Helsinki Center of Economic Research

Discussion Papers

Analysis and Synthesis of Wage Determination in Heterogeneous Cross-sections

Antti Suoperä
Statistics Finland

and

Yrjö Vartia
University of Helsinki and HECER

Discussion Paper No. 331
June 2011

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

HECER
Discussion Paper No. 331

Analysis and Synthesis of Wage Determination in Heterogeneous Cross-sections*

Abstract

The aggregation problem of heterogeneous micro behaviours is first discussed on a general level using analysis and synthesis operators. These show in detail what kind of components appears in the aggregation problem. In the analysis stage employee specific wage equations are allowed and estimated. The wage equations for log-wages are specified as non-linear with respect to education and experience and their interaction. Intercept heterogeneity arises from a very detailed micro partition and slope heterogeneity from different OLS regressions in several occupation groups.

These thousands different behaviours are summed together into a single representative behaviour, which is used in the synthesis stage to describe the macro behaviour. In this way, the macro behaviour is derived by minimal assumptions. Heterogeneous micro equations are written as a sum of representative and deviation behaviours (heterogeneity effects), which reproduces the original thousands of regressions by a single equation. Its estimation by OLS reproduces the coefficients of the representative behaviour and produces also their standard errors, which are hard to derive otherwise.

JEL Classification: B41, C02, C43, C81, C82, E01

Keywords: aggregation, micro foundations, methodology of economics, organizing micro- and macroeconomic data, fine labour partition, occupation specific models, wage equations, heterogeneous behaviour, standard errors of the representative behaviour, analysis and synthesis of wage determination, cross section estimation, large sample methods, labour economics, Mundlak critique.

Antti Suoperä

Statistics Finland
Työpajankatu 13, Helsinki
FI-00022 Statistics Finland
FINLAND

e-mail: antti.suopera@stat.fi

Yrjö Vartia

Department of Political and Economic Studies,
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014
FINLAND

e-mail: yrjo.vartia@helsinki.fi

* We thank Ph.D. Antti Kauhanen and prof. Pentti Vartia for useful comments on a previous version of the paper.

1. Introduction

We examine two problems - estimation of wage equations and aggregation of them into a representative/average equation. We call the first part as the analysis stage and the second as the synthesis stage.

The analysis stage consists of two parts: first, the partition of employees into K disjoints subsets or micro classes and second, the regression analysis stage. The definition of basic micro partition may in principle be decided in extremely many different ways. In our case the solution of the partition depends on the use of labour input as production factor and is based on very detailed classification of actual jobs and plants. In the regression analysis stage wage equations are specified to belong into the family of flexible functions and they are estimated by the OLS –method. The wage equations for log-wages are specified as non-linear with respect to quantitative variables and their interaction term.

In the synthesis stage the micro equations are aggregated together into representative behaviour. In this stage we use standard results of the OLS –method and write, say J, estimated wage equations only as one wage equation, which consists of two parts: The representative behaviour for all micro units and the heterogeneous behaviour, which measures individual specific behaviour as deviation of the representative behaviour.

We approximate the exact micro relations by flexible functional forms and accept the hypothesis of heterogeneously behaving agents. Our mathematical analysis is based on Vartia's (1979, 2008a) paper 'On the Aggregation of Quadratic Micro Equations'. Vartia shows that aggregation of heterogeneous quadratic micro equations leads to macro model including covariances between exogenous independent variables and their parameters. Covariance terms will appear in the macro model already in the linear but also in the quadratic case, see also van Dahl and Merkies, 1984. Since Theil, these covariance terms have been regarded as 'nuisance parameters' in unbiased estimation of macro parameters. We show in section four by the two stage OLS-method (Suoperä, 2003), that covariance terms are not 'nuisance parameters'. On the contrary, they include necessary input information for unbiased estimation of macro parameters. Similar methods of analysis and synthesis have been used in several important applications, see Suoperä (2003, 2004a,b, 2009a,b).

We partition the study in the following way: The second section describes shortly the solution of aggregation problem for heterogeneously behaving micro units. In section three the micro equations for the log-wage are specified to be linear with respect to parameters. In section four we derive the representative behaviour for heterogeneously behaving cross-sections. In section five we give an empirical example of wage determination in heterogeneous cross-sections in Finnish labour markets. Section 6 concludes.

2. *Analysis and synthesis: Micro foundations of macro behaviour*

First we outline shortly, how macro level behaviour can be inferred from any hypothetical micro behaviours and micro data using Analysis[1] and Synthesis[2] in the following way. Deeper treatment of the applied aggregation[3] is presented in Vartia (2008a, 2008b, 2009) and Lintunen et al 2009.

Analysis. Consider the finite set $A=\{a_1,a_2,\dots,a_n\}$ of economic agents. Suppose that each economic agent a_i has a possibly non-linear behavioural function $f_i : R^K \to R$, which maps its inputs x_i to its output y_i:

$$(2.1)\qquad y_i = f_i(x_i) = f_i(x_{1i},\dots,x_{Ki}),\ y = BC(\tilde{y},\alpha) = \tfrac{1}{\alpha}(\tilde{y}^{\alpha}-1) = \tfrac{1}{\alpha}(e^{\alpha\log\tilde{y}}-1).$$

These functions determine the Box-Cox transformation of the original ratio scale economic variable $\tilde{y}$, which is implicitly defined by $\tilde{y}=(\alpha y+1)^{1/\alpha}$. For $\alpha=0$, we have $y = BC(\tilde{y},0)=\log\tilde{y}$, $\tilde{y}=e^{y}$. In our estimated wage equations the explained variable is the logarithmic wage $y=\log\tilde{y}=\log w$. We consider all these *n* nonlinear functions $f_i : R^K \to R$ as known. This assumed information defines in (2.4) the analysis operator $y = A(f,x)$. In actual estimations these functions are taken as systematic parts of the estimated equations. From the knowledge of f_i´s, we can define their mean as follows:

$$(2.2)\qquad \bar{f} = \tfrac{1}{n}(f_1+f_2+\dots+f_n) = \tfrac{1}{n}\sum f_i .$$

The mean function $\bar{f}$ assigns to any fixed input *the mean* of the individual outputs at that input: $\bar{f}(x) = \tfrac{1}{n}(f_1(x)+f_2(x)+\dots+f_n(x)) = \tfrac{1}{n}\sum f_i(x) = \tfrac{1}{n}\sum f_i(x_1,\dots,x_K)$. This is a standard definition in the function theory. (Think e.g. what the sum of two functions means in the formulae *D(f+g) = Df + Dg* or $E(X+Y)=EX+EY$. It means the same there as here.) Then decompose all behaviours as representative and deviation behaviours as follows:

$$(2.3)\qquad f_i = \bar{f} + (f_i - \bar{f}) = \bar{f} + \delta f_i \quad \text{or as column vectors} \quad f = (f_1,\dots,f_n)' = \bar{f}1 + \delta f .$$

[1] The main meaning of analysis is
a. The separation of an intellectual or material whole into its constituent parts for individual study.
b. The study of such constituent parts and their interrelationships in making up a whole.
c. A spoken or written presentation of such study: published an analysis of poetic meter.

2 The main meaning of synthesis is
a. The process of combining objects or ideas into a complex whole.
b. The combination or whole produced by such a process.

[3] The main meaning of aggregation is: a collection of parts of a whole

Like for regression functions, the mean of deviation functions is zero for all input vectors: $\langle \delta f(x)\rangle = \langle f(x)\rangle - \langle \bar{f}(x)\rangle = 0$. In our most advanced estimated models there are more than 20000 different agent functions f_i (for the private sector and one year). Their reporting would require thousands of pages, while the representative behaviour $\bar{f}$ fits on one page. If actual agent-wise behaviours are needed, they are best described via deviations δf_i from the representative behaviour. Because agents have their own behavioural functions and they have different agents-specific input situations, we need a notation that distinguishes these. This notation is provided by analysis and synthesis operators. The whole micro system can be represented in an exact and compact notation as

(2.4) $$\begin{bmatrix} y_1 \\ \\ y_n \end{bmatrix} = \begin{bmatrix} f_1(x_1) \\ \\ f_n(x_n) \end{bmatrix} = A\begin{bmatrix} f_1, x_1 \\ \\ f_n, x_n \end{bmatrix}$$ or more compactly $y = A(f, x)$. *"Eight symbol formula"*

We call $A(f,x)$ the *Analysis-operator* (shortly *A-operator*). It provides the separation of an intellectual or material whole (the set of agents and their behaviours) into its constituent parts for individual study. The vector of functions *f* and all the input situations are needed to know the corresponding vector of outputs. Straightforward calculation shows that $A(f,x)$ is *linear* in its f -arguments: (i) $A(f+g,x) = A(f,x) + A(g,x)$ and (ii) $A(\lambda f, x) = \lambda A(f,x)$ for all possible choices of arguments. Thus decomposition (2.3) leads to the fundamental formula

(2.5) $$y = A(f,x) = A(\bar{f}1, x) + A(\delta f, x), \text{ where}$$

$$A(\bar{f}1, x) = A\begin{bmatrix} \bar{f}, x_1 \\ \\ \bar{f}, x_n \end{bmatrix} = \begin{bmatrix} \bar{f}(x_1) \\ \\ \bar{f}(x_n) \end{bmatrix}$$ = vector of *representative behaviours* at actual inputs and

$$A(\delta f, x) = A\begin{bmatrix} \delta f_1, x_1 \\ \\ \delta f_n, x_n \end{bmatrix} = \begin{bmatrix} \delta f_1(x_1) \\ \\ \delta f_n(x_n) \end{bmatrix}$$ = vector of *deviation behaviours* at actual inputs.

Individual components satisfy identically $f_i(x_i) = \bar{f}(x_i) + \delta f_i(x_i)$ for all the agents. Representative functions $\bar{f}(x_i)$ appear here naturally as explanatory factors in the micro level. They show how agents would behave if they had the same average behaviour together with their actual inputs. This decomposition is applied for heterogeneous regressions in chapter 5. The representative function $\bar{f}(x)$ describes naturally the overall behaviour. It reflects the common features of the micro behaviours. In the synthetic macro level, the representative functions reappear in $CB\langle x\rangle = \frac{1}{n}\sum \bar{f}\langle x_i\rangle$ and especially in $RB\langle x\rangle = \bar{f}\langle x\rangle$, where instead of micro inputs the average inputs are the arguments. As will be shown, the function $RB\langle x\rangle = \bar{f}\langle x\rangle$ is the main term of the macro behaviour.

Synthesis. What kind of macro behaviour arises from these micro behaviours? The representative behaviour $\bar{f}$ is a synthetic concept already in the analysis stage. In economics, it corresponds to a fictitious representative agent. It is useful both in the concept level and when explaining average outputs (concrete macro level). At the macro level the output is taken as the *point of inertia* or its mean

(2.6) $$\langle y \rangle = \bar{y} = \tfrac{1}{n}\sum y_i \,.$$

In terms of the Box-Cox transformation[4] this is $\langle y \rangle = \langle BC(\tilde{y},\alpha) \rangle = \langle \tfrac{1}{\alpha}(\tilde{y}^{\alpha} - 1) \rangle$. It is a very complicated function of the micro characteristics, namely the following micro-to-macro (MicMac) function

(2.7) $$\langle y \rangle = \tfrac{1}{n}\sum A_i(f,x) = \langle A(f,x) \rangle \,.$$

It depends on huge number $K \times n$ (say 5*600 000) of micro inputs and all n micro functions. We denote this dependency using the *Synthesis-operator* defined by

(2.8) $$\langle y \rangle = \langle A(f,x) \rangle = S(f,x)$$

or in component form $S(f,x) = \tfrac{1}{n}\sum f_i(x_i)$. The synthesis-operator shows explicitly that both the behaviours f and the inputs $x = (x_1,\ldots,x_n)'$ affect the macro output. The linearity of $A(f,x)$ in its f –argument implies the similar linearity of $S(f,x)$:

(2.9) $$S(\lambda_1 f + \lambda_2 g, x) = \lambda_1 S(\lambda_1 f, x) + \lambda_2 S(g, x) \,.$$

This leads to the fundamental decomposition of the macro behaviour

(2.10) $$S(f,x) = S(\bar{f}1, x) + S(\delta f, x)$$ meaning

Macro Behaviour $MB(x)$ = *Common Behaviour* $CB(x)$ + *Heterogeneity Effects* $HE(x)$.

