# Use of regression trees in the study of nonparametric wage structures

by

Ricardo Mora

Thesis submitted for the degree of doctor of Philosophy (PhD)

on February, 1998

at

The London School of Economics and Political Science

University of London

## Acknowledgements

My foremost thanks are due to Danny Quah for his supervision and encouragement over the last four years. My thanks should also go to Irini Moustaki and Pedro Delicado not only for their advice and helpful comments but also for their kind readiness to help whenever asked. I want to express my thanks to all members of the Economics Department at the London School of Economics where I have found a very friendly and encouraging environment. Financial support from the Bank of Spain is also greatly acknowledged.

However, my deepest gratitude is owed to my parents, Hortensia and Ricardo, and my partner, Adelheid. Their love, help and understanding has taken me where I am now. I would like to dedicate this thesis to them.

# Abstract

This study is concerned with the application of multivariate nonparametric models known as regression trees to the analysis of the U.S. wage structure. In Chapter 1, I first review regression trees and other available multivariate nonparametric techniques, highlighting their differences and common features. In the second part of Chapter 1, I look at the literature on the U.S. wage structure in connection with the issue of functional specification and argue that regression trees is particularly well suited for analyzing wage structures. In Chapter 2, I implement regression trees on U.S. wages for white male workers to estimate experience-wage profiles and unveil local sudden breaks in the profiles at the end of the working life. For 1980, these breaks account for about 50% of the negative average differential between the last two experience groups. This effect decreases continuously until 1995. In Chapter 3 I propose a simple extension of the Oaxaca-type average wage gap decompositions between any two groups of workers. This procedure can be carried out without any compromise in the interpretation using a nonparametric wage structure. I then study wage gap decompositions for Mexican workers in the U.S. labor market. Finally, in Chapter 4 I apply regression trees to study both the relative growth performance of workers' real wages and the sources of wage dispersion and its evolution in the U.S. from 1980 onwards. On trends, the technique uncovers a linear structure for the growth experience of white workers with less than forty years of experience. On dispersion, at least 10% of the increase in observed variance came from changes in the structure of wages itself.

## Contents

## List of tables

## List of figures

# Chapter 1

## Introduction: Regression trees and nonparametric wage structures

In the following chapters I will be concerned with the application of multivariate nonparametric models known as regression trees to the analysis of the U.S. wage structure. In this chapter, I first review neural networks, projection pursuit and regression trees -all of them available multivariate nonparametric techniques-. Then, I present a brief survey on empirical results on the wage structure in the U.S. labor market. The chapter ends with an outline of the rest of the thesis.

### 1.1 Regression trees and predictive learning

Mathematics, statistics, engineering, artificial intelligence, and neural networks all study the problem of predictive learning. A simple example of a predictive learning problem involves one set of variables, sometimes called inputs, sometimes called explanatory variables, and other times called independent variables, and just one variable called either output, response, or dependent variable. This variable is defined over a subset of the real line[1].

The problem can consist of designing a system with which interpolations of the output can be obtained from the inputs. This is a mathematical problem called function approximation. If some of the inputs are not observable, the mathematical model is of a statistical nature. Let the model be

$$y = f(x_1, \cdots, x_n) + e = f(\mathbf{x}) + e. \qquad \textbf{(1.1)}$$

In **(1.1)** I assume an additive relation for the two sets of independent variables. The residual e is the effect of all the unobserved variables on the dependent variable. A fixed effect can be thought both as an observed constant effect and as the effect of a constant unobserved variable. This makes model **(1.1)** ambiguous. To solve this ambiguity, the unobserved residual e can be defined by $E[e|\mathbf{x}]=0$ where E[] is the expectation operator.

The objective of predictive learning is to obtain a useful

[1] In statistics, when the response is categorical, the problem is one of classification or discriminant analysis.

approximation to $f(\mathbf{x})=E[y|\mathbf{x}]$.

Neural networks use finite learning samples $\{y_i,\mathbf{x}_i\}_i$ and learning algorithms to produce outputs f* in response to inputs **x*** and adapt its outputs f* according to residuals $\{(y_i-f^*(\mathbf{x}_i))\}_i$. Statistics seeks to obtain true approximations of f on the local area defined by the finite sample $\{y_i,\mathbf{x}_i\}_i$. Thus, the stress in statistics is on validating results while the stress in neural networks is on learning.

Predictive learning is implemented through an optimization problem on a finite sample. The objective function could be

$$\hat{f}(\mathbf{x}) = \underset{g(\mathbf{x})}{\operatorname{argmin}} \sum_{i=1}^{N} \left[y_i - g(\mathbf{x}_i)\right]^2 \quad . \qquad \textbf{(1.2)}$$

An infinite number of functions can yield the minimum value, zero, for the solution in **(1.2)**. In order to obtain a well defined problem in finite samples[2], further assumptions must be

---

[2] The similar problem for a "sample of infinite size" is well defined in the sense that only f(·) will be the solution of the problem. Thus,

$$f(\mathbf{x}) = \underset{g}{\operatorname{argmin}} \int E_\varepsilon[\, f(\mathbf{x}) + e - g(\mathbf{x})]^2\, p(\mathbf{x})d\mathbf{x}.$$

See e.g. Friedman (1994).

added. Every predictive learning method can be characterized as a set of constraints on problem **(1.2)**.

There are two general forms of restrictions. The first way to restrict the problem is to place constraints on the class of eligible functions $g(\mathbf{x})$. The second alternative consists of selecting a small number of input variables, so that the dimension of the variable space is small relative to the finite sample.

For high dimensional spaces, the second alternative is in many practical cases not feasible. The reasons are generally known as the curse of dimensionality. Firstly, there is the problem of minimum required number of observations to have a densely packed sample. If 100 observations represents a dense sample for a single input system, then for $\mathbf{x} \in R^N$, the required sample would be $100^N$. For 10 input variables we would need $10^{20}$ observations. Secondly, all observations are close to an edge of the sample[3]. Consider a sample of three observations in a one-dimensional problem. Two of the observations must be at the edge of the sample. If you add a new dimension (input) to the problem, all three observations will be at the edge, making interpolation impossible.

[3] That is, there are no more observations between that observation and the edge of the space along one dimension.

The only way to overcome these problems in a high dimensional setup is by incorporating outside knowledge -knowledge not related to the sample- in the way of functional restrictions. A direct way of doing this is by incorporating a penalty to criterion **(1.2)**:

$$\hat{f}(\mathbf{x}) = \underset{g(\mathbf{x})}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left[ y_i - g(\mathbf{x}_i) \right]^2 + \lambda\, \phi[g(\mathbf{x})] \right\}. \qquad \textbf{(1.3)}$$

The best penalty $\phi \geq 0$ is one that has small values for those $g(\mathbf{x})$ that are close to $f(\mathbf{x})$ and large values otherwise. The strength parameter $\lambda$ is inserted so that the penalty effect can be adjusted independently from its form.

If no restriction is imposed on $\lambda$, then we might want to solve the two-step problem:

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmin}} \sum_{i=1}^{N} \left[ y_i - \hat{f}(\mathbf{x}_i \mid \lambda) \right]^2 \ . \qquad \textbf{(1.4)}$$

However, this is not very interesting since **(1.4)** always yields

$$\hat{\lambda} = 0. \qquad \textbf{(1.5)}$$

The best fit is given by the solution without penalty. This fit, however, may be very poor for any other independent sample, i.e., the estimated surface may have very low predictive power. A common approach to overcome overfitting is to divide the sample into a learning and a test sample. The learning sample pursuit, neural networks, and regression trees[5]. The objective is to highlight what is common between these methods and what characterizes regression trees more neatly.

### 1.1.1 Least squares parametric regression

Penalty function $\phi$ takes the form

$$\begin{aligned} \Phi[g(\mathbf{x})] &= \infty \quad \text{if} \quad g(\mathbf{x}) \neq h(\mathbf{x} \mid \theta_1, \ldots, \theta_k) \\ \Phi[g(\mathbf{x})] &= 0 \quad \text{if} \quad g(\mathbf{x}) = h(\mathbf{x} \mid \theta_1, \ldots, \theta_k), \end{aligned} \qquad \textbf{(1.6)}$$

where $h(\cdot)$ is a continuous, doubly differentiable function with respect to $\theta$ that is completely specified but for a finite number of unknown parameters in $\theta$.

Problem **(1.3)** collapses to

$$\left\{\hat{\theta}_1, \ldots, \hat{\theta}_k\right\} = \underset{\theta_1, \ldots, \theta_k}{\operatorname{argmin}} \left\{\sum_{i=1}^{N} [y_i - h(\mathbf{x}_i \mid \theta_1, \ldots, \theta_k)]^2\right\}. \qquad \textbf{(1.7)}$$

The choice of $\lambda$ becomes irrelevant and there is no need for cross-validation.

---

[5] For an introduction of neural networks, see e.g. Kuan and White (1994). See Huber (1984) for a review on projection pursuit. The basic reference for regression trees is Breiman et alia (1984). These are the most popular predictive learning models.

Assuming certain statistical properties for the error term, e, leads to a complete inferential theory.

The function h(·) needs not be linear, as it is the case of the binary logit probability model, where

$$h(\mathbf{x} \mid \theta) = G(\mathbf{x},\theta) = \frac{1}{1 + \exp(-\mathbf{x}'\theta)}. \qquad \textbf{(1.8)}$$

This is a multivariate nonlinear model and can be thought of as a three-layer system. The first layer corresponds to the inputs, $\mathbf{x}$. The second layer is the linear index of the inputs, $\mathbf{x}'\theta$. The third layer is the output h(·).

The model is flexible in the sense of giving nonlinear responses because of the nonlinear relation between the second and the third layer. This fundamental source of flexibility is fully exploited in artificial neural networks, projection pursuit, and regression trees.

1.1.2 Single hidden layer artificial neural networks

Penalty function $\phi$ now takes the form

$$\begin{aligned} \Phi[g(\mathbf{x})] &= \infty \quad \text{if } g(\mathbf{x}) \notin \left\{ h_T(\mathbf{x} \mid \theta) \right\}_{T=1}^{\infty} \\ \Phi[g(\mathbf{x})] &= \mu(\theta) < \infty \quad \text{else} \end{aligned} \qquad \textbf{(1.9)}$$

where the family of functions $\{h_T\}$ are

$$h_T(\mathbf{x}|\mathbf{a},\gamma) = \sum_{t=1}^{T} a_t \cdot s(\mathbf{x}'\gamma_t) \qquad \textbf{(1.10)}$$

and $s(\cdot)$ is any smooth sigmoid function such that $0 \le s(\cdot) \le 1$. Thus, $G(\cdot)$ in **(1.8)** is one of such functions[6].

The indicator function

$$I(\mu) = \begin{cases} 1 & \text{if } \mu \text{ true} \\ 0 & \text{if } \mu \text{ false} \end{cases} \qquad \textbf{(1.11)}$$

was chosen as $s(\cdot)$ in the first published articles on neural networks. The indicator function is an activation devise. It incorporates a new structure to model **(1.10)** when a threshold is overcome in an index or hidden layer.

For some applications, smooth functions as activation devices seem more useful than step functions. In biological neural systems there is a tendency of certain types of neurons to be idle in the presence of low levels of observed input activity, and to become active only after input activity passes a threshold. However, in empirical studies the threshold cannot

---

[6] Another common example is the arctangent function: $s(z) = \frac{1}{\pi}\tan^{-1}z + \frac{1}{2}$.

be properly detected when the problem is highly complex, as when there are millions of switches. Therefore, it is probably a better modelling strategy to bet on smooth activation functions, where the threshold is blurred.

There are three ways in which **(1.10)** generalizes over **(1.8)**. Although model **(1.10)** has also three layers, its second layer is more complex than the second layer in model **(1.8)**. This layer is usually called the hidden layer in the neural network literature. The number of elements in the hidden layer is not fixed to one, but chosen in the optimization. The second way in which model **(1.10)** is more general is that each component of the second layer affects the output independently. The extent to which it does so, $a_m$, must also be estimated. The smooth activation function, $s(\cdot)$, can be chosen from several alternatives. Finally, cross-validation is crucial and the selection of $\mu(\cdot\ )$ will influence the complexity of the estimated model.

The role of $s(\cdot)$ is pivotal in understanding the potential applicability of neural networks. When $s(\cdot)$ is near 1, then the corresponding $a_m$ intensively affects the output. If it is near zero, the output is almost not affected by that element. One can think of the elements of the hidden layers as rules, which will specially apply under particular conditions in the inputs. This feature is known in artificial neural networks as context sensitivity and in regression trees as nonhomogeneity. It is correctly regarded as a feature of the model's flexibility.

When the number of elements in the hidden layer is assumed to be known, consistency results are obtained for backpropagation estimates of model **(1.10)**. See for example Kuan and White (1994) for consistency and asymptotic normality results.

1.1.3 Projection pursuit and multiple hidden layer artificial neural networks

In these two cases, penalty functions $\phi$ take the same general form as in **(1.9)**. In projection pursuit the family of functions $\{h_T\}$ are

$$h_T(\mathbf{x}|\mathbf{a},\gamma) = \sum_{t=1}^{T} s_t(\,\mathbf{x}'\gamma_t \mid \mathbf{a}_t) \qquad \mathbf{(1.12)}$$

so that projection pursuit can be seen as a generalization from single hidden layer artificial neural networks. The functions $s_t(\cdot)$ need not have a sigmoid form and, for example, smoothers on local linear fits can be chosen to estimate their shape.

Multiple hidden layer artificial neural networks are also more general models than single hidden layer models. Four layer neural networks have two hidden layers. The outputs from the first hidden layer are taken as the inputs for the second hidden layer. In practice, $h_T(\cdot)$ will have a more flexible structure than a single hidden layer with the same number of elements in each layer.

Whether adding more hidden layers improves prediction will depend on how well the resulting effect of the functions $h_T(\cdot)$ matches $f(\cdot)$.

### 1.1.4 Tree structures

Tree structured models have penalty functions similar to that in **(1.9)**. The family $h_T(\cdot)$ includes all functions with the generic form

$$h_T(\mathbf{x}|\mathbf{a}) = \sum_{t \in T} a_t \, I\left(\mathbf{x} \in t\right), \qquad \mathbf{(1.13)}$$

where $I(\cdot)$ is the indicator function and $T$ is a partition of the space of all possible values of $\mathbf{x}$, $R^M$.

Usually, the subsets $t \in T$ are restricted to be hyper-rectangles parallel to the coordinate axis, so that

$$I\left(\mathbf{x} \in t\right) = \prod_{j=1}^{M} I(u_j < x_j \le v_j), \qquad \mathbf{(1.14)}$$

where the parameters $\{u_j, v_j\}$ are the respective lower and upper limit of the region on each axis.

Thus, $h_T(\cdot)$ can also be expressed as

$$h_T(\mathbf{x}|\mathbf{x} \in t) = a_t \ . \qquad \mathbf{(1.15)}$$

These models borrow their name from the fact that they can be represented by two-dimensional binary trees. This is not a small advantage in a high dimensional nonparametric context.

Consider, as an example, the following tree structure defined on the set $C^2 \equiv \{[0,x_{i2}],\ x_{i2} \in R,\ i=1,2\}$

$$f(x_1\ ,\ x_2) = \begin{cases} c_1 & \text{if } x_2 \le x_{21} \\ c_2 & \text{if } x_2 > x_{21}\ ,\ x_1 \le x_{11} \\ c_3 & \text{if } x_2 > x_{21}\ ,\ x_1 > x_{11} \end{cases} . \tag{1.16}$$

Figure 1.1 in Appendix A shows a two-dimensional representation of this three-dimensional surface. If the structure had more inputs, it would not be possible to draw this graph.

The same structure can be represented by Figure 1.2 in Appendix A. This figure is a binary tree diagram. Each brand represents a split of the input space. Each node of the tree represents a subregion of the space. The root node represents the entire space. The terminal nodes represent the regions, t, associated with partition T.

Another example of a tree structured model is shown in Figure 1.3 in Appendix A. The dimension of the input space is four. Thus, it is impossible to draw something like Figure 1.1. Nonetheless, all the relevant information that can be obtained from graphs such as Figure 1.1 can also be obtained from tree diagrams when the surfaces are smooth but for a limited number of edges.

The most important assumption in tree models such as **(1.16)** is therefore the existence of sudden changes in the flat surfaces. This is a very bad assumption if the surface changes smoothly. Another problem is the possibility that the surface does not change along parallel lines from the axis. Figure 1.4 represents a structure with these two features. These shortcomings suggest generalizing **(1.15)** to propose

$$h_T(\mathbf{x}|\mathbf{x} \in t) = f_t(\mathbf{x}, \mathbf{a}_t). \qquad \textbf{(1.17)}$$

General models along this line have been implemented. They mitigate the rigidity of the activation function by looking at a very large number of splitting criteria and a large number of surfaces. The cost comes both in terms of the complexity of the efficient algorithms and the interpretation of the results.

