at least output $\overline{y}$.

Minimising Equation 4 subject to the restriction results in the equilibrium wage w^*:

$$w* = h(w)/h'(w) \tag{5}$$

Equation 5 shows that the employer pays a wage which minimises the average costs for an efficiency hour. This wage is called efficiency wage and is indicated as w* in Figure 1. w^* is defined by the point where the tangent to the capacity curve goes through the origin. Because of the structure of the work capacity curve there can be unemployment without driving the wage downwards to an equilibrium without unemployment in the labour market.

2.1.2 Second variant of the model: landless workers and workers with land, two separate labour markets

In this variant, there are both workers owning land (Type 0) and landless workers (Type 1). Furthermore, there exist two employers, each of whom can employ workers either with or without land.

Type 1 workers will split their time between working on their own land and for the employer so that their consumption (welfare) is maximised. This is the case when marginal productivity of efficiency hours on their own land and opportunity costs of efficiency hours being worked for the employer are

equal. Labour supply is equal to labour demand, because Type 1 worker's maximisation of consumption corresponds to the employer's maximisation of the productivity of Type 1 workers.

An employer offers $(1 - \lambda)$ hours of work to a Type 1 worker, the Type 1 worker will then work the remaining λ hours or h_1 efficiency hours on his own land and produce an output $g(h_1)$. $g(.)$ is concave and $g(0)$ is zero. For Type 1 workers, the consumption c is not any more equal to w as in the first variant, because they receive a further consumption from their land. Thus, consumption of Type 1 workers is greater than their wage. The idea is that labour markets are for example situated in two different regions, one only with Type 0 and the other only with Type 1 workers, so employers cannot choose between them.

The minimisation problem of the employer for Type 1 workers is again subject to the restriction that he produces at least $\overline{y}$ and in addition that the workers' consumption is at least their wage plus what they receive from their land. The first constraint is simplified by a normalisation in the equation below to get an easier solution (see Bliss and Stern (1978, p.336) for a detailed explanation), the second constraint comes from the wish of the employer to get a certain amount of productivity that can only be guaranteed with consumption c.

$\min w \cdot (1 - \lambda) \cdot l$

subject to:

$(1 - \lambda) \cdot h(c) \cdot l \geq 1$

and

$c \leq (1 - \lambda)w + g[\lambda \cdot h(c)]$

Minimisation leads to an equilibrium wage level:

$$w^* = g'[\lambda \cdot h(c)] \cdot h(c) \tag{6}$$

The wage per efficiency hour for Type 1 workers is equal to the marginal product that the worker gets from his own land. Wages and output (or consumption) of Type 0 workers are lower than that of Type 1 workers. The tangent of the Type 1 worker that goes through the origin is situated on a point at the h(c) curve to the right of the tangent of type 0 workers. The equilibria for Type 0 and Type 1 workers are shown in Figure 2.

The employer of landless workers is facing the same minimisation problem as in the first variant in the model. Comparing equilibrium wages in equation 5 and 6, the wage-costs of the employer who employs only landless workers

are higher than the wage-costs of an employer employing only workers owning land. Thus, the employer of type 0 workers makes less profit. The comparison between the two equilibria is further discussed in Bliss and Stern (1978)

2.1.3 Third variant of the model: landless workers and workers with land, one labour market

The third variant is the least restricted variant among the models with demand monopsony. Both types of workers (with land and without) compete for work in the same labour market. I need two assumptions to get a result from the model. Type 1 workers with land can generate a fixed amount of consumption $\overline{c}$ from their land without work. $\overline{c}$ arises without work and cannot be increased. Furthermore, the supply of type 1 workers is restricted.

The employer minimises his costs, given that at least $\overline{y}$ is produced, that he pays a wage that is high enough to dispose of labour freely ($\overline{w}$ which is in this case the same for both types of workers) and that there is a restricted supply of Type 1 workers.

$\min w_0 l_0 + w_1 l_1$

subject to

$l_0 \cdot h(w_0) + l_1 \cdot h(\overline{c} + w_1) \geq 1$

$w_0 \geq \overline{w},\ w_1 \geq \overline{w},\ l_1 \leq \bar{l}_1$

The result of variant three is that all Type 1 workers are employed and only a part of the Type 0 workers. The diminution of the wage for Type 1 workers by $\overline{c}$ guarantees that both types of workers realise the same level of consumption.

$w_0 = w*$

$h'(w_0) = h'(\overline{c} + w_1)$

$w_0 = \overline{c} + w_1$

The equilibrium can be seen in Figure 2. Type 1 workers with land are able to work for lower wages, thus their r_1 is lower than the r_0 of Type 0 workers. The tangents to the capacity curves indicate the equilibrium wages w_0 and w_1, the difference between the two amounts to $\overline{c}$.

2.1.4 Competitive labour demand side

With a competitive labour demand side, the wage per efficiency unit should be equal for all workers. Type 1 workers with land profit from the unique

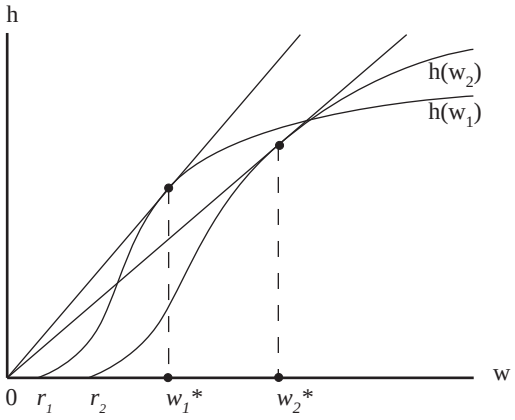

Figure 2: Wage capacity curves of two individuals, variant 3

wage level because they have a second source of consumption at their disposal. They receive a higher wage per hour because of their higher productivity and can consume (again) more. All workers have the same chance to get employed; wage differences between type 0 and type 1 workers only exist because of productivity differences.

2.2 Dynamic Variant of Dasgupta

Dasgupta (1997) argues that a static model is not able to explain the existence of nutritional poverty traps, because the long-run effects of nutrition are not taken into account. He also criticises the existing empirical papers on nutritional efficiency wages, because they had taken the model too literally. His model is based on the deterministic differential equation of a person's health stock (or nutrition stock for the consistence of the models)

$$dm_t/dt = w(m_t) - h(m_t), \tag{7}$$

$\overline{m} > m > \underline{m}$

with m_t being an index of a person's health or nutrition at time t and $h(m_t)$ and $w(m_t)$ being a person's food requirements and income.

w(m) and h(m) are drawn in Figure 3 so that they intersect at a person's health $\widehat{m}$.