This crucial *Functional Equation* holds identically for all possible inputs. All of its components are MicMac-functions, whose inputs $x = (x_1,\ldots,x_n)'$ of huge dimension $K \times n$ come from the micro level but its three outputs are a macro statistics. It is the macro equivalent of the fundamental formula (2.5). Our interpretation of *the ag-*

gregation problem is how also the input side of (2.5) can be expressed in terms of macro statistics. In *Macro Behaviour* $MB(x) = S(f, x)$ the functional arguments f are taken from agent level. In *Common Behaviour* $CB(x) = S(\bar{f}1, x)$ the functional arguments are restricted to be the average ones for everyone, but input variables remain at their actual values. *Heterogeneity Effects* $HE(x) = S(\delta f, x)$ gives the macro effects of heterogeneous behaviour over individuals.

The Common Behaviour $CB(x) = S(\bar{f}1, x)$ may be further decomposed as follows

(2.11) $\quad CB(x) = RB\langle x\rangle + NLE(x) =$ *Representative Behaviour + Non-Linear Effects*,

where $RB\langle x\rangle = CB(\langle x\rangle 1) = \bar{f}\langle x\rangle$ and $NLE(x) = CB(\langle x\rangle + \delta x) - CB\langle x\rangle$. We have omitted unnecessary ordinary brackets in combination to arrow brackets. For instance, macro economics deals with Representative Behaviour $RB\langle x\rangle = S(\bar{f}1, \langle x\rangle 1) = \bar{f}\langle x\rangle$, where both the output and input $\langle x\rangle$ are averages. $RB\langle x\rangle = \bar{f}\langle x\rangle$ is the natural point of comparison where the other components are related. By Taylor expansion the non-linearity effects depends mainly on the variances and covariances of the inputs and on the second derivatives $NLE(x) \approx \frac{1}{2}\sum\sum \bar{f}_{kl}\langle x\rangle \mathrm{cov}(x_k, x_l)$. It vanishes identically if the average function is affine, $\bar{f}(x) = \bar{a} + \bar{b}x$, or all inputs are the same, $x = \langle x\rangle 1 + \delta x = \langle x\rangle 1$. This essentially solves our aggregation problem. The average macro output depends on the representative behaviour $\bar{f}\langle x\rangle$ = average behaviour at K average inputs, possibly on variances and covariances of inputs and on contingent but probably tiny heterogeneity effects (how inputs correlate with agents having deviating behaviours). Important explanatory variables in addition to the averages are the *squares* of the standard deviations (and not the standard deviations as such) of the variables. This shows that macro behaviour is more uniform than commonly expected. The result is a general non-parametric one holding on any twice differentiable agent behaviours, on their average output, K input averages, K input variances and even more input covariances. Our powerful notation hides many of the details and complications.

For *affine behaviours* $f_i(x_i) = a_i + \sum_{k=1}^{K} b_{ki}x_{ki} = a_i + b_i'x_i$ we have simply $\bar{f}(x_i) = \bar{a} + \bar{b}'x_i$ and $\delta f_i(x_i) = \delta a_i + \sum_{k=1}^{K}\delta b_{ki}x_{ki} = \delta a_i + \delta b_i'x_i$. The Common Behaviour becomes $CB(x) = S(\bar{f}1, x) = \bar{a} + \bar{b}'\bar{x} = RB\langle x\rangle$, and the non-linearity effect vanishes identically, $NLE(x) \equiv 0$. Heterogeneity Effects are given by $HE(x) = S(\delta f, x) = \frac{1}{n}\sum_{i=1}^{n}(\delta a_i + \sum_{k=1}^{K}\delta b_{ki}x_{ki}) = \sum_{k=1}^{K}\frac{1}{n}\sum_{I=1}^{n}\delta b_{ki}x_{ki} = \sum_{k=1}^{K}\mathrm{cov}(b_k, x_k)$, a sum of K covariances over different variables. Instead of $K \times n$ arguments, *CB*-function depends on only K averages $\bar{x} = (\bar{x}_1, ..., \bar{x}_K)$ and of nothing else, say of the variances or the log-variances of the distributions of the input

[4] Box-Cox transformation defines the moment mean as follows: $\left\langle \frac{1}{\alpha}(\tilde{y}^{\alpha} - 1)\right\rangle = \frac{1}{\alpha}(K^{\alpha} - 1) \Leftrightarrow \left\langle \tilde{y}^{\alpha}\right\rangle = K^{\alpha} \Leftrightarrow$

$K = \left\langle \tilde{y}^{\alpha}\right\rangle^{1/\alpha} = M(\tilde{y}, \alpha)$

variables. Similarly *HE* is a sum of only *K* covariances. Further $\bar{y} = \frac{1}{n}\sum_{i=1}^{n}(a_i + \sum_{k=1}^{K} b_{ik}x_{ik})) =$ $MB(x) = CB(x) + HE(x) = \bar{a} + \bar{b}'\bar{x} + \sum_{k=1}^{K}\mathrm{cov}(b_k, x_k)$ is an algebraic identity, which holds for all possible values of its variables and parameters. The reduction in dimensions ($K \times n$ versus $K + K$) and on concept level is maximally great, cf. Vartia (1979, 2008a). The next paragraph outlines a more general way to proceed. The philosophy behind the System of National Accounts *SNA* is that only the averages (or totals) of the variables matter. This may be called the Single Statistics Paradigm *SSP* which will be changed in the future e.g. by including measures of variability in *SNA*.

For *parameter linear behaviours* $f_i(x_i) = a_i + \sum_{l=1}^{L} b_{li}z_{li}$, where $z_{li} = g_l(x_i) = g_l(x_{1i},...,x_{Ki}),\ g_l : R^K \to R$ are given linearly independent functions, the representative behaviour becomes $\bar{f}(x) = \bar{a} + \sum_{l=1}^{L}\bar{b}_l z_l = \bar{a} + \sum_{l=1}^{L}\bar{b}_l g_l(x)$. The Common Behaviour can be written as in the affine case $CB(x) = S(\bar{f}1, x) = \ \bar{a} + \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L}\bar{b}_l z_{li} = \bar{a} + \sum_{l=1}^{L}\bar{b}_l\bar{z}_l$, where $\bar{z}_l = \frac{1}{n}\sum g_l(x_i) = \langle g_l(x)\rangle$ are the statistics[5] needed in the input side. Usually this is in variance to *SSP*. Heterogeneity Effects are $HE(x) = S(\delta f, x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L}(b_{li} - \bar{b}_l)g_l(x_i) = \quad \sum_{l=1}^{L}\frac{1}{n}\sum_{i=1}^{n}(b_{li} - \bar{b}_l)g_l(x_i) = \quad \sum_{l=1}^{L}\mathrm{cov}(b_l, g_l(x))$ $= \sum_{l=1}^{L}\mathrm{cov}(b_l, z_l)$, again a sum of covariances but now over different transformations $g_l : R^K \to R,\quad l = 1,...,L$ of the variables. Again $MB(x) = RB\langle x\rangle + NLE(x) + HE(x)$ is an identity valid for all values of its terms. In our estimations we have $L = 8$ genuine variables in the private sector together with more than 20000 jobs indicators, which determine the agent-specific constants. The reduction from dimensions $K \times n = 5 \times 600000$ of the variables in (2.10) is dramatic.

Estimation of heterogeneous micro behaviours.

(2.12) $\qquad \hat{y}_i = f_i(x_i) = a_i + \sum_{l=1}^{L} z_{li}b_{li} = a_i + z_i'b_i,$ for all $(\hat{y}_i, z_i)$-vectors

where $z_{li} = g_l(x_i)$. Arguments z_i and b_i are *K*-dimensional vectors of input variables and parameters, which are allowed to vary not from one agent to another but according to 9-183 (in private sector) ISCO occupations. These occupations specific parameters and the constants a_i (depending on more than 20 000 micro classes of jobs) are estimated by OLS (fixed effect or covariance model). Behaviour functions are specified to be linear functions with respect to parameters. The simplest example of this kind function is an affine function, but they

[5] The transformation $g(x) = x^{\alpha}$ of a positive variable x is concave for $0 < \alpha < 1$ and convex for $1 < \alpha$ which can be inferred from its second derivative. In the concave case we have the Jenssen inequality $\langle g(x)\rangle < g\langle x\rangle$ and $\langle g(x)\rangle = g\langle x\rangle - [g\langle x\rangle - \langle g(x)\rangle]$. This contributes to the terms $\bar{b}g\langle x\rangle$ in $RB\langle x\rangle = \bar{f}\langle x\rangle$ and $-\bar{b}[g\langle x\rangle - \langle g(x)\rangle]$ in $NLE(x)$.

are easily specified to belong into the family of flexible functions (quadratic ones here). Table 5.1 shows the definition of the variables, which are: $z_1 = x_1 =$ Female indicator,…, $z_4 = x_4 =$ Education in years, $z_5 = g_5(x) = 0.5x_4^2$, $z_6 = x_5 =$ Experience in years, $z_7 = g_7(x) = 0.5x_5^2$, $z_8 = g_8(x) = x_4 x_5$. Averaging (2.12) over individual agents and utilising the results above or the basic lemma of aggregation[6] (Vartia, 1979, 2008a) we get the macro function $G: R^{8+8} \to R$ such that

(2.13) $$\langle \hat{y} \rangle = MB(x) = G(\bar{z}, \mathrm{cov}(z,b)) = \bar{a} + \bar{z}'\bar{b} + \sum_{l=1}^{8} \mathrm{cov}(z_l, b_l)$$, for all values of $\langle \hat{y} \rangle, \bar{z}, \mathrm{cov}(z,b)$.

We see that the aggregate output variable $\bar{y} = \langle \hat{y} \rangle$ does not depend only on *8*-dimensional aggregate input variable $\bar{z}$ but also on the *8*-dimensional vector of the covariances $\mathrm{cov}(z,b)$. The reduction from the number of variables $K \times n = 5 \times 600000$ in *MicMac*-function (2.10) is dramatic. From now on the derived variables $z_l = g_l(x)$ are denoted by customary symbols x_l.

3. The Analysis stage

The analysis stage consists of classification of labour input and regression analysis stages. These two stages are closely related. To avoid confusion about what are we measuring, the basic micro classification should be a partition of the micro units or agents. That is, the finite set of employees should be partitioned into disjoint sets with union of all employees. The partition of labour input follows a typical micro economic textbook and is based on a single plant and its production factors. Regression analysis combined with the partition is operational especially in construction of hedonic index numbers, wage differentials and their consistent decompositions (Koev, 2003; Suoperä, 2003, 2004, 2007, 2009, 2010).