### 1.1.5 Universal approximators

In the artificial neural networks literature, the models reviewed in the previous sections are often interpreted as approximations of the true model. A universal approximator is a flexible functional form that can approximate an arbitrary function to a particular level of accuracy. Neural networks, projection pursuit and trees all share the functional form

$$h_T(\mathbf{x}|\mathbf{a},\gamma) = \sum_{t=1}^{T} a_t \cdot b(\mathbf{x},\gamma_t). \qquad \textbf{(1.18)}$$

These models are also universal approximators for the class of all continuous functions in the sense that any arbitrary continuous function $f(\cdot)$ admits the representation

$$f(\mathbf{x}) = \sum_{t=1}^{\infty} a_t \cdot b(\mathbf{x}, \gamma_t) \qquad \textbf{(1.19)}$$

for some set of sequence coefficient values[7] $\{a_m\}^{\infty}$. Therefore, these nonparametric models have similar spanning properties as, for example, polynomials.

In spite of this result, it is reasonable to expect that in small samples selection of the method becomes critical.

**1.2 Estimation of regression trees**

The original problem of estimating **(1.15)** is rather trivial if we know the tree structure. Since the constant c which minimizes $E[(y-c)^2|\mathbf{x}\in t]$ is $E[y|\mathbf{x}\in t]$, then the least squares estimator, LS, is the sample average within each terminal node.

If we do not know the structure of the tree, then LS will not be in general implementable for even not-too-high dimensional problems. The reason is that LS will be a combinatorial -not an analytical- problem in this context. In order to minimize **(1.3)** one should evaluate all possible structures. This is an enormous

[7] See Friedman (1994).

task. Consider LS on a sample with 50 different cells assuming that the structure has at most two terminal nodes. The number of possible models is near $6x10^{14}$. This is a very big number. If we could process every iteration in a millionth of a second, we would obtain LS only after 17 years of uninterrupted computations. Given the state of computer technology other alternatives must be considered.

A second best solution of the problem is recursive partition. The initial region is divided into two regions according to a splitting criterion. Then, recursive partition is carried out on each region. This strategy is implementable because, as partitioning takes place, the corresponding regions -nodes in tree terminology- include smaller and smaller subregions of the original input space. The number of nodes increases at every step, but each node becomes ever more local. An essential feature in the procedure is that splitting of each region is assessed only by studying a limited number of possible splits.

Let us now describe the splitting or tree-growing algorithm. Assume that we have already several nodes and a tree structure T and that we want to split node t*, one terminal node in T.

Define d(**x**) as the tree structure projections from T. The average residual sum of squares for tree T is

$$R(T) = \frac{1}{N} \sum_{i=1}^{N} ( y_i - d(\mathbf{x}_i) )^2, \qquad \textbf{(1.20)}$$

where N is the number of observations in our estimation sample. We can also express R(T) explicitly as a function of the residual sum of squares within each terminal node of the tree. Then, **(1.20)** becomes

$$R(T) = \frac{1}{N} \sum_{t \in T} \sum_{\mathbf{x}_i \in t} ( y_i - d(\mathbf{x}_i))^2 \qquad \textbf{(1.21)}$$

where $d(\mathbf{x}_i)=(1/N_t)\cdot\sum_{\mathbf{x}\in t} y_i$.

Recursive partitioning is defined through recursive optimization. Heuristically, when a variable has a strong contribution to the true tree structure, a split based on that variable will likely improve the fit and thus reduce R(T) greatly. Thus, recursive partitioning can be defined by choosing the split at each step of the algorithm such that the reduction in R(T) is maximized. The split chosen at each step of the algorithm, s*, satisfies

$$\Delta R(s^*,T) = \max_{s \in S} \quad R(T) - R(T_s), \qquad \textbf{(1.22)}$$

where $T_s$ is any tree obtained by splitting a terminal node, t, into a left node, $t_L$, and a right node, $t_R$. Since

$$R(T) - R(T_s) = \frac{1}{N} \sum_{\mathbf{x}_i \in t} ( y_i - d(\mathbf{x}_i))^2 - \frac{1}{N} \sum_{\mathbf{x}_i \in t_L} ( y_i - d(\mathbf{x}_i))^2 - \frac{1}{N} \sum_{\mathbf{x}_i \in t_R} ( y_i - d(\mathbf{x}_i))^2 \qquad \textbf{(1.23)}$$

the algorithm involves at each step the evaluation of the optimal split at each two new terminal nodes from the previous split.

We can keep splitting the estimation sample until there are no nodes with elements with different characteristics or the number of observations reach a lower limit. This lower limit can be fixed by the researcher according to the problem and is called the splitting rule. Splitting ends when we obtain the largest possible tree, $T_{MAX}$. Often, the result will be equivalent to dividing the sample into all possible cells and computing within cell averages, a standard nonparametric analysis.

Growing the tree until no further partitioning is possible helps avoiding having to select a rule to stop splitting. Usually, however, $T_{MAX}$ will be too complex in the sense that some terminal nodes could be aggregated into one terminal node. A more simplified structure will normally lead to more accurate within node estimates since the number of observations in each terminal node grows as aggregation takes place. It is also intuitive to see that if aggregation goes too far, aggregation bias will become a serious problem.

In order to aggregate from $T_{MAX}$ we can use a clustering algorithm procedure[8]. Breiman et alia (1984) propose to compute the

[8] See e.g. Gordon (1993) and Hartigan (1975) for introductions to clustering algorithms. Piccolo (1990) proposes

error-complexity measure $R(\alpha,T)=R(T)+\alpha|T|$ for all possible trees obtained from simplifying the structure by cutting, or pruning, tree $T_{MAX}$. Here, $|T|$ denotes the number of terminal nodes in T, that is, the complexity of the tree, and $\alpha$ is a given parameter. Note that $R(\alpha,T)$ is an error-complexity function by which a model is selected trading variance with complexity.

The tree structured estimate for a given $\alpha$ is the value that minimizes $R(\alpha,T)$ for the set of subtrees of $T_{MAX}$. The resulting tree belongs to a much broader set of trees than the sequence of all trees obtained in the recursive partition algorithm. Heuristically, part of the harm done by recursive partition is reduced. Thus, regression trees are much more powerful pattern recognition tools than ordinary clustering algorithms.

Optimization of the error-complexity function for *all* possible values of $\alpha$ leads to an increasing finite sequence of real values $0=\alpha_1<\alpha_2<...<\alpha_Q$ and a decreasing finite sequence of subtrees $T_1 \succ T_2 \succ ... \succ \{root\}$, such that for any real value $\alpha \in [\alpha_k, \alpha_{k+1})$, $T_k$ is the smallest subtree of $T_{MAX}$ minimizing $R(\alpha,T)$. See Breiman et alia, (1984, p.289) for a proof of this result. Implementing cost-complexity minimization for all $\alpha$ is then possible through a weakest-link algorithm.

---

an alternative use to clustering algorithms in time series modelling. See Scott and Symons (1971) and Bryant and Williamson (1978) for the statistical analysis of these techniques.

Any branch, $T_t$, spanning from a nonterminal node t of a tree T, is cut only if

$$\frac{1}{N}\sum_{\mathbf{x}_i \in t}(y_i - d(\mathbf{x}_i))^2 + \alpha \le \frac{1}{N}\sum_{t' \in T_t}\sum_{\mathbf{x}_i \in t'}(y_i - d(\mathbf{x}_i))^2 + \alpha|T_t|, \quad \textbf{(1.24)}$$

so that the first branch to be cut minimizes

$$\frac{\sum_{\mathbf{x}_i \in t}(y_i - d(\mathbf{x}_i))^2 - \sum_{t' \in T_t}\sum_{\mathbf{x}_i \in t'}(y_i - d(\mathbf{x}_i))^2}{|T_t| - 1}. \quad \textbf{(1.25)}$$

The initial intractable problem is thus reduced to one of selecting an optimum-size tree from a decreasing sequence of subtrees.

At each step of the pruning algorithm R(T) increases so that $R(T_1)$ is the lowest value of the sequence $\{T_1 \succ T_2 \succ \ldots \succ \{root\}\}$. For the learning sample, our estimates $d(\mathbf{x})$ from $T_1$ are therefore least squares estimates among the sequence. This property is satisfied in the estimation sample by definition, but it does not have to do so in an independent sample. Choosing $R(T_1)$ as our fit of the tree structured model may lead to overoptimistic results for $R(\cdot)$ and the model will be overfitted.

There are three strategies to obtain unbiased estimates of $R(\cdot)$. The first one is the use of an independent test sample. This is most appropriate, due to its simplicity, when the data set has many observations. It simply consists of randomly dividing the entire sample into a learning and a test sample. The tree is grown and pruned with the learning sample, while unbiased estimates of R(T), $R^{ts}$, can be obtained with the observations of the test sample and the estimates of the learning sample.

The other two alternatives to obtain unbiased estimates of $R(\cdot)$ are K-fold cross-validation and the bootstrap method. Since they will not be implemented in the following chapters, I refer the reader to Breiman et alia (1984) for an introduction to the application of these methods in regression trees.

It is possible to compute standard errors for $R^{ts}$ from the test sample, $SE(R^{ts})$. $R^{ts}$ may be very flat along the sequence only to increase at the last, coarser subtrees. When this happens it may be difficult to justify the least squares subtree and it is probably better to study several alternatives. Breiman et alia (1984) suggest the 1 SE rule, which consists of selecting the simpler tree whose $R^{ts}$ is not larger than the minimum $R^{ts}$ plus 1 standard error. Using these corrections may greatly reduce the number of terminal nodes of the tree. In the regression trees literature, sometimes this is referred to as the goal of obtaining parsimonious models. In a more general context parsimony and complexity are, however, different concepts. For example, in a simple linear model with one continuous independent variable, complexity is infinity, whilst we can still talk of a simple parsimonious model. In parametric structures, the goal of parsimony can be obtained not only by reducing the number of projected values for the dependent variable -complexity-, but also through the estimation of simple relations between the dependent variable and the independent variables. In the following, I will nonetheless use the concepts of parsimony and complexity as interchangeable.

The small sample statistical properties of the estimator just described are not known. This problem is not trivial because the technique involves partitions of the input space based on the learning sample. Thus, the estimates are the results of random partitions.

Nonetheless it is possible to know something about the behavior of the recursive estimates as the sample becomes larger and larger. The fundamental consistency conditions for random partitions are surprisingly general. All we need is an ever more dense sample at all n-dimensional balls of the input space in order to approximate in a q-square sense the nonparametric surface. If the partition guarantees this, then the estimates should converge to the true function. Cost-complexity minimization together with test sample unbiased estimates of $R(\cdot)$ guarantee that such condition is satisfied by regression tree partitions. The basic results can be found in Breiman et alia (1984, chapter 12).

A word of caution is nonetheless necessary. For small samples, high correlation in the explanatory variables will induce instability in the tree topology so that slight changes in the learning sample may cause splits to be made on different variables. In this case, the interpretation of the contribution of each variable will become problematic.

### 1.3 Regression trees in economics

Although regression trees has been used in several scientific fields such as medical diagnosis, automatic identification of chemical spectra and pollution level predictors in urban areas, its implementation in economics has been to date rather small.

Nonetheless, we have seen some examples of implementation of the technique in economics during the 90s. Here I present a brief summary of three of them.

They highlight in my view its potential applicability to study economic issues. In the following section I will argue that the study of wages is a field where this technique can be interesting to apply.

#### 1.3.1 A classification algorithm

Cotterman and Perachi (1992) describe a method for deciding how to aggregate a set of elementary U.S. industries. The method is based on regression trees methodology and it is an alternative procedure to standard clustering algorithms that allows for a much broader set of aggregation alternatives.

The fundamental difference with respect to standard regression trees is that their methodology simplifies the first algorithm, growing-the-tree, into a simple clustering algorithm. The second step corresponds to pruning-the-tree. Finally, since they are only interested in reporting alternative levels of industry aggregation, they do not implement cross-validation, but present the differences between their aggregation techniques and the 15-industry level proposed be the U.S. Bureau of the Census.

In order to grow the tree, they propose two measures of closeness between the elementary industries. First, using data from the Current Population Survey prepared by the Bureau of the Census, CPS, and a matching algorithm, it is possible to have for some individuals two independent codings of the industry of employment. For some workers, however, these codings do not coincide. Further, they seem to affect some pairs of industries more than others. The authors attribute them to three potential causes: data collecting errors, errors in the matching algorithm or, finally, ambiguity in the definition of the elementary industries. They assume that the last cause is the relevant one and propose mismatch rates between industry codings as measures of similarity. The second proposed measure of industry closeness is the workers' transitions between industries. The measure is reasonable when individuals move more frequently between similar industries. A pooled sample from 1977 to 1982 was used in the algorithms. Both measures led to different results for the hierarchical tree. It is interesting that the authors depart from common practice in clustering when they propose these measures. The usual procedure is to define closeness through a distance based on a vector of characteristics. They tried to overcome the arbitrary step of defining this vector.

Optimal pruning using log-wage data for the years 1971-1982 again from the CPS was carried out. Residual sum of squares were obtained from regressions of weekly wages on years of schooling, years of experience and its square. The authors find important differences in the various aggregation schemes, and conclude[9] that since "(...) the outcome of much applied work may hinge on the aggregates employed", then " (...) procedures for classification and aggregation are legitimate and important subjects of inquiry".

### 1.3.2 Melon prices

Russel Tronstad (1995) applies regression trees to estimate discounts and premiums due to various characteristics of wholesale melons. Characteristics considered are melon type, size, grade, shipping container, week, and year.

Melons are highly perishable products, so that supply of melons

[9] Cotterman and Perachi (1992, p.50).

can be assumed to be perfectly inelastic. After correcting from season, differences in prices between different types of melons thus show differences in demands that are due to differences in the characteristics, not in relative supplies.

The model was estimated with weekly price data from 3 January 1990 through 28 December 1993. Twelve different melon types were considered. The data source was the *Los Angeles Wholesale Fruit and Vegetable Report*, published by the U.S. Department of Agriculture.

The results compared favorably to standard parametric regression in the sense of a higher coefficient of determination. The author also considers that regression trees performed better since it allowed for interaction between discrete variables. Allowing for these interactions on the OLS regression would have required a very large number of dummy variables. Efficient estimation would have then demanded the implementation of some model selection algorithm.

### 1.3.3 Multiple growth regimes

Durlauf and Johnson (1995) use regression trees to identify national economies with different laws of growth.

They argue that a cross-section linear regression applied to growth data generated by economies converging to multiple steady states can produce a negative initial income coefficient. Thus, a negative sign in this coefficient cannot be taken as evidence of convergence in income per capita for all countries.

The data source is Summers and Heston (1988) and the World Bank's *World Tables* and *World Development Report*. The authors initially carry out ad hoc splits of the countries into two, three and four groups based on their initial per capita output and literacy rates. They then test for the existence of a common growth path for these groups. They reject the null hypothesis against the alternative of multiple regimes in a human capital specification and in an augmented version that tries to incorporate social and political factors. In both cases, they reject the hypothesis of a single growth regime against the hypothesis of several regimes.

Regression trees allows for endogenously finding the number and specification of growth regimes. The splitting criteria in the tree are based on initial literacy rates and income per capita. This is consistent with the multiple regime framework since if economies are concentrated around several steady states, then their initial values for these variables will cluster for each group. Within nodes sum of squares are computed from the residuals of the growth equations.

The algorithm partitions the world economy into four groups and the estimates are consistent with the view that different economies have access to different aggregate technologies.

In the following section I will consider the use of regression trees in the study of the wage structure. I will present empirical results that highlight the strong context-sensitivity feature that wages present in the U.S.. I will then give an outline of the rest of this volume and will end with a description of the data and programming to be used in the following empirical applications.

## 1.4 Nonparametric Wage Structures

### 1.4.1 Wage structures

The concept of the wage structure is fundamental in the empirical analysis of the characteristics of the wage distribution. It refers to the vector of prices set for various labor market skills and the rents received for employment in particular sectors of the economy.

The labor market is seen as a complex structure that consists of interrelated local markets with different market equilibrium wages. A description of this structure and its evolution is of clear interest to study problems as varied as the effects of technological change, the sources of wage inequality, and discrimination in the labor market.

We can start by simply assuming that the logarithm of the market-clearing wage for any worker, $w_i$, depends on observed and unobserved characteristics that position the worker in a segment of the labor market.

It is customary to assume a linear relation between the observed and unobserved effects, where the unobserved effect term will have zero expected value and small variance $\sigma^2$. The general specification for the econometric model in this literature is then simply

$$w_i = f(\mathbf{x}_i) + v_i \quad . \tag{1.26}$$

The simplest and commonest specification for **(1.26)** is a polynomial parametric relation between the explanatory variables. For example, a squared term for experience on top of a linear model is usually included since the publication of the seminal work of Mincer (1974)[10].

The linear parametric approach implies that each variable's additional contribution to wages is constant or follows a

---

[10] It is customary to refer to a wage equation that is linear on the education level, experience, and experience$^2$ as a Mincer equation. The conection between human capital models and this simple specification was one of the main contributions of Mincer. This is nowadays sometimes recognized by referring to Mincer equations just as human capital specifications.

be solved by adding just a few quadratic terms in the right-hand side of the equation? My intuition is that it cannot. By postulating a rigid structure, the researcher may distort the information available in the data set so that the results will not be useful.

In the next section, I will review the empirical evidence on context sensitivity in the U.S. wage structure.

### 1.4.2 A survey on context sensitivity in the U.S. wage structure

There is a remarkable consensus in the economic literature on the recent general trends of relative wages. It seems that the fundamental dynamic features of wages are sufficiently well described along just three or four dimensions.

