# Wage inequality and gender wage gap in Hungary,

1992-1997

By: Zsuzsanna Gulybán

Submitted to

Central European University Department of Economics

In partial fulfillment of the requirements for the degree of Masters Arts

Supervisor: John E. Sutherland

Budapest, Hungary

2007

## Abstract

The aim of the paper is to investigate how wage inequality evolved in Hungary between 1992 and 1997 and what effect the composition of change in inequality may have on the estimation of gender wage gap. The change in the composition of the work force may lead to some bias in the estimation of the overall residual variance (spurious growth) and the residual variance free of this bias measured with Lemieux's method (2006) is lower for both genders. According to my results this effect is stronger for women, which makes it probable that some components of gender wage gap decomposed with the Juhn-Murphy and Pierce (1993) method have different weights as usually assumed.

# Table of contents

Abstract	ii
Table of contents	iii
Introduction	1
Wage inequality	2
Tendencies in Hungary	4
Accounting for composition effects á la Lemieux	7
Data and trends in within-group inequality by skill groups	11
Change in the gender wage gap	23
Juhn-Murphy-Pierce method	23
Results	27
Conclusion	29
Literature review	32

## Introduction

This paper has two aims. First is to investigate how wage inequality evolved after transition in Hungary and which factors dominated this process. Second is to estimate gender wage gap considering that what possible biases may emerge from the estimation method used. In the most recent literature (Lemieux, 2006) both the effect of changes in the price of unobserved skills and the effect of the change in the composition of the work force are considered when residual wage inequality is estimated. The author claims that a better measure of inequality can be attained by omitting the composition effect from the overall residual inequality.

The importance of finding a more precise estimation method of residual inequality comes from its role in estimating gender wage gap. Juhn, Murphy and Pierce (1993) control for the effects of possible changes in residual inequality but they estimate it with a less advanced method: they do not control for the effects of changes in the composition of the work force; they consider only the measured change in inequality. Using the residual variances obtained by Lemieux (2006) may lead to different measures of gender wage gap.

In this paper the residual wage inequality is estimated and then the change in the gender wage gap in Hungary between 1992 and 1997 with a consideration that how residual variance obtained by controlling for composition effect would change our results.

The outline of the paper is as follows: first the inequality related tendencies in Hungary are reviewed, and then follows the examination of the theory of accounting for composition effects, after comes the data investigation. The second part of the paper focus on the gender wage differences, with a short review of tendencies, the discussion of the decomposition method of Juhn, Murphy and Pierce(1993) and finally the results. The last section focuses on the common points on the two fields and on the possible effects of incorporating the inequality estimating method into the gender wage gap estimating one.

1

## Wage inequality

Most Central European countries experienced a sharp increase in both income and wage inequality during and after the transition from socialism to capitalism (Newell, 2001). With market liberalization, Hungary has gone through the same process (Tóth István György, 2003). However, there was some debate over its scale compared to other transition countries. World Development Report of World Bank (1996) claimed that Hungary had the smallest income inequality increase in the region, but local researchers did not agree (Andorka, Ferge, Tóth, 1997). Nevell (2001) also supports Andorka et al.

Explaining wage with human capital related variables like education and experience has very developed traditions since Mincer (1974); consequently it is straightforward to base investigation of inequality in wages on these factors. These variables are not able to explain all the variance in wages and variance of the residuals (inequality in the residuals) is the subject of several studies.

According to Juhn, Murphy and Pierce (1993) and Katz and Autor (1999) even within narrowly defined experience and education groups there was increasing inequality from the 70s till the 90s in the US. This residual inequality is supposed to be caused by different factors: Juhn, Murphy and Pierce (1993) interpret it as the result of increased returns to the unobservable skill components which is the result of increase in the demand for skills. Others argue that the role of minimum wage is considerable also (DiNardo, Fortin, Lemieux (1996), Lee (1999) and Teuling (2002)).

Lemieux (2006) emphasizes two additional factors. On the one hand he claims that change in the composition of work force had a major impact. Even Mincer (1974) discusses that the conditional distribution of the error term in the mincerian equation is not necessarily homoscedastic; residual wage dispersion generally increases both in experience and in education. The reason is that different level of investments into on-the-job training of individuals makes different level of wages probable. Those who are willing to devote more time and effort to on-the job training are used to accept lower earnings initially and later get much higher earnings than the average. This leads to steeper experience-earnings profile than the average. With mincerian equation the average is measured, so higher experience can lead to higher dispersion of residual wages (Mincer (1974), more recently Chay and Lee (2000)). One other reason for increasing inequality in experience can be the learning ability of market: more experience means more available information for the market on the productivity of the worker (Farber and Gibbons, 1996).

Moreover, increase in the education level causes higher dispersion because of the self selection of those who invest into schooling: they should have higher marginal returns to education. That is why log-wage-schooling relationship is generally convex (Mincer, 1974, 1997, Rosen 1977) which means that the labor market price of schooling will be higher at higher levels of education. This may lead to higher level of dispersion also (Lemieux, 2002).

However Lemieux (2006) claims that the increasing average age of employees and more years spent in school in average must be the main reason of increasing residual inequality in the U.S. in the last 30 years. This is what he calls *composition effect*.

The other factor that he emphasizes is the possible increase in measurement error.

In this paper I will investigate how residual inequality evolved in Hungary after transition, from 1992 to 1997. The analogy for the first factor is more or less straightforward: the question is what kind of composition effect can be identified in Hungary during these years. The variables of interest changed in opposite direction: average education of employees slightly increased while average age strongly decreased.

Unfortunately there is not much possibility for estimating the change in the measurement error. Lemieux (2006) uses 2 different databases and makes the conclusion of increasing measurement error by comparison. For carrying out the same research work for

3

Hungary, a database with hourly wages measured would be inevitable. Unfortunately such a database is not publicly available at the present.