Of the three equilibria, the one at $\widehat{m}$ is not stable, because once a person moves ϵ away from $\widehat{m}$, the dynamic leads him to either the stable equilibrium $\underline{m}$ or $\overline{m}$, depending on the direction of the deviation from $\widehat{m}$.

In this model, people exist who have a health status around $\widehat{m}$ and thus are almost equal in their health status, but they can then move in the opposite

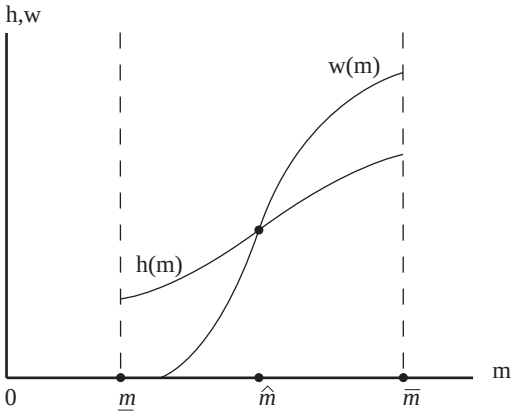

Figure 3: Dynamic model of Dasgupta

direction, dependent on the direction of the deviation from $\widehat{m}$.

If one discounts the future well-being of all people, which is denoted as $u(m)$, horizontal equity could be shown even in the presence of poverty traps. This is the reason why Dasgupta (1997) stresses that it is necessary that all life history be taken into account when looking at poverty traps.

As a stationary state, Dasgupta (1997) takes the energy balance condition explained in the introduction gets the same result as the static model:

$\beta = \Phi(w - r^*)$

where β is the number of tasks the person completes in each period, w is consumption and r^* is the maintenance expenditure plus energy expenditure (i.e. the whole energy needed to maintain the BMI status). $\Phi(w - r^*)$ is then the same nutrition-productivity curve as in the static models before.

2.3 Testable Hypotheses

The first obvious fact that one should see from the data is that individuals who are worse nourished than others, have a higher propensity to be unemployed. This is the prerequisite which has to hold if nutritional efficiency wages or poverty traps exist. The problem with this descriptive observation is that the direction of the relation is not clear. One has to adopt more sophisticated econometric methods to avoid the problem of endogeneity. "The incidence of both malnutrition and involuntary unemployment need to be endogenous. The model should identify which category of people will suffer from unemployment, particularly it should identify those people who will be denied access to work that pays enough to enable them to produce enough for an employer to wish to hire them in the first place." (Dasgupta and Ray, 1986).

To summarise the ideas coming out of the described theories I can formulate the following hypotheses.

- If there is a monopsony on the labour demand side, type 0 (landless) workers should receive higher wages than workers owning land. In return workers with land should be employed preferentially.
- If the demand side is competitive, the wage per efficiency hour should be the same for all workers. But workers owning land have a further possibility of consumption and thus they can generate a higher productivity, a higher hourly wage and again a higher consumption. The difference between hourly wages of landless workers and workers with land responds to productivity differences. Both types of workers are equally employed.
- Despite the structure of the demand side it is always true that no employer will pay other than the efficiency wage to a landless labourer. Theory states that there exists one and only one wage level w^* which is paid to type 0 workers without land. The reason is that labour supply of type 0 workers is always abundant, in the case of a monopsony or competition. w^* should even be similar over regions as long as type 0 labour is abundant.
- When the relationship between productivity and consumption is greater in the long run than in the short run, theory suggests that wages for long-term employed should be higher than that for short-term employed. Long-term contracts should exist more often than short-term contracts, because they are also favourable for the employers. A stable contract means for the employee a better nourishment that leads to a higher productivity.
- In a dynamic view, long term nutritional indicators such a height should lead to a higher productivity and in the steady state to a higher hourly wage.

2.4 Policy conclusions

The existence of nutritional efficiency wages gives rise to discussions about different policy measures on how to avoid them. As one policy action, a land reform that contains a redistribution can be carried out. Dasgupta and Ray (1986, 1987) show with Figure 4 how a land reform influences workers with and without land. Along the x-axis the workers are aligned with rising amount of land. The efficiency wage w* is a horizontal line because it is the same for all workers. $\underline{w}$ is the reservation wage for the workers which

is too high for very poor workers to get a job and also too high for very rich workers. The first are involuntarily unemployed, the last voluntarily. Type 1 workers with land must cede land to type 0 workers without land. Individuals to the left of A can generate a (higher) non-labour income and can thus supply labour to a lower wage. Individuals to the right of B generate a lower non-labour income after the reform, their reservation wage declines, too. Labour supply and output will rise altogether, because more individuals are able or willing to work. Workers without land that have also worked before the reform will possibly be more productive ceteris paribus because of the allocation of land. The equilibrium wage will sink so that individuals who worked before the reform and who did not profit from the land reform are now worse off than before the reform.

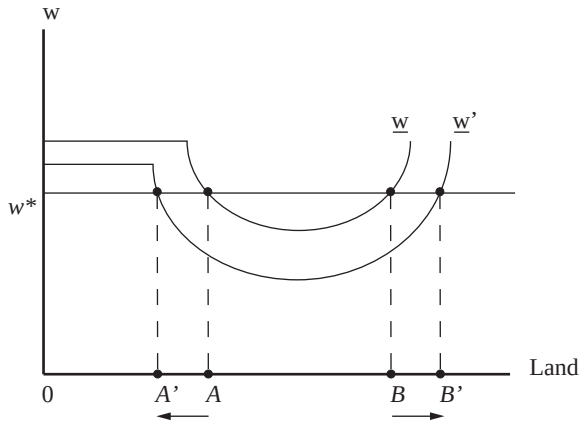

Figure 4: Land refom

A second policy reform can be a food program. This reform shifts point A so that workers who could not work before, can now supply work. Reservation wages and involuntary unemployment will fall while voluntary unemployment of the rich workers with much land will not be changed.

3 Existing empirical work

Existing empirical investigations assume diverse directions of the correlation of nutritional and labour market status and they also come to different conclusions from their estimations.

Many of the investigations aim at explaining undernutrition. As explanatory variables personal characteristics, the labour market status and the wage of the individuals are used (see for example Sahn (1994) for Côte d'Ivoire or Cigno and Rosati (1999) for India). Sahn (1994) finds in his investigation on the contribution of income on improved nutrition that raising expenditure

levels are a key factor to reduce chronic malnutrition of pre-school age children. Income is not a very important factor in reducing acute malnutrition of children. Cigno and Rosati (1999) investigate if the nutritional status of children depends on their work status. They find that the nutritional status of children working full time is better than of children not working. A review on the estimations of calorie-income elasticities is done by Bouis and Haddad (1992). They show that the relation has not been determined unequivocally because of the different approaches of the different investigations.