First, there was a general trend of wage dispersion and a slow down of growth in real wages during the eighties. Second, there have been increases in the wage differentials between workers with college and high school education for all demographic groups defined by gender and age. Experience differentials have continued a long-term increasing trend. On the other hand, gender differentials narrowed further while race differentials

---

reported that the parameters of these terms were not significantly different from zero when wages instead of earnings were used in the regression.

remained stable in the last twenty years. Industry differentials have also remained stable.

Compared with the trends observed in the sixties and seventies, the growing inequality in the eighties was not a unique phenomenon. Wage dispersion within groups and increases in experience differentials were not new trends. However, the reductions in the College premium during the 1970's and the narrowing of the race differentials before 1975 did not occur later[13].

On top of these general stylized facts we can find many reported local cases of context sensitivity or nonhomogeneity in wage differentials. Probably the best known of these is the different behavior of the experience differential between high school and college graduates. It seems that the combination of college attendance and job experience was an unbeatable one during the

---

[13] See, for example, Levy and Murname (1992) and Buschinski (1994). Allen (1995) analyzes changes in the wage structure across manufacturing over the years 1890-1990. He concludes that interindustry wage differentials were highly stable over the entire period for production workers. Interindustry wage differentials were stable for all workers from 1958 onwards.

eighties[14]. Welch (1979) argued that college graduates are more imperfect substitutes for more experienced graduates than is the case for workers with less education.

Similar asymmetries can be found along other demographic dimensions. Let us consider the interaction between education and race. While the college-going rate for 18-24 year old, white, high-school graduates increased from 31.2% to 38.1% between 1979 and 1987, the college going rate for black high school graduates in this age group fell from 29.5% to 28.1%, suggesting that the college premium has a race story inside[15].

The race differential for women almost disappeared in the 1970s, while it remained stable for male[16]. Thus, either the wage structure is becoming nonhomogeneous with respect to the race differential, or it never was.

Some authors have studied the relationships between sector of employment and race. Greene and Rogers (1994), for example, find important differences between the private and public sectors with respect to earnings of college-educated black and white

---

[14] See, amongst others, Bound and Johnson (1992), Katz and Murphy (1992), and Murphy and Welch (1992).

[15] See Levy and Murnane (1992) and also Ashraf (1995).

[16] See, for example, Murphy and Welch (1992) and Blau and Beller (1992).

professionals.

Bound and Holzer (1993) estimate the effects of industrial shifts in the 1970s on the wages and employment of black and white males and find that while the magnitudes of these effects are fairly small for many groups, they can account for about 40-50 percent of the employment decline for less-educated young blacks.

Firm and industry effects have also been compared on several occasions. Davis and Haltiwanger (1991) find that steady growth in wage differentials among plants in the manufacturing sector between 1975 and 1986 accounts for half of the growth in wage dispersion within this sector[17].

Bound and Johnson (1992) find that when 45 instead of 17 industries were used for "all men" and "all women" groups, most

---

[17] It is unclear whether this pattern extends to other industries, particularly after considering that international competition may have increased the pressure on firms to choose between quality improvements or cost reductions in the labor force. Levy and Murnane (1992) argue that the eighties may just have been a period of adjustment for the manufacturing sector that will end when those firms which chose the losing strategy disappear.

of the industry effects were picked up by the dummy variables for 17 industries. The interesting exceptions were non-college men during the 1980's, for whom the use of detailed industry dummies increased the total industry wage effects by up to one-half.

Blackburn (1990) shows that approximately 15 percent of the increase in within-group variation for men stems from the movement of workers from goods producing industries to services.

Other variables of potential interest have also been studied and their interactions locally analyzed. A few examples follow.

There is no clear consensus on regional variations and their effects on the wage structure. Eberts and Schweitzer (1994) find that the trend in regional variation can be traced to declining differences in labor market valuations of worker attributes rather than to shifts in the regional composition of the workforce. See also Gyimah and Fichtenbaum (1994) for an investigation on the regional differences in labor market gender and race discrimination. Larger differences do not imply larger discrimination.

Immigrants with lower initial wages were assimilated in the U.S. market faster than those with higher initial wages (LaLonde and Topel, 1991).

Social status has also been considered: Among males, growth in the proportion of males in the labor force who are unmarried has affected the married status differential[18].

Finally, Adamson (1993) study differences in union affiliation relative wages across gender and race and finds that the female union effect declined over the 1970-1982 period whilst the size of the male union effect remained stable.

The list of cases is not at all exhaustive. The examples nevertheless transmit the message that local analysis unveils context sensitivity. This is done by focusing on the interactions of at most three variables. Note that in order to observe context sensitivity using parametric techniques we must include at least quadratic effects in the set of explanatory variables or, more generally, estimate different wage equations in different segments of the labor market. This, in effect, eliminates the possibility of a global context-sensitive parametric approach.

From the extensive literature on wage premiums we must conclude that any global analysis of wages may suffer from aggregation bias and that even local studies should take account of nonhomogeneous features in the structure.

---

[18] See Blackburn (1990) and Blackburn and Korenman (1994).

It seems therefore desirable to keep the empirical analysis to the highest level of flexibility. A nonparametric approach over all possible points of the input space cannot be used for practical reasons. A simple example will fix ideas. Bound and Johnson (1992) did not try a very detailed specification of the input space. They took three periods (1973-1974, 1979, and 1988) and for each of them the population was divided into 32 subsamples -according to education, potential labor market experience and gender- on which wage regressions were carried out with dummies for the following characteristics: Educational attainment, nonwhite, part-time employment, residence in an SMSA, four major regions and employment in 17 major industries. The most complex nonparametric surface would involve 15,000 different labor groups. In order to have at least 50 observations for each type of worker, the research can only be carried out with samples of at least 750,000 observations.

Thus, it seems that ad hoc searches for the best functional local parametric specification is the only available strategy. It is not. Parsimonious nonparametric econometric models such as regression trees allow for simple nonparametric structures in the sense of a low number of different expected equilibrium wages[19]. They thus provide a very useful tool in the study of local segments of the labor market.

---

[19] This is the result when the splitting rule constraint is activated in the splitting algorithm.

## 1.5 Outline of the thesis

### 1.5.1 Outline of the thesis

The fundamental econometric problem that model **(2.1)** may have is related to the statistical relation between the observed variables and a subset of the unobserved variables. The general idea is that if economic agents take decisions based on an opportunity set not observable to the researcher, and if opportunities vary across agents, then observable data will be censored and the error term may not be independent of one of the regressors.

In the context of measuring the returns to education, this self-selection/omitted variables problem would imply that OLS estimates of the education coefficient may be upward biased. Intuitively, individuals with higher ability, an unobservable variable, will normally choose higher levels of schooling because they can benefit most from it (e.g. Griliches (1977)). The interpretation of the importance of the education variable in the estimated tree will also present the same problem.

In the context of gender differentials, participation decisions in the labor market censors especially female data. Again, LS estimates will be biased if the decision to participate is affected by some observable factors in the wage equation **(2.1)** or an ability dimension related to observable variables, such as education. The same problems affect again the interpretation of the gender estimate in regression trees.

It is possible to define regression trees so that these problems can be addressed. The main idea is to use standard parametric regression techniques designed to cope with these problems at each node instead of simple within node averages in order to estimate the impurity at each node. This is, no doubt, a very interesting direction to enlarge this study.

In the rest of the thesis, I will present, however, the results of applying the simplest regression trees algorithms to samples of populations for whom, I will argue, the econometric problems just mentioned are minimized.

Three applications of regression trees on the study of the wage structure are implemented in the following three chapters of this dissertation:

a.- I will first estimate experience-wage profiles for white male full-time employed workers.

b.- Secondly, I will decompose average wage differentials of different groups using nonparametric structures estimated by regression trees.

c.- Finally, I will look at trends and inequality using nonparametric wage structures estimated by regression trees.

Obviously, the choice of the subjects probably reflects mostly the interests of the author and I would not like to suggest that these fields are the most promising for the application of nonparametric multivariate techniques to the study of wage structures.

Before turning to the results of the empirical applications, I would like to comment on the data set and the programming language used.

### 1.5.2 Data and programming

In the following chapters I will use the outgoing rotation groups of the Current Population Survey (CPS). The Current Population Survey is a monthly survey of now about 60,000 households prepared by the Bureau of Labor Statistics, BLS. An adult (the reference person) at each household is asked to report on the activities of all other persons in the household. There is a record in the file for each adult person. The universe is the adult noninstitutional population.

Each household entering the CPS is administered four monthly interviews, then ignored for eight months, then interviewed again for four more months. If the occupants of a dwelling unit move, they are not followed, rather the new occupants of the unit are interviewed. Since 1979 only households in months fourth and eighth have been asked their usual weekly earnings/usual weekly hours of work. These are the outgoing rotation groups, and each year the BLS gathers all these interviews together into a single Merged Outgoing Rotation Group. A consequence of this construction is that an individual appears only once in any file year, but may reappear in the following year. The National Bureau of Economic Research, NBER, has prepared a CD-ROM with extracts of the files.

These data have, however, some limitations as a data set for studying the evolution of wages across different groups[20]. I will comment on four problems that may distort results.

First, the definition of income does not include fringe benefits, which have constituted a rising proportion of income compensation. Levy and Murnane (1992) argue that after adjusting for fringe benefits the difference between the rates of growth of real wages for the sixties and eighties diminishes, but the eighties value is still well below the pre-1973 period.

Second, to preserve confidentiality in the upper tail of the income distribution, the statistics reported are top-coded at $50,000 from 1968-1981, $75,000 from 1982 to 1984 and $99,000

[20] For a more detailed discussion, see Levy and Murnane (1992).

from 1985. This problem can actually be lessened by avoiding all together the tails of the distribution although more sophisticated procedures have been proposed[21].

The fact that the CPS lacks information on firm specific activities can be important if workers have heterogeneous characteristics and production potential across firms.

Finally, the CPS data may overstate the rate of growth during the 1980's in the proportion of new labor market entrants who were college educated and underestimate the earnings of workers who completed a normal high school programme. Bishop (1991) finds an unprecedented mismatch in the 1980's between the CPS data on new entrants who had accomplished college education and the number of degrees awarded in the U.S.. With respect to high school graduates, before 1988 the CPS treated both holders of the General Educational Development exam and traditional high school graduates as having completed 12 years of schooling. However, Cameron and Heckman (1991) find, with data from the National Longitudinal Study of Youth data set, that the first group earnings patterns are indistinguishable from high school dropouts.

---

[21] Truncation corrections for top coding normally assume a gamma distribution for the upper tail of the yearly income distribution.

I implemented regression trees algorithms on large data sets of wages. The computations were carried out using the author's procedures programmed in GAUSS for regression trees on ordered variables. An interesting feature of the chosen programming language was the possibility of using a simple matrix programming language together with large data sets. In particular, the procedures were able to process sets with more than 60,000 observations with speed in a personal computer and limited memory. Available commercial software would not do the job. The procedures are available upon request to the author.

Appendix A

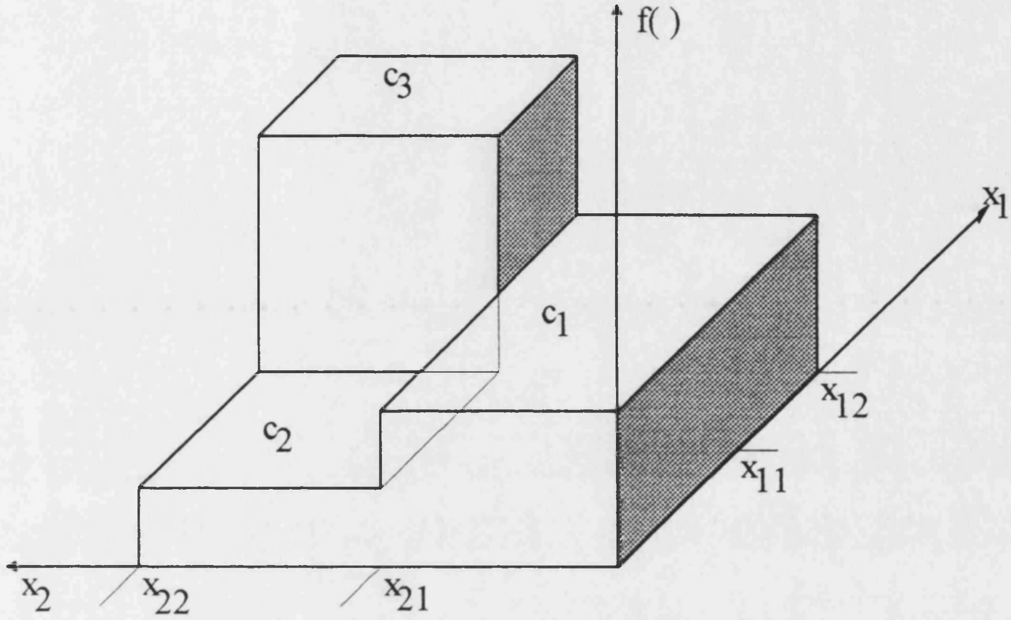


**Figure 1.1** 2D representation of structure

$$f(x_1\ ,\ x_2) = \begin{cases} c_1 & \text{if } x_2 \le x_{21} \\ c_2 & \text{if } x_2 > x_{21}\ ,\ x_1 \le x_{11} \\ c_3 & \text{if } x_2 > x_{21}\ ,\ x_1 > x_{11} \end{cases}$$

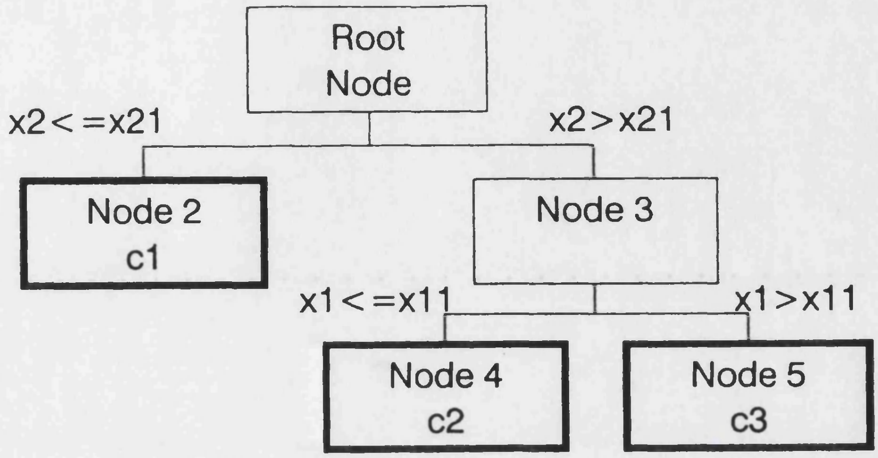


**Figure 1.2** Binary tree representation of structure

$$f(x_1, x_2) = \begin{cases} c_1 & \text{if } x_2 \le x_{21} \\ c_2 & \text{if } x_2 > x_{21}, x_1 \le x_{11} \\ c_3 & \text{if } x_2 > x_{21}, x_1 > x_{11} \end{cases}$$

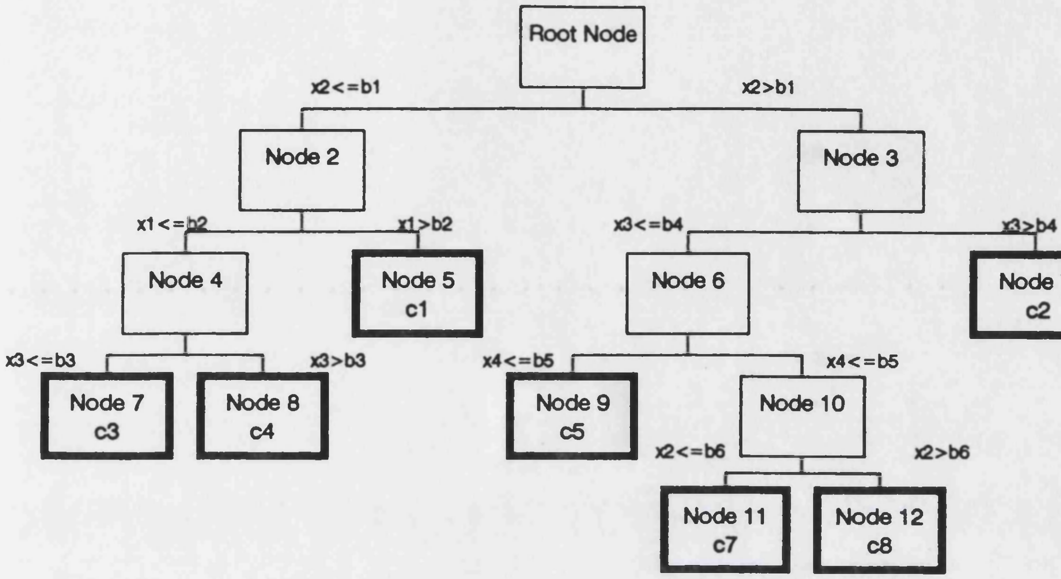


**Figure 1.3** Binary tree representation of a 4-dimensional tree structure

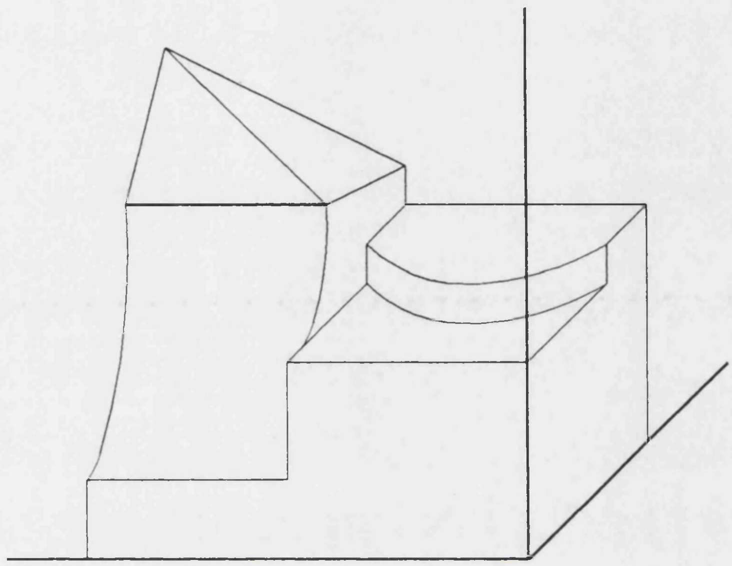

**Figure 1.4** 2D representation of a general tree structure

Chapter 2

# Tree estimation of experience-wage profiles

## 2.1 Introduction

The theory of human capital in its general form has no functional specification. The observed inverted-u shape in the experience- wage profile[1] is explained as the result of two effects: (a) increases in the wage as the worker gains experience and spends less time in training, and (b) decreases in wages due to human capital losses or depreciation. For convenience, the depreciation rate is often taken to be constant. However, if it is not, multiple human capital accumulation paths can occur in a cross-section of workers.