3.1 Partition of labour input

We examine time periods $t = 1,\ldots,T$ and the finite set $A = \{a_1, a_2, \ldots, a_n\}$ of employees of every year. Employees are divided into hourly and monthly paid, whenever feasible. They are considered separately in government, municipal and private sectors. We define our partition according to the International Standard Classification of Occupations (i.e. ISCO-88, International Labour Office, Geneva, 1991) together with actual jobs (duties) and plants. The ISCO forms a hierarchical structure of occupations starting from the 1-digit (main groups like, Managers, Professionals, Technicians and Associate professionals, Clerks...) and ending into the 4-digit occupations, which has been divided if necessary for the national requirements further into the 5-digit occupations. The finest classification of the ISCO forms a partition, but it is not fine enough to take into account actual jobs (or duties) used in separate plants.

---

[6] The basic lemma of aggregation (BLA): $\mathrm{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x} \cdot \bar{y}$ is equivalent to

To make more accurate partition we take each ISCO class separately and divide it into disjoint sets by following steps: First, the ISCO occupations consist of actual jobs (or duties), which form a partition $P = \{A(1), A(2), \ldots, A(K)\} = \{A(k)\}$ with union $A = \bigcup_{k=1}^{K} A(k)$. In addition to *P*-partition, all individuals are classified into plants they are working in. Let also these disjoint sets $A'(1), A'(2), \ldots, A'(K')$ form a partition $P' = \{A'(k')\}$ of the basic set $A = \bigcup_{k'=1}^{K'} A'(k')$. The Cartesian product $P \times P'$ of two such partitions $P = \{A(k)\}$ and $P' = \{A'(k')\}$ is defined in the usual way as the union of all the sets $A(k) \times A'(k') = \{a \mid a \in A(k) \wedge a \in A'(k')\} = A(k,k')$. Clearly every individual belongs to exactly one class in the both subclassifications. In statistical terminology $P \times P'$ is just the cross-classification between the two classifications, which are the fine job and the plant classifications in our case. The Cartesian product $P \times P'$ is illustrated in Figure 3.1. The shaded cells are classified two dimensionally and are removed. In the second stage (Figure 3.2) the partition is carried out first vertically according to the plants only and then horizontally according to the jobs for the remaining cells.

Intuitively, the arising partition P* uses information on cross-classification of both the plant and the job of the employee whenever the number of observations allows that in the cell. Then it uses either the plant or the job information only, in this order. Those marginally important cells, where the marginal job frequencies are less than five, are omitted as unclassified. The micro partition P* is essentially based on cross-classification of both the plant and the job information and it allows effective homogenisation of these factors. Those two-dimensional cells, which contain five or more observations in the cross-classifications, define the Cartesian product part of it (Figure 3.1). The remaining small cells (Figure 3.2) are classified lexicographically first according to plants only (if this vertical strip contains at least five observations) and after that horizontally according to jobs only in the similar way. Those remaining strips having less than five observations (about 1 % of all observations) are omitted from further calculations.

Figure 3.1: The first stage: the *Cartesian product part of the partition P**:

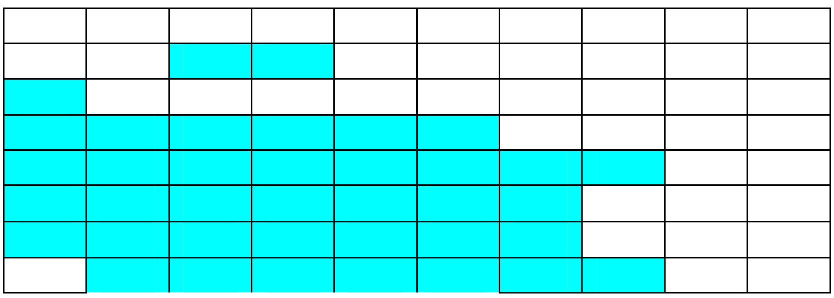
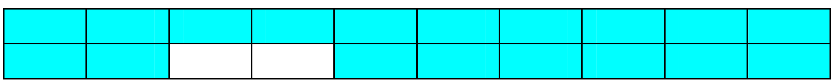
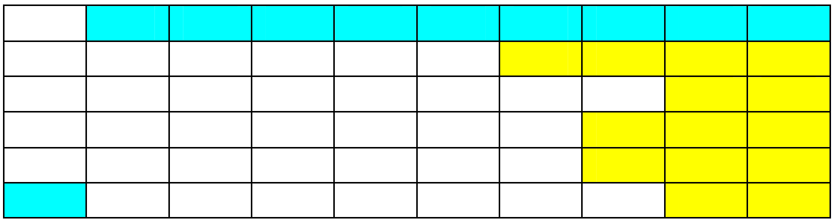


Figure 3.2: The classification of remaining employees first according to plants only and then jobs only.



$\frac{1}{n} \sum_i x_i y_i = \bar{x} \cdot \bar{y} + \mathrm{cov}(x, y)$ which is called the basic lemma of aggregation (BLA).



In order to eliminate quality differences accurately in the final labour cost index (see Suoperä, 2003), we have designed a fine micro classification *P** within crude international occupational classes (ISCO). Following the terminology of index number calculations, we refer to it as micro index classification or micro classes. In regression analysis stage, wage models are calculated separately within these classes (ISCO) at one-digit or more detailed level. This modelling partition is thus kept as rather coarse. On the other hand, the basic micro partition within each ISCO –class is designed as fine as is practically feasible.

*3.2 Definition of indicator variables for partition P**

Within each estimation sector (government, municipal, private) $A(k)=\{a_1(k),a_2(k),\dots,a_{n(k)}(k)\}$ having *n(k)* employees, we may define the indicator $1(a,B)$ for any subset $B \subset A(k)$:

(3.1) $$1(a,B)=T(a\in B)=\begin{cases}1,\text{if } a\in B\\ 0,\text{if } a\notin B\end{cases}.$$

The indicator $1(a,B)$ just tells whether *a* belongs to *B*, when it attains its value 1 (and equals zero otherwise). Typical indicators used are based on micro partition *P**. They attain value 1 for all employees being in that plant and special job, etc. These indicators partition employees into homogeneous subsets (in the private sector in 2000 there are 20876 such micro classes) and control wage differences between them. All indicators appear as additive terms in the equations or as dummy variables affecting only the intercept. This is a standard OLS modelling technique referred as parallel, fixed effect or dummy regression model. Its different names stem from alternative mathematical representations of the same analysis-of-covariance model. Especially its formulation as within-estimator is computationally extremely efficient (Hsiao, 1986, p. 25-32, 128-140).

We refer to this detailed partition *P** as micro classes *B(k)* and use the sub-index *k* for it.

3.3 Regression modelling stage

Let's examine the data generating process of wages for a given ISCO occupation group *j*. Each ISCO occupation group is stratified into disjoint strata according to the *P**-partition. The wage equation is specified as semi-logarithmic regression model, which is generally called to fixed-effects dummy-variable approach (Hsiao, 1986, s.29-32). We specify the model as linear in respect to parameters

(3.2) $$y_{ijt} = \alpha_{jkt} + \underline{x}'_{ijt}\underline{\beta}_{jt} + \varepsilon_{ijt}$$

where $y_{ijt} = log(w_{ijt})$ represents employee *i* specific logarithmic wage per hour in some rather large (varying from 1, 9,…, 183 in the private sector in 2000) ISCO occupation *j* in period *t*. Parameters β_{jt} in the regression model *j* are allowed to vary according to this occupational grouping and time. (Government, municipal and private sectors are not shown in the notation. Parameters α_{jkt} represent wage effect for individual employees belonging to the micro class *B*(*k*) in the equation *j* in period *t*. The K vector $\underline{x}'_{ijt}$ consists of exogenous independent variables typically used in empirical analysis of labour markets (constant, sex, part-time and non-permanent employment indicators and especially education and experience). The equation (3.2) has non-linear quadratic terms in experience and in education and also their interaction term. This lowers the number of distinct variables to $K - 3 = 9 - 3 = 6$ (including the constant). The wage equations have flexible functional form and all reaction parameters are time and sector specific.

The term ε_{ijt} is random error term, which does not contain systematic information about the data generating process of wages. It is assumed, that $E(\varepsilon_{ijt}|x_{ijt}) = 0$ and $Var(\varepsilon_{ijt}|x_{ijt}) = \sigma^2_{jt} < \infty$. In our model specification the error covariance matrix is diagonal – a most natural situation for cross-sectional data. Given the assumed properties of ε_{ijt}, the best linear unbiased prediction for y_{ijt} is the conditional expectation conditional on classification of labour inputs and given exogenous explanatory *x*-variables. The solution reveals many useful properties we utilise in the synthesis stage, especially in aggregation of the parameters.

The *P**-partition is very detailed including about 5 000, 10 000 and 20 000 indicator variables in government, municipal and private sectors respectively. Because of large number of observations it is necessary to transform observations as deviation of means with respect to the *P**-partition (Davidson & MacKinnon, 1993, p. 19-25). Then the OLS estimator for $\underline{\beta}_{jt}$ is

(3.3) $$\hat{\underline{\beta}}_{jt} = \left[\sum_i \sum_k \left(\underline{x}_{ijkt} - \overline{\underline{x}}_{jkt}\right)\left(\underline{x}_{ijkt} - \overline{\underline{x}}_{jkt}\right)'\right]^{-1} \sum_i \sum_k \left(\underline{x}_{ijkt} - \overline{\underline{x}}_{jkt}\right)\left(y_{ijkt} - \overline{y}_{jkt}\right).$$

It is called the covariance estimator by established practice in our analysis of covariance model. The micro partition-specific wage effects α_{jkt} are estimated in the second step as follows:

(3.4) $$\hat{\alpha}_{jkt} = \overline{y}_{jkt} - \overline{\underline{x}}'_{jkt}\hat{\underline{\beta}}_{jt},$$

where $\overline{y}_{jkt}$ is the arithmetic average of logarithmic wages for class *j*, *k* and period *t*. The elements of vector $\overline{\underline{x}}'_{jkt}$ are the arithmetic averages of explanatory variables in the same classes and the parameter vector $\hat{\underline{\beta}}_{jt}$ is the OLS-estimators for equation j. According to the Frisch, Waugh and Lovell -theorem (Davidson & MacKinnon, 1993), OLS –estimation of the slopes can always be carried out via centralised variables. The constant term is estimated by forcing the regression plane through the point of averages. This method is computationally extremely effective.

Under the assumed properties of random error term, OLS –estimator's (2.3) and (2.4) are the best linear unbiased estimators (BLUE). Other fundamental algebraic and extremely operational aspects for the OLS –solution are: First, the least squares residuals sum up to zero and is orthogonal to all exogenous independent variables. Second, the regression hyperplane passes through the point of averages of input and output variables. Third, the average of the fitted values (conditional averages) from the regression equals the average of the actual values. All the three properties will always be satisfied for each micro class $B(k)$ and for each single equation j.

4. *The synthesis stage*

In the synthesis stage we aggregate the agent specific behaviours into the representative/average one. Also some connections between the micro and macro equations are shown. In this stage, we show some very useful and operational results used previously for the hedonic quality adjustment.

In the analysis stage, we estimated J separate wage equations each having very detailed classification of labour input. These models are specified according to Becker's human capital theory by allowing the characteristics and behaviours to vary freely from group to group. For each of these subgroups j semi-logarithmic quadratic regression models with several controlling "dummies" are estimated. The estimated regression models are

(4.1) $$y_{ij\,t} = \log w_{ij\,t} = \hat{\alpha}_{jkt} + \underline{x}'_{ij\,t}\hat{\underline{\beta}}_{jt} + e_{ijt}, \text{ or } y_{ijt} = \hat{y}_{ijt} + e_{ijt}$$

where $\hat{\alpha}_{jkt}$ is the estimated wage effect for the micro class $B(k)$ in the wage equation j in time period t. K vector $\hat{\underline{\beta}}_{jt}$ consists of the wage effects for the explanatory variables including quadratic terms of education, experience and their interaction.