Another direction of exploring the issue at hand is to explain the influence of nutrition and health on income. I will follow this approach in my empirical investigation. Most of the articles concentrate on the estimation of several measures of health on wages, productivity and labour supply. As indicators for health, morbidities, anthropometric measures and food intake are used. Each of these indicators do have their own characteristics. Morbidities are often self-reported and contain reporting errors. These errors can be correlated with income (education) and lead to biased estimates. The property of anthropometric measures is that, for example, the BMI varies in the short run and BMI is correlated with maximal physical capacity independent from energy intake. Food intake is a good indicator to measure productivity in physically exhausting jobs. In the long run, the energy use of a body adjusts so that energy intake does not correlate immediately with productivity. The influence of health status on income depends on the kind of labour. Therefore one should explain employment and health status simultaneously.

There exist several empirical papers trying to explain the relation between nutrition and income. Behrmann and Deolalikar (1989) and Strauss and Thomas (1995) have written comprehensive surveys on the existing studies on nutritional poverty traps. I will only mention some studies that have taken into account the most common shortcomings, as they are mentioned by Deolalikar (1988): First, most empirical tests are made as simple correlation tests without controlling for other variables; second, only average, not marginal productivity is used as dependent variable; and third, unobserved individual-specific effects are not taken into account.

Strauss (1986) estimates an agricultural production function with instrumental variables estimation using data from a rural Sierra Leone cross-section survey of 1974-1975. The calorie availability of the family is one of the endogenous variables, as is hours of family and hired labour. Instruments are hired labour calories, several price indices, wages, household size and other variables. The estimated output elasticity of family caloric intake amounts to 0.34 at the mean and is statistically significant. The interesting point is the falling elasticity with rising family calorie consumption: at a very low caloric intake (1500 calories is the mean intake of the lower third of the sample) the

elasticity is 0.49, at the mean of the upper third at 4500 calories per day, it is only 0.12. Deolalikar (1988) criticises Strauss (1986) in using average calorie intake per consumer equivalent, because intra-household allocation of food could be correlated with the used instruments as for example price indices. Furthermore, calorie intake is described by Deolalikar (1988) as a bad proxy for energy availability.

Deolalikar (1988) uses ICRISAT VLS data of 1976/77 and 1977/78 to construct a panel of south Indian rural individuals and families. He estimates an individual-level wage equation and a household-level farm production function. For my purpose, the equation of interest is the wage equation. The wage equation is estimated in a semi-logarithmic form, taking as explanatory variables the average daily energy intake, the weight-for height and other common variables. The regression is made as panel regression, the author decides for the random effects specification after using the Hausman test . The main result is that there is no difference between the parameters for males and females, that daily calorie intake is insignificant and that weight-for-height as an indicator for medium-term nutritional status is significant and positive with a wage elasticity of 0.28 in the random effects specification. The endogeneity problem is solved by the introduction of a panel estimation, making the untested hypothesis that the endogeneity stems from the covariance of the nutritional status variable with unobserved time-invariant factors.

Haddad and Bouis (1991) and Foster and Rosenzweig (2000) use a stratified random panel of 448 farming households in the Philippines in the years 1984-85. Haddad and Bouis (1991) estimate a market wage function with long run indicators of nutritional status (height), medium run (BMI) and short run indicators (individual level calorie intake). The last two do not effect wages, but height possibly does, because it is correlated with education. The different estimates of the elasticity of height on wages center around one percent. Foster and Rosenzweig (2000) investigate the effect of calorie intake, BMI and height on market wages of farm work. They estimate firstly a BMI production function and then a wage regression with the dependent variable being piece rate minus time wages. They find that calorie intake is more rewarded under piece rates, while BMI is more rewarded under time wage. Because they suspect BMI and calorie intake to be endogenous, they use instrumental variables (owned land, household's sex- and age-composition, round dummies and land interacted with round variables).

A fundamental critique against using calorie availability as a measure for productivity comes from Swamy (1997). He calculates in his paper the calories that could be bought with different wage levels and finds that even the lowest wages would be enough to enable the workers to buy enough calories for their nutrition. He concludes that it can not be due to low

wages that the workers are unemployed. I will take BMI instead of calorie availability and argue that with BMI as the outcome of nutrition and health, it can be accounted to the productivity of the worker, even, if the worker could maintain a higher nutritional status with his wages.

To resume the results of the different studies, it seems to be important to take into account the endogeneity problem and also use panel data to account for individual-specific effects. The results of studies that investigate the influence of BMI on wages indicate that there is a positive relationship while studies which investigate the influence of wages on BMI indicate that there is more or less no influence.

4 Data

My study is based on the Côte d'Ivoire Living Standard Survey (CILSS). The survey has been conducted by the Direction de la Statistique in Côte d'Ivoire in cooperation with the World Bank. The data set is documented in Grootaert (1986). The Survey consists of a household, a community and a price survey. The household survey collected information on expenditures, income, employment, assets, basic needs and other socio-economic characteristics of the households and the individuals living in these households. The community survey collected data on economic and demographic characteristics of the rural communities to which each cluster of households belongs. The price survey collected data on prices at the nearest market to each cluster.

Each round of the survey covers a period of 12 months and is accessible from the World Bank for the years 1985-1988. The observations exist in a rotating panel, so that in every year half of the households (and most of the individuals living in these households) can be followed up into the following year. The sample size is 1600 households (15000 individuals) in each year. Thus, there are overlapping panels of 800 households each. With this construction of the data set, it is possible to use lagged variables and thus to take a first step in avoiding the problem of endogeneity of the relationship between nutritional status and labour market outcomes, or to use the whole panel of individuals to get a higher amount of observations and thus more precise estimation results. In this case, the endogeneity problem will have to be solved by instrumental variable estimation.

I first describe the data set I used. Variable names and their meaning are specified in Table 8 in the Appendix, statistics of the variables can be seen in Table 3. Variables to be explained in my later investigation are wages and a dummy variable which indicates whether a person is employed or not.