[1] See Figure 2.1 in appendix A.

While the general human capital model predicts wages that eventually decline, alternative theories of wage growth over the life cycle do not necessarily predict declining wages at older ages. So, for example, the shirking model which views wage growth as a worker discipline device, can have rising wages if the future disutility of effort is not too high (See Johnson and Neumark, 1996). Declines in wages for older workers are then related to negative shifts in the labour demand for these workers.

In this chapter I will assume that sudden losses of human capital can affect individuals at the late stages of their working lives. This may induce different human capital accumulation regimes in a cross-section that will result in structural breaks in the observed experience-wage profiles. Although I use the human capital explanation to fix ideas, it is interesting to note that the empirical analysis could also be carried out trying to isolate sudden drops in observed wages at the end of the working life for some local types of workers and study whether these falls are demand or supply driven. Alternatively we can see this study as a way of characterizing local types of workers that suffered sudden losses in their wages at the end of their working lives.

A parametric fit of profiles with late structural breaks will wrongly exploit losses of human capital for the most experienced workers as information on the concavity of the profile. For example, if we fit a smooth quadratic function, a sudden loss will be understood not as a step movement but as a smooth decline and the fit may still be reasonably good[2].

Moreover, parametric regressions will smooth differentials in the experience-wage profiles among workers with different characteristics. Thus, local losses or breaks in the experience-wage profiles will pass unnoticed to the researcher.

In section 4 of this chapter I present the results from estimating a multivariate nonparametric human capital surface for 1980, 1985, 1990, and 1995 with regression trees. This econometric model allows for local breaks of the experience-wage profiles in a multivariate context. We can therefore uncover sudden losses of wages for old workers by defining local smoothness applying a kernel estimator on the projections obtained from the nonparametric regression.

This exercise will allow me to answer the question: Are there any breaks in the experience-wage profiles? If there are, can they explain the declining average tendency of wages at the end of the working life as in Figure 2.1 in Appendix A?

---

[2] Murphy and Welch (1990) argue that "two-thirds of the late career decline implied by the quadratic is an artifact of specification" (Murphy and Welch, 1990, page 204).

## 2.2 The theoretical framework

Each worker works a fixed amount of time and has a certain amount of productive capacity. We can call this potential earnings or human capital. Human capital is not transferable and it can be augmented by learning at school, at college, as an apprentice, and on the job. Let us denote human capital by k and its time derivative by $\dot{k}$.

I will assume that on-the-job learning is not firm-specific and is provided by firms freely as a by-product of the production process. They do so because they cannot reap the returns of learning from the worker. If they tried, the worker would move to another firm at a wage reflecting the full value of the human capital embodied in her ( Becker, 1993, pages 30-40).

Training is however not costless for the worker since firms will implicitly charge her with its cost. We can assume, for example, that firms offer jobs with different learning and production intensities. Take $0 \leq x \leq 1$ as the proportion of training of a job. If the loss of output for the firm due to training is associated solely with the shift of the worker's own time from work to training, then we can think of x as the proportion of time spent on training and (1-x) as the proportion of time spent on production. Each worker's contribution to the production process will depend on $(1-x) \cdot k$.

Competitive firms will offer to the worker a contract specifying the wage and the training content. Profit maximization implies that wages will equal marginal revenues for the firms. If r is the rental rate of human capital and if workers are perfect substitutes in the production process, r must equal the marginal productivity of total human capital. Then, wages for workers with k human capital in a job with x learning intensity are equal to $w=r(1-x)k$. The costs of training are rxk.

Working under this contract will produce new human capital to the worker as a by-product. We can denote this by $\dot{k} = f(k,x)$. The problem the worker faces is dynamic. Training today raises tomorrow's wage through an increase in human capital, but decreases today's wages.

This problem[3] can be formally stated as

$$\begin{aligned}&\max_{\{x\}}\int_0^T rk(1-x)e^{-it}\,dt\\&\text{s.t.}\\&\dot{k} = f(k,x), \quad k(0)=k_0,\\&0\le x\le 1\end{aligned} \qquad \textbf{(2.1)}$$

where T is the worker's fixed end of the working life and i is the interest rate in the capital market.

---

[3] See Ben-Porath (1967) for a similar formulation and an introduction to the early literature. See Weiss (1986) for a survey on functional specifications.

This is a standard control problem and can be solved by maximizing full earnings of the worker, that is, his current earnings and the value of the additional human capital. This is effectively done by the Maximum Principle of Pontryagin. Different human capital production functions, $f(\cdot)$, will render different optimal paths for $x(t)$, $k(t)$, and $w(t)$.

Since $\dot{k}$ is net increase of human capital, $f(\cdot)$ must take account of the fact that human capital can also depreciate. Some examples are of interest. General technological improvements will make part of all workers' human capital obsolete. It is reasonable to assume that this type of destruction of human capital will be proportional to the amount of human capital the worker has. The proportion will also be equal among workers, reflecting the generality of the process. This case justifies the assumption of a constant rate of depreciation among workers[4]. Workers suffering certain progressive illnesses will

---

[4] Growing old is a source of physical deterioration that may have a similar property. Fair (1994) studies the rate at which people phisically deteriorate using data on race and field records. The results show that the depreciation rate remains fairly constant until athletes are in their fifties. Mincer (1974, page 22) already suggests that a very low rate of depreciation during most of the working life, beginning to rise only at the end, could be assumed on health and psychological grounds.

have increasing destruction of their productive capacity. Here, the best model is a positive time trend for the human capital depreciation rate. Finally, specialists working in a technologically changing sector of the economy may suffer sudden losses of human capital with a positive probability. Here, human capital depreciation is probably best modelled as a random process.

Therefore, as Mincer (1974, page 20) argues, "[...] the finiteness of life, the increasing incidence of illness at older ages, and the secular progress of knowledge, which makes older education and skill vintages obsolescent, are compelling facts suggesting that as age advances, effects of depreciation eventually begin to outstrip gross investment".

In the next subsections, I will review several specifications for $f(k,x)$, commenting on the properties of the optimal paths. I will stress the effects of the assumption of a constant rate of depreciation of human capital on the evolution of optimal human capital accumulation and wages.

There are of course many other specifications available in the literature. The purpose here is to show the effects of the assumptions on the depreciation of human capital. I begin with a simplified version of the benchmark model developed in Ben-Porath.

### 2.2.1 Ben-Porath (1967)

We assume that $f(k,x)=(kx)^{1-a}-\delta k$. This specification incorporates a constant rate of depreciation that is not necessarily different from zero. A positive constant rate of depreciation means that all individuals suffer proportional losses of human capital at all ages. One can think of this as the effect of general technological progress. When gross investment in human capital is small enough, this depreciation can be the driving force behind falls in potential earnings. A priori, we can expect this to happen at the end of the working life, when it simply does not pay to fight against the loss of knowledge by sacrificing current earnings.

The Hamiltonian of the control problem **(2.1)** takes on the form

$$H = rk(1-x)e^{-it} + v\Big((kx)^{1-a} - \delta k\Big), \qquad \textbf{(2.2)}$$

where v is the discounted marginal value of an additional unit of human capital.

Necessary conditions for the optimal control path for this problem are

$$\begin{aligned} &H_k = -\dot{v} \\ &H_v = 0 \\ &k(0) = k \\ &v(T) = 0 \\ &H(k^*,v^*,x^*,t) = H(k^*,v^*,x,t), \quad \forall t \in [0,T] \end{aligned} \qquad \textbf{(2.3)}$$

Let us define current marginal values as $u=v \cdot e^{it}$. From now on, I omit for simplicity asterisks in the optimal paths. From **(2.3)** we have three groups of necessary conditions:

*Dynamic conditions*

$$\dot{u} = \left( (i + \delta) - \frac{(1-a)(kx)^{1-a}}{k} \right) u - r(1-x),$$
$$\dot{k} = (kx)^{1-a} - \delta k. \qquad \textbf{(2.4)}$$

*Initial and terminal conditions*

$$k(0) = k_0$$
$$u(T) = 0 \qquad \textbf{(2.5)}$$

*Static conditions*

$$\text{If } u(1-a)k^{1-a} > rk \rightarrow x = 1$$
$$\text{If } u(1-a)\frac{(kx)^{1-a}}{x} = rk \rightarrow 0 < x \leq 1 \qquad \textbf{(2.6)}$$

Conditions **(2.6)** show that the worker will either devote all time to learning, schooling, or will learn on the job, but at no point in the worker's working life will on-the-job learning cease[5].

---

[5] This result is contingent on the functional specification. When the marginal productivity of training on production of human capital is linear, there must be a period at the end of the working life in which no training takes place. In the absence of depreciation this will lead to a flat experience-wage profile in the last stages of the working life.

Terminal condition $u(T)=0$ secures no schooling at the end of the worker's life. Dynamics during the schooling period are

$$\frac{\dot{u}}{u} = (i + \delta) - \frac{(1-a)}{k^{a}},$$
$$\frac{\dot{k}}{k} = \frac{1}{k^{a}} - \delta, \tag{2.7}$$

with $k(0)=k_0$. The terminal condition for u, $u(t_s)=u^s$, can be fully characterized once the solution for the optimal path in the following phase, on-the-job learning, has been obtained.

We can solve this system by a change of variable. Note that human capital growth depends on its value only. Taking $z=1/k^a$ we get the separable differential equation

$$\frac{dz}{(\delta - z)z} = a dt \tag{2.8}$$

and solving back for $k^a$ we obtain the optimal path for k(t),

$$k(t)^{a} = \frac{1}{\delta}\left(1 - (1 - \delta k_0^{a})\, e^{-a\delta t}\right). \tag{2.9}$$

Taking into account the differential equation for u -the value of human capital investment- **(2.9)** implies that u will decrease during this period. As the worker ages, investment in human capital becomes less valuable as future life becomes shorter.

---

If there is depreciation, the plateau happens in the middle of the working life. See, for example, Sheshinski (1968).

As u decreases and k increases, the system moves closer to the interior solution and on-the-job learning.

From the terminal condition on u(T) and the static conditions we see that there must always be a phase of on-the-job learning. The existence of schooling will depend on the life span of the worker. The shorter the working life is, the lower the discounted marginal value of investing in human capital, and the more likely it is that it is optimal for the worker to start working at t=0.

I will now study the interior solution for x. The static condition for the internal solution is

$$(1 - a)\ u\ (kx)^{1-a} = rkx, \quad 0 < x \leq 1. \qquad \textbf{(2.10)}$$

Consider for simplicity the case when a=1/2. If we substitute **(2.10)** into the differential equation for k in **(2.4)** we obtain:

$$\frac{\dot{k}}{k} = \frac{u}{2rk} - \delta. \qquad \textbf{(2.11)}$$

If the marginal value of training is positive, u>0, then from **(2.10)** we see that there will be some amount of training, x>0, and the rate of growth of the human capital stock will tend to be positive. However, when the training effort is so small that the first term of the right-hand side of **(2.11)** is smaller than δ, then the stock of human capital will fall. It will be shortly shown that u also follows a declining path during the working phase. Thus, human capital will fall at the late stages of the worker's life, when u/2rk is smaller than the constant rate of depreciation.

The constant rate of depreciation is therefore the driving force behind the decline of potential and observed wages.

If no depreciation affected human capital, average wages would increase always. Then, since human capital investment has always a positive although decreasing marginal value, potential earnings will always increase, and the training effort will gradually be reduced.

Let us now study the system for $0<a<1$. If we substitute the dynamic condition for human capital into the static condition, we get

$$\frac{\dot{k}}{k} = (1-a)^{-1}\left(\frac{rx}{u}\right) - \delta. \qquad \textbf{(2.12)}$$

On the other hand, from the two dynamic conditions we have

$$\frac{\dot{u}}{u} = \left((i+\delta) - (1-a)\left(\frac{\dot{k}}{k} + \delta\right)\right)u - r(1-x) \qquad \textbf{(2.13)}$$

so that by substituting **(2.12)** into **(2.13)** and rearranging we obtain that the marginal value of investment depends only on time:

$$\dot{u} = (i + \delta)u - r, \quad u(T) = 0 \tag{2.14}$$

so that

$$u(t) = \left(\frac{r}{(i+\delta)}\right)\left(1 - e^{-(i+\delta)(T-t)}\right) \tag{2.15}$$

which is a decreasing function of time[6]. From this and the static condition we have that gross human capital investment always increases:

$$(kx)^a = \left(\frac{1-a}{r}\right) u \tag{2.16}$$

and human capital will decrease when

$$\left(\frac{(1-a)u}{r}\right)^{\frac{a}{1-a}} < \delta k. \tag{2.17}$$

A simple analytical solution for the optimal path of k can be obtained if we further assume that a=1/2.

The differential equation for k takes then the form:

$$\dot{k} = \frac{u}{2r} - \delta k, \qquad k(t_s) = k^s. \tag{2.18}$$

---

[6] This is, again, a feature of the Ben-Porath model that makes solving the model a simple task. We will see a similar result once we introduce a random shock.

The solution for k is

$$k(t) = k^s e^{\delta(t_s-t)} + A(1 - e^{\delta(t_s-t)}) + B\ (e^{(i+\delta)t} - e^{\delta(t_s-t)}) \qquad \textbf{(2.19)}$$

where $A=1/(2(i+\delta)\delta)$ and $B=-(e^{-(i+\delta)T})/(2(i+\delta)(i+2\delta)$.

Since $w=r\cdot k\cdot(1-x)$ and $(kx)=(u/2r)^2$, we have that wages increase only if human capital increases since the marginal value of human capital investment is always decreasing:

$$\dot{w} = r\dot{k} - r\,\frac{d}{dt}(kx)\,. \qquad \textbf{(2.20)}$$

Obviously, wages will peak later than potential wages due to the cost of training.

To sum up, when the depreciation rate is positive, this model generates concavity of the experience-wage profile during the working period with increasing wages at the beginning and decreasing wages at the end[7].

If there is no depreciation, this model will generate concavity of the experience-wage profile during the working period but cannot account for the observed decline in wages during the last years of the working life.

---

[7] As already stated, a linear model for the effect on human capital acumulation of training with depreciation, as in Sheshinski (1968), will generate concave experience-wage profiles with a plateau in the middle of the working life.

There may be, however, an alternative explanation for the reduction in average wages in a cross-section. Consider the effect of an unexpected drop in the stock of human capital for an old worker.

To use the results of the model, we simply have to consider what happens when T is very small and there is a sudden loss in $k_0$. If the loss is sufficiently big, there will be not enough time left in the working life to recover from the initial loss, and the wage will be lower that at t=0. Thus, we can see declining wages for workers who suffer unexpected losses of human capital. This argument is clearly not a satisfactory explanation if all wages tended to decline at the last stages of the working life in a cross-section, since if there was a tendency for losses of human capital then, they would not be unexpected.

It may be however a good explanation for part of the decline of average wages at the last stages of the working life. The argument would be that although not everyone suffers depreciation of human capital, those who do, do in such a scale, that push average wages down because all incentives for further investment have been exhausted.

If we introduce in the model different types of workers and observe that different old workers are affected in different periods by losses of human capital, we may compute the effect of these losses on average wages[8]. It is however then more realistic to assume that this uncertainty in the losses is nevertheless taken into account by the workers in their plans for human capital accumulation. I will follow this line of modelling at the end of this section.

### 2.2.2 Uncertainty and human capital

The model just reviewed includes the two fundamental aspects of the human capital explanation to the inverted-u shape of average wages. First, finiteness of life is crucial since human capital is embodied in each person and cannot be transferred. This becomes an irresistible factor to decrease investment effort along time since there is a trade off between investment and current earnings. For the same reason, on-the-job learning will tend to take place at the beginning of the working life. This accounts for concavity.

Secondly, losses of human capital will induce decreases in wages at all ages. However, since the investment effort decreases with experience, the overall effect of the losses plus learning will tend to be negative only at the end of the working life.