The least squares solution implies the following three basic implications: First, the sum of residuals equals zero for any micro class $B(k)$ and the least squares residuals are orthogonal to x-variables, i.e. $\sum_i e_{ijt}\underline{x}'_{ijt} = \underline{0}$ for all $a_i \in B(k)$. Second, the regression hyperplane passes through the averages of input and output variables, $\bar{y}_{jk\,t} = \hat{\alpha}_{jkt} + \underline{\bar{x}}'_{jk\,t}\hat{\underline{\beta}}_{jt}$. Third, the average of predicted values equals the average of actual values, i.e. $\bar{y}_{jk\,t} = \bar{\hat{y}}_{jk\,t}$, which follows form the property of the residual term. These three properties tell, that dependent y-variable is decomposed into two orthogonal components, where the first one is written as a linear combination of the x-variables and the other as an error component that is orthogonal to all x-variables.

The first aggregation result follows from the algebraic results of OLS solution: For any micro class $B(k)$ aggregation over i $(a_i \in B(k))$ equals $\bar{y}_{jkt} = \hat{\alpha}_{jkt} + \underline{\bar{x}}'_{jkt} \underline{\hat{\beta}}_{jt}$. The second aggregation result follows from the first one: Aggregation over all observations for any equation j equals $\bar{y}_{jt} = \hat{\hat{\alpha}}_{jt} + \underline{\bar{x}}'_{jt} \underline{\hat{\beta}}_{jt}$, where $\hat{\hat{\alpha}}_{jt} = \sum_k f_{jk} \cdot \hat{\alpha}_{jkt}$, $\bar{y}_{jt} = \sum_k f_{jk} \cdot \bar{y}_{jkt}$ and $\underline{\bar{x}}'_{jt} = \sum_k f_{jk} \cdot \underline{\bar{x}}'_{jkt}$, where f_{jk} is the fixed relative frequency for the micro class j, k. This aggregation result satisfies also the condition, that the regression hyperplane passes trough the point of averages of input and output data for all equations separately.

When aggregating all micro relations into the macro level some preliminary results are needed. First, because $\underline{\hat{\beta}}_{jt}$ is the best linear unbiased estimator for $\underline{\beta}_{jt}$, then $\underline{\hat{\hat{\beta}}}_t = \sum_j f_j \cdot \underline{\hat{\beta}}_{jt}$, where f_j is the fixed relative frequency for equation j, form the best linear unbiased estimators for the population parameters. Similarly, the aggregation of constant terms over the micro classes, we get the estimator for the constant term in the macro level $\hat{\hat{\alpha}}_t^* = \sum_j \sum_k f_{jk} \cdot \hat{\alpha}_{jkt}$. Averaging (4.1) over all the observations and using the basic lemma of aggregation[7] we get

$$(4.2) \qquad \bar{y}_t = \hat{\hat{\alpha}}_t^* + \underline{\bar{x}}'_t \underline{\hat{\hat{\boldsymbol{\beta}}}}_t + \sum_{j=1}^{K} \text{cov}(x_{jt}, \hat{\beta}_{jt})$$

Terms $\left(\hat{\alpha}_{jkt} - \hat{\hat{\alpha}}_t^*\right)$ sum up to zero. The last term consists of covariance terms between the exogenous independent variables and their parameters. When aggregating deterministic micro equations the relation (4.2) is the final macro equation.

The solution (4.2) is easy to interpret. Let's look it once more: First, the basic micro partition P^* is designed as fine as practically feasible to describe adequately the use of labour inputs as production factors. Second, the extensive heterogeneous behaviour of economic agents is included. Third, the wage equations have specified as linear functions with respect to parameters and they belong into the family of flexible functions (quadratic in quantitative exogenous independent variables including interactions between them). Forth, only two assumptions is needed; $E(\varepsilon_{ijt} | x_{ijt}) = 0$ and $Var(\varepsilon_{ijt} | x_{ijt}) = \sigma^2_{jt} < \infty$. This condition may be weakened by including heteroscedastic random error terms over the basic micro partition P^* (i.e. the GLS –method). This will only make the estimation more efficient and do not change our analysis and synthesis stages in any way. Fifth, the parameterisation of the wage equations have been defined in a very detailed way including 'time-varying' parameters for each strata (i.e. $\hat{\alpha}_{jkt} \neq \hat{\alpha}_{j'k't'}$, $\forall j \neq j'$, $k \neq k'$, $t \neq t'$) and separate micro equations (i.e. $\underline{\hat{\beta}}_{jt} \neq \underline{\hat{\beta}}_{j't'}$, $\forall j \neq j'$, $k \neq k'$, $t \neq t'$). Sixth, for any arbitrarily chosen level of aggregation predictions of y-variable are best linear unbiased pre-

[7] The Basic Lemma of Aggregation: $\frac{1}{n} \sum_i x_i y_i = A(x) A(y) + cov(x, y)$. This follows from the identity $cov(x, y) = \frac{1}{n} \sum_i (x_i - A(x)) y_i = \frac{1}{n} \sum_i x_i y_i - A(x) A(y)$, where $A(x) = \frac{1}{n} \sum_i x_i$ and $A(y) = \frac{1}{n} \sum_i y_i$ are arithmetic means (Vartia, 1979).

dictions. The regression hyperplane passes through the point of averages of input and output variables[2] for any aggregation level.

The above properties taken together imply extremely operational and useful results for the analysis and synthesis of micro and macro relations. They simply say that the macro equation may be defined easily for any level of aggregation by averaging input and output variables and equation specific beta parameters. This is done under fairly general conditions that accept non-linearity's in the *x*-variables and heterogeneous behaviour of micro agents. The functional form may be generalised further to be some other flexible functional form, which is linear with respect to parameters (i.e. for example the polynomial functions are accepted).

The final form of the macro equation (4.2) is analogous to all *J* micro equations. The average of *y*-variable depends on averages of *x*-variables having the same non-linearity in *x*-variables as in micro equations (4.1). Even both the micro and macro equations belong to the same family of functions and have analogous form. Theil (1954) was right about the average macro parameters – they cannot be estimated unbiasedly using only the average macro variables. The macro equation (4.2) tells, that dependent aggregate variable depends not only on aggregate explanatory variables but also on the covariance terms; the covariance terms have appeared in the synthesis stage.

Since Theil (1954), these covariance terms have been considered as 'nuisance parameters' in estimating macro models by aggregate variables. We do not regard covariance terms any more as such but as important information in the specification of the models. Their inclusion in the models is made possible by the rapid growth of computer capacity. They have fundamental roles both in formulating the macro behaviour and in estimating its parameters. In fact, the covariance terms includes necessary information for unbiased estimation of macro parameters. To understand this, we suggest using the process of breaking a macro model down to observation level so that the role of covariances is displayed. This is an intermediate equation between the original micro equation (4.1) and the macro equation (4.2). We call such a method as being a solution backwards. What will this mean as a principle for our analysis? We break a complex macro model (4.2), including its covariance terms (previously taken as "nuisance parameters") and its average *x*-variables, down into simpler elements – into the observation level resulting

$$(4.3) \qquad y_{ijt} = \hat{\bar{\alpha}}_t^* + \underline{x}'_{ijt}\hat{\bar{\underline{\beta}}}_t + \left(\hat{\alpha}_{jkt} - \hat{\bar{\alpha}}_t^*\right) + \underline{x}'_{ijt}\left(\hat{\underline{\beta}}_{jt} - \hat{\bar{\underline{\beta}}}_t\right) + e_{ijt} \,.$$

This decomposes the regression model into two parts: A representative behaviour for all individuals $\hat{\bar{\alpha}}_t^* + \underline{x}'_{ijt}\hat{\bar{\underline{\beta}}}_t$ and two terms describing individual behaviour as deviation from the representative one $\left(\hat{\alpha}_{jkt} - \hat{\bar{\alpha}}_t^*\right)$

[2] For the quadratic term, we use $\frac{1}{n}\ \sum_i x_i^2 = A(x^2)$, where $A(x^2)$ is arithmetic mean of x_i^2. This may be expressed equivalently by $\frac{1}{n}\ \sum_i x_i^2 = \{A(x)\}^2 + var(x)$= square of the arithmetic mean plus the variance of *x*.

and $\underline{x}'_{ijt}\left(\underline{\hat{\beta}}_{jt}-\underline{\hat{\bar{\beta}}}_{t}\right)$. Equation (4.3) is a special case of the fundamental formula (2.5). The equation (4.3) is a reparametrized version[3] of (4.1) – only their arguments are decomposed differently. In the former the heterogeneity is distributed among all micro units and in the latter this is separated into its own terms. The wage equation consists of two sets of variables: The first set includes the exogenous independent variables $\left(1; \underline{x}'_{ij\,t}\right)$, and the other all the 'covariates' $\left(\hat{\alpha}_{jkt}-\hat{\bar{\alpha}}_{t}^{*}\right)$, $\underline{x}'_{ijt}\left(\underline{\hat{\beta}}_{jt}-\underline{\hat{\bar{\beta}}}_{t}\right)$, which are the microelements of the covariance terms[4] distributed element by element in the observation level. Dimensions for the both sets of variables are K+1. Next we free all the parameters of (4.3), including the unities of the heterogeneity terms, and form the second stage estimation equation:

(4.3b) $$\mathrm{y}_{ij\,t}=\bar{\alpha}_{t}^{*}+\underline{x}'_{ij\,t}\underline{\bar{\beta}}_{t}+\left(\hat{\alpha}_{jkt}-\hat{\bar{\alpha}}_{t}^{*}\right)\gamma_{t}+\underline{x}'_{ij\,t}\left(\underline{\hat{\beta}}_{jt}-\underline{\hat{\bar{\beta}}}_{t}\right)\delta_{t}+\mathrm{e}_{ijt}\,.$$

Because (4.3b) includes all the information needed to calculate all the wage equations, estimation of it by OLS regenerates these OLS-solutions or reduces to (4.4). Especially we have for its OLS-estimates $(\hat{\gamma}_{t},\hat{\delta}_{t})=(1{,}1)$ because the minimum of the least square of residuals is attained. The explicit form of (4.3b) includes an interesting property: it provides also the *reliability* or *standard errors* of the macro estimates (average estimates) or the variance-covariance matrices of the macro parameters. This seems to be a new powerful result for heterogeneous micro equations used in Suoperä (2003, 2004a,b, 2009a,b). The equation (4.3) includes all information that is needed – the knowledge of input and output variables in the observation level. Collecting that information observation by observation we may write the equation (4.3b) fully consistently as follows

(4.4) $$\underline{y}_{t}=\underline{\mathrm{X}}_{1t}\underline{\beta}_{1t}+\underline{X}_{2t}\underline{\beta}_{2t}+\underline{e}_{t}\,.$$

Here the first column of matrix $\underline{X}_{1t}$ is a vector of ones (i.e. constant) and the other columns are correspondingly micro explanatory variables of the wage model. The variables of the matrix $\underline{X}_{2t}$ are the covariates $\left(\hat{\alpha}_{jkt}-\hat{\bar{\alpha}}_{t}^{*}\right)$ and $\underline{x}'_{ij\,t}\left(\underline{\hat{\beta}}_{jt}-\underline{\hat{\bar{\beta}}}_{t}\right)$. The two estimates of the (K+1) dimensional vectors of parameters are $\underline{\hat{\beta}}'_{1t}=\left(\hat{\bar{\alpha}}_{t}^{*};\underline{\hat{\bar{\beta}}}'_{t}\right)$ and $\underline{\hat{\beta}}'_{2t}=\left(1;\underline{1}'\right)=$ vector of ones. The element i in the equation (4.4) is exactly the observation i in the equation (4.3). For example, the element i of the residual vector $\underline{e}_{t}$ is exactly the OLS residual i estimated in the analysis stage. So, the equation (4.4) is based purely on 'bookkeeping', because it is formed of known equations and their estimated parameters and is rewritten in the formulae (4.3-4). In the equation the first term on the right, $\underline{X}_{1t}\underline{\hat{\beta}}_{1t}$, indicates the common behaviour of all observations, while the term $\underline{X}_{2t}\underline{\hat{\beta}}_{2t}$ contains observa-

---

[3] That is $y_{ij\,t}=\hat{\alpha}_{jkt}+\underline{x}'_{ij\,t}\underline{\hat{\beta}}_{jt}+e_{ijt}=\hat{\bar{\alpha}}_{t}^{*}+\underline{x}'_{ij\,t}\underline{\hat{\bar{\beta}}}_{t}+\left(\hat{\alpha}_{jkt}-\hat{\bar{\alpha}}_{t}^{*}\right)+\underline{x}'_{ijt}\left(\underline{\hat{\beta}}_{jt}-\underline{\hat{\bar{\beta}}}_{t}\right)+\mathrm{e}_{ijt}$, cf. (2,4).