Table 3: Summary statistics

	Wage regression		Probit regression	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Lnwage	3.981	1.605		
Work			0.488	0.5
BMI	21.572	2.432	21.818	2.984
BMI squared	471.257	108.447	484.916	142.88
Height	1.637	0.084	1.623	0.086
Age	45.872	14.139	40.97	16.255
Age squared	2304.079	1322.25	1942.691	1447.931
Male	0.692	0.462	0.481	0.5
Married	0.854	0.353	0.734	0.442
Nationality	0.9	0.3	0.884	0.32
Other Cities	0.013	0.113	0.256	0.436
East Forest	0.377	0.485	0.277	0.447
West Forest	0.216	0.411	0.167	0.373
Savannah	0.395	0.489	0.3	0.458
Agriculture	0.912	0.283		
Ind. and Constr.	0.014	0.119		
Services, Transport	0.031	0.173		
Trade	0.043	0.202		
Distance to doctor	29.287	21.001		
Rain	124.3	36.96		
Land in Hectares	8.827	8.885	6.945	12.917
Lnhhincpc	11.64	0.766	11.646	0.808
Reading skills	0.13		0.205	0.403
Math skills	0.15		0.23	0.421
year 85	0.063	0.242	0.084	0.277
year 86	0.315	0.465	0.283	0.451
year 87	0.465	0.499	0.424	0.494
year 88	0.158	0.364	0.209	0.407
N	1787		6372	

The dummy variable whether a person is employed is constructed from the wage variable. Whenever an individual stated how much wages he earned, the variable "work" is set to one. As Table 3 shows, almost 50% of the individuals state to get wages.

The log of hourly wages is taken from the Question: "How much money did you get for your job?" which was posed after the question: "During the last 7 days, have you worked for anybody who is not a household member, for example an owner, a company, the state or another person who is not a household member?"

The wage variable is called Lnwage and exists for all individuals who have earned wages in their job. The problem of the wage variable is that

individuals stated the amount of money they got from the employer and the frequency so that I had to calculate hourly wages. Hourly wages are thus not accurate numbers, but can only be seen as an estimate of the wages. There is a huge variation in the wages stated in the different frequency cells, but the aggregated numbers fit well to other studies using these data and also to official numbers of Côte d'Ivoire (Table 7). Quite a lot of the wages are below the minimum wage which amounts to 1431.20 per day at the time of the interviews. The average sample wage which amounts to approx. 53.50 CFAF or 0.16 $ per hour is above the minimum wage. The approximate wage per day would be 2,380 CFAF, taking the average 7 hours per day as working time. But still in 1996, the minimum wage, which was only valid for the official sector, has not been enough to live from (World Bank, 1997).

Non-labour income Lnhhincpc is calculated as the log of the weekly income of the household excluding the wage of the observed individual divided by the number of all household members.

BMI means the Body mass index of the individual. The BMI is calculated as $(weight\ in\ kilos)/(height\ in\ meters)^2$. To account for nonlinearities, I include also BMI squared and Height in the regressions. Height is included as a long-term indicator for health, which should influence productivity positively (as mentioned as important in Dasgupta (1997)). The anthropometric measures were directly measured by the interviewer and the measurement was repeated during the second interview after one week whenever the numbers were not plausible. Also some individuals with plausible results were remeasured at random. The data taken for this investigation stem from the first round and were completed by the data of the second round whenever data of the second round were available and data of the first round were not plausible.

To measure the education of the individuals only very rough measures can be used if I want to have a consistent measure over the whole time span. More sensitive measures were either not available for all years or the number of observations was very small. Finally, I chose two dummies indicating whether the individual is able to read a newspaper (Reading skills) and whether he was able to do written calculations (Math skills).

Age is measured in years and individuals below 16 and above 65 years were excluded from the sample, because of problems in the measurement of the labour market status or the BMI. This results in a mean age of over 40 years in both samples.

Regional dummies were introduced to account for differences between urban and rural areas and between areas at the seaside and landlocked areas. Abidjan as the biggest city is excluded from the regressions. It has a unique

position because, for example, an investigation on land availability makes no sense in a large city. The other cities are still in my sample, because there exist many individuals owning land. An interesting observation is that almost 26% of observations stem from other cities in the regression on the probability to work, but only 1% of the individuals from other cities earn wages. As can be seen, individuals in the wage regression own at an average 2 hectares more land than individuals in the whole sample (6.9 to 8.8 hectares).

Further variables are the distance of the cluster the individual lives in to the next doctor and the land that household has to its disposal.

Because the study includes expenditure and wage data over time, I correct them with the Consumer Price Index of the IMF as presented in Grootaert (1993) and displayed in Table 7. The underlying assumption when using the CPI for the whole country is that regional price differences remain stable over the time period considered.

The CILSS data contain price variables for the clusters. But as Grootaert and Kanbur (1993b) demonstrate, the price index included is not appropriate. The problems encountered are described in more detail in Glewwe (1990, p.43, Appendix A). Prices in the CILSS dataset were collected for 18 food items and 4 non-food items. As the prices of the four non-food items vary more than could be expected (due to transportation costs), Glewwe (1990) uses the prices of canned tomato paste as a proxy for non-food prices. The price of canned tomato paste is the only price for a good which was "relatively non-perishable and found throughout the country" (Glewwe, 1990, p.43). Grootaert and Kanbur (1993b) argue that neither the price index of the CILSS dataset is appropriate nor the price index of Glewwe (1990) and they combine the CILSS price index with the data from the International Comparisons Project. The ICP collected prices in Côte d'Ivoire in 1985 and it contains around 20000 prices covering 912 product codes that have been harmonised with the CILSS categories by Grootaert and Kanbur (1993b). They arrive in a regional price index for 1985, which is prolonged over the whole time span until 1988 by Grootaert (1993). They had to assume "that regional price differences remained steady for the other years" (Grootaert, 1993, p.15), but they updated the expenditure shares for each year so that the Index changes from year to year.

5 Results

Following theory, two different outcomes are expected. First, nutritional status should influence the probability to work, because involuntary unemployment exists in the model. Unemployment is due to low productivity

levels corresponding to the undernourishment of individuals. Second, looking at employed workers, nutritional status should still influence the market by influencing productivity. Workers with better nutrition will be more productive and will thus receive higher wages.

5.1 Instruments for BMI

Theory tells that the influence of nutrition on labour market participation and wages is endogenous: Whenever a person is unemployed he or she will suffer from malnutrition and thus will have a harder time to find work than before. Thus I need instruments to account for the endogeneity problem of BMI and BMI^2. I use the distance of a doctor from the cluster the individual lives in, the rainfall over the year in the cluster and the land in hectares the individual has at his disposal as instruments. The doctor's distance is a proxy for the health status of the individual that influences wages only over the overall nutritional measure BMI. The rainfall and the land variables are used as instruments, because it is known from theory that land influences productivity via the nutritional status of the individual. The reasoning is that individuals with land are able to generate more consumption of food (from their land) than landless workers. Rainfall influences land quality and thus should also influence the nutritional status of the individual.