We have seen that even when no unexpected vintage effects in a

[8] See Willis (1986) for a simple model of heterogenous human capital and schooling.

cross-section are present, average wages may first go up and then down depending on our assumptions on human capital destruction. To account for this behavior in a model without a constant rate of depreciation, we need that old workers suffer losses of human capital.

Uncertain losses of human capital have been modelled by Williams (1979) in a context of uncertainty for a level parameter in f(.) and the rental rate of human capital. He further assumes a positive correlation for the two shocks and concludes that workers may hedge against obsolescence by greater investment.

Williams (1978) uses a two period model to conclude that investment in human capital is encouraged when risk increases. Nonetheless, the fundamental result on investment effort still holds: investment in human capital is declining monotonically with the worker's age.

In the following section I present a simple model of human capital accumulation when there is uncertainty in the depreciation of human capital.

### 2.2.3 A simple model of uncertainty in the depreciation of human capital

In this section I will develop a model of human capital accumulation using the theoretical setting introduced at the beginning of this section. For simplicity, I will present the model in discrete time.

I assume a Ben-Porath technology with parameter a=1/2. Depreciation of human capital has two sources:

- overall technological improvements lead to global obsolescence in human capital. This effect is assumed to be proportional to the amount of human capital and will be called overall depreciation.

- workers may suffer a sudden loss of human capital[9]. The degree of the loss will be a multiple of a certain constant level. The probability of this event follows a Poisson distribution. This effect will be called sudden depreciation.

The timing of information and actions in each period is as follows. At the beginning of period t the current value of the sudden loss of human capital is realized. The worker chooses human capital investment based on this realization and the human

---

[9] Following Williams (1979), one could assume that workers may suffer sudden stochastic losses in their rental rate of human capital due to changes in supply and demand factors. The model I present is, I believe, simpler and makes the point clearer, but interpretations in terms of more complex models are also of obvious interest.

capital for the beginning of next period is obtained.

The problem any worker faces can be formally stated as

$$\begin{aligned} v_1(k_1,u_1) &= \max_{\{x_t\}} E_1\left[\sum_{t=1}^{T} rk_t(1-x_t)\right] \\ &\text{s.t.} \\ &k_{t+1} = (1-\delta)k_t + A(k_t x_t)^{1/2} - u_t k, \\ &k_0 \text{ given} \\ &0 \leq x_t \leq 1, \quad t=1,\dots,T-1 \end{aligned} \tag{2.21}$$

where T is the worker's last period and $u_t$ is a random variable that follows a Poisson distribution with coefficient $\lambda_t$. Bellman's Principle of Optimality implies that the worker must assume that he will also optimize her decisions in future periods[10]:

$$v_1(k_1,u_1) = \max_{x_1} rk_1(1-x_1) + E_1\left[v_2(k_2,u_2)\right], \tag{2.22}$$

and, for any period t=1,...,T-1,

$$v_t(k_t,u_t) = \max_{x_t} rk_t(1-x_t) + E_t\left[v_{t+1}(k_{t+1},u_{t+1})\right]. \tag{2.23}$$

A simple way to solve this problem is by backward recursion. I will see what happens at T, and T-1 and propose a recursive solution to the problem. At t=T, the worker has no further incentive for training, so $x_T=0, v_T=rk_T$ so that $E_{T-1}[v_T(k_T,u_T)]=rk_T$.

---

[10] See, for example, Stokey and Lucas (1989).

At t=T-1,

$$v_{T-1}(k_{T-1},u_{T-1}) = \max_{x_{T-1}} \; rk_{T-1}(1-x_{T-1}) + rk_T. \qquad (2.24)$$

Taking into account the difference equation for $k_t$, it is simple to see that an internal solution of the problem at T-1 satisfies the first order condition:

$$-rk_{T-1} + \left(\frac{A}{2}\right) rk_{T-1} \left(k_{T-1}x_{T-1}\right)^{-1/2} = 0 \qquad (2.25)$$

so that

$$\left(k_{T-1}x_{T-1}\right)^{1/2} = \left(\frac{A}{2}\right). \qquad (2.26)$$

We can immediately see that

$$k_T = (1-\delta)k_{T-1} + \left(\frac{A^2}{2}\right) - ku_{T-1}. \qquad (2.27)$$

Note that if $\lambda_{T-1}<1$ and $\delta=0$, the most likely event is that human capital will increase in the last period by $A^2/2$. Nevertheless, if we study a cross-section of workers with independent and identically distributed shocks, then the average level of human capital may decrease since

$$E[k_T] = k_{T-1} + \left(\frac{A^2}{2}\right) - k\lambda_{T-1}. \qquad (2.28)$$

We finally obtain that $v_{T-1}=rk_{T-1}+rk_T-r(A^2/4)$.

To sum up, we have the following results:

$$v_T = rk_T, \quad v_{T-1} = rk_{T-1} + rk_T - r\left(\frac{A^2}{4}\right), \tag{2.29}$$

and

$$k_T = (1-\delta)k_{T-1} + (1 + (1-\delta))\left(\frac{A^2}{2}\right) - ku_{T-1}. \tag{2.30}$$

Suppose that

$$v_t = rk_t + ra_t E_t[k_{t+1}] + f_t, \tag{2.31}$$

and

$$k_{t+1} = (1-\delta)k_t + a_t\left(\frac{A^2}{2}\right) - ku_t, \tag{2.32}$$

where $f_t$ is a function only of the parameters of the model and

$$a_{t-1} = 1 + (1-\delta)a_t, \quad a_T = 0. \tag{2.33}$$

Note that these conditions hold for t=T-1. Under **(2.32)-(2.33)** the problem at t-1 takes the form

$$v_{t-1} = \max_{x_{t-1}} \; rk_{t-1} - rk_{t-1}x_{t-1} + E_{t-1}[\,v_t\,], \tag{2.34}$$

and the first order condition for the internal solution[11] is

$$\left(k_{t-1}x_{t-1}\right)^{1/2} = a_{t-1}\left(\frac{A}{2}\right). \qquad (2.35)$$

By substituting this last equation in the equation in differences for k, we see that $k_t$ takes the form that is equivalent to **(2.32)** one period in advance:

$$k_t = (1-\delta)k_{t-1} + a_{t-1}\left(\frac{A^2}{2}\right) - ku_{t-1}. \qquad (2.36)$$

On the other hand, substituting **(2.35)** into **(2.34)** and applying the expectation operator $E_{t-1}$ on $k_{t+1}$, we have:

$$v_{t-1} = rk_{t-1} + ra_{t-1}E_{t-1}[k_t] + f_{t-1}, \qquad (2.37)$$

where

$$f_{t-1} = -r\left(\frac{a_{t-1}A}{2}\right)^2 + r\frac{(a_tA)^2}{2} - ra_t\lambda_t k + f_t. \qquad (2.38)$$

Thus, **(2.36)** is indeed the optimal solution of human capital accumulation contingent on the stochastic shocks. From **(2.35)** and **(2.36)** and taking into account that $a_t$ is a decreasing sequence we obtain that wages will decrease only if human

[11] Thus, as in the perfect foresight model, gross investment is a decreasing function of time.

capital decreases[12] since:

$$w_t - w_{t-1} = r(k_t - k_{t-1}) + (a_{t-1}^2 - a_t^2)\frac{A^2}{4}. \qquad \textbf{(2.39)}$$

From **(2.36)** we see that wages will tend to fall when depreciation of human capital takes place. Uncertain depreciation only at the end of the working life can be modelled by assuming that $\lambda_t=0$ for all $t<T$. Further assuming that $0<\lambda_T<1$ will lead to rare falls in human capital in the sense that most workers will probably not experience them. However, average wages across workers will likely show a decline even when $\delta=0$ since $\lambda_T>0$.

Are these effects relevant to explain observed data? To answer the problem, we need an econometric specification that can encompass on one hand accumulation paths with sudden local losses of capital and on the other hand accumulation paths which are smooth.

## 2.3 The econometric model

The theory of human capital predicts declining levels of on-the-job training intensity. Alternative human capital models predict

---

[12] Note that wages may increase even if human capital decreases when the effect of less training compensates in wages the effect of destruction in human capital.

either constant experience-wage profiles after a plateau is reached, increasing but concave experience-wage profiles, or inverted-u shaped experience-wage profiles.

The existence of heterogeneous human capital that can be subjected to local sudden losses induce different human capital accumulation paths in the same experience-wage profile. Old local workers will give up net investing after the loss. Young local workers will still find heavy investment profitable. Furthermore, if losses concentrate on older local workers, a structural break in the local profile will appear only at the end of the profiles. The inverted-u shape will be the result of the overlapping of different accumulation regimes for young and old local workers. Mincer (1974) justifies the use of a quadratic function for the experience returns over a log linear schooling model for the analysis of the wage structure on theoretical and empirical grounds. It is instructive to replicate the basic arguments.

Following human capital arguments, schooling takes place only if it gives a positive return $\rho$. Potential earnings when leaving school thus are $k(s)=k_0 \cdot e^{\rho s}$ where s is the number of years at school. This is the log-linear schooling model. After leaving school, individuals will take on-the-job training. If investment in human capital declines linearly so that $k(x)=k_0(1- x/T)$ then potential wages after x years of experience are

$$E(x) = k(s)\ \exp\left\{\rho\int_0^x k(0)\left(1 - \frac{t}{T}\right) dt\right\} \qquad \textbf{(2.40)}$$

and observed log wages then take the form

$$\log w = \log k_0 + \rho s + \rho k_0 x - \left(\frac{\rho k_0}{2T}\right) x^2 + \log(1 - k(x)). \qquad \textbf{(2.41)}$$

The quadratic specification is a convenient simplification of **(2.41)**. There is some empirical evidence in favor of this compromise[13].

Smooth simple specifications -like that of equation **(2.41)**, or the quadratic simplification that results from dropping the last term, or even the Ben-Porath specification- do impose restrictions on the wage structure that might render the results useless in studying local breaks in the experience surfaces.

Interaction terms between education, experience, or other variables in parametric forms may also prove of little value because we are looking precisely for nonhomogeneous behavior in the experience-wage profiles at the end of the working life.

---

[13] Heckman (1976) was not able to reject a quadratic specification against the alternative hypothesis that earnings were generated by the Ben-Porath model. For more sceptical views, see Willis (1986, footnote 2), Murphy and Welch (1990) and Yuengert (1994).

In order to study the problem of sudden losses in the wage during the last stages of the working life I carry out the following two-step empirical analysis:

- I estimate a nonparametric surface for log wages for white male nonfarm full-time employed workers. This model does not restrict local behavior in the experience-wage profiles, but obtains a simplification from the general nonparametric model in a multivariate context. This simplification is a product of the structure observed in the data, so that no theoretical considerations must be introduced. Since the theoretical model of human capital in its general form needs not to be conclusive about the evolution of the marginal value of training investment and the structure of depreciation to render the fundamental results, this model can be used to learn about depreciation and human capital accumulation from the data.

- I then postulate smooth local expected values based on the estimated surfaces for workers in their last stage of the working life by levels of education, type of job, and area. Finally, I test whether local log wages are statistically below the predicted smooth value of the surface.

This two-step procedure is fully justified if we assume flat wage experience profiles after a certain level of experience is reached. It is a conservative approach when profiles are always increasing. Only when overall depreciation is substantial may this strategy lead to misleading results. Assuming possible changes in the rental rates of human capital for different workers, one can interpret the results as evidence of local drops in the willingness of employers to employ old workers -assuming the labor supply constant-.

The analysis is replicated in four different years to check whether the breaks in the surfaces have persistence over the decades. I present the results in the following section.

## 2.4 Empirical results

The data that I will use in this section corresponds to the extracts of the 1980, 1985, 1990, and 1995 Annual Earnings File of the Current Population Survey (CPS) prepared by the NBER.

Before identifying any loss in observed wages as losses in human capital, we should try to control for other possible causes. Gustman and Steinmeier (1984) show that failing to account for partial retirement may result in an overestimate of the decline in wages at old ages. As Johnson and Neumark (1996) point out, if higher-wage workers tend to retire before lower-wage ones, a cross-section analysis will show a spurious decline in wages at older ages. Mincer and Ofek (1982) study the effects of interruption in work careers on human capital and wages. Their study is an investigation of depreciation of human capital in immigrants or returnees to the job market. Using longitudinal data they show that the longer the interruption the greater the decline in wages. Thus, demographic groups that are more likely to interrupt their job careers may present spurious declines in their wages in a cross-section.

In order to minimize the magnitude of these problems, the sample I study consists of wages for white full-time employed male workers with no more than forty years of experience[14].

Wages are the logarithms of earnings per week divided upon hours per week on the job. The variable specification includes education, potential experience, type of job, and region. Here, education is divided upon six categories: Less than six years of education, between six and 12 years of education, high school completed, some college, college completed, and postgraduate studies. Potential experience is divided into eight categories of five years ranging from 1 to 40 years of experience. Potential experience is computed from age minus education minus six. Type of job is a binary variable relating to whether the occupation can be classified as white or blue collar[15]. Region

---

[14] An interesting topic beyond the scope of this study is whether wage decline also implies decline in total compensation. See Johnson and Neumark (1996, footnote 4).

[15] White-collar workers includes managerial, executive, technical, sales and administrative occupations. Blue-collar workers includes craft, precision, and repair operations,

refers to three main areas: east, central and west[16].

For each year, the sample is randomly divided into two samples of sizes of 2/3 and 1/3 of the total, respectively. The first sample is used to carry out the estimation of the wage structure, the second to obtain an honest size tree. The stop-splitting rule is 100 observations. The honest size tree is obtained each year with a zero-SE rule[17].

Aside from the nonparametric analysis, I carry out "Mincer" quadratic equations extended with type of job and region . Table 2.1 shows the fit and complexity results of the regressions for

---

operatives, and transport and service workers. For the 1970 and 1980 occupation codes, white-collar workers would range from codes 1 to 400.

[16] East comprises New England, Middle Atlantic, and South Atlantic. West is Mountain and Pacific. The rest is Central.

[17] The k-SE tree is the simplest tree with the test sample average residual sum of squares smaller than $R(T_{LS})+k\cdot SE(T_{LS})$ where $SE(T_{LS})$ is the estimated standard error for $R(T_{LS})$ and $T_{LS}$ is the test sample LS partition in the sequence. If observations are independent, Breiman et alia (1984,p.306) show that

$$SE(T) = \left( \frac{1}{\sqrt{N}} \right) \left[ \left( \frac{1}{N} \right) \sum \left( w_i - f(\mathbf{x}_i) \right)^4 - R(T)^2 \right]^{\frac{1}{2}}.$$

the nonparametric and for the parametric functions for all years.

The results show that simple nonparametric structures have at least as good a descriptive power as the parametric analysis carried out. For the second sample, this advantage does not disappear. The more interesting fact in Table 2.1 has to do with the last column. Complexity of the model is described by the number of different logwage expected values the model gives. The parametric structures give always 288 different values[18]. Trees will tend to be simpler due to the interaction between the stop-splitting rule and pruning[19]. In Table 2.1 we see that in the more complex tree, complexity drops 44% with respect to the unrestricted nonparametric model or the parametric model.

---

[18] This is the number of cells in the independent variables space: (8 levels of experience) x (6 levels of education) x (3 main areas) x (2 types of employed workers).

[19] However, note how complexity in the tree structures seems to change widely in the years. This is in my opinion an artifact of the 0-SE rule in the algorithm and the flat surface for $R^{ts}$ along the sequence of optimal trees. Differences in complexities for the nonparametric surfaces should be taken with caution.

**Table 2.1** Goodness-of-fit results for logwage regressions. Human capital specification.

Data: Labour extracts-CPS. Male, white, full-time employed workers.

Human capital variable specification + Region + Type of Worker

| | No.observations | | Goodness-of-fit[1] | | Complexity[2] |
|---|---|---|---|---|---|
| | Learning | Test | RE | $R^2$ | |
| **1980** | | | | | |
| Tree | 28,650 | 14,239 | 21.11% | 21.72% | 161 |
| Quadratic | 28,650 | 14,239 | | 18.4% | 288 |
| **1985** | | | | | |
| Tree | 24,795 | 12,323 | 23.34% | 25.26% | 50 |
| Quadratic | 24,795 | 12,323 | | 21.7% | 288 |
| **1990** | | | | | |
| Tree | 46,827 | 23,273 | 31.11% | 31.18% | 130 |
| Quadratic | 46,827 | 23,273 | | 29.2% | 288 |
| **1995** | | | | | |
| Tree | 39,121 | 19,443 | 31.71% | 30.48% | 33 |
| Quadratic | 39,121 | 19,443 | | 29.4% | 288 |

Note: Tree refers to surfaces estimated with regression trees using the 0-SE rule and 100 as the stop-splitting rule in the algorithms. Quadratic refers to "Mincer" quadratic equations extended with type of job and region.

[1]For the tree: RE=1-R(T)/R(root) in the second sample, whilst $R^2$ is for the first sample. For the quadratic specification, $R^2$ is computed for the entire sample.

[2]Complexity is the number of terminal nodes in the trees whilst for the parametric regressions it consists of the number of cells in the independent variables space.

To give a graphical description of the wage structures predicted by the optimal size trees, I plot average predicted values by experience and education for all years in Figures 2.2 to 2.5 in Appendix A. Appendix B shows the results of the parametric analysis. From the graphs in Appendix A we see that concavity is a clear feature of the surface. Plateaus are quite common, so that there are many experience-wage profiles that show a linear behavior with no positive slope.