[4] Covariance is a translation invariant statistics, because cov(x+a,y+b)= $cov(x, y)$ for all constants (or shifts) a and b.

tion by observation the heterogeneous behaviour differing from the common behaviour. The equation (4.4) is just a rewritten original regression model including J separately estimated former wage equations (4.1). In addition to duplicating previous parameter estimates, we also get their standard errors. It may come first as a surprise that this large OLS-model must exactly replicate all the previous average parameter values and give the unity coefficients for the covariates. The reason for this is purely algebraic in character. We have in the large OLS-estimation all the sufficient information to produce OLS-solution not only for the combined large sector (say private one, see Table 5.1), but also for all its separate wage equations. Because (4.4) is capable of producing the previous OLS-solution with its overall minimum sum of squares, this actually is its OLS-solution. All other parameter estimates would give a larger sum of squares. The large OLS-estimation (4.4) replicates in this way all the previous groupwise regressions of the analysis stage: even the residuals are identical in them.

Dependent variable y depends on two conditional information sets - the set of exogenous independent variables and the set of covariates. This decomposition is extremely helpful for understanding the consequences of excluded heterogeneous behaviour in econometric modelling. A model where factors measuring heterogeneous behaviour are excluded produces in an unknown way biased estimates for the average parameters. This is special case of omitted variable bias (Judge et al, p. 839-844, 1982; Amemiya, p. 12-15, 1986, Greene, p. 401-404, 245, 1997. This bias will vanish if the set of covariates $\underline{X}_{2t}$ measuring heterogeneous behaviour is orthogonal with all the remaining variables $\underline{X}_{1t}$ (Greene, p. 246, 1997). This condition is so strong that one should never assume it in empirical analysis.

To conclude, the re-parameterised equation (4.1) in terms of (4.3) - (4.4) has a more central goal of producing the reliability of estimated vectors $\underline{\hat{\beta}}_{1t}, \underline{\hat{\beta}}_{2t}$ or the variance-covariance matrices of vectors $\underline{\hat{\beta}}_{1t}, \underline{\hat{\beta}}_{2t}$. The macro parameter estimates $\underline{\hat{\beta}}_{1t}$ turn out to be very accurate, see Table 5.1.

5. *First Analysis then Synthesis: Empirical Example of the Data Generating Process of Wages in Finland*

In estimation of wage models we use the SES data (the structure of earnings and salaries data) constructed by Statistics Finland. The private sector SES data contain hourly and monthly paid employees in manufacturing (TT -organisation), services (PT-organisation), automobiles and transport (AKL -organisation), church and unincorporated government enterprises. It covers about 60 to 70 presents of employees in the private sector. The private sector data has been fulfilled by sample of employees working in firms that do not belong to these organisations. The small firms (fewer than five employees) are excluded from the private sector data. In the government and municipal labour markets the SES data contains all employees working in these markets in the measurement period.

5.1 Model specifications

In the empirical part we examine the following specifications for wage equations:

1. *The Freakish Model*, where the micro partition *P** is excluded and only one wage equation, the same for all employees, is specified for each large sectors.
2. The second specification is called *the Parallel Model*, where we generalise the freakish model by including the micro partition *P**.
3. In the third specification, which we call *the Fair Model*, we exclude the one equation case and estimate the wage equations for the one-digit ISCO occupations (nine equations for each sector).
4. The estimation part ends to *the Good and the Best Models*, where the micro partitions *P** with fine estimation classes of wage equations is included.

The wage equations are specified as 'nested models', which makes possible to test different parameterisation of the wage models for example by the usual F –statistic.

The wage equations are specified as non-linear in experience and in education and log-linear in the wages. Other controlled variables include the indicators for women ("gender"), part time employment and non-permanent employment relationship. In the semi-micro level many additional effects have been controlled by the "dummy-technique" actually applied by centralising (i.e. by partition *P**) the applied models by concentrating on their "within variation" in major plants and jobs. Wage equations are estimated separately for the municipal, government and private sectors. The models are actually specified according to a familiar econometric modelling of labour markets in accordance with Becker's human capital theory.

We analyse hourly wages for regular working hours when ever feasible. Practically it means that we select every year a sample covering about 60 to 80 percent of employees from the SES data. The sample sizes will be about 100000, 280000 and 600000 in the government, municipal and private sectors respectively. Accuracy of the estimated wage equations can be evaluated using basic results of mathematical statistics, estimation and sampling. Results concerning these topics are well known for sampling surveys with moderate sizes of samples from finite populations. Only some of its most elementary results are referred here. Consider the mean of the population of some variable y and its standard estimator from a random sample. The mean is calculated from a random sample of n observations and its accuracy is measured by the standard error of the mean, which decreases towards zero with the sampling size n. To be concrete, let the standard error of mean be 10 log-percent for a single observation. For a given stratum $A(k)$ of size 100, the variance of the mean is one percent of the original variance and its standard error is one tenth of the standard error of one observation (i.e. one log-percent). For mean of size 10 000 the variance of mean is 0.001 percent of the original variance and its standard error is one hundredth part (i.e. 0.1 log-percent) of the standard error of single observation. Generally as the variance of mean goes zero at the speed of $1/n$, its standard error goes to zero at the speed of $1/\sqrt{n}$, the inverse of the square root of the number of observations. It is natural that the mean is estimated more accurately than its observations. The same statistical fact is transformed in the paper also to modelling. Aggregation of the models (and their parameters) from the sector to the macro level (or to the representative agent), almost trivially produces very accurate results.

This is quite a remarkable non-parametric result and we do not need to assume e.g. the normality of the variables. The similar increase in accuracy happens also when averages of regression coefficients are concerned. Their variances are also essentially inversely proportional to the number of observations. Based on these results, one should not be surprised any more, that the accuracy of our regression coefficients (measured in standard errors) is 50 – 100 times greater than usually in (macro) econometric papers. This is no wonder, because we have e.g. in the municipal sector about 280000 observations, which is 2800 times more than in a typical macro econometric time series study. Therefore, the standard errors of the estimates should diminish by the factor $\sqrt{2800} \approx 53$ and the t-values increase by the same amount (say from t = 2 to t = 53*2 = 106). Real effects become certainly ”significant for any p-values” as known in large sample theory for at least 80 years. If the t-value in a sample of n = 100 is of size 0,1, the corresponding value for n = 280 000 is t = 53*0,1 = 5,3. The macro parameter is ”certainly statistically significant” although ”the real effect” of the corresponding parameter value is usually of no actual relevance! Even irrelevant details from the point of view of ”real importance” becomes often statistically significant in really large samples, because infinite information reveals all non-zero effects.

Let us summarise the main points of our empirical study. First wage equations, starting from the Freakish Model and ending to large number of wage equations, are considered separately in different wage categories in different sectors. In the final estimation stage the classification is rather detailed – namely 323 separate wage equations together with the exceptionally detailed micro partition P^*. The large database of almost a million officially registered employees allows a detailed analysis of wage behaviour. For each wage equation, their typical characteristics and behaviours are allowed to vary freely from one group to another. For these subgroups semi-logarithmic non-linear regression models with several controlling “dummies” are estimated first. The models are specified according to Becker’s human capital theory. After the estimation the wage equations have been aggregated as averages which produce the macro equation (4.2). A backward solution of the macro equation results the equation (4.3) in the observation level, which may be represented equivalently by the equation (4.4) for all employees taken together. The standard errors for the macro parameters have derived from the equation (4.3) and are estimated by (4.4). The estimation results for different model specifications in the private sector in 2000 are presented in Table 5.1. Other similar estimation results for other sectors are given in Table 5.3.

Table 5.1: Estimation results for the wage equation (4.4) in the private sector in the year 2000. Estimates and their standard errors are multiplied by 100 for easier interpretations (except the constants, coefficients for covariates (unity's) and their standard errors). The reactions parameters of the inputs to log-wage are expressed in log-percentages, essentially the percentage change of wage caused by 1 per cent increase of the input, see L. Törnqvist, P. Vartia and Y. Vartia, 1985).

Statistics and variables	Freakish	Parallel	Fair	Good	Best
Observations	638495	638495	638495	638495	638495
Equations	1	1	9	142	183
Micro classes	1	20876	20876	20876	20876
Number of estimates	9	20884	20948	22012	22340
Adj. R^2	42.0683	80.678	81.1288	81.6561	81.7003
RMSE	26.3518	15.2187	15.0401	14.8285	14.8107
SSE	44337.7	14304.48	13969.07	13552.46	13511.72
Constant (α*)	4.278735	4.603768	4.064752	4.095357	4.083748
	(0.013473)	(0.007786)	(0.007908)	(0.007674)	(0.007668)
Female indicator (x_1)	-15.943	-8.664	-8.458	-8.364	-8.344
	(0.0678)	(0.0397)	(0.0412)	(0.0389)	(0.0388)
Part-time indicator (x_2)	-10.828	0.5165	1.7919	1.5105	1.4558
	(0.1428)	(0.0831)	(0.0824)	(0.0811)	(0.081)
Non-permanent (x_3)	-13.136	-6.918	-7.674	-8.309	-8.342
	(0.1217)	(0.0705)	(0.0696)	(0.0686)	(0.0685)
Education in years (x_4)	-10.24	-8.906	-1.98	-2.478	-2.318
	(0.1799)	(0.1039)	(0.107)	(0.104)	(0.104)
$x_5 = 0.5\, x_4\, x_4$	1.344	0.778	0.3318	0.3681	0.3569
	(0.0118)	(0.0068)	(0.0071)	(0.0069)	(0.0069)
Experience in years (x_6)	1.407	0.6039	1.7973	1.8286	1.8429
	(0.0272)	(0.0157)	(0.0151)	(0.0148)	(0.0148)
$x_7 = 0.5\, x_6\, x_6$	-0.062	-0.036	-0.044	-0.044	-0.044
	(5.778E-4)	(3.345E-4)	(3.314E-4)	(3.269E-4)	(3.264E-4)
Interaction ($x_8 = x_4\, x_6$)	0.0384	0.0461	-0.036	-0.038	-0.039
	(0.0018)	(0.001)	(9.783E-4)	(9.598E-4	(9.588E-4)
$He(\alpha) = \left(\hat{\alpha}_{jkt} - \hat{\bar{\alpha}}_t^*\right)$		1	1	1	1
		(0.000878)	(0.000917)	(0.000899)	(0.000897)
$He(x\beta) = x'_{ij\,t}\left(\hat{\beta}_{jt} - \hat{\bar{\beta}}_t\right)$			1	1	1
			(0.001076)	(0.000915)	(0.00091)
Average of covariates (=covariances)			1.5324	1.8445	1.9456