5.2 The probability to work

For a first look at the relation between nutrition and employment status, I show descriptive probabilities to be employed in the second year of each of the three panels dependent on employment and nutritional status in the first year of the panel in Table 9. Looking for a relationship over time takes into account the poverty trap reasoning of Dasgupta (1997) that it is harder for unemployed or undernourished (or both) to find work because of their destitution. Variance of the indicators is relatively high so that the differences are not statistically significant, but the shares show nevertheless a theoretically interesting picture.

If the sample is split into individuals who were well-nourished and undernourished (BMI above or equal and below 18.5) in the first year, it can be seen that the probability of well-nourished workers to be employed in the second year goes from 41% in the first round to 57.1% in the third round. Interestingly, the difference to workers who have been undernourished in the first year is reducing over the years. While in the first round, only 23.6% of undernourished were employed in the second year (which is half of the share of well-nourished that found a job), almost the same share of undernourished

as well-nourished had a job in the second year of the last round (51.3%).

Splitting the sample into individuals who were employed and those who were unemployed in the first year leads also to highly different values. Only between 19.5 and 27.5% of unemployed are able to get a job in the following year, while 72.3 to 86.4% of the employed are still employed one year later.

I can further split the groups into employment and nutritional status combined. There is not a big difference in finding a job for the unemployed whether they are well-nourished or not. The probability to find a job for well-nourished lies between 20 and 28%, while it is only 2 percentage points lower for undernourished. When looking at the unemployed, the difference in the nutritional status leads to a difference of almost 30 percentage points in the employment probability in the second year of the first round (74% for well-nourished vs. 46.7% for undernourished). In the last round, the probability to find a job is almost equal at around 85% regardless of nutritional status.

It can be stated that the probability to work is always higher for well-nourished workers relative to undernourished workers, regardless of their labour market status in the first year, but first year employed workers will also more likely work in the second year than in the first unemployed year. The employment effect is even higher than the nutrition effect, finding a first hint to the existence of long-term contracts that have been found to be productivity-pushing in the theoretical model.

Table 4: Probability to work - Probit results

	(1)	**(2)**	**(3)**	**(4)**
Variable	**work**	**work**	**BMI**	**BMI squ.**
BMI	0.124**	3.279		
	(2.62)	(0.73)		
BMI squared	-0.002*	-0.070		
	(2.26)	(0.64)		
Height	-0.071	-1.046	-1.526*	-87.876**
	(0.23)	(0.38)	(2.42)	(3.01)
Lnhhincpc	0.275**	0.181**	0.130*	5.602*
	(10.46)	(3.21)	(2.19)	(2.04)
Age	0.147**	0.129**	0.023	1.023
	(18.28)	(8.86)	(1.56)	(1.51)
Age squared	-0.002**	-0.001**	-0.001**	-0.024**
	(16.95)	(6.20)	(3.51)	(3.25)
Male	1.010**	0.992**	0.442**	18.037**
	(18.42)	(7.44)	(4.06)	(3.58)

Continued on next page...

... table 4 continued

	(1)	(2)	(3)	(4)
Variable	**work**	**work**	**BMI**	**BMI squ.**
Married	0.433**	0.398*	0.435**	19.081**
	(7.84)	(2.34)	(4.05)	(3.84)
Nationality	-0.083	-0.260*	0.178	7.792
	(1.26)	(2.04)	(1.15)	(1.08)
Read Newspaper	-0.072	0.084	0.053	1.873
	(0.58)	(0.42)	(0.19)	(0.15)
Math skills	0.003	-0.175	0.216	10.576
	(0.03)	(0.65)	(0.81)	(0.86)
Other Cities	0.116+	0.632*	-0.436	-18.065
	(1.86)	(2.40)	(1.21)	(1.08)
East Forest	0.101+	0.157	0.186+	9.946+
	(1.73)	(0.62)	(1.67)	(1.93)
West Forest	-0.409**	-0.503	0.555**	25.373**
	(6.49)	(1.52)	(4.49)	(4.43)
year 86	0.079	0.044	0.497**	21.285**
	(1.28)	(0.26)	(3.28)	(3.03)
year 87	0.229**	0.319**	-0.032	-1.808
	(3.50)	(2.86)	(0.21)	(0.26)
year 88	0.543**	0.670**	-0.295	-14.205
	(7.47)	(2.74)	(1.58)	(1.64)
Land	0.002		0.006	0.243
	(1.45)		(1.39)	(1.16)
Rainfall			0.002+	0.100+
			(1.74)	(1.66)
Distance to doctor			-0.004+	-0.145
			(1.79)	(1.56)
Constant	-8.787**	-42.071	21.434**	498.824**
	(10.99)	(0.99)	(17.19)	(8.63)
N	6372	3884	3884	3884
Log-likelihood	-3383.7			
$\chi^2_{(18)}$	1191			
(Pseudo-)R-squared	0.2326		0.0608	0.0526

Std. errors corrected for correlation between obs. of the same individual

Column: (1) Probit, (2) Probit-IV, (3) and (4) OLS

Significance levels : + : 10% * : 5% ** : 1%

For the explanation of the probability to be employed in the sense of the

static idea of nutritional efficiency wages with involuntary unemployment, the whole sample has been pooled, so that for most individuals, two subsequent observations can be used. Standard errors are corrected for the correlation between observations of the same individual. To explain the employment probability, a probit estimation is used. Explanatory variables in addition to the BMI and BMI squared are personal characteristics such as age, marital status, sex and a proxy for education, time dummies and regional dummies to control for differences between urban and rural areas and between coast and landlocked areas. To control for productivity differences, I further introduce the availability of land as an additional explanatory variable. Per capita household income of the household members other than the observed person is used as another measure influencing productivity and thus labour market chances. Table 4 shows the results of the pure probit estimation in column (1), probit with instrumented BMI and BMI squared in column (2) and the two instrumental regressions in columns (3) and (4). I only included cities other than Abidjan and the rural regions as described in the data section.

Overall, the estimation results show the expected result, the indicators for the goodness of the estimation show a relatively high explanatory power of the model (with an $Pseudo - R^2$ of 23% in the case of column (1)). Also, many single parameters show a significant effect with the expected direction.

Looking at the influence of the body mass index, which is our most important factor of influence, it can be seen that it shows a significant impact on the probability to be employed in the probit model, but only, when the endogeneity problem is not taken into account. The instrumented anthropometric measures loose their influence on the probability to work. Height as a long term nutritional indicator does not show a significant influence. From the dynamic model of poverty traps (Dasgupta, 1997), we would have expected to find a positive sign of height on the probability to work.

Regarding the personal characteristics, men's work probability is higher than that of women. One further can observe, that the probability to work increases with age with diminishing intensity and that married individuals are also more probable to work than singles. Nationality only becomes significant in the instrumented regression in column (2) and indicates that foreign workers work more likely than Ivorien. The educational variables Reading and math skills do not show a significant impact in the estimations.