## Tendencies in Hungary

This period is rather special compared to the US. Most transition countries coped with increasing unemployment in the beginning of 90s, and mainly older and less educated people lost their job. Following the mid-nineties job destruction stagnated and job creation started to increase very slowly. Meanwhile, returns to education increased a lot and experience acquired in socialism became less worthy (Campos and Jolliffe, 2004a, Galasi and Varga, 2005). This increase in the demand for more educated people induced an expansion of higher education, but the effect of the jump in the number of more educated persons is negligible for the whole work force. Altogether, employers strongly preferred young and highly educated employees (Galasi, Varga, 2005).

Considering wages instead of employment status these tendencies are not so obvious because the wage of those who lost their jobs cannot be measured. Although focusing on the changes in the demographic characteristics of those who are employed reflect these processes. This is the case in my data also.

On Table 1A the average age of workers and average years spent in school can be seen in the starting and ending periods under investigation, and also the change between them. The average age of men employed decreased by 1.2 years during these 7 years. For women the change is not so huge: a little less than a year, 0.9 year. This can be due to the preference of younger employees by employers (the supply of workers did not become younger as in Hungary the society is ageing). Education increased for both groups, but only slightly. The possible reasons are that the employers prefer more educated fork force, or the increase of the proportion of more educated labor supply. For women the increase in the years spent in school is 3 times the increase for men. This might be due to the lower opportunity cost of learning or to the stronger discrimination for less educated female work force.

#### Table 1A.

#### Average age and years spent in school for workers.

	men			women		
	1992-93	1996-97	change	1992-93	1996-97	change
age	38.519	37.284	-1.235	38.272	37.339	-0.933
years in school	11.649	11.757	0.108	11.679	11.980	0.301

On Table 1B the percentage distribution of workers by education and experience groups can be seen. The education and experience categories used here are described more thoroughly later. We can see that the percentage of workers with the lowest level of education (primary school or less) decreased by a quarter and all other categories above have extended for both male and female. The only exception is bachelor's or master's degree owner male workers. The percentage share of this group dropped, which is a puzzle. For men, the percentage of those increased the most that have already finished high school and for women, the most educated group extended the most.

#### Table 1B.

	men		W	omen
	1992-93	1996-97	1992-93	1996-97
A. education categories				
primary school	16.53	12.69	22.42	16.86
vocational school	39.37	40.87	19.09	20.02
high school	25.44	28.48	36.37	38.22
ba and ma degree	18.65	17.96	22.12	24.90
B. years of experience				
0-10	20.01	25.28	20.84	23.56
10-20	28.93	27.86	26.27	27.68
20-30	31.05	29.72	33.10	35.34
30+	24.14	17.13	19.79	13.41

### Percentage distribution of workers by education and experience groups.

Experience shows similar pattern: the share of most experienced group decreased a lot in favor of those with less experience for both genders. For men only the category with less than 10 years of experience expanded, and all other experience groups shrunk. Women at the edges of the experience range smooth in the tendencies: more younger, much less older people employed, but the share of women in between also increased.

These tendencies are very different from those phenomena in the US. on which Lemieux (2006) focuses. In the US. the share of more experienced and more educated workers increased from years to year. The fact that both variables changed in the same direction and the residual wage inequality is positively related to both of them makes the estimation of composition effect more straightforward. In Hungary the changes in the characteristics happened in the opposite direction, opposite to each other. Presumably their composition effect on residual wage inequality is opposite also: lower average age should decrease while higher average education should increase the composition effect. By Lemieux's approach which is shown in the following, only the overall composition effect can be estimated, without any regard to the individual causal relationship with any of the 2 variables.

### Accounting for composition effects á la Lemieux

In the following a there is a brief discussion of how to control for composition effects according to Lemieux (2006). He explains two methods of measurement.

Both of them focus on the change in the variance of the residual. The first is about dividing the sample into a given number of cells and controlling for both the change in the within group variance and the change in the composition of the work force along these cells. The other is about estimating an alternative residual variance where the effect of change in the composition of work force is omitted by calculating an appropriate weight for each residual.

A short discussion of these approaches will follow later on.

The basis of the estimation is the Mincer-type wage equation:

$$w_{it} = x_{it}b_t + \varepsilon_{it} \quad (1)$$

where  $w_{ii}$  is the natural logarithm of the hourly wage rate, of individual *i* at time *t*,  $x_{ii}$  is a vector of observed skills, *b* is the return to observed skills,  $\varepsilon_{ii}$  is the standard residual. The inequality measured in this residual is what Lemieux calls "residual wage inequality". This residual is the product of some unobserved skills,  $e_{ii}$  and its price<sup>1</sup>:

 $\varepsilon_{it} = p_t e_{it}$  (2)

$$Var(\varepsilon_{it}) = p_t^2 Var(e_{it}) \quad (3)$$

<sup>&</sup>lt;sup>1</sup> And the measurement error added which is not included here as it is not used in the rest of the model. For the original model see Lemieux (2006).

From (3) it is obvious that changes in the residual variance can be due to changes in the price of unobserved skills or change in unobserved skills themselves.