Table 5.1 coincides precisely with the macro equation (4.4). For the Freakish Model, the heterogeneous effects $He(\alpha)$ and $He(x\beta)$ will vanish from the macro model by definition - this specification excludes all heterogeneous behaviour. The Parallel Model includes the micro partition P^*, but assumes the same slope coefficients for all employees. In this specification or its suitable variations (see for example Bayard, Hellerstein and Troske, 2003; Korkeamäki, and Kyyrä, 2002; Korkeamäki, and Kyyrä, 2003; Korkeamäki, Kyyrä and Luukkonen, 2004; Mundlak, 1978), the beta heterogeneity is excluded by assumption (or is assumed to be random coefficients, which are independent of the x-variables). The heterogeneity effects, $He(x\beta)$, will vanish by definition. This is probably the most used specification in modelling the data generating process of wage behaviour in econometric studies of labour economics. It would suffer here badly from omitted variable bias. However, the standard errors of its biased parameter estimates closely approximate the SE's of the more flexible models. Other model specifications (Fair, Good and Best Models) include extensive α - and β - heterogeneity.

From the micro partition P^* and the equation (4.1) we see, that even in a small Finnish economy the number of estimated parameters will be high. In our case (Finland) the number of estimated parameters for the Best Model in private sector is 183×8+20 876 = 22 340 in year 2000. For government and municipal sectors the number of estimated parameters are 5 099 and 8 930, respectively (Best Model, Table 5.3). Extremely high number of observations (degrees of freedom) allows estimation of this many (and even more) parameters. It is clear, that it is not feasible to consider them in detail – we make synthesis of them by the principle described in the synthesis stage. We are able to summarise the large number of estimated parameters by mere eleven macro parameters (including the unity coefficient for the sum of covariates) represented in Table 5.1. Most apparent empirical outcomes concerning the labour markets in Finland are the following. First, the estimates of macro parameters converge always towards the estimates of the macro Best Model in all three sectors (see Table 5.2).

Their sizes correspond to our a priori expectations. Second, the SE's are small and t-values will be high for each model because of large sample properties discussed earlier. They all are always highly significant. Remarkable, the largest t –values (roughly 1000) always appear for the covariates measuring the α - and β - heterogeneity.

Interpretation of the coefficients of the macro models need some explanation. The Freakish Model treats all the wage earners as homogeneous and this extremely stiff nine parameter model excludes both the classification of labour inputs and the heterogeneity's of betas. Its slope estimators are based purely on the total sum of squares and cross products around the overall averages of input and output variables. As will be shown (Table 5.2), this specification is rejected always in all risk levels. All other estimators of different wage equations are the 'within-groups' estimators of the fixed effects model (FE-model). They have been estimated according to the Frisch, Waugh and Lovell –theorem first by eliminating the wage effects of the micro partition P^*. In statistical terms, the macro estimates are averages of the within-groups estimates. We refer to them as pure effects, because the classification disturbances have been removed in their estimation.

5.2 RE-model and Mundlak critique

An alternative model for the single equation FE-model (Parallel Model) including the micro partition is the 'random effects' model (RE-model) (Balestra and Nerlove, 1966; Wallace and Hussain, 1969). The use of the RE-model is justified by two arguments: First, the gain in efficiency because it utilise the 'between-groups' estimator in addition to the 'within-groups' estimator. Second, it is commonly argued that economic effects are indeed random and not fixed (Maddala, 1971). According to Mundlak (p. 70, 1978), these two arguments for deciding whether to use RE- or FE-model are inadequate. More important is the argument, that the RE-model has completely neglected the consequences of the dependencies (linear or more complicate) between the parameters and the explanatory variables. Mundlak (1978) shows, that the correlation between the effects and the explanatory variables leads to a biased estimator. We noticed a similar problem in the case of several wage equations - the dependencies between the heterogeneous beta parameters and the explanatory variables. By aggregating all the covariates, $\underline{x}'_{ijt}\left(\hat{\underline{\beta}}_{jt}-\hat{\hat{\underline{\beta}}}_{t}\right)$, in the equation (4.3) produces the covariance terms $cov\left(\underline{x}_{t},\hat{\underline{\beta}}_{t}\right)$ into the macro level. The observations of the covariates are elements of the covariance term and they are literally linear combinations of the deviation parameters (which sum and average to zeroes) and explanatory variables. Hildreth and Houck (1968), Swamy (1970, 1971, 1974) and Hsiao (1975) assumes the deviation parameters to be random and that the expectation of covariates, $E\left\{\underline{x}'_{ijt}\left(\underline{\beta}_{jt}-\underline{\beta}_{t}\right)\right\}$, are zeros (Greene, p. 669, 1997). They replace $\underline{x}'_{ijt}\left(\underline{\beta}_{jt}-\underline{\beta}_{t}\right)$ by the sample estimate $\underline{x}'_{ijt}\left(\hat{\underline{\beta}}_{jt}-\hat{\hat{\underline{\beta}}}_{t}\right)$, where $\hat{\underline{\beta}}_{jt}$ is the OLS estimate vector for equation *j* and the $\hat{\hat{\underline{\beta}}}_{t}$ is the average of them. The values of the $\underline{x}'_{ijt}\left(\hat{\underline{\beta}}_{jt}-\hat{\hat{\underline{\beta}}}_{t}\right)$ all vary around zero getting both positive and negative values. The averages of these must thus concentrate around some value - usually near zero, because positive and negative variable-specific contributions tend to annihilate each other. In fact, our covariance terms or their average provides a sample estimate for the correspondent population covariance's, which the proponents of the RE –model have assumed to vanished by a assumption. We can test this rather bold assumption of the RE –modellers simply using the observed values of the covariance terms, which should all be zeros. For the labour markets in Finland the averages of covariates are about 1.5 to 2 log-present (i.e. the last line of Tables 5.1 and 5.3) indicating dependencies between groupwise heterogeneous slope coefficients and their explanatory variables. They clearly are non-zeros in our case and the RE –modelling is unrealistic here. In the RE-model these dependencies have completely neglected – 'assuming them zero' is out of question in our case. The assumption to be satisfied in our case needs imposing, $E\left\{\underline{x}'_{ijt}\left(\underline{\beta}_{jt}-\underline{\beta}_{t}\right)\right\}=\underline{0}$, which would only complicate the analysis and lead to restricted estimators. However, unless the restrictions are correct in the population, the restricted estimator is biased (Greene, p. 669-670, p. 405, 1997) because of omitted variables.

5.3 Testing the nested models

In the synthesis stage, because of the algebraic aspects of the OLS solution, we didn't keep covariates, $\underline{x}'_{ijt}\left(\hat{\underline{\beta}}_{jt}-\hat{\hat{\underline{\beta}}}_{t}\right)$, as elements of error term similarly as in the RE-model, but as the set of ordinary explanatory variables. In each model including the α - and/or β - heterogeneity the (sum of) covariates are highly significant having exceptionally high *t*-values roughly 500, 700 and 1200 in the government, municipality and private sectors. Now in addition to well-motivated estimates and their standard errors, also tests of homogeneity of coefficients between the subgroups may be tested by standard statistical procedures. The residuals in the equation (4.4) coincide exactly with OLS residuals in (4.1) for all estimated models. Because the OLS residuals minimise the sum of squares within all separate wage equations (4.1) by definition, also the minimum sum of squares for them all in (4.4) must be the same. In fact, the model (4.4) replicates in this way all the previous groupwise regressions of the analysis stage: even the residuals are identical in them. Moreover, the models are nested with each other so that they can be obtained from each other by imposing suitable linear restrictions on parameters. For example, the Freakish Model can be derived from the Parallel Model by restricting all wage effects, $\hat{\alpha}_{jkt}$, in the Parallel Model equal to α estimated in the Freakish Model. Under these restrictions the residuals and their sum of squares

(SSE) in the Parallel Model will be changed equal to the residuals and their sum of squares in the Freakish Model. Similarly, the Parallel Model can be derived from the Fair Model by restricting all nine separately estimated β -vectors to a single β -vector estimated in the Parallel Model. Again imposing suitable linear restrictions on parameters in the Fair Model changed the residuals and their sum of squares equal to that of the Parallel Model. In general, the classification of wage equations follows the international standard classification of occupations (ISCO-88COM), which forms a hierarchical 'tree' system of occupations – all the employees are first divided into the disjoint main groups, which in turn are divided into the disjoint 2 digit occupation groups and so on. The classification of wage equations in the Fair Model is based to the ISCO main groups (9 groups). This means for example in the private sector case (see Table 4.1), that we need 72 (i.e. 9*8) linear restrictions for the Fair Model consisting of the 9 separately estimated wage equations, whose β -vectors are to be restricted into the β -vector estimated in the Parallel Model. Imposing these linear restrictions, the Fair Model reduces into the Parallel Model. Similarly, by imposing 1136 (i.e. 142×8) linear restrictions for the Good Model having 142 wage equations they will reduce into the nine wage equation estimated in the Fair Model. Even though, all the models have been estimated without these linear restrictions on parameters, the models are nested in sense that they can be obtained from each other's by making suitable linear restrictions on parameters of the more general model. This makes easy to test significance of α - and/or β - homogeneity hypothesis for example by the usual F -statistic[8]. In the Table 4.2 we show the values of the *F* test statistic for different hypothesis. In the first column, we have test statistics for the hypothesis that all the wage effects of the partition *P** are equal i.e. the α -homogeneity such that the Freakish and the Parallel Models are statistically equal. In the second column we have the values of the test statistics for the hypothesis of the β - homogeneity between the Fair and Parallel Models and so on.

Table 5.2: The values of the F –test statistics in testing the hypothesis of the α - or β - homogeneity between the Freakish, Parallel, Fair, Good and Best Models in the year 2000 (Sum of Squared Errors from the Table 5.1 and similar information in Table 5.3). The number of linear restrictions (*NR*) in parenthesis.

Sectors	Testing the α-homogeneity (Freakish vs. Parallel)	Testing the β -homogeneity: (Parallel vs. Fair)	Testing the β -homogeneity: (Fair vs. Good)	Testing the β -homogeneity[9]: (Good vs. Best)
Private	61.12 (*NR* = 20 876)	205.9 (*NR* = 72)	16.7 (*NR* = 1136)	2.8 (*NR* = 656)
Government	57.98 (*NR* = 4675)	33.9 (*NR* = 72)	8.10 (*NR* = 296)	2.7 (*NR* = 256)
Municipal	71.29 (*NR* = 8234)	58.5 (*NR* = 72)	16.0 (*NR* = 496)	3.4 (*NR* = 400)

The critical values of the F-test statistics reduce here to the critical values of χ^2_{NR}/NR being slightly greater than 1 for large values of *NR* (e.g. 1% critical value of $\chi^2_{60}/60$ is 1.46 and closer to one when *NR* > 60) because its nominator has essentially infinite degrees of freedom, see Greene (1997, p. 344 and p. 657). Even the F-statistics 2.7-3.4 of the fourth column have p-values smaller than 0.0001.