Regarding the regions, Savannah is the left out category. Only West Forest shows a lower opportunity to find work, in Other Cities and East Forest, more people work.

Finally, I introduce two variables that measure the non-wage income of the workers: the availability of land and the per capita household income of

the remaining household members. The land variable shows no direct effect on the probability to work and it also is not significant in the regressions in columns (3) and (4), which are used to calculate the instrumented variables BMI and BMI squared. Theory suggests that workers owning land should generate a productivity advantage from their land and should then be able to find work more easily. This result cannot be confirmed by the estimations. But, as a more general indicator of non-labour income, the log of per capita household income shows a significant positive effect on the probability to work and on the BMI. The positive effect on the probability to work is a strong result, as in many studies on labour market participation, a substitution effect has been found: With a higher income of the partner, the necessity to earn own income is reduced. My result suggests that the income effect of labour market participation is higher than the substitution effect of non-labour income.

As the results of Table 9 have shown a huge (though not statistically significant) difference among the groups of undernourished and unemployed in the different combinations, I try to find a relation between the labour market status of today and the employment and nourishment status of the year before. The result of the probit regression is shown in Table 5. $Work_{t-1}$ is now the lagged employment status (being one if the individual was employed in the year before), BMI_{t-1} is the lagged BMI and $Work * BMI_{t-1}$ is the interaction between the two. One can see as before in Table 9, that the most important factor explaining employment today is employment last year. The probability to work is also higher with rising BMI, but not in any economically important way. The joint effect is not significant. The rest of the parameters do not show new insights into the relation between the explanatories and the probability to work, so I will not discuss the parameters further.

5.3 The wage equation

To look at the problem of efficiency wages I estimate a wage regression on the log of hourly wages. The idea is that workers who are better nourished are more productive and will get higher wages. Table 6 contains two wage regressions in column (1) and (2), explaining the log hourly wages of the individuals with most individuals being observed twice. Standard errors are corrected for the correlation between observations of the same individual. Column (1) shows the OLS results without respecting endogeneity of BMI and BMI^2, column (2) contains an instrumental variable estimation with the instrument equations shown in columns (3) and (4).

As instruments, the distance of a doctor from the cluster the individual

Table 5: Probability to work - Probit results with lag

Variable	Coefficient	(t-value)
$Work_{t-1}$	1.311**	(3.49)
BMI_{t-1}	0.019+	(1.68)
$Work * BMI_{t-1}$	-0.011	(0.63)
Height in meters	-0.520	(1.29)
Lnhhincpc	0.204**	(5.38)
Land in Hectars	0.002*	(1.99)
Age	0.104**	(10.45)
Age squared	-0.001**	(10.08)
Male	0.826**	(11.45)
Married	0.325**	(4.95)
Nationality	-0.079	(0.89)
Read newspaper	0.049	(0.31)
Math skills	-0.023	(0.15)
Other Cities	-0.041	(0.52)
East Forest	-0.104	(1.38)
West Forest	-0.375**	(4.29)
year 86	-0.433**	(5.83)
year 87	-0.384**	(6.42)
Constant	-4.785**	(5.79)
N	3196	
Log-likelihood	-1515.8	
$\chi^2_{(18)}$	1398.5	
Pseudo-Rsqu.	0.3157	
Significance levels :	† : 10%	∗ : 5% ∗∗ : 1%

lives in, the rainfall over the year in the cluster and the land in hectares the individual has at his disposal are used as before. As column (3) and (4) show, that only rainfall and Land in Hectares do have a positive and significant impact on BMI; rainfall has also a significant effect on BMI^2. The distance to the next doctor does not influence BMI, but it shows the correct (negative) sign.

BMI and BMI^2 only influence the earnings in the instrumented equation and then only at the 10 percent significance level. The economic effect of BMI shows a significant influence on wages in the IV regression. The influence of BMI on hourly wages (not in logarithms) is displayed in Figure 5 with the other explanatories set to the mean (dummies are set to zero): Severe undernutrition leads to a wage of approximately zero that cannot be accepted by any worker. With a rising BMI-level, the wage increases exponentially. From a BMI of 22 upwards, wages fall again. This is perfectly in line with theory which predicts high increases in productivity at a low level and decreasing returns at a higher level of nutrition (Graph 1). Differences

between men and women are not significant.

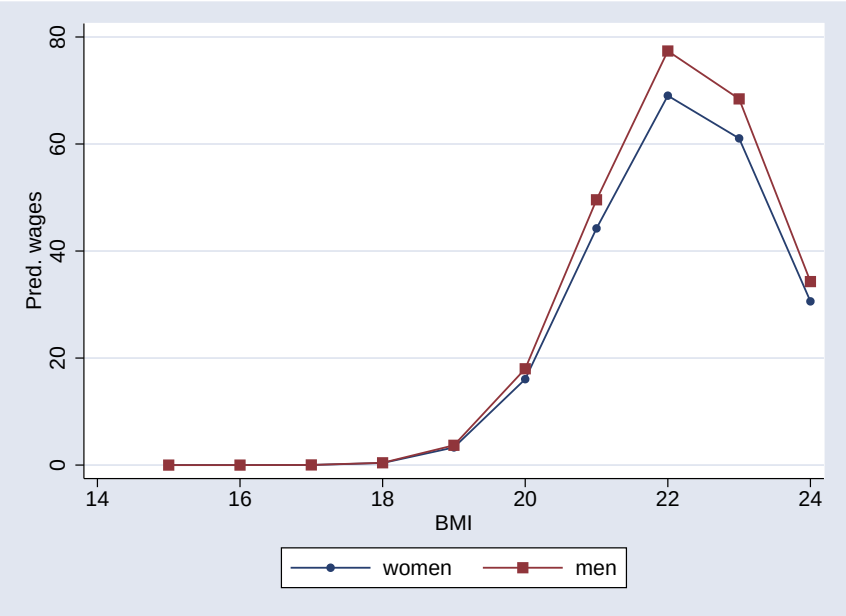

Figure 5: Predicted log wages dependent on BMI level

In the description of the results of the remaining coefficients, I only rely on the IV estimation. Height as a long run indicator of nourishment shows a positive and significant influence on wages as expected from the theory on poverty traps (Dasgupta, 1997). Older workers earn more than younger, but only with an economically small effect and without diminishing returns as in standard results. Marital status and nationality do not matter for income. The sectors trade and services and transport pay higher wages than the reference sector agriculture while wages in industry and construction are not different from agricultural wages. This result is in line with standard results on wage determinants. Other Cities, West Forest and Savanna (the left-out category), the latter two being the more landlocked areas, which do not include major cities, are doing worse than East Forest which includes some major cities. The education measures were again not significant, only in the OLS regression in column (1) reading skills are positive as could be expected.