Composition effects can be accounted by the following way. Consider a case where observed skills  $x_{ii}$  are divided into a finite number of cells, *j*. The unconditional variance of unobserved skills  $Var(e_{ii})$  is linked to the conditional variance:

$$Var(e_{it}) = \sum_{j} \theta_{jt} \sigma_{jt}^{2} \quad (4)$$

Where  $\sigma_{jt}^2 = Var(e_{it} | x_{it} \in j)$  and  $\theta_{jt}$  is the share of workers in experience-education group *j* at time *t*. The conditional variance in wages  $V_{jt}$  is linked to the conditional variance of unobserved skills by the equation (5):

$$V_{jt} = p_t^2 \sigma_{jt}^2 \tag{5}$$

To identify the effects of changes in skill prices Lemieux (2006) impose the following restriction: the distribution of unobserved skills among workers with the same level of experience and education is the same over time.

$$\sigma_{jt}^2 = \sigma_j^2 \quad \forall t \quad (6)$$

Substituting equation (4) and (6) into (3) leads to:

$$Var(\varepsilon_{it}) = p_t^2 \sum_j \theta_{jt} \sigma_j^2 \quad (7)$$

From (7) it is straightforward that if the skill composition of the work force is held constant at some counterfactual  $\theta_j^*$  shares, an increase in the residual variance is due to an increase in skill prices. Plugging (5) into (7) leads to:

$$Var(\varepsilon_{it}) = \sum_{j} \theta_{jt} V_{jt} \quad (8),$$

Which shows how to compute counterfactual residual variance,  $V_t^*$  (as  $V_{jt}$  can be computed from the sample):

$$V_t^* = \sum_j \theta_j^* V_{jt} \tag{9}$$

This relationship makes the decomposition of the change in the residual variance possible between 2 periods into 2 terms in the following way:

$$V_{t} - V_{s} = \sum_{j} (\theta_{jt} V_{jt} - \theta_{js} V_{js}) = \sum_{j} \theta_{js} (V_{jt} - V_{js}) + \sum_{j} (\theta_{jt} - \theta_{js}) V_{jt}$$
(10)

The first term on the right hand side of (10) is a weighted average of changes in the within group variance. As  $V_{jt} = p_t^2 \sigma_j^2$ , rising prices of unobserved skills can be checked by this term. It can be also interpreted as the change in the counterfactual variance,  $V_t^*$  if the counterfactual weights are set at base period  $s: \theta_j^* = \theta_{js}$ 

The second term is the composition effect. Note that when changes in the weights are positively correlated with the within group variances, then there is a spurious growth in the residual variance.

In Table 2 some basic trends in residual and within group inequality is presented with reference in notations to equation (10), and with estimation on the overall residual inequality, on the composition effect and on the effect of changing in the price of unobserved skills for both genders. Overall inequality proved to be negative and prices must have been decreasing also. The composition effect is positive, although its size is very small.

Instead of using cells, another approach might be estimating a logit model to reweigh the data such that the distribution of skills remains constant over time (Lemieux, 2002, 2006, DiNardo et al. 1996).

Residual variance can be computed directly from the individual level data:

$$V_t = \sum_i w_{it} r_{it}^2 \quad (11)$$

Where  $r_{it}$  is estimated wage residual and  $w_{it}$  is the sample weight, for worker *i* at time *t* (note analogy with eq. (8) for grouped data). The counterfactual variance is (analogy with eq. (9)):

$$V_t^* = \sum_i w_{it}^* r_{it}^2 \quad (12)$$

The task is to find the counterfactual weight  $w_{it}^*$  that makes the counterfactual distribution of skills at time *t* the same as in a base year. This can be obtained by estimating a probability,  $P_i$  to be in year *t* relative to the base year by a logit model (on a pooled sample for the base year and year *t* and with the same regressors as the mincerian equation) and calculating the counterfactual weight in this way:

$$w_{it}^* = [(1 - P_i) / P_i] w_{it}$$
 (13)

In case of Hungary this means that younger and more educated people are more likely to be observed in period *t*, which means a larger value for  $P_i$ , a lower value for  $(1 - P_i)/P_i$ , so they are "down weighted" by  $w_{ii}^*$ .

Unfortunately there was some confusion about how to normalize the new weights. In Lemieux (2002) the ratio of estimated probability of being in year s / the estimated probability of being in year t are multiplied by the "unconditional probability that an observation is in period t (the weighted share of the pooled sample that is in period t)". This gives the proportion of sample size of year t divided by the overall sample size of the 2 years.

In Lemieux (2006) the same approach can be found which in the theoretical review of this paper, suggesting a weight of one over the sample size of only year *t*. Lemieux (2002) also remarks in a footnote that the correction factor is "of little importance, since it changes the re-weighting factor only in a proportional way", but there is much difference between the estimated counterfactual residual variance of the 2 approaches sketched above, because the proportion itself affects the size of counterfactual variance. That is why I did not report results of this method.

### Data and trends in within-group inequality by skill groups

The data used for the analysis is the Hungarian Household Panel Database from 1992 to 1997 of TÁRKI. It contains 8043 observations for most years, except for 1996 with 8211 and 1997 with 8311 observations. It can be regarded as a representative sample from the contemporaneous Hungarian population. It contains many observations irrelevant to the particular topic of this paper (children, unemployed, retired). The sample size decreases because of the separation of the two genders. Even then the smallest sample size used for a regression throughout the study included at least 200 observations, but usually much above.

In the paper only the wage of employees (who claim to be an employee) who are working age is analyzed, which is set to range from 16 years to 60 years. Defining working age is not straightforward, as the retirement age of men and women is different and changed from year to year in this period. Besides many people have chosen early retirement after transition (instead of potentially facing unemployment). Setting the upper bound at 60 seems plausible, as barely anyone claimed to be employed above this age.

The database asks questions about the last month wages of the main job and the average hours worked a week. Hourly wage is calculated from there raw data. Lemieux (2006) stresses the role of database where workers paid by the hour are asked directly about their wages instead of using such a transformed data, because in this way measurement errors and also biases coming from the changes of measurement error can be diminished.

Few employees reported very unlikely wage-hours worked combination; they are omitted (1 person in 1996 and 5 persons in1997). They might seem to be of minor importance because they are few but leaving them in the sample leads to jumps in the variances.

There is no reference in Lemieux (2006) whether he used real or nominal wages. Following the practice of Katz and Autor (1999), who always use real wages, I did the same.

11

The high inflation of this period also suggests preferring real wage. The CPI used in the calculation is available in the Statistical Yearbook of Hungary 2004, the publication of the Hungarian Central Statistical Office.