Note that predictions for workers with the lowest levels of experience and education for 1985, 1990, and 1995 are the same as for workers with a higher level of education. This is the result of lack of observations for these cells. As Tables 2.7 to 2.11 in Appendix A show, there are no workers for these characteristics during those years. The parametric approach will just use the structure as the only information to predict the values for the wages of these workers. If we do not want any structure imposed, the best we can say here is that we do not know what their predicted wages are.

Experience-wage profiles in a cross-section inevitably mix several effects with the increase in human capital. When we compare different years' experience profiles, as in Figure 2.1 in Appendix A, inflation will shift the profiles upwards when we work with nominal log wages. Therefore, differentials along different experience levels will be the same.

We also have vintage effects in cross-sections. These are related to differences in the quality of the workers due to changes in the quality of education prior to work and other factors. Continuous improvements in education levels lead to an underestimation of true experience differentials, since younger workers are of higher quality than older workers. When the same age group is considered along several years, a continuous improvement of the quality of the workers will show up in a decrease in positive experience differentials and an increase in negative experience differentials[20]. Thus, if we regress differentials for the same age group for different years on a constant and a time trend, a positive coefficient on the trend would signal a gradual deterioration in the quality of the newcomers, while a negative coefficient would signal a gradual improvement of the workers' quality. If changes in quality stabilize, the trend parameter will be zero. These step-improvements in the quality of the workers are observationally analogous to sudden losses of human capital in a generation and will not be considered here.

If we work with experience groups instead of age groups, the analysis of the previous paragraph must be changed to account for the education effect. In experience groups there are workers of different ages because there are workers with different levels of education. Thus, under gradual improvements of the quality of the workers, differentials are bigger the higher the education of the workers, since these workers are older. Conversely, if quality is decreasing, differentials are smaller the higher the education of the workers. Thus, by introducing

[20] See, for example, Neuman and Weiss (1995).

education on our regression on experience levels we can correct for the education bias.

To sum up, positive signs in the trend coefficient together with negative signs in the education coefficient are signals of a decreasing quality in the newcomers to the labor market. Negative signs in the trend coefficient together with positive education coefficients along the same experience levels signal increasing quality in the newcomers.

In order to test for the vintage effect, I use the projections from the tree. These are simply within groups averages, so this analysis can be seen as "between groups" were the groups have been endogenously identified by the data. The results of these regressions are summarized in Table 2.2. The regressions were carried out for individuals with more than five years of education since, as already mentioned, the projections given by the tree for some workers with less than five years of education were obtained without observations. Differentials are changes with respect to the least experienced groups, so that all differentials are positive.

Trend coefficients are positive for most experience groups. The only non-significant coefficients occur for the least experienced groups. Education is generally significant, and corroborates the hypothesis of a general continuous decline in the quality of newcomers during the period with respect to

**Table 2.2** Experience wage differentials in the nonparametric surfaces and vintage effects

Data: Projections from Trees at all points of the independent variables space.

Independent variable: experience wage differential with respect to group with less experience.

| | Model 1: Contant + trend | | | Model 2: Constant+trend+education | | | |
|---|---|---|---|---|---|---|---|
| Experience | Const | Trend | $R^2$ | Const | Trend | Educa | $R^2$ |
| 6:10 years | 0.284 | -0.012 | 0.013 | 0.418 | -0.012 | -0.033 | 0.174 |
| | (0.026) | (0.009) | | (0.037) | (0.009) | (0.007) | |
| 11:15 years | 0.341 | -0.009 | 0.008 | 0.470 | 0.009 | -0.032 | 0.172 |
| | (0.025) | (0.009) | | (0.035) | (0.008) | (0.006) | |
| 16:20 years | 0.348 | 0.025 | 0.036 | 0.527 | 0.025 | -0.045 | 0.226 |
| | (0.032) | (0.012) | | (0.044) | (0.011) | (0.008) | |
| 21:25 years | 0.326 | 0.042 | 0.090 | 0.577 | 0.042 | -0.063 | 0.413 |
| | (0.034) | (0.012) | | (0.041) | (0.010) | (0.008) | |
| 26:30 years | 0.354 | 0.040 | 0.068 | 0.632 | 0.040 | -0.070 | 0.396 |
| | (0.037) | (0.014) | | (0.046) | (0.011) | (0.009) | |
| 31:35 years | 0.363 | 0.035 | 0.054 | 0.676 | 0.035 | -0.078 | 0.492 |
| | (0.037) | (0.013) | | (0.041) | (0.010) | (0.008) | |
| 36:40 years | 0.348 | 0.036 | 0.046 | 0.618 | 0.036 | -0.068 | 0.313 |
| | (0.041) | (0.015) | | (0.053) | (0.013) | (0.010) | |

Note:
Standard Errors in parenthesis. The regressions were carried out for individuals with more than five years of education. There were 120 observations in each regression:4 years x 5 levels of education x 2 types of workers x 3 regions.

workers with more than 15 years of experience.

How may these results affect our analysis of sudden losses of human capital at the last stages of working lives? If this deterioration of quality was a long started process, then observed experience differentials in the cross-sample actually underestimate the absolute value of the real experience differentials for equal-quality workers between the most experienced and the previous group when this differential is negative. In other words, our observed experience differentials by education may undervalue the real loss of human capital for these workers, because they are of intrinsically higher quality than workers of the following generation. In the following, I will assume that there was no deterioration in the quality of labor market entrants.

Can we see "general" sudden losses of human capital for the most experienced workers in the surfaces? We can try to uncover these effects by fitting parametric surfaces on the nonparametric surfaces and see whether a dummy variable for the most experienced workers has a significantly negative coefficient.

The main results of the analysis are presented in Table 2.3. Based on the surface graphs in Appendix A, I try two simple parametric specifications. The first one is the quadratic experience profile. The second specification is a linear profile with a single spline in one of the interior experience groups[21]. This group is chosen by selecting the spline model which minimizes the errors sum of squares. This is therefore a LS estimate of the spline model when the point of the spline is

---

[21] See Poirier (1976) for a general discussion of polynomial splines.

**Table 2.3** General breaks in accumulation paths and experience profiles

Data: Projection experience differentials from Trees at all points of the independent variables space.

| | 1st Coefficient[1] | 2nd Coefficient[1] | Dummy | $R^2$ |
|---|---|---|---|---|
| **1980** | | | | |
| Quadratic | 0.077 (0.02) | -0.005 (0.002) | • | 0.547 |
| Quadratic+Dummy | 0.076 (0.03) | -0.005 (0.003) | -0.002 (0.04) | 0.547 |
| Spline 26:30ys | 0.033 (0.006) | -0.038 (0.02) | • | 0.544 |
| Spline+Dummy | 0.033 (0.006) | -0.038 (0.02) | -0.002 (0.05) | 0.544 |
| **1985** | | | | |
| Quadratic | 0.074 (0.02) | -0.004 (0.02) | • | 0.732 |
| Quadratic+Dummy | 0.099 (0.02) | -0.008 (0.003) | 0.052 (0.03) | 0.734 |
| Spline 11:15ys | 0.094 (0.02) | -0.073 (0.02) | • | 0.736 |
| Spline+Dummy | 0.095 (0.02) | -0.074 (0.02) | 0.006 (0.02) | 0.737 |
| **1990** | | | | |
| Quadratic | 0.155 (0.02) | -0.013 (0.002) | • | 0.784 |
| Quadratic+Dummy | 0.156 (0.03) | -0.013 (0.003) | 0.0009 (0.04) | 0.784 |
| Spline 21:25ys | 0.071 (0.008) | -0.081 (0.01) | • | 0.781 |
| Spline+Dummy | 0.068 (0.01) | -0.069 (0.02) | -0.033 (0.03) | 0.782 |
| **1995** | | | | |
| Quadratic | 0.127 (0.02) | -0.008 (0.002) | • | 0.873 |
| Quadratic+Dummy | 0.119 (0.02) | -0.008 (0.002) | -0.014 (0.03) | 0.873 |
| Spline 21:25ys | 0.071 (0.006) | -0.061 (0.01) | • | 0.875 |
| Spline+Dummy | 0.070 (0.006) | -0.053 (0.01) | -0.023 (0.03) | 0.875 |

Note:
Standard Errors in parenthesis. There were 288 observations in each regression:8 levels of experience x 6 levels of education x 2 types of workers x 3 regions.
[1]For quadratic functions, the first coefficient refers to the linear effect and the second to the quadratic effect.For Linear functions with splines, the first coefficient is the first slope, whilst the sum of the first and the second coefficient is the second slope.

unknown. If errors are assumed normal, this is a ML estimate. Since the relevant information is the shape of the differentials' profile, all projections are used here.

The quadratic model fits the surfaces better than linear splines in all years but 1985. Inference is not valid if model specification has been implemented with the same data. We can see, nonetheless, results in Table 2.3 as results coming from two different researchers with different prior beliefs regarding functional especification. What Table 2.3 says is that none of them would reject the hypothesis of absence of general breaks in the accumulation paths for the most experienced workers. This could be interpreted as evidence that workers smooth their profiles by taking into account the general obsolescence effect that technological progress induces on human capital.

I am sceptical about how Table 2.3 can answer questions on sudden losses of human capital. The reason is that general parametric specifications impose restrictions on human capital destruction. In particular, the restriction that it is not locally affecting workers of a certain type, that is, workers with a particular class of human capital. A model including different types of workers and knowledge should allow for possible different depreciation experiences. Therefore, we could try to observe breaks in parametric profiles for each particular type of worker fitted from the tree projections.

**Table 2.4** Parametric especification searches for each type of worker along the nonparametric surfaces

Data: Projection experience differentials from Trees at all points of the independent variables space.

| | Number of Cases | Constant better | $R^2$ -range |
|---|---|---|---|
| Quadratic | 37 | 2 | 0 - 0.941 |
| Spline at 11:15 years experience | 87 | 19 | 0 - 1 |
| Spline at 16:20 years experience | 12 | 0 | 0.359 - 0.978 |
| Spline at 21:25 years experience | 4 | 0 | 0.299 - 0.828 |
| Spline at 26:30 years experience | 4 | 0 | 0.197 - 0.849 |

Note:
For each type of worker, the model with the best fit in terms of $R^2$ was chosen. There were two simple parametric models evaluated: the quadratic model included a constant and a polinomial of second order for experience. A single spline was allowed in a linear model for experience. There were 144 types of workers (4 years x 6 levels of education x 2 types of workers x 3 regions)~and 7 observations in each regression. Regressions were fitted without a constant and shifting the experience codes one level to the left, so that the differential between the group with least experience and itself should be zero.

A summary of specification searches for each worker's experience profile is presented in Table 2.4. There are 144 different types of workers depending on the year, the area, the education level, and the occupation type. Again, I consider the quadratic specification and linear splines. Models were chosen using the LS principle.

Table 2.4 can be read in two different ways. It can be seen as a convenient way of summarizing the general characteristics of the surfaces induced by the tree. From this point of view, we learn from Table 2.4 that the quadratic form is not necessarily the best description of most experience profiles when we carry out the analysis at a more disaggregated level. Mostly, linear specifications are better descriptions of the overall shape of experience profiles for workers. The last column puts a question mark on the validity of these parametric approximations. Some of them are not even better than the constant differentials model[22], as the second column shows.

Inference on the models is invalidated if the same data is used both for model searches and testing. So, testing for a break in the wages of the workers is not valid here. Traditional inference is valid only when either the model function is known or when a new data set is available for inference.

The practical consequence of not following this rule is that a test for the absence of a negative break in the experience profiles for each type of worker under the best model will more frequently be not rejected than as predicted by theory. Here, only one type of worker was accepted to suffer a significant drop at the 10% significance level. Namely, 1980 most experienced workers who had completed College, were white-collar, and worked in the west. It could be argued that a case had arisen in the disaggregated analysis of a break, contradicting the general rejection of such a situation in Table 2.3. However, should not we expect from inferential theory as

[22] Regressions were fitted without a constant and shifting the experience codes one level to the left, so that the differential between the group with least experience and itself should be zero.

likely a rejection between 144 cases even when no drop took actually place?

The way I choose to study sudden losses on wages is by defining a smooth local value of log-wages for most experienced workers based on a nonparametric smoother on the surface projections of workers with no less than 20 years of experience[23] for each type of worker.

I carry out a simple t-student test to see whether the expected value, estimated with the average of the workers' logwages, for each type of worker was significantly below the smooth value. The null hypothesis is that the expected value equals the smooth value. The alternative is that it is below the smooth value. Table 2.5 presents the categories of those workers for which the null was rejected at the 5 percent level of significance.

Perhaps the most surprising result is the fact that most cases concentrate[24] on just two years, 1980 and 1990. This is an

[23] The Nadaraya-Watson kernel estimator was chosen to smooth the proyections. The Gaussian kernel was taken. The kernel estimator is then a simple weighted average of the three nearest values with weigths approximately equal to 0.80, 0.18 and 0.02. See, for example, Fan and Gijbels (1996).

[24] If we assume that each affected group represents a case of one in the Poisson distribution, then naive first moment

**Table 2.5** Tests for sudden losses of human capital

Workers for which the null hipothesis $H_0$: wage = smooth value was rejected against the alternative $H_1$: wage < smooth value at the 5 percent confidence region.

Type of Worker

| Year | Reg | Type[1] | Educ[2] | Smooth Value | Average[3] | St. Dv.[4] | No.[5] |
|---|---|---|---|---|---|---|---|
| 1980 | East | W-C | H-S | 2.0567631 | 1.9993380 | 0.32727646 | 380 |
| 1980 | West | W-C | H-S | 2.2069504 | 2.1534498 | 0.38769741 | 194 |
| 1980 | East | B-C | H-S | 2.0779702 | 1.9554500 | 0.37040496 | 101 |
| 1980 | West | B-C | H-S | 2.1500409 | 2.0501431 | 0.39944996 | 48 |
| 1980 | West | W-C | C | 2.2267310 | 1.9033250 | 0.39612297 | 7 |
| 1985 | Cent | B-C | S-C | 2.3042300 | 2.0418850 | 0.48328715 | 21 |
| 1990 | East | W-C | 6-12 | 2.3832221 | 2.3560687 | 0.41897359 | 815 |
| 1990 | West | W-C | S-C | 2.5997430 | 2.4490954 | 0.45836199 | 54 |
| 1990 | East | B-C | C | 3.0015800 | 2.7901184 | 0.46135565 | 19 |
| 1990 | Cent | B-C | P | 2.9400951 | 2.6934876 | 0.78710996 | 77 |
| 1990 | West | B-C | P | 3.0356000 | 2.8906128 | 0.55138070 | 51 |
| 1995 | Cent | W-C | C | 2.6722500 | 2.4806190 | 0.37476633 | 32 |

Note: The smooth-value is obtained with the Nadaraya-Watson kernel estimator. The Gaussian kernel was taken on the three nearest experience levels tree projections.

[1]W-C: white collar B-C: blue collar

[2]6-12: between 6 and 12 years of education; H-S: High school completed; S-C: Some colleged done; C: College completed; P: Postgraduate

[3] Average logwage for all workers in the sample.

[4] Standard deviation of logwage for all workers in the sample.

[5] Number of workers in the sample.

indication that these falls are erratic in time. Also note that no type of worker suffers these losses during several years. For

estimates of $\lambda_8$ are 0.035 for 1980 and 1990 amd 0.007 for 1985 and 1990. The estimates are much lower if we use individual data.

the 1980 sample, we have predominantly High School, white collar workers in the list. In 1990, most of the workers belong to a single group, east white collar workers with 6 to 12 years of education. However, several groups of College and Postgraduate Studies also appear. Causal inspection on the list suggests that Higher Education and Blue-collar occupation was a dangerous combination.

It is tempting to give explanations for these results. However, "ex-post" explanations are unsatisfactory in the sense that they may be reasonable, but just like a number of many other reasonable explanations. To properly explain these results, we would have to go far beyond the scope of this analysis[25].

How big is the combined effect of the fall in wages for all these workers on average wages of the entire sample? Can it explain declining wages at the end of the working life?

To answer this question, I computed the average fall in the sample for each year at the last stage of the working life first

[25] But simple explanations could be treated as initial hypothesis. Results in Table 3.5 may be the result of geographical, technological and institutional factors such as regional shocks, the introduction of computers in the work place affecting educated workers with administrative tasks, or short term effects of trade liberalization.

**Table 2.6** Effects on average wages at each level of experience of human capital

Data: Individual observations.

| | No. workers | Sample Average | No losses[1] | Substituted[2] |
|---|---|---|---|---|
| **1980** | | | | |
| 31:35 years | 3234 | 2.03 | . | . |
| 36:40 years | 2934 | 1.99 | 1.98 | 2.01 |
| **1985** | | | | |
| 31:35 years | 2384 | 2.26 | . | . |
| 36:40 years | 2157 | 2.23 | 2.24 | 2.24 |
| **1990** | | | | |
| 31:35 years | 5218 | 2.59 | . | . |
| 36:40 years | 3925 | 2.53 | 2.57 | 2.54 |
| **1995** | | | | |
| 31:35 years | 4915 | 2.72 | . | . |
| 36:40 years | 3286 | 2.68 | 2.68 | 2.68 |

Note:
[1] Average logwages when workers from types listed in table 2.5 are not included.
[2] Average logwages with smooth values for workers of types listed in table 2.5.

for all workers and then for the subsample of workers who do not appear on the list of Table 2.5. I also computed average wages if workers from the list in Table 2.5 had the smooth wages instead of the actual wages so that the comparison between averages is done on the same population. The results are presented in Table 2.6.