Table 6: Determinants of Wages

	(1)	(2)	(3)	(4)
Variable	**lnwage**	**lnwage**	**BMI**	**BMI squ.**
Lnhhincpc	0.594**	0.556**	0.022	1.436
	(9.42)	(4.66)	(0.26)	(0.38)
BMI	0.030	12.890+		
	(0.33)	(1.66)		
BMI squared	-0.000	-0.289+		
	(0.00)	(1.67)		
Height	0.730	3.617*	-0.148	1.723
	(1.09)	(2.06)	(0.17)	(0.04)
Age	0.042*	0.100*	-0.006	-0.138
	(2.47)	(2.38)	(0.25)	(0.13)
Age squared	-0.000*	-0.001	-0.000	-0.010
	(2.31)	(1.50)	(0.86)	(0.86)
Male	1.319**	0.056	0.590**	21.599**
	(10.67)	(0.07)	(3.64)	(2.95)
Married	0.264*	0.214	0.187	8.055
	(2.18)	(0.79)	(1.09)	(1.04)
Nationality	0.336*	0.480	0.253	12.283
	(2.44)	(1.27)	(1.24)	(1.33)
Reading skills	0.585*	-0.051	-0.487	-21.435
	(2.44)	(0.11)	(1.24)	(1.20)
Math skills	-0.164	0.230	0.437	19.659
	(0.77)	(0.53)	(1.17)	(1.16)
Other Cities	0.213	0.321	-0.811+	-37.896+
	(1.45)	(0.36)	(1.72)	(1.78)
East Forest	0.250**	0.414+	0.185	8.865
	(2.60)	(1.90)	(1.22)	(1.30)
West Forest	0.302**	0.512+	0.559**	24.972**
	(3.05)	(1.70)	(3.08)	(3.04)
Ind. and Constr.	0.403	1.252	-0.494	-14.920
	(1.33)	(0.74)	(0.94)	(0.63)
Services, Transport	1.575**	3.883*	1.766**	85.669**
	(7.16)	(2.41)	(5.41)	(5.80)
Trade	1.734**	2.514**	1.136**	53.147**
	(9.85)	(3.38)	(4.15)	(4.29)
year 86			0.670**	30.377**
			(2.86)	(2.86)
year 87			-0.051	-2.008
			(0.22)	(0.19)

Continued on next page...

... table 6 continued

	(1)	(2)	(3)	(4)
Variable	**lnwage**	**lnwage**	**BMI**	**BMI squ.**
year 88			-0.294	-12.418
			(1.07)	(1.00)
Rainfall			0.005**	0.246**
			(2.84)	(2.89)
Land in Hectares			0.010+	0.425
			(1.72)	(1.57)
Distance to doctor			-0.003	-0.149
			(1.19)	(1.17)
Constant	-7.404**	-154.381+	20.567**	403.011**
	(4.33)	(1.74)	(11.57)	(5.02)
N	3033	1787	1787	1787
R^2	0.27		0.09	0.09
F	33.76	7.16	8.37	8.10

Reference categories: year 85, Savannah, not married, female, foreign, agriculture
Std. errors corrected for correlation between obs. of the same individual
Column: (1), (3) and (4) OLS, (2) IV
Significance levels: + : 10% * : 5% ** : 1%

What can be done to improve the status of undernourished individuals in Côte d'Ivoire? The policy to distribute land to the poorest to help them increase their productivity in the labour market was mentioned at the end of the theory chapter. Being the correct measure when looking at the theory, I cannot fully agree on basis of the regression results. Non-wage income in general influences BMI and also the labour market indicators directly, but the indirect influence on the labour market outcome over BMI is weak. With the results at hand, anthropometric measures seem to matter more in determining the productivity of a worker, once he found a job, less in helping him to find employment.

6 Conclusions

The aim of the paper was to find nutritional efficiency wages and a poverty trap in a country of Sub-saharan Africa - Côte d'Ivoire. While Côte d'Ivoire is relatively rich in comparison to its neighbours, poverty is still very high. During the period under consideration from 1985 to 1988, government spending was cut down and the prices in the agricultural sector rose. Adjustments

were further undertaken to save money in the public sector: Wages of employees were frozen so that they decreased with inflation and many workers were laid off. But looking at the probit results with lagged explanatories, unemployment is one of the worst things that can happen to a worker. Once a worker is laid off it is very hard for him to get into a new job. The worsening labour market situation of the unemployed comes about because of the destitution of the workers that cannot buy enough food. With respect to the model, all the mentioned actions undertaken by the government to enhance growth and lower poverty are not helpful to generate workers who are able to help themselves out of poverty.

Regarding the policy question on how to achieve the goal to reduce poverty and hunger as formulated in the International Development Goals, the answer from my regression is that ex ante well-nourished workers do have an advantage in their wage negotiations, but they do not have an advantage in finding a job. Finding a job and increasing wages are both achieved best if the non-labour income of the individual is increased. The ownership of land does not per se help to increase the BMI, so a land reform as explained in the theory would only help to improve the situation of the landless if non-labour income can be effectively increased.

References

Behrmann, J. R., and A. B. Deolalikar (1989): "Health and Nutrition," in *Handbook of Development Economics, Vol. 1*, ed. by H. Chenery, and T. N. Srinivasan, chap. 14. North-Holland, Amsterdam.

Bliss, C., and N. Stern (1978): "Productivity, wages and nutrition Part I: The theory," *Journal of Development Economics*, 5, 331–362.

Bouis, H. E., and L. J. Haddad (1992): "Are estimates of calorie-income elasticities too high? A recalibration of the plausible range," *Journal of Development Economics*, 39, 333–364.

Cigno, A., and F. C. Rosati (1999): "Why do Indian children work and is it bad for them?," mimeo.

Dasgupta, P. (1997): "Nutritional status, the capacity for work, and poverty traps," *Journal of Econometrics*, 77, 5–37.

Dasgupta, P., and D. Ray (1986): "Inequality as a determinant of malnutrition and unemployment: theory," *Economic Journal*, 96, 1011–1034.

——— (1987): "Inequality as a determinant of malnutrition and unemployment: policy," *Economic Journal*, 97, 177–188.

Deaton, A. (1997): *The analysis of household surveys - A microeconometric approach to development policy*. John Hopkins University Press, Baltimore and London.

Deolalikar, A. B. (1988): "Nutrition and labor productivity in agriculture: Estimates for rural south India," *The Review of Economics and Statistics*, 70, 406–413.