The TARKI household panel database asks questions about the highest degree acquired; the education related variables are constructed with the help of this. Answers are given in 9 categories by which approximation is possible about how many years the employee spent in school. During this approximation the official minimum time required to get that degree is considered. It should be noted that there may be divergences from the "real" time spent in school because either the categories are not as flexible as the education system, or the meaning of the categories could change over time. Kertesi and Varga (2005) shed light on the dimensions of the first problem. Appendix reports the years related to the different categories by this study.

The measures of residual wage inequality are computed from the residuals of a classical mincerian equation. The log of hourly wages is regressed on age, years spent in school, age squared (and a constant) separately for men and women.

Lemieux (2006) uses a quite special regression in his analysis: "a regression of log wages on an unrestricted set of dummies for age, years of schooling and interactions between nine schooling dummies and a quadratic in age" (p 469). He argues that the advantage of this regression is its flexibility. There are some reasons for ignoring his regression form also. The fewer dummies are made, the more information is lost. The more dummies are made, the harder it is to handle the regression during estimation. There must be some optimal choice in this trade-off, but using the variables themselves instead can be an alternative solution. Then the more information is kept on the expense of not allowing for break points in the regression. An extension of this study can be testing for which functional form is better.

To increase the sample size I will follow Lemieux (2006) in pooling 2 years in the beginning (1992-1993) and at the end (1996-1997) of the reference period.

To analyze the basic trends in within group variances, the work force is divided into 20 education-experience skill groups. On the education dimension the categories are: finished primary school or less, vocational school, finished high school, and finished college or university (or even more). Each of the experience group categories include ten years, and calculated as a potential experience: age minus years spent in school minus seven, the age of compulsory school enrollment.

Since the group of workers with more than 40 years of experience is empty or almost empty in many cases, I do not include these into the chart. For bachelor's or master's degree owners it is not a surprise considering that the retirement age is around 60 years. In 1996-97 it is empty for less educated women also which may be due to large unemployment and early retirement among the older female population. The reason to include this group is the fact that it provides some information about less educated men.

Within-group variances of men can be seen on Table 2A. Lemieux could conclude for the US. and for 30 years that variance increases in age and also in experience. Unfortunately we have a less obvious pattern.

Looking at a particular schooling category it is not obvious that residual variance increases in experience. It is true for example for high school graduates in 1992-93, but for most of the schooling categories variance shows concave shape with a peak somewhere between 10 and 40 or even 30 years of experience. Both for 1992-93 and 1996-97 primary school and vocational school finishers have concave variance with a maximum at 10-20 years of experience, in 1996-97 variance is also concave with peak of 20-30 years of experience. It would be a nice experiment to compare the tendencies with Lemieux's multiple-dummies regression to check for the effect of functional form specification.

Looking at a particular experience group at different education levels behaves much nicer. There are some strong exceptions, like the variance of workers with 10-20 years of experience for both years, but overall we can say that the variance increases for most of experience groups in education.

### Table 2A:

## Within group variance of wages by experience-education cell for men, 1992-93 and

### 1996-97

A. by education and	withir	n group vai	riance	work force share		
experience	1992-93	1996-97	change	1992-93	1996-97	change
•	V <sub>is</sub>	V <sub>it</sub>	V <sub>it</sub> - V <sub>is</sub>	$\Theta_{is}$	Θ <sub>it</sub>	$\Theta_{it} - \Theta_{is}$
primary school	<b>,</b> -	,	, , , , , , , , , , , , , , , , , , ,	,-		, , , , , , , , , , , , , , , , , , ,
0-10	0,0969	0,1096	0,0127	0,0260	0,0134	-0,0126
10-20	0,2657	0,1616	-0,1040	0,0348	0,0372	0,0023
20-30	0.1552	0.1120	-0.0431	0.0455	0.0444	-0.0011
30-40	0.0731	0.1025	0.0294	0.0348	0.0196	-0.0152
40+	0.1754	0.1254	-0.0500	0.0242	0.0124	-0.0118
vocational school						
0-10	0.1299	0.1659	0.0360	0.0838	0.1187	0.0349
10-20	0.1997	0.1978	-0.0020	0.1163	0.1166	0.0003
20-30	0.1793	0.1400	-0.0393	0.1157	0.1197	0.0040
30-40	0.1100	0.1396	0.0296	0.0667	0.0475	-0.0192
40+	0.1299	0.0080	-0.1219	0.0112	0.0062	-0.0050
high school						
0-10	0.1918	0.1109	-0.0810	0.0549	0.0867	0.0318
10-20	0.1717	0.1245	-0.0471	0.0767	0.0722	-0.0045
20-30	0.1523	0.1740	0.0217	0.0803	0.0743	-0.0060
30-40	0.1584	0.1419	-0.0166	0.0366	0.0495	0.0129
40+	0.1281	0.0098	-0.1183	0.0059	0.0021	-0.0038
ba and ma degree						
0-10	0.2440	0.2871	0.0431	0.0354	0.0341	-0.0014
10-20	0.1950	0.1027	-0.0923	0.0614	0.0526	-0.0088
20-30	0.2908	0.2356	-0.0552	0.0691	0.0588	-0.0102
30-40	0.2068	0.1330	-0.0738	0.0207	0.0341	0.0134
40+						
B. weighted average						
actual shares	0.1767	0.1543	-0.0224			
1. period shares	0.1767	0.1530	-0.0237			
2. period shares	0.1790	0.1543	-0.0247			

The change of within group variance can also be seen. The most salient pattern is the fact that in most groups within group variance decreased, which lead to overall decrease of variance as well.

Results for women can be seen on Table 2B. Tendencies are harder to recognize (if there are any). Looking at particular education groups only two of them show some pattern: in 1992-93 the vocational school finishers have concave shaped variance and the residual variance of college or university finishers is increasing in experience. All other groups are random, no systematic rules can be observed in them. Checking a particular experience group across different education groups gives shows some pattern also. In only 1996-97 10-20 years of experience is slightly increasing in education, in both years 20-30 years of experience does the same and in 1992-93 the category of 30-40 years is almost well-behaving. The changes in the variances are mainly negative here also.