For all years but 1995, average wages would be 1% higher if no breaks had occurred. If we consider the effect on average wages of these workers, the value of 1990 shows the importance of the effect of the loss in workers with low wages. Average wages without these workers would be 4% higher and the drop in wages for the last experience group would have been of only 2% instead of the reported 6%. However, this result is overestimating the effect because many low-wage workers are excluded in the hypothetical situation. A more realistic measure of that would be average wages with smooth values for the workers with a loss. Then, the wage differential would still be 5%.

For 1980, the sudden losses help reducing the negative experience differential by about 50%. For 1985, by around 33%. The value for 1990 is of only 16.6%. Finally, the reduction in the negative differential is null for the last year of the analysis.

Thus, it seems that sudden losses may have been rather important for some workers in 1980 and 1990. However, their overall impact on the experience-wage profiles is rather limited for 1990 and 1995. On the other hand, these local falls were as important in 1980 as smooth falls in all wages.

## 2.5 Conclusions

In this chapter I have estimated a nonparametric experience-wage profile in a multivariate environment to search for local workers who suffered a sudden loss in wages in their last stage of their working life. I estimated the model with regression trees for individual observations of white male workers for 1980,1985,1990, and 1995.

The nonparametric approach to the estimation of the experience profile fitted the sample better than a simple quadratic specification. When the projections of the tree were carefully studied, linear splines appeared as reasonable alternatives to the quadratic function for many workers.

With respect to sudden losses in wages, there were not many groups affected in two years, 1985 and 1995. For 1980 and 1990 I found drops in log wages for groups of blue-collar, educated workers.

With respect to the extent these losses explain the decline in average wages, I had mixed results. For 1980, 1985, and 1990, average wages would be 1% higher if no breaks had occurred. For 1980, the sudden losses help reducing the negative experience differential by about 50%. For 1985, by around 33%. The value for 1990 is of only 16.6%. Finally, the reduction in the negative differential is null for the last year of the analysis.

Appendix A

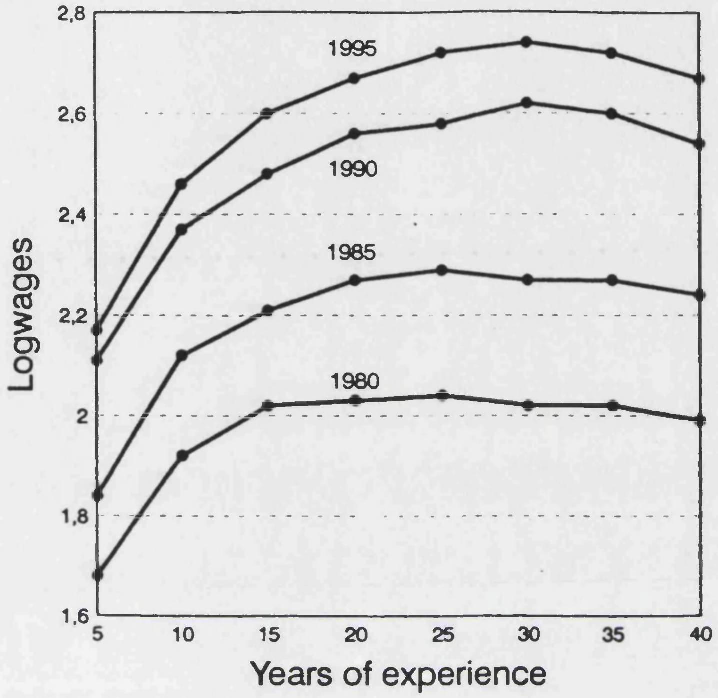


**Figure 2.1** Average wages of white, male, nonfarm, full-time employed workers.

Source: CPS.

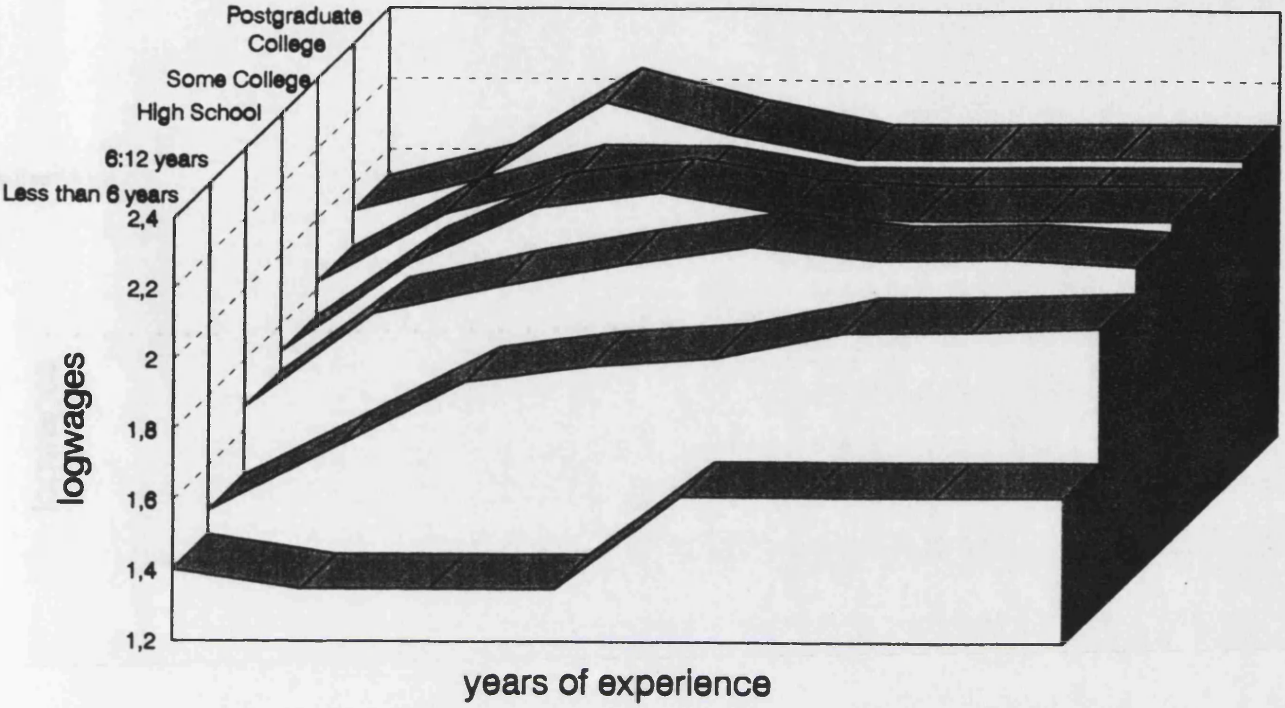


**Figure 2.2** Experience-wage profiles. Unweighted averages of nonparametric projections. 1980.

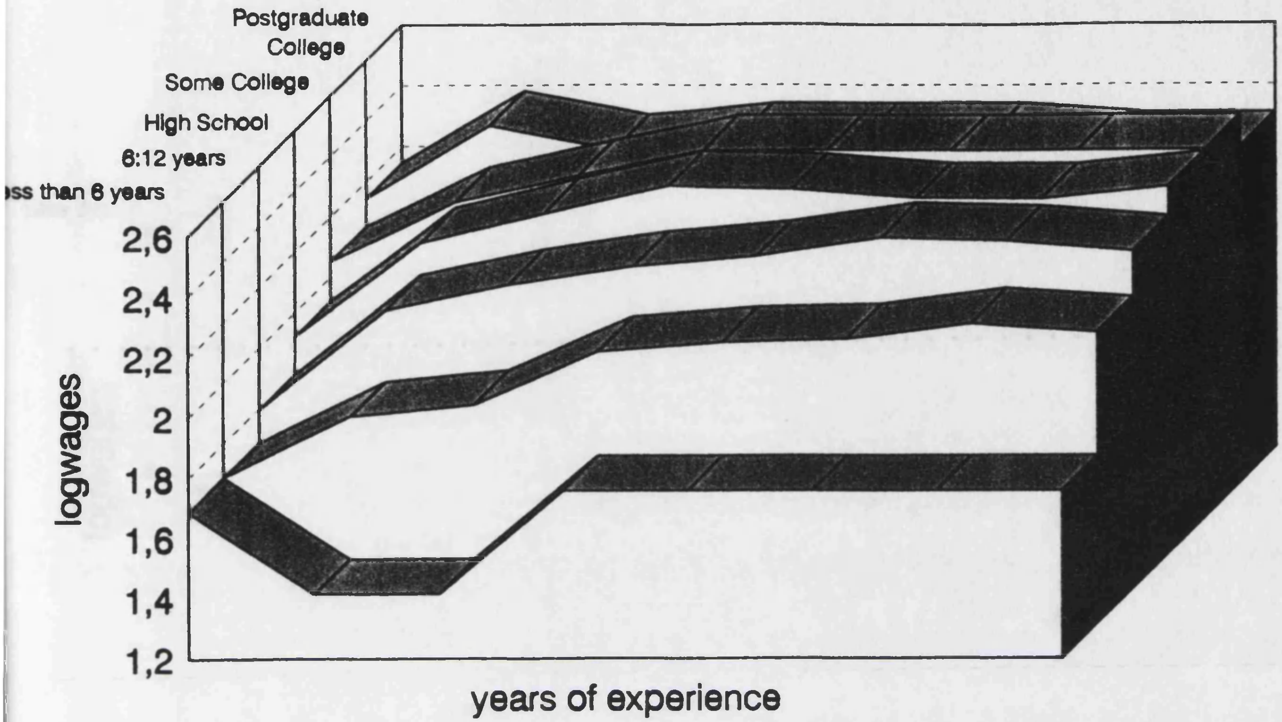


**Figure 2.3** Experience-wage profiles. Unweighted averages of nonparametric projections. 1985.

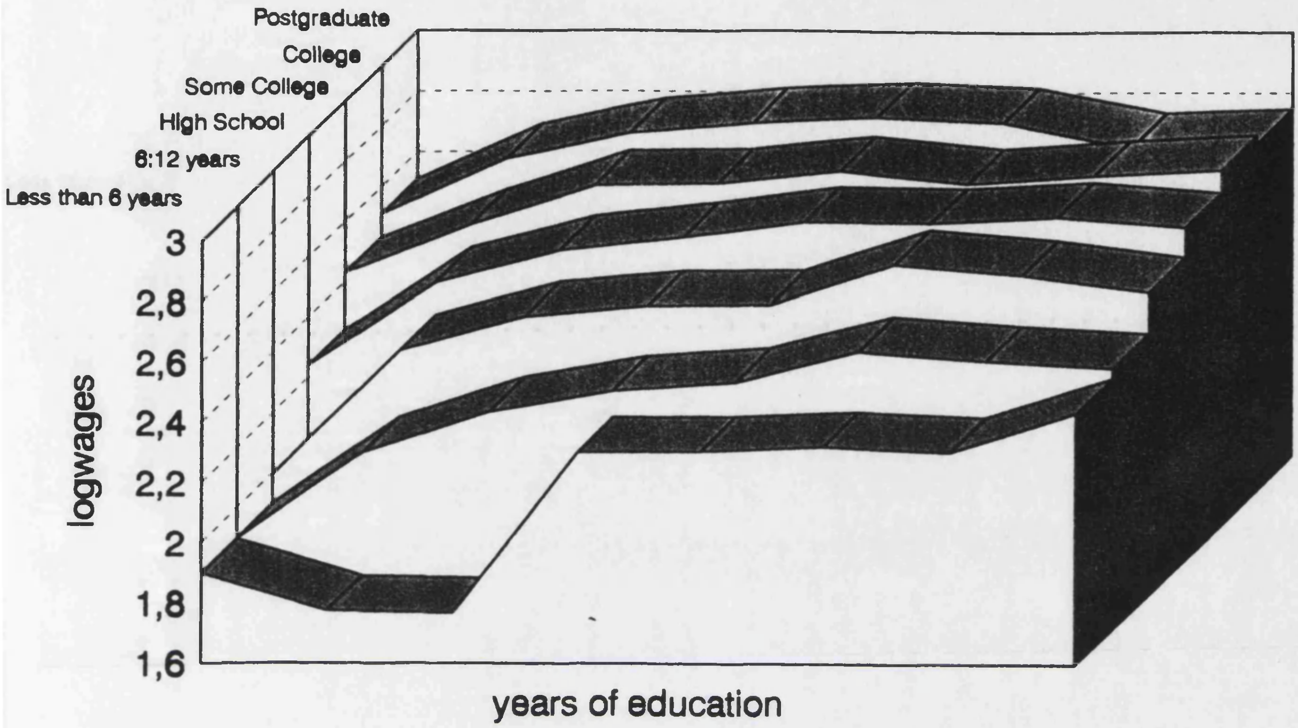


**Figure 2.4** Experience-wage profiles. Unweighted averages of nonparametric projections. 1990.

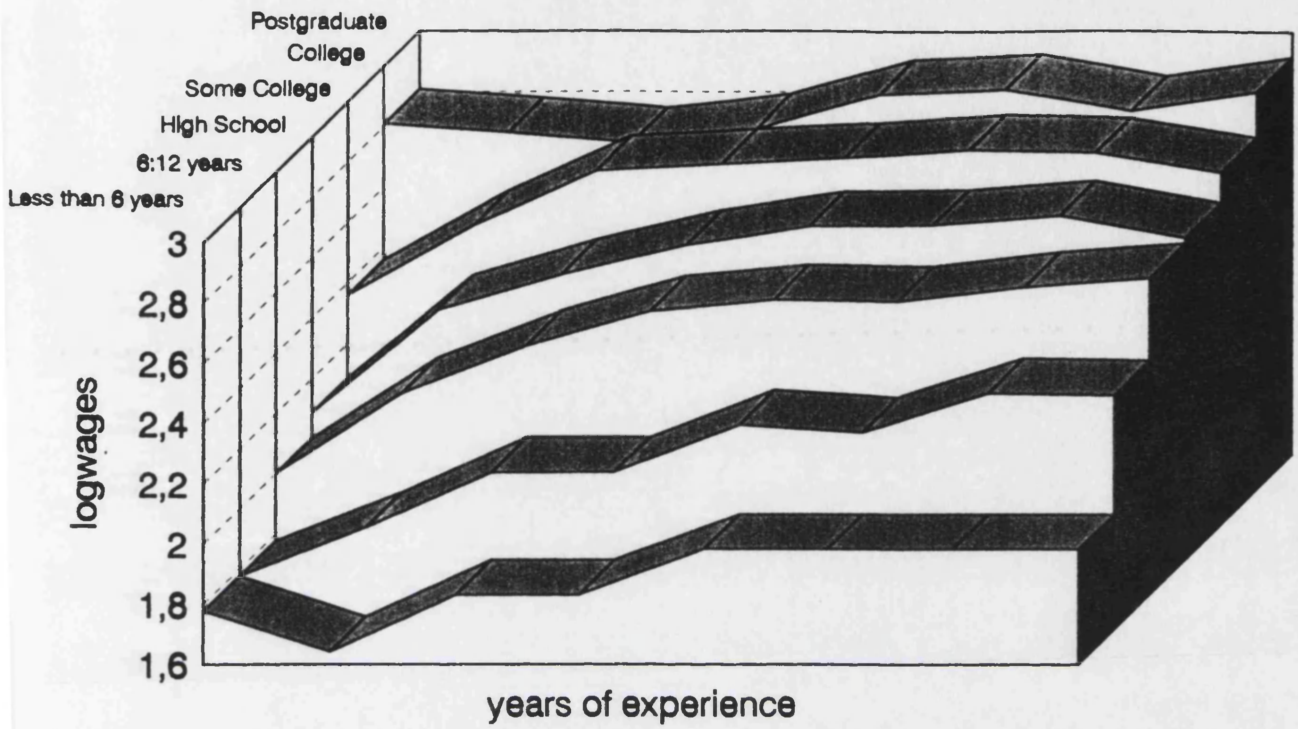


**Figure 2.5** Experience-wage profiles. Unweighted averages of nonparametric projections. 1995.

**Table 2.7** No. of observations in each cell.1980.

| | Education | | | | | |
|---|---|---|---|---|---|---|
| Exp | (1) | (2) | (3) | (4) | (5) | (6) |
| 5 | 2 | 1660 | 4812 | 2020 | 648 | 160 |
| 10 | 28 | 764 | 2358 | 1042 | 310 | 102 |
| 15 | 38 | 664 | 1742 | 578 | 112 | 72 |
| 20 | 38 | 700 | 1458 | 394 | 106 | 28 |
| 25 | 52 | 740 | 1190 | 264 | 48 | 24 |
| 30 | 46 | 792 | 1080 | 220 | 58 | 26 |
| 35 | 58 | 948 | 926 | 192 | 54 | 20 |
| 40 | 72 | 946 | 738 | 118 | 22 | 8 |

(1) Less than five years of education; (2) From six to twelve years; (3) High School; (4) Some college; (5) College; (6) Postgraduate.