[8] $F = \{(SSE_0-SSE_1)/NR\}/\{SSE_1/(N_t-JR-NR)\}$, where SSE_0 is the sum of squared errors of the restricted model, SSE_1 is the sum of squared errors of the free model, $(N_t\text{-}JR\text{-}NR)$ is the degrees of freedom of the free model (see equation (5.7)) and *NR* is the number of linear restrictions (White, 1984).

The hypothesis of the α -homogeneity is rejected always in all sectors. Technically speaking, the Freakish Model suffers in our case from omitted biases of more than 20 000, 4500 and 8200 variables in the private, government and municipal sectors. As the Freakish Model clearly demonstrates, the micro partition of the production factors - mostly based on the Cartesian product between actual jobs and plants - must be included and their exclusions would generate serious omitted variable biases.

In the second column we have the test statistics for the hypothesis of the β - homogeneity between the Fair and Parallel Models. The hypothesis is rejected in all sectors. Similarly all other hypothesis of the β - homogeneity between different models are rejected. The hypothesis of α - and/or β - homogeneity are always rejected in all sectors. Typical for these test statistics are, that they decline step by step about forth part of the test statistics compared with the test statistics computed on the previous step. In the private sector the tests statistics decline even faster towards the critical values of the F –statistic.

Quite surprisingly even the average macro estimates are almost equal between the Fair, Good and Best models in all sectors (Table 5.1 , Table 5.3), the hypothesis of β - homogeneity between them will be rejected. Technically already quite crude classification of wage equations (i.e. the ISCO main groups) produces the macro estimates converging near to their true values even though the micro estimates are biased. This means for example for the Oaxaca (1973) decompositions (quality corrections in index number calculations), that the crude classification of regression equations leads unfortunately first into the biased quality corrections and second into the biased quality adjusted indexes because of biased estimates of micro equations. This of course holds also for the Parallel Models – because the estimates of quality characteristics are biased, the Oaxaca decomposition (i.e. for example the male-female quality adjusted wage differential and second their quality correction) will be biased in an unknown way for any subgroups of employees.

5.4 Sectoral Best Models in 1998-2000

Table 5.3 represents the estimation results for the Best Model in the year 1998, 1999 and 2000 in the government, municipal and private sector. In the government, municipal and private sector we estimate at least 52, 87 and 182 wage equations separately. Practically this means the estimation of 110 610 unknown parameters for these three years. Table 5.3 (i.e. only one page) represents the synthesis of these estimates, their standard errors, the adjusted coefficient of determination R^2, the sum of squared errors (SSE) and the root of mean squared errors (log-%). The estimates and their standard errors are multiplied by 100 and are expressed in more simple form as log-percentages, see L. Törnqvist, P. Vartia and Y. Vartia, 1985.

Constants and coefficients for covariates (ones) and standard errors of covariates are not expressed in log-%.

---

[9] F test statistics have been calculated by the maximum number of linear restrictions for the wage equations in the Best Model (i.e. F attain it's minimum in testing the β -homogeneity between the Good and Best Models.

Table 5.3: Estimation results for the wage equations (4.2) in the government and municipal sectors in the years 1998, 1999 and 2000, log-% (estimates and their standard errors are multiplied by 100 (except the constants, coefficients for covariates (unity's) and their standard errors) and are expressed in log-percentages, see L. Törnqvist, P. Vartia and Y. Vartia, 1985).

Statistics and variables	Government sector			Municipal sector			Private sector		
Year	1998	1999	2000	1998	1999	2000	1998	1999	2000
Observations	101336	103216	102147	246296	254210	266467	614500	626173	638495
Equations	53	52	53	89	88	87	183	182	183
Micro classes	4993	4725	4675	7864	11398	8234	19650	20443	20876
Number of estimates	5414	5141	5099	8576	12102	8930	21114	21899	22340
Adj. R^2	88.8619	87.562	87.4047	87.5137	87.8828	87.6395	82.2682	81.4843	81.7003
RMSE	10.3735	11.0514	11.0896	9.4771	9.2413	9.475	14.0916	14.6109	14.8107
SSE	1031.611	1197.195	1192.838	2134.305	2066.889	2311.29	11779.46	12896	13511.72
Constant (α*)	4.0214	3.929591	3.992516	4.244964	4.149977	4.022095	3.975635	4.08905	4.083748
	(0.01775)	(0.012266)	(0.012261)	(0.009056)	(0.008592)	(0.008397)	(0.007387)	(0.007604)	(0.007668)
Female indicator (x_1)	-1.549	-1.943	-2.124	-1.195	-1.014	-1.307	-8.108	-8.019	-8.344
	(0.0704)	(0.0749)	(0.075)	(0.0502)	(0.0482)	(0.049)	(0.0376)	(0.0387)	(0.0388)
Part-time indicator (x_2)	-1.129	-0.371	-1.394	3.492	2.784	2.1398	-0.839	-0.835	1.4558
	(0.1836)	(0.1819)	(0.173)	(0.0861)	(0.0772)	(0.0741)	(0.0805)	(0.0798)	(0.081)
Non-permanent (x_3)	-6.913	-7.068	-6.05	-6.343	-6.645	-7.2	-7.286	-7.289	-8.342
	(0.0847)	(0.0891)	(0.0915)	(0.055)	(0.0518)	(0.0505)	(0.071)	(0.0757)	(0.0685)
Education in years (x_4)	-2.11	-0.033	-0.771	-7.285	-5.358	-2.349	-1.889	-3.286	-2.318
	(0.1463)	(0.1535)	(0.1529)	(0.1166)	(0.1112)	(0.1088)	(0.1003)	(0.1029)	(0.104)
$x_5 = 0.5\, x_4\, x_4$	0.2771	0.1041	0.1822	0.6941	0.5349	0.2671	0.3148	0.433	0.3569
	(0.0092)	(0.0097)	(0.0096)	(0.0075)	(0.0072)	(0.0071)	(0.0067)	(0.0068)	(0.0069)
Experience in years (x_6)	1.7126	1.7304	1.9108	1.2412	1.2346	1.1411	1.8034	1.7844	1.8429
	(0.0286)	(0.0292)	(0.029)	(0.0198)	(0.0184)	(0.0178)	(0.0148)	(0.0151)	(0.0148)
$x_7 = 0.5\, x_6\, x_6$	-0.046	-0.047	-0.049	-0.034	-0.034	-0.032	-0.044	-0.045	-0.044
	(6.929E-4)	(7.155E-4)	(7.121E-4)	(4.153E-4)	(3.869E-4)	(3.719E-4)	(3.233E-4)	(3.297E-4)	(3.264E-4)
Interaction ($x_8 = x_4\, x_6$)	-0.008	-0.006	-0.017	-0.011	-0.011	-0.006	-0.035	-0.033	-0.039
	(0.0014)	(0.0014)	(0.0014)	(0.0012)	(0.0011)	(0.0011)	(9.744E-4)	(9.842E-4)	(9.588E-4)
$He(\alpha)=\left(\hat{\alpha}_{jkt}-\hat{\hat{\alpha}}_t^*\right)$	1	1	1	1	1	1	1	1	1
	(0.001965)	(0.002013)	(0.00206)	(0.001457)	(0.001393)	(0.001389)	(0.000906)	(0.00092)	(0.000897)
$He(x\beta)=x'_{ijt}\left(\hat{\beta}_{jt}-\hat{\hat{\beta}}_t\right)$	1	1	1	1	1	1	1	1	1
	(0.001944)	(0.002084)	(0.00203)	(0.001457)	(0.001392)	(0.001393)	(0.00092)	(0.000922)	(0.00091)
Average of covariates	1.32	2.08	1.51	0.99	1.16	1.49	2.10	1.68	1.95

The most important slope coefficients having the highest t-values (the female indicator (x_1), the non-permanent employees (x_3), the experience and it's squared form multiplied by half (x_6 and x_7) and even the interaction between education and experience (x_8)) are stable in time. Looking at the standard errors of the slope coefficients of these variables, even though they are closely related in time, homogeneities between them within each sector will be surely rejected. The part-time indicator (x_2) varies between 1 to 2.3 log-percent in different employee markets, whereas the slope coefficients of the education seem to be unstable between successive years. Table 5.4 shows the marginal effect of education and experience on wages on the average points of these variables in the years 1998, 1999 and 2000.

Table 5.4: The marginal effects of education and experience on wages in the average points (in parenthesis) in the government, municipal and private sector in 1998, 1999 and 2000, log-%.

Education: $\hat{\beta}_{4t} + \bar{x}_{4t}\hat{\beta}_{5t} + \bar{x}_{6t}\hat{\beta}_{8t}$	Government	Municipal	Private
1998	1.59 ($\bar{x}_4 = 13.98$)	1.19 ($\bar{x}_4 = 12.60$)	1.27 ($\bar{x}_4 = 12.34$)
1999	1.30 ($\bar{x}_4 = 14.01$)	1.12 ($\bar{x}_4 = 12.62$)	1.38 ($\bar{x}_4 = 12.38$)
2000	1.43 ($\bar{x}_4 = 14.07$)	0.88 ($\bar{x}_4 = 12.65$)	1.31 ($\bar{x}_4 = 12.45$)
Experience: $\hat{\beta}_{6t} + \bar{x}_{6t}\hat{\beta}_{7t} + \bar{x}_{4t}\hat{\beta}_{8t}$	Government	Municipal	Private
1998	0.62 ($\bar{x}_6 = 21.40$)	0.27 ($\bar{x}_6 = 24.57$)	0.46 ($\bar{x}_6 = 20.84$)
1999	0.64 ($\bar{x}_6 = 21.40$)	0.25 ($\bar{x}_6 = 24.81$)	0.43 ($\bar{x}_6 = 20.98$)
2000	0.62 ($\bar{x}_6 = 21.47$)	0.27 ($\bar{x}_6 = 24.83$)	0.43 ($\bar{x}_6 = 20.98$)

The marginal effects of education and experience estimated in the average points of these variables varies slightly within sectors, but are surely significantly different between sectors being highest in the government and lowest in the municipal sector. Marginal effects of experience are extremely systematic and stable. As an example, one additional year of education (or experience) in private sector increases wages only by 1.27-1.38 % (0,43-0.46%). These are rather small additional effects of non-typical education (or experience) within micro classes of jobs. Wages are not paid on education or experience but on the job done, which requires a certain education and experience. Non-typical additional education is valued positively but not as strongly as is usually anticipated.

6. Conclusions

The study is divided into the analysis and synthesis stages. The analysis stage consists of the partition of labour input and the estimation of wage equations. The partitions of employees into disjoint sets are done for the ISCO occupations in the 4- or 5-digit level in three separate stages described in section two. The partition is performed mostly in accordance with the basic textbooks of the production analysis (see for example Chambers, 1988) – a minor part of data (i.e. small jobs or duties in plants) is partitioned by actual jobs only. The wage equations are specified in five different ways starting from the one wage equation without partition and ending to fine classification of wage equations including the micro partition of the labour inputs. Different models are always specified as nested such that they can be obtained from each others by imposing suitable linear restrictions on parameters of the more parameterised model. The wage equations are specified to belong into the family of flexible functions and the explanatory variables used in regressions are in accordance with the human capital theory of Becker. About 5 000, 10 000 and 20 000 unknown parameters are estimated by the OLS-method in government, municipal and private sectors every year (see Table 5.1 and Table 5.3). Somebody may feel now, that it is impossible to digest so large, detailed and complicated information. Some explanation why we are doing so is needed. Our intent on doing so is based on the following facts: Detailed information of wage behaviour is necessary, first in testing the homogeneity hypothesis of partition effects and second the homogeneity hypothesis of equal slope coefficients of different wage equations. We do not lay the foundation of our analysis on commonly used habits, like ad hoc selection of the parallel specification (i.e. slope coefficients equal for all employees), but more likely select the model specification after testing them in pairs.