#### Table 2B:

Within group	variance of	wages by	experience-	education	cell for	women,	1992-93	and
			1996-97					

A. by education	education within group variance		wo	work force share		
and experience	1992-93	1996-97	change	1992-93	1996-97	change
•	V <sub>js</sub>	V <sub>it</sub>	V <sub>it</sub> - V <sub>is</sub>	$\Theta_{is}$	Θ <sub>it</sub>	Θ <sub>jt</sub> - Θ <sub>js</sub>
primary school	·					
0-10	0.1937	0.1171	-0.0766	0.0187	0.0144	-0.0043
10-20	0.3052	0.0809	-0.2244	0.0304	0.0220	-0.0083
20-30	0.0892	0.1249	0.0357	0.0759	0.0728	-0.0031
30-40	0.1579	0.0626	-0.0953	0.0829	0.0565	-0.0264
40+	0.1308	0.0480	-0.0827	0.0163	0.0029	-0.0135
vocational school						
0-10	0.1268	0.1110	-0.0158	0.0549	0.0728	0.0179
10-20	0.1420	0.0821	-0.0599	0.0520	0.0517	-0.0002
20-30	0.1144	0.1505	0.0361	0.0555	0.0613	0.0058
30-40	0.0762	0.0525	-0.0238	0.0286	0.0144	-0.0142
high school						
0-10	0.1703	0.1872	0.0169	0.0893	0.0920	0.0026
10-20	0.1165	0.1520	0.0355	0.0934	0.1159	0.0225
20-30	0.1672	0.1527	-0.0145	0.1238	0.1274	0.0036
30-40	0.1510	0.1884	0.0374	0.0566	0.0469	-0.0097
ba and ma degree						
0-10	0.1390	0.1502	0.0112	0.0455	0.0565	0.0110
10-20	0.1596	0.1597	0.0002	0.0870	0.0872	0.0002
20-30	0.1754	0.1527	-0.0227	0.0759	0.0920	0.0161
30-40	0.2062	0.1187	-0.0875	0.0128	0.0134	0.0006
B. weighted average						
actual shares	0.1493	0.1397	-0.0095			
1. period shares	0.1493	0.1346	-0.0146			
2. period shares	0.1482	0.1397	-0.0085			

The share of different skill groups can be seen on the left hand side of the tables. The tendencies mentioned above can be traced. The share of least educated decreased for both groups, while the share of more experienced decreased also. Difference of genders in the selection can be noticed only at the most educated group: while the increase of the share of women with college or university is independent of their experience, the same if true for men, but in the other direction: their share decreased (except for those who have more than 30 years of experience).

In the header of Table 2A and Table 2B we can see the counterpart parameters from equation (10). Lemieux (2006) remarks that composition effect, which is captured by the second part of the right hand side of equation (10) results in a spurious growth in the residual variance when the two factors are positively correlated. That is why he calculates the correlation coefficient between the within group variance for the last period ( $V_{it}$ ) and changes in the shares ( $\Theta_{it}$ -  $\Theta_{is}$ ). In our case we have fairly different results for the two genders: for men it is very small, it is 0,077512 and for women it is considerable, it is 0,558687. This suggest that for women bigger spurious growth can be expected as higher is the correlation, which seems a bit contradictional as in the overall tendencies not much pattern could be realized for men. (Panel B of Table 2Bconfirms that for women composition effect is higher).

In the lower part of Table 2A and 2B the overall residual variances can be seen, calculated for different skill group shares also.

In terms of equation (10) the change in the actual shares can be also described by:  $\sum_{j} (\theta_{ji} V_{ji} - \theta_{js} V_{js})$ (a). The difference in the weighted average using 1. period shares is  $\sum_{j} (\theta_{js} V_{ji} - \theta_{js} V_{js})$ (b), which is exactly the first term on the right hand side of equation (10), and which is the term that can separate the effect of changing skill prices. The difference in the weighted average using 2. period shares is  $\sum_{j} (\theta_{ji} V_{ji} - \theta_{ji} V_{js})$ (c). This sheds light on the measure of the composition effect we have to deal with, as it is obvious that the difference between the change in the weighted average with actual shares and the change in the weighted average with 1. period shares (shortly: (a)-(b)) is exactly the composition effect  $\theta_{ji} V_{ji} - \theta_{js} V_{ji}$ .

For men the change of the overall residual variance is -0.0224. The second row shows us that the change in the residual variance is not much smaller (-0.0237) when the shares are hold at their 1992-93 level. This shows us that most part of the change in the residual variance is due to the decrease in the within group variances, due to the possible decrease in the price of unobserved skills. The composition effect is responsible for only 0.0013.

The results for women in table 2B seem rather different: the change in the overall residual variance is only half of the men's, -0.0095. Keeping the shares of period 1992-93 gives residual variance change of -0.0146, which shows that the majority of the changes is due to the decrease in the within group variance, in price of unobserved skills. The difference between the two gives the composition effect; it is 0.0051, which is almost 5 times the composition effect of male workers.

We can conclude that the overall residual variance decreased for both groups. The price of unobserved skills must have decreased, as the effect of within group variance is negative. The changes in the composition of the work force had positive effects on the residual variance, which, assuming that variance increases in both education and experience, suggests that the positive effect of the increase in the schooling of workforce dominated the negative effect of decrease in age.

Men had a sharper drop in the price of unobserved skills, while women had bigger variance because of the change of their work force.

As the scale of within-group variance came into the foreground, in the following the detailed year by year evolution of within group variance can be seen for each of the education groups for men (Figure 1A) and women (Figure 1B).

As there are changes in the experience distribution of the work force which must be controlled, the variance for each education group is defined as the simple averages of the within group variances over the experience groups. For example the within group variance for workers who finished primary school is the average of within group variance for primary school finishers with 0-10, 11-20, 21-30, 31-40 and more than 40 years of experience (with

18

this the exact share of employees in the particular experience group is ignored). In terms of equation (10), we are dealing with the evolution of the components of the first term on the right hand side.