Foster, A. D., and M. R. Rosenzweig (2000): "A test for moral hazard in the labor market: Contractual arrangements, effort, and health," in *Readings in Development Economics, Vol. II, Empirical Microeconomics*, ed. by P. Bardhan, and C. Udry, vol. 2, chap. 6, pp. 171–196. The MIT Press, Cambridge, Mass.

Foster, J., J. Greer, and E. Thorbecke (1984): "A class of decomposable poverty measures," *Econometrica*, 52, 761–65.

Glewwe, P. (1990): "Investigating the Determinants of Household Welfare in Côte d'Ivoire," *LSMS Working Paper, No. 71*.

Grootaert, C. (1986): "Measuring and Analysing Levels of Living in Developing Countries - An Annotated Questionnaire," *LSMS Working Papers, No. 24*.

——— (1993): "The Evolution of Welfare and Poverty Under Structural Change and Economic Recession in Côte d'Ivoire, 1985-88," *Policy Research Working Paper, No. 1078, Africa Technical Department, The World Bank*.

Grootaert, C., and R. Kanbur (1993a): "The Lucky Few Amidst Economic Decline - Distributional Change in Côte d'Ivoire As Seen Through Panel Data Sets, 1985-88," *Policy Research Working Papers, No. 1113, Country Department, The World Bank*.

——— (1993b): "A New Regional Price Index for Côte d'Ivoire Using Data from the International Comparisons Project," *Policy Research Working Paper, No. 1080, Africa Technical Department, The World Bank*.

Haddad, L. J., and H. E. Bouis (1991): "The impact of nutritional status on agricultural productivity: wage evidence from the Philippines," *Oxford Bulletin of Economics and Statistics*, 53, 45–68.

IMF, OECD, United Nations, and World Bank (2000): "A Better World for All," http://www.paris21.org/betterworld.

Leibenstein, H. (1957): *Economic Backwardness and Economic Growth: Studies in the Theory of Economic Development*. John Wiley, New York.

Mirrlees, J. A. (1975): "A pure theory of underdeveloped economies," in *Agriculture in development theory*, ed. by L. G. Reynolds. Yale University Press, New Haven and London.

Ray, D. (1998): *Development Economics*. Princeton University Press, Princeton, New Jersey.

Sahn, D. E. (1994): "The contribution of income to improved nutrition in Côte d'Ivoire," *Journal of African Economies*, 3(1), 29–61.

Stiglitz, J. E. (1976): "The efficiency wage hypothesis, surplus labour and the distribution of income in LDC's," *Oxford Economic Papers*, 28, 185–207.

Strauss, J. (1986): "Does better nutrition raise farm productivity?," *Journal of Political Economy*, 94(2), 297–320.

Strauss, J., and D. Thomas (1995): "Human Resources: Empirical Modeling of Household and Family Decisions," in *Handbook of Development Economics*, ed. by J. Behrmann, and T. Srinivasan, chap. 34, pp. 1883–2023. Elsevier, Amsterdam.

Swamy, A. V. (1997): "A simple test of the nutritionbased efficiency wage model," *Journal of Development Economics*, 53, 85–98.

United Nations (2000): "United Nations Millenium Declaration," http://www.un.org/millennium/declaration/ares552e.pdf.

World Bank (1997): "Poverty in Côte d'Ivoire - A Framework for Action," *Report No. 15640-IVC, Human Development Technical Group III, Côte d'Ivoire Country Department, Africa Region.*

7 Appendix

Table 7: Macro Data for Côte d'Ivoire

	1985	1986	1987	1988
HH cons. exp., billions of CFA Francs	1836.4	2007.8	2041	1971.1
Ch. in cons. prices	1.86	9.68	6.94	6.93
Exports of Goods and Services	1466.3	1252.7	1013.5	788.7
GDP	3137	3244	3118	3067
Pop. in mill.	10.46	10.88	11.29	11.7
GDP per capita	299.90	298.16	276.17	262.14
Exchange Rate	449.3	346.3	300.5	297.9
GDP Growth	4.9	3.4	-1.6	-2
Cons. Growth	4.1	13.1	-7.1	-16.6

Source: IMF, World Bank (1997); Grootaert (1993)

Table 8: Description of the variables used

Name of the variable	**Description**
Bmi	Body Mass Index measured as $weight/height^2$
Height	Height in meters
Regional Dummies	Dummies for Other cities, East Forest, West Forest and Savanna
Married	One, if person is married
Age	Age in years
Age^2	Age in years squared
Male	One, if person is male
Nationality	One, if person is Ivorien
Education	Dummies for: Being able to read a Newspaper and Being able to do written calculations
Land in ha	Hectares of land used by the household during the last 12 months
Lnhhincpc	Log of weekly household income (from Survey) per capita (excluded the observed worker)
Lnwage	Log of wages earned, transformed to hourly wages (including non-monetary payments)
Work	One, if person has stated wages
Sector dummies	Dummies for: agriculture; industry and construction; services, transport and hotel; trade
Rainfall	Rainfall in cm during the last year in the area (areas can include several clusters)
Distance to doctor	Distance from the cluster the family lives in to the next doctor

Table 9: Mean Probability to be employed in the second year

First year status	**First round**	**no. of cases**	**2nd round**	**no. of cases**	**3rd round**	**no. of cases**
well-nourished	41.0	553	47.7	1266	57.1	1112
undernourished	23.6	72	40.4	89	51.3	115
unemployed	19.5	394	24.6	724	27.8	625
employed	72.3	231	73.2	631	86.4	602
unempl. and well-nourished	19.9	337	24.7	671	28.2	561
unempl and undernourished	17.5	57	22.6	53	25.0	64
empl. and well-nourished	74.1	216	73.6	595	86.6	551
empl and undernourished	46.7	15	66.7	36	84.3	51

Differences are not statistically significant

Curriculum Vitae

Employment

- April 2005 - today: Lecturer, Department of Economics, University of Erlangen-Nuremberg
- July 2004 - April 2005: Research Assistant, Dep. of Economics, University of Technology Darmstadt
- September 2003 - November 2003: Marie-Curie Fellow, University College Dublin
- March 2001 - July 2004: Research Assistant, Dep. of Economics, Ludwig-Maximilians University Munich
- February 1999 - March 2001: Economist, DIW Berlin, Department International Economics

Education

- 2005 Ph.D. in Economics (Dr. oec. publ.), University of Munich
- 2001 - 2005 Ph.D. Programme in Economics, University of Munich
- 1998 Diplom, University of Munich
- 1993 Abitur, Ignaz-Günther Gymnasium Rosenheim