In statistical inference of the wage models, we first test the homogeneity hypothesis of wage effects for partition of production factors. The homogeneity hypotheses are always rejected. Second, we test the hypothesis of homogeneity of slope coefficients (i.e. hypothesis of the parallel models) and similarly they are always rejected. The values of usual F test statistics converges quite rapidly towards their critical values suggesting to use detailed partition of labour input together with detailed classification of wage equations (see Table 5.2).

The overwhelming amount of details arouses the need to make the synthesis of the estimated wage models to get a better overview of their relevant aspects. Utilising the basic lemma of aggregation (Vartia, 1979, 2008a) the aggregation over all employees leads to the macro model (4.2). The parameters in the macro model are all known once the micro parameters are estimated and estimation of them is not necessarily needed. They are calculated by aggregating OLS-estimates from the micro level. The standard errors of the macro parameters are instead not easily found. For the estimation of variance-covariance matrix for the average macro parameters, we suggest a backward solution of the macro model (4.2) into the observation level. Practically this means that the covariance terms appeared in the synthesis stage are broken down to its elements in the observation level, to the covariates. These covariates are literally linear combinations of the deviation parameters (i.e. $\underline{\hat{\beta}}_{jt} - \underline{\hat{\bar{\beta}}}_{t}$) multiplied with appropriate explanatory variables. The method reproduces the estimated wage equations exactly in the observation level, but now in the mean-deviation re-parameterised form (4.3). The first part of it consists of the common behaviour described by the mean parameter part of the equation and the second part the heterogeneity effects described by the covariates. Putting all the observations together, we get the equation (4.4), where, of course, the first part in the right describes the common and the second part the heterogeneous behaviours. The model (4.4), based on the mean-deviation re-parameterisation, is mathematically exactly equal in all arguments compared with the estimated wage equation (4.1) together taken – even the residuals are equal observation by observation. This is a known result mentioned shortly e.g. by Balestra and Nerlove in their introduction in Matyás and Sevestre (1996). They just simply state that the total sum of squares of one large SUR-model reduces to the sum of squares summed over the equations. This means, that the separately estimated wage equations by the OLS method are in fact equivalent to one large SUR estimation with diagonal covariance matrix. Therefore, minimising the sum of squared residual first in the equation level is equivalent to the minimising all of them at the same time in the mean-deviation re-parameterised form for sector as a whole. So, the estimation of the wage equation (4.4) separately for sectors as a whole reproduces exactly the average OLS-estimates and the unity coefficients for the covariates. Because the estimates for the macro model are already known, they need not necessarily be estimated once more. The re-parameterisation has a more central goal – the model (4.4) can be used to estimate the variance-covariance matrix for the macro estimate vectors by well-known partitioned regression method.

The re-parameterised model (4.3) has similarities with the RE –model (see seminal papers Hildreth and Houck, 1968, Swamy, 1970, 1971, 1974 and Hsiao, 1975). The covariates measuring the heterogeneous behaviour of wages have exactly the same form - differences arouses in the use of this conditional information contained in the covariates. RE-modellers keep the covariates as components of error term and assume them to be orthogonal to the set of explanatory variables with mean zeros and constant variances, whereas we keep the covariates as commonly used explanatory variables in the OLS-method. The equation (4.4) reveals an interesting property – if two sets of variables are orthogonal, (i.e. $\underline{X}'_{2t}\underline{X}_{1t} \equiv \underline{X}'_{1t}\underline{X}_{2t} \equiv \underline{0}$), then two separate coefficient vectors and their variance-covariance matrix can be obtained by separate OLS regressions – $\underline{y}_t$ on $\underline{X}_{1t}$ alone and $\underline{y}_t$ on $\underline{X}_{2t}$ alone (W. Greene, p. 245-250, 1997). Similarly, as Mundlak stressed (1978), the RE –model completely neglects the dependencies between heterogeneous slope coefficients and their appropriate explanatory variables. Moreover the sample estimates of the averages of covariates (i.e. covariance terms) are never near to zeros, but varies mostly between 1.5 to 2.1 log-% being always positive in sign.

In aggregating wage equations (empirical OLS –solutions) from the observation level into the macro level, two covariance terms appears in the macro wage equation – the first one in aggregation of partition specific wage effects (i.e. constant terms) and an other one in aggregation of systematic parts of wage equations. Unfortunately, the empirical macro relation (4.4) cannot be found without empirical microanalysis and synthesis of the estimation results found in the estimation stage. Even for a large number of time periods, when empirical data includes heterogeneous behaviour and macro parameters are estimated by regressing $\bar{y}_t$ on $\bar{x}'_t$- variables, macro parameters will be possibly biased (excluding the case of two orthogonal sets of variables). Regressing $\bar{y}_t$ on $\bar{x}'_t$- variables (or between the totals) is very inefficient compared to aggregating the estimated micro equations. We simply propose the following order: first analysis and then synthesis of this conditional information. The mainstream macro econometrics uses the reverse order. First micro variables are aggregated into the macro series. After that OLS and the aggregated time series are used to estimate the macro economic relation. This common macro time series methodology leads to very inefficient and possibly biased results. It is not necessarily badly biased but will be a highly inaccurate because of serious loss of information.

References

Amemiya, T. “Advanced Econometrics”, Basil Blackwell Ltd, Oxford, 1986.

Baldwin, T. “Routledge Encyclopedia of Philosophy”, under ‘Analytical Philosophy’, 1998.

Balestra, P. and Nerlove, M. “Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand of Natural Gas”, Econometrica, vol. 34, p.585-612, 1966.

Bayard, Hellerstein and Troske, “New Evidence on Sex Segregation and Sex differences in Wages from Matched Employee-employer Data”, Journal of Labour Economics, vol. 21, no. 4, 887-922, 2003.

Blackburn, S. “Oxford Dictionary of Philosophy”, under ‘Analysis’, 1996.

Chambers, R.G. “Applied Production Analysis: a Dual Approach”, New York: Cambridge University Press, 1988.

van Dahl J., Merkies, A.H.Q.M., "Aggregation in economic research", D. Reidel Publishing Company, Dordrecht-Boston-Lancaster, 1984

Davidson, R. and MacKinnon, J. “Estimation and Inference in Economics”, New York: Oxford University Press, 1993.

Greene, W. “Econometric Analysis”, Prentice-Hall International, Inc. (third ed.), 1997.

Hicks, J. R. “Value and Capital”, Oxford, Clarendon Press, 1939.

Hildreth, C. and Houck, C. “Some Estimators for a Linear Model with Random Coefficients”, Journal of the American Statistical Association, vol. 63, p.584-595, 1968.

Hintikka, J. and Remes, U. “The Method of Analysis”, Dordrecht, 1974.

Hsiao, C. “Analysis of Panel Data”, New York: Cambridge University Press, 1986.

Hsiao, C. “Some Estimation Methods for a Random Coefficient Model”, Econometrica, vol. 43, p.305-325, 1975.

Judge, Hill, Griffiths, Lutkepol and Lee, “An Introduction to the Theory and Practice of Econometrics, New York: John Wiley and Sons, 1982.

Koev, E. “Combining classification and hedonic quality adjustment”, Licentiate thesis, University of Helsinki, 2003.

Koev, E. and Suoperä, A. “Price Indices for Detached Houses and Detached Housing Plots, 1985=100”, Statistics Finland, 2004.

Korkeamäki, and Kyyrä, “The Gender Wage Gap and Sex Segregation in Finland”, VATT Discussion Papers no. 288, 2002.

Korkeamäki, and Kyyrä, “Explaining Gender Wage Differentials: Findings from a Random Effects Model”, VATT Discussion Papers no. 320, 2003.

Korkeamäki, Kyyrä and Luukkonen, “Miesten ja naisten palkkaerot yksityisellä sektorilla”, VATT Discussion Papers no. 327, 2004.

Lintunen, J.- Ropponen, O. -Vartia, Y. ”Micro meets Macro via Aggregation”. Hecer Discussion Paper No 259 / 2009.

Maddala, G. “The Use of Variance Components Models in Pooling Cross Section and Time Series Data”, Econometrica, vol. 39, p.341-358, 1971.

Matyás, L. and Sevestre, P., Eds. "The Econometrics of Panel Data: Handbook of Theory and Applications, 2nd ed. Dordrecht: Kluwer-Nijoff, 1996.

Mundlak, Y. "On the Pooling of Time Series and Gross Sectional Data", Econometrica, vol. 46, p.69-86, 1978.

Oaxaca, R. "Male-Female Wage Differentials in Urban Labour Markets", International Economic Review, vol. 14, p. 693-709, 1973.

Suoperä, A. "Analysis of the EU's LCI –formula: Empirical example of the construction of the LC index by imputed statistical data", Unpublished manuscript in EU's LCI project, 2003.

Suoperä, A. "Muutama uusi näkökulma hinta-aggregoinnista ja hedonisista indeksimenetelmistä: Empiirinen esimerkki toimisto- ja liiketilojen vuokriin", Statistics Finland, Unpublished manuscript, Helsinki, 2004a.

Suoperä, A. "Survey on the rents of office and shop premises: Some new perspectives on price aggregation and hedonic index methods", Statistics Finland, Unpublished manuscript, Helsinki, 2004b.

Suoperä, A. "Naisten ja miesten palkkaerot valtiosektorilla ja yleisillä työmarkkinoilla", Valtiovarainministeriön toimeksianto, Unpublished manuscript, Helsinki, 2009a.

Suoperä, A. "Kunnan ja yleisten työmarkkinoiden palkkaerot: Hedoninen menetelmä", Kunnallinen työmarkkinalaitos, Unpublished manuscript, Helsinki, 2009b.

Swamy, P. "Efficient Inference in Random Coefficient Regression Models", Econometrica, vol. 38, p.311-323, 1970.

Swamy, P. "Statistical Inference in Random Coefficient Regression Models", New York: Springer-Verlag, 1971.

Swamy, P. "Linear Models with Random Coefficients", In P. Zarembka, ed., Frontiers in Econometrics, New York: Academic Press, 1974.

Theil, H. "Linear Aggregation of Economic Relations", Amsterdam, North-Holland Pub. Co., 1954.

Theil, H. "Linear Aggregation in Input-Output Analysis", Econometrica, vol. 25, p.111-122, 1957.

Törnqvist, L., Vartia, P. and Vartia, Y. "How should Relative Changes be Measured?", The American Statistics, vol. 39, No. 1, p.43-46, 1985.

Vartia, Y.O. "Kvadraattisten mikroyhtälöiden aggregoinnista", ETLA Discussion Papers No 25, Helsinki, 1979.

Vartia, Y.O. "On the Aggregation of Quadratic Micro Equations", Hecer Discussion Papers No 248, 2008a.

Vartia, Y.O. "Integration of Micro and Macro Explanations". Hecer Discussion Papers No 239 / 2008b.

Vartia, Y.O. "Whole and its Parts: Micro Foundations of Macro Behaviour". Hecer Discussion Paper No 257 / 2009.

Wallace, T. and Hussain, A. "The Use of Error Components in Combining Cross Section with Time Series Data", Econometrica, vol. 37, p.55-72, 1969.