On Figure 1A we can see that variance is really higher for male workers with more years of education: after 1994 the variance is the smallest for primary school finishers and highest for high school graduates and ba or ma degree owners. The college and university finishers have outstandingly high variance, but it decreased by the time. The variance for primary school finishers decreased also. Within group variance did not change much systematically for male workers during this period, it slightly decreased for all groups.





On Figure 1B the evolution of within-group variance can be seen for each experience group, with variances averaged over education groups. In this case this means that for example the within group variance for workers who have 0-10 years for experience as the average of the variances of primary, vocational and high school finishers and college and university graduates. Four experience groups move together, the within group variance evolved rather the same for them; systematical pattern can not be seen. For the group above

40 years of experience the variance dropped a lot, suggesting a fall in the price of the unobserved skills of this group. The variance of group 20-30 and entrants behaves rather steady, while that of group 10-20 has dropped some.

On Figure 1C the within-group variance can be seen for women by education groups. Women also show the classic features: the more educated have the higher variance and less



Figure 1B

education means less variance. However there is some volatility during this period: college and university finishers have survived a drop in their unobserved-skill prices by the end of the period. Even then no trend is obvious.

Figure 1C



Figure 1D presents the within group variance graph of women for experience groups. The variation of most experience groups is steady, only slight changes happened. Slight decrease for entrants, 10-20 and 30-40 group and slight increase for group 20-30.

The variance of most experienced group (40+) for women has begun its drop from the same level as for men and had arrived at the same level also, but have reached it much earlier.



Figure 1D

Taken Figure 1A, 1B, 1C and 1D suggest that there is not much change in the within group variance for most groups. The only clear tendency in the drop in the unobserved skill prices of the most experienced group, for workers of more than 40 years of education.

Table 2A and 2B also showed that the overall variance is negative for both groups and the composition effect is even smaller than the unobserved price effect. For women, it is on one third of the unobserved price effect, for men it is the 1/18 part.

These numbers seem rather small; it is a valid question whether composition effect is really important? When does it matter after all? Composition effect matters when it is large enough to cause biases in estimation of residual variance. In Lemieux (2006) <sup>3</sup>/<sub>4</sub> of the overall residual variance was composition effect which means that <sup>3</sup>/<sub>4</sub> of the growth in the residual variance is a spurious consequence of composition effects. According to the results of Table 2, we may think that the composition effect has different effect in Hungary for men and women. For men 6% of the overall residual variance in composition effect (which is very small), for women the ratio is much bigger, it is a little more than 50%. This suggests that for women omitting the possibility of the composition effect may cause a much higher bias in the estimation of the residual variance than for men.

Figure 1 shown that within-group variances are rather volatile, but altogether only few tendencies can be discovered. This makes harder to separate the tendencies in composition effect also. In spite of the expectations not much difference can be discovered between men and women in within group variances. As for men only 6% of variance change is due to composition effect, one may expect that as the within group variances are responsible for the 94% of changes, some tendencies may be discovered. Although for women within group variances are responsible only for the 50% of changes in the variance, there within-group variance graphs for men and women are rather similar.

## Change in the gender wage gap

There are several methods to estimate gender wage gap. The Oaxaca's decomposition is one approach; it is applied by an article of Campos and Jolliffe (2004b) for Hungary between 1986 and 1998. Another method may be the Juhn-Murphy-Pierce decomposition, which is an advanced form of the former, as it takes into consideration the effect changes in inequality also. My focus in on their way of controlling for changes in inequality and on incorporating Lemieux's inequality-estimation method.

In Hungary the gender gap considerably decreased after transition (Campos and Jolliffe (2004), which the authors interpret as the effect of an "extraordinary improvement of women's relative situation". Elizabeth Brainerd (2000) found in 7 transition countries that although increasing wage inequality depressed relative wages of women, if the widening is not tremendous (like in Russia or Ukraine), the losses may be offset by gains from returns to observed skills, and "an apparent decline in discrimination". In terms of the key-equation of the JMP decomposition, equation (4), losses from term (D) can be offset by gains from (B) and (C).

We have already seen that residual inequality decreased between 1992 and 1997 for Hungary.

## Juhn-Murphy-Pierce method

To explore the reasons for the change in female-male relative wages, one approach may be the framework given by Juhn, Murphy and Pierce (1993). Their method makes the decomposition of different effects possible and controls for inequality changes also.

First, consider wage equation for male individual *M* and period *t*:

$$W_{Mt} = X_{Mt}\beta_t + e_{Mt} \quad (14)$$

where  $W_{Mt}$  is the log of monthly wages,  $X_{Mt}$  is the vector of explanatory variables,  $\beta_t$  is a vector of coefficients and  $e_{Mt}$  is the residual, the component of wages accounted for by unobservables.

According to Juhn, Murphy and Pierce (1993) this residual consists of two components: an individual's percentile in the residual distribution,  $\theta_{Mt}$  and the distribution function of the wage equation residuals,  $F_t(.)$  in the following way:  $e_{Mt} = F_t^{-1}(\theta_{Mt}|X_M)$ , where  $F_t^{-1}(|X_M|)$  is the inverse cumulative residual distribution for workers with characteristics  $X_{Mt}$  in year t.

Blau and Kahn (1997) and Brainerd (2000) defines  $\theta$  differently, they standardize the residual:  $\theta_{Mt} \coloneqq e_{Mt} / \sigma_t$ , where  $\sigma_t$  is the residual standard deviation of male wages for that year. It shows the unexplained level of male residual wage inequality, and has a mean 0 and variance 1. Brainerd (2000) remarks that this shows the percentile the individual occupies in the residual distribution which is a little bit misleading<sup>2</sup>. The value  $\theta$  defined this way can be negative also, so it is not the value of this standardized residual that shows the percentile, but the accompanying cumulative distribution function values (which can be obtained from the sample). By this simplification the male wage equation becomes the following:

$$W_{Mt} = X_{Mt}\beta_t + \theta_{Mt}\sigma_t \quad (15)$$

The male-female wage gap for period *t* is this:

$$D_t \equiv W_{Mt} - W_{Ft} = \Delta X_t \beta_t + \Delta \theta_t \sigma_t \quad (16)$$

where M and F refer to male and female averages, and  $\Delta$  is for the "average male-female difference for the variable immediately following" (Blau and Kahn, 1997). I interpret this explanation phrasing as the difference of the average male and average female variable values (as they did not mention paired sample requirements).

 $<sup>^{2}</sup>$  In the article she mentions the role of ranks given to individuals in the distribution, but not really connects to the formulas.

For the last term,  $\theta_{Ft} = (W_{Ft} - X_{Ft}\beta_t)/\sigma_t$  is needed, where  $\beta_t$  is the coefficient of the male regression (14). This reflects the difference between the wage a woman receives and she would receive if her skills were rewarded at the same rate at which men's skills are rewarded (Brainerd, 2000).

According to the equation (16), the gap in a given period consists of the differences in observed skills weighted by the return received by men to these skills and the differences in the standardized residual, weighted by residual male inequality.

The difference in the gender gap between 2 periods is this:

$$D_{t} - D_{s} = (\Delta X_{t} - \Delta X_{s})\beta_{t} + \Delta X_{s}(\beta_{t} - \beta_{s}) + (\Delta \theta_{t} - \Delta \theta_{s})\sigma_{t} + \Delta \theta_{s}(\sigma_{t} - \sigma_{s})$$
(17)  
(A) (B) (C) (D)

The first term (A) is the "observed X's effect", the changes in gender wage differential that comes from changes in male-female differences in observed labor market skills. The next term (B), the "observed prices effect" is the change in the price that the labor market attaches to the observed skills of men. The third term (C), the "gap effect" is the effect of the change in the relative position of women in the male residual wage distribution when male wage distribution is held constant. Women move upwards if their unobserved labor market skills improve relative to men's or if labor market discrimination against women decline.

The fourth term (D) is the "unobserved prices effect" that measures the change in the gender gap due to the widening or the narrowing of the residual male wage inequality, holding the gap in male-female unmeasured skills constant (Brainerd, 2000). Assume that deficits in unmeasured relative skills or discrimination lower women's position in the male wage residual distribution. Then in case of wider distribution, or with other words in higher

inequality women have to suffer from higher wage gap as this inequality imposes larger penalty on being below average in the distribution.

The calculation of the third and fourth component of equation (17) is the most tedious part of the model. According to Blau and Kahn (1997) it should be done in the following way considering year *s* and *t*. First give each woman in year *s* a percentile number based on the ranking of her wage residual (from the male wage regression for year *s*) in the *s* year distribution of male wage residuals. It shows her position in the year *s* male wage distribution. Then match her with the residual in male distribution of year *t* which has the same percentile that she had in the residual male distribution in year *s*. The average of these residuals (multiplied by -1 as the mean male residual is always 0) is the estimate for  $\Delta \theta_s \sigma_t$ . For  $\Delta \theta_t \sigma_t$  the average female residual from male wage regression of 1996-97 should be considered.

Altogether we can say that the effect of gender specific factors is reflected in the sum of the first and third terms: the effect of different observable skills and gender differences in wage rankings at a given level of observables. The second and fourth term reflects the wage structure, the effect of changing returns to observed and unobserved characteristics (Blau and Kahn, 1997).

The same data is used as at Lemieux's decomposition to make the results comparable. The log of real hourly wages is considered, and the sample consists of employees. The pooled sample from the beginning and the end of the period (1992-93 and 1996-97) should be handled with more care as in the JMP model the coefficients are interpreted also.

## Results

The particular values for decomposed changes in gender wage differences can be seen in Table 3 for Hungary, between 1992 and 1997. Observed change in the gender gap can be calculated according to this:  $\Delta \ln W_t - \Delta \ln W_s$ , the average male-female difference for the log real wages, (which is the left hand side of equation (17)). This difference in differences is negative which tells us that the gender wage gap has closed from 1992-93 to 1996-96. Besides each component of the equation is negative which shows that the gender differences decreased in the components separately.

### Table 3.

#### Decomposition of the change in the gender wage differential

Observed chang	e in the gender gap		-0,05669
	Observed X's	(A)	-0,02602
	Observed prices	(B)	-4,3E-06
	Gap	(C)	-0,01843
	Unobserved		
Of which	prices	(D)	-0,01363

We have seen in the first chapter (Tendencies in Hungary) that the observable characteristics of employees changed a lot. The term denoted by (A) reports the wage effect of these changes. This effect seems to the bigger and it is negative, which shows that the two genders became more similar. The second term controls for the changes in the differences in the returns to skills, it decreased also, but only a little bit. The third term controls for the differences in the wage distributions (keeping inequality constant), and according to the results this gap between male and female wage-distribution decreased also. The last term checks for the change in unobserved prices, this is negative and considerable also.

The aim of estimating the gender gap was to show that the residual variance is used to measure inequality, but it is also essential in measuring wage gaps.

According to my results both male and female variances change, both became smaller. The effect is stronger for women, which makes it probable that components (C) and (D) would be smaller and bigger accordingly. It is not hard to see that (D) should increase as the female residual standard deviation is in it with negative sign. Effect on (C) is less obvious, but as the left hand side of (17) must remain the same, it must decrease.

## Conclusion

I have found that both residual wage inequality and gender wage gap decreased between 1992 and 1997 in Hungary. For men, wage inequality decreased by 0.0224, twice as for women (0.0095) and the gender wage gap

After controlling for changes in the composition of the work force we see that residual wage inequality decreased by even more. This decrease is bigger for women, but the change in the residual variance for women is smaller than for men even after cleaning the composition effect.

Two factors are considered when checking for composition effect, education and age; both are positively related to residual variance. The work force became more educated and less old, the former is increasing the residual variance, and the latter is decreasing. The decrease in the cleaned residual variance, in the 'composite effect'- free variance may be due to the dominance of the effect of lower average age.

This alternative residual variance estimating method can affect estimates for gender wage gap. It is common to control for changes in inequality when decomposing gender wage gap and variance in one of the main inequality measures. It would alter the estimates of the "gap effect" in Juhn, Murphy and Pierce decomposition (1993), which is the effect of the change in the relative position of women in the male residual wage distribution. It would also change the "unobserved prices effect" that measures the change in the gender gap due to the widening or the narrowing of the residual male wage inequality, holding the gap in malefemale unmeasured skills constant.

A possible extension of this paper would be to estimate the exact size of the changes of the 'gap' and 'unobserved prices effect' by using the alternative residual inequality measure proposed by Lemieux (2006). It is rather easy it do this in practice in case of the unobserved price effect as standard deviation of the residuals is directly in the formulas. In case of the 'gap effect' the interpretation of composition effect becomes much more complicated.

# Appendix

The Questionnaire is the same for each year. The following years spent in school can be related to the categories used by the questionnaire:

Cat.	Definition in the questionnaire	Def. in English	years
1.	nem járt iskolába	not attend to school	0
2.	1-3 osztály	1-3 classes	2
3.	4-5 osztáy	4-5 classes	4
4.	6-7 osztály	6-7 classes	6
5.	8 általános	primary schhol	8
6.	szakmunkásképző	vocational school	11
7.	befejezett középiskola	finished high school	12
8.	befejezett főiskola	finished college	15
9.	befejezett egyetem	finished university	17

## Literature review

Andorka Rudolf, Ferge Zsuzsa, Tóth I. György. 1997. Valóban Magyarországon a legkisebbek az egyenlőtlenségek? *Közgazdasági Szemle, 2. sz. 89–112. o.* 

Blau, Francine D. and Lawrence M. Kahn . 1997. Swimming Upstream: Trends in the Gender Wage Differential in the 1980s. *Journal of Labor Economics, Vol. 15, No. 1, Part 1. (Jan), pp. 1-42.* 

Blau, Francine D. and Lawrence M. Kahn. 1992. The Gender Earnings Gap: Learning from International Comparisons. *The American Economic Review, Vol. 82, No. 2, Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association. (May, 1992), pp. 533-538.* 

Brainerd, Elizabeth. 2000. Women in Transition: Changes in Gender Wage Differentials in Eastern Europe and the Former Soviet Union. *Industrial and Labor Relations Review, Vol.* 54(1), 138-62, October 2000.

Campos, Nauro, F. and Dean Jolliffe 2004.a. After, Before and During: Returns to Education in Hungary (1986-1998). *Discussion Paper 4215, Centre for Economic Policy Research, February* 

Campos, Nauro and Dean Jolliffe. 2004.b. Does Market Liberalization Reduce Gender Discrimination? Econometric Evidence from Hungary, 1986-1998. WDI Working Papers

Chay, Kenneth and David Lee. 2000. Changes in Relative Wages in the 1980s: Returns to Observed and Unobserved Skills and Black-White Wage Differentials. *Journal of Economeirics*, 2000, 99(1), pp. 1-38.

DiNardo, John, Nicole M. Fortin, Thomas Lemieux. 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach *Econometrica, Vol. 64, No. 5. (Sep.), pp. 1001-1044.* 

Farber, Henry and Robert Gibbons. 1996. Learning and Wage Dynamics, *Quarterly Journal of Economics*, 1996, 111(4), pp. 1007-1047.

Galasi Péter, Varga Júlia. 2005. Munkaerőpiac és oktatás, MTA Magyar Közgazdaságtudományi Intézet

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy, Vol. 101(3), 410-442.* 

Katz, Lawrence, David Autor. 1999. Changes in the Wage Structure and Earnings Inequality Handbook of Labor Economics, Vol3.A, North-Holland, 1999.

Kertesi Gábor, Varga Júlia. 2005. Foglalkoztatottság és iskolázottság Magyarországon. Közgazdasági Szemle, july-aug, pp633-662.

KSH (Hungarian Central Statistical Office). 2004. Statistical Yearbook of Hungary 2004.

Lee, David. 1999. Wage Inequality in the U.S. during the 1980s: Rising Dispersion or falling minimum wage? *Quarterly Journal of Economics 114, 977-1023*.

Lemieux, Thomas. 2002. Decomposing Changes in wage Distributions: A Unified Approach, *The Canadian Journal of Economics, Vol 35, No 4. (Nov) pp. 646-688.* 

Lemieux, Thomas. 2006. Increasing Residual Wage Inequality: Composition Effects, Noisy data or Rising Demand for Skill? *American Economic Review, Vol. 96(3), 461-498, June* 

Mincer, Jacob. 1974. Schooling, experience and earnings. New York: National Bureau of Economic Research, Inc.

Mincer, Jacob. 1997. Changes in Wage Inequality. 1970-1990. Research in Labor Economics, 16, 1-18.

Newell, Andrew. 2001. The Gender Pay Gap in the Transition from Communism: Some Empirical Evidence. *Discussion Paper No 267, March IZA* 

Rosen, Sherwin. 1977. Human capital, a survey of empirical research. *Research in Labor Economics, vol.1., ed. R. Ehrenberg (Greenwich).* 

Tóth István György. 2003. Jövedelemegyenlőtlenségek: tényleg növekszenek, vagy csak úgy látjuk? *Közgazdasági Szemle, L. évf. 3. sz. 209–234. p.* 

World Bank (1996): From plan to market. World Development Report 1996. Published for the World Bank. Oxford University Press, Oxford.