**Table 2.8** No. of observations in each cell.1985.

| | Education | | | | | |
|---|---|---|---|---|---|---|
| Exp | (1) | (2) | (3) | (4) | (5) | (6) |
| 5 | 0 | 950 | 3590 | 1724 | 516 | 182 |
| 10 | 12 | 732 | 2518 | 878 | 328 | 138 |
| 15 | 14 | 580 | 1854 | 832 | 258 | 76 |
| 20 | 28 | 480 | 1412 | 488 | 110 | 44 |
| 25 | 32 | 484 | 1162 | 342 | 70 | 52 |
| 30 | 28 | 494 | 866 | 220 | 50 | 26 |
| 35 | 20 | 522 | 788 | 166 | 42 | 14 |
| 40 | 44 | 628 | 696 | 120 | 24 | 12 |

(1) Less than five years of education; (2) From six to twelve years; (3) High School; (4) Some college; (5) College; (6) Postgraduate.

**Table 2.9** No. of observations in each cell.1990.

| | Education | | | | | |
|---|---|---|---|---|---|---|
| Exp | (1) | (2) | (3) | (4) | (5) | (6) |
| 5 | 0 | 3734 | 678 | 2806 | 228 | 476 |
| 10 | 16 | 3946 | 556 | 2286 | 194 | 628 |
| 15 | 66 | 4080 | 628 | 2376 | 220 | 792 |
| 20 | 58 | 3202 | 548 | 2378 | 200 | 794 |
| 25 | 86 | 2564 | 420 | 1500 | 162 | 594 |
| 30 | 54 | 2402 | 246 | 1050 | 88 | 454 |
| 35 | 62 | 1960 | 198 | 824 | 72 | 362 |
| 40 | 56 | 1650 | 124 | 512 | 40 | 176 |

(1) Less than five years of education; (2) From six to twelve years;(3) High School;(4) Some college; (5) College; (6) Postgraduate;

**Table 2.10** No. of observations in each cell.1995.

| | Education | | | | | |
|---|---|---|---|---|---|---|
| Exp | (1) | (2) | (3) | (4) | (5) | (6) |
| 5 | 0 | 456 | 1272 | 1336 | 1662 | 246 |
| 10 | 2 | 426 | 1720 | 986 | 1802 | 494 |
| 15 | 10 | 470 | 2014 | 1140 | 1870 | 600 |
| 20 | 14 | 498 | 2262 | 1094 | 1788 | 762 |
| 25 | 26 | 406 | 1942 | 1106 | 1680 | 834 |
| 30 | 24 | 346 | 1514 | 964 | 1082 | 588 |
| 35 | 50 | 340 | 1194 | 670 | 682 | 292 |
| 40 | 22 | 310 | 956 | 388 | 370 | 176 |

(1) Less than five years of education;(2) From six to twelve years;(3) High School;(4)Some college;(5) College;(6) Postgraduate;

## Appendix B

PARAMETRIC ESTIMATION: 1980:

| | | | |
|---|---|---|---|
| Valid cases: | 42889 | Dependent variable: | LOGWAGE |
| Missing cases: | 0 | Deletion method: | None |
| Total SS: | 7644.148 | Degrees of freedom: | 42883 |
| R-squared: | 0.184 | Rbar-squared: | 0.184 |
| Residual SS: | 6234.926 | Std error of est: | 0.381 |
| F(5,42883): | 1938.489 | Probability of F: | 0.000 |

| Variable | Estimate | Standard Error | t-value | Prob >\|t\| | Standardized Estimate | Cor with Dep Var |
|---|---|---|---|---|---|---|
| CONSTANT | 1.070337 | 0.010627 | 100.722457 | 0.000 | --- | --- |
| EXP | 0.218934 | 0.003500 | 62.545170 | 0.000 | 1.204839 | 0.271903 |
| EXPSQ | -0.019240 | 0.000417 | -46.122679 | 0.000 | -0.887876 | 0.217968 |
| EDUCA | 0.110929 | 0.002259 | 49.099929 | 0.000 | 0.234894 | 0.147801 |
| OCUPA | -0.046345 | 0.005003 | -9.263719 | 0.000 | -0.042722 | 0.020096 |
| REGION | 0.071170 | 0.002425 | 29.352184 | 0.000 | 0.128683 | 0.127628 |

PARAMETRIC ESTIMATION: 1985:

| | | | |
|---|---|---|---|
| Valid cases: | 37118 | Dependent variable: | LOGWAGE |
| Missing cases: | 0 | Deletion method: | None |
| Total SS: | 7699.400 | Degrees of freedom: | 37112 |
| R-squared: | 0.217 | Rbar-squared: | 0.217 |
| Residual SS: | 6026.651 | Std error of est: | 0.403 |
| F(5,37112): | 2060.151 | Probability of F: | 0.000 |

| Variable | Estimate | Standard Error | t-value | Prob >\|t\| | Standardized Estimate | Cor with Dep Var |
|---|---|---|---|---|---|---|
| CONSTANT | 1.131831 | 0.012019 | 94.173997 | 0.000 | --- | --- |
| EXP | 0.257480 | 0.004022 | 64.018482 | 0.000 | 1.240128 | 0.312701 |
| EXPSQ | -0.022294 | 0.000481 | -46.329343 | 0.000 | -0.898808 | 0.249725 |
| EDUCA | 0.127588 | 0.002531 | 50.417168 | 0.000 | 0.251709 | 0.185025 |
| OCUPA | -0.017138 | 0.005432 | -3.155097 | 0.002 | -0.015402 | 0.046476 |
| REGION | 0.052103 | 0.002778 | 18.757760 | 0.000 | 0.086334 | 0.093425 |

PARAMETRIC ESTIMATION: 1990:

| | | | |
|---|---|---|---|
| Valid cases: | 70100 | Dependent variable: | LOGWAGE |
| Missing cases: | 0 | Deletion method: | None |
| Total SS: | 19603.098 | Degrees of freedom: | 70094 |
| R-squared: | 0.292 | Rbar-squared: | 0.292 |
| Residual SS: | 13873.564 | Std error of est: | 0.445 |
| F(5,70094): | 5789.514 | Probability of F: | 0.000 |

| Variable | Estimate | Standard Error | t-value | Prob >\|t\| | Standardized Estimate | Cor with Dep Var |
|---|---|---|---|---|---|---|
| CONSTANT | 1.310001 | 0.008837 | 148.240205 | 0.000 | --- | --- |
| EXP | 0.235094 | 0.003359 | 69.999573 | 0.000 | 0.926566 | 0.261332 |
| EXPSQ | -0.019718 | 0.000391 | -50.444566 | 0.000 | -0.668458 | 0.208301 |
| EDUCA | 0.133183 | 0.001512 | 88.060455 | 0.000 | 0.333783 | 0.412387 |
| OCUPA | 0.159974 | 0.004001 | 39.979573 | 0.000 | 0.151054 | 0.338184 |
| REGION | -0.014856 | 0.002190 | -6.783754 | 0.000 | -0.021573 | -0.029885 |

PARAMETRIC ESTIMATION: 1995:

| | | | |
|---|---|---|---|
| Valid cases: | 58564 | Dependent variable: | LOGWAGE |
| Missing cases: | 0 | Deletion method: | None |
| Total SS: | 18194.644 | Degrees of freedom: | 58558 |
| R-squared: | 0.294 | Rbar-squared: | 0.294 |
| Residual SS: | 12843.937 | Std error of est: | 0.468 |
| F(5,58558): | 4878.982 | Probability of F: | 0.000 |

| Variable | Estimate | Standard Error | t-value | Prob >\|t\| | Standardized Estimate | Cor with Dep Var |
|---|---|---|---|---|---|---|
| CONSTANT | 1.159559 | 0.011309 | 102.538437 | 0.000 | --- | --- |
| EXP | 0.250461 | 0.004009 | 62.474000 | 0.000 | 0.915443 | 0.274595 |
| EXPSQ | -0.020126 | 0.000459 | -43.811163 | 0.000 | -0.642602 | 0.223619 |
| EDUCA | 0.169929 | 0.001961 | 86.657689 | 0.000 | 0.359592 | 0.415750 |
| OCUPA | 0.124692 | 0.004612 | 27.039037 | 0.000 | 0.111853 | 0.318421 |
| REGION | -0.021175 | 0.002550 | -8.303805 | 0.000 | -0.028850 | -0.044604 |

Chapter 3

# Decomposition of average wage differentials for nonparametric wage structures: An application to Mexican workers in the U.S.

## 3.1 Introduction

Wage gaps between two groups exist because of differences in the characteristics of workers of each group and differences in the value the market assigns to the characteristics in each group. The wage gap reflects skills' differences, both observed and unobserved, and differences in the premiums that the two groups have. The basic method to decompose wage gaps can be found in Oaxaca (1973).

If the two groups are equally productive, the extent of the second source of the wage gap is frequently interpreted[1] as wage discrimination. To measure the 'non-discriminatory' wage structure, one must make some assumptions on how the market would behave in such a situation. There have been some attempts to solve this difficulty, notably Cotton (1988), Neumark (1988) and Oaxaca and Ransom (1994).

This paper extends these decomposition techniques to nonparametric tree surfaces. Although it is impossible to talk of simple gender, race, or ethnicity differentials in tree structures, we can still decompose average wage gaps in observed and unobserved components. We can also obtain measures of discrimination and/or sample effects, etc., by simply applying the existing procedures to the nonparametric case.

In short, the model gives flexibility to the wage structure at no cost in the scope of the analysis.

In the next section I present the standard decompositions

---

[1] Unobservable factors such as motivation and cultural background may be related to some observable factors and affect the individual's productivity. Kim and Polachek (1994), Neumark and Korenman (1994) and Polachek and Kim (1994) use panel data to solve endogeneity and heterogeneity problems. Heckman (1979) proposes an adjustment in OLS techniques to account for sample selection bias.

carried out in the literature. Then, I carry out the same decompositions for nonparametric structures and interpret them. In the empirical section of this chapter, I carry out two empirical applications of these decompositions to U.S. micro data on wages of Mexican workers. The chapter ends with some conclusions.

## 3.2 Decomposition of average wage differentials

### 3.2.1 The parametric approach

The problem is to measure the shares in the average wage gap between any two groups, say group 1 and group 2 due to different workers' characteristics and different wage premiums. The commonest approach consists of fitting a wage function to each group and then computing the decomposition.

Let us first assume that we know the values of the wage premiums, $\mathbf{b}_1$ and $\mathbf{b}_2$, so that the wage for any worker i belonging to group 1 is

$$w_i = \mathbf{x}_i' \mathbf{b}_1 + e_i \qquad \textbf{(3.1)}$$

and the wage for any worker i from group 2 is

$$w_i = \mathbf{x}_i' \mathbf{b}_2 + e_i. \qquad \textbf{(3.2)}$$

Average wages within each group are simply

$$\overline{w}_1 = \overline{\mathbf{x}}_1'\mathbf{b}_1 + \overline{e}_1$$
$$\overline{w}_2 = \overline{\mathbf{x}}_2'\mathbf{b}_2 + \overline{e}_2 \quad , \tag{3.3}$$

where the bar over a variable stands for the average operator. If we denote **b** as the wage premiums that would exist if workers from groups 1 and 2 were indistinguishable, then it is straightforward to decompose the average wage gap between groups 1 and 2 as the sum of three conceptually different components:

$$\overline{w}_1 - \overline{w}_2 = \overline{\mathbf{x}}_1'(\mathbf{b}_1 - \mathbf{b}) + \overline{\mathbf{x}}_2'(\mathbf{b} - \mathbf{b}_2) + (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'\mathbf{b} + \left(\overline{e}_1 - \overline{e}_2\right). \tag{3.4}$$

The first two components in the right hand-side of the equation measure the importance of different wage premiums for the two groups. The first one could be understood to be a prize to group number 1 if it is positive. Also, the second term could be seen as the effect of discrimination against group number 2. If we expect that workers from the two groups have different productivity levels, then these first two terms show the extent to which the wage premiums are affected by the productivity differentials for the two groups.

The third component in the right-hand side of **(3.4)** measures the effect on the wage gap of the differences in the characteristics for the two groups.

Finally, the last component of equation **(3.4)** reflects the importance of unobservable factors in the wage gap.

If $\mathbf{b}_1=\mathbf{b}$, that is, if we consider the wage structure for group 1 as the one that would prevail when workers from the two groups were, holding everything else constant, indistinguishable, then the decomposition simplifies to

$$\overline{w}_1 - \overline{w}_2 = \overline{\mathbf{x}}_2'(\mathbf{b}_1 - \mathbf{b}_2) + (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'\mathbf{b}_1 + \left(\overline{e}_1 - \overline{e}_2\right) \quad \textbf{(3.5)}$$

This is the well-known Oaxaca decomposition, and consists of only three components: the structural term, the sample term and the rest. A similar expression can be obtained by assuming that $\mathbf{b}_2=\mathbf{b}$.

Of course, there is no reason why sample and structural components should be the same in the two decompositions. Thus, these decompositions depend upon the assumption on $\mathbf{b}$.

Following Oaxaca and Ransom (1994), we can assume that

$$\mathbf{b} = A\cdot\mathbf{b}_1 + (I-A)\cdot\mathbf{b}_2 \quad \textbf{(3.6)}$$

where A is a weighting square matrix, not necessarily diagonal. Obviously, to make these decompositions operational we need to have estimates for $\mathbf{b}_1$, $\mathbf{b}_2$, and A.

A simpler decomposition can be obtained after making further assumptions on the way $\mathbf{b}_1$ and $\mathbf{b}_2$ relate. Suppose, without loss of generality, that the first element of this vector corresponds to the constant term in the wage equation. Suppose that $\mathbf{b}_1$ is the same as $\mathbf{b}_2$ except for the value of the parameter of the constant term. Thus, $b_{1i}=b_{2i}$ for all $i\neq1$, $b_{11}\neq b_{21}$. and we can set

$\mathbf{b}_1$=( $a_1$+d , $a_2$ , $a_3$ ,..., $a_n$ )' and $\mathbf{b}_2$=$\mathbf{a}$=( $a_1$ , $a_2$ , $a_3$,...,$a_n$)'. Then, the wage for any worker i belonging to group 1 is

$$w_i = \mathbf{x}_i' \mathbf{a} + d + e_i \qquad \textbf{(3.7)}$$

while the wage for any worker i from group 2 is

$$w_i = \mathbf{x}_i' \mathbf{a} + e_i \qquad \textbf{(3.8)}$$

and we can again decompose the average wage differential in several components:

$$\overline{w}_1 - \overline{w}_2 = d + \overline{\mathbf{x}}_1'(\mathbf{a} - \mathbf{b}) + \overline{\mathbf{x}}_2'(\mathbf{b} - \mathbf{a}) + (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'\mathbf{b} + \left(\overline{e}_1 - \overline{e}_2\right). \qquad \textbf{(3.9)}$$

Now, the first three components measure structural components in the average wage gap. Note that although this expression now simplifies irrespectively of **b** into

$$\overline{w}_1 - \overline{w}_2 = d + (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'\mathbf{a} + \overline{e}_1 - \overline{e}_2 \qquad \textbf{(3.10)}$$

the second term in the last equation will measure the sample effect only under the assumption that **b=a.** Although this is again arbitrary and other specifications for **b** are potentially valid, the advantage of **(3.10)** is that it can be implemented in a single regression with a dummy variable. All previous expressions are not implementable in the sense that $\mathbf{b}_1$, $\mathbf{b}_2$, and **b** are not known. The empirical counterpart of **(3.4)** is:

$$\overline{w}_1 - \overline{w}_2 = \overline{\mathbf{x}}_1'(\hat{\mathbf{b}}_1 - \hat{\mathbf{b}}) + \overline{\mathbf{x}}_2'(\hat{\mathbf{b}} - \hat{\mathbf{b}}_2) + (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'\hat{\mathbf{b}} + \overline{\hat{e}}_1 - \overline{\hat{e}}_2 \qquad \textbf{(3.11)}$$

where the hat superscript denotes estimated values. So, if we assume $\mathbf{b}$=$\mathbf{b}_1$ and estimate the wage equations within each group with OLS we obtain the following decomposition:

$$\overline{w}_1 - \overline{w}_2 = \overline{\mathbf{x}}_2'(\hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2) + (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'\hat{\mathbf{b}}_1 \ . \tag{3.12}$$

As suggested above, other possibilities are available. Reimers (1983) chooses

$$\mathbf{b} = \tfrac{1}{2}\mathbf{b}_1 + \tfrac{1}{2}\mathbf{b}_2 \tag{3.13}$$

whilst Cotton (1988) takes $A=s_1 \cdot I$, where $s_1$ is the fraction of workers from group 1. Neumark (1988) proposes a least square criterion to estimate the nondiscriminatory wage structure from the pool sample of workers[2].

Again, if estimates are obtained with OLS, then the fourth term in the decomposition in **(3.4)** disappears, and the resulting decomposition has the form:

$$\overline{w}_1 - \overline{w}_2 = \overline{\mathbf{x}}_1'(\hat{\mathbf{b}}_1 - \hat{\mathbf{b}}) + \overline{\mathbf{x}}_2'(\hat{\mathbf{b}} - \hat{\mathbf{b}}_2) + (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'\hat{\mathbf{b}} \ . \tag{3.14}$$

There are two criticisms to this approach. First, it is unwarranted that the unexplained term in the wage gap must be zero. This is a result of the estimation technique, which fully exploits the null covariance between the error term and the

---

[2] Oaxaca and Ransom (1994) show this is equivalent to a weighting matrix estimated by premultiplying the inverse of the moment matrix of the vector of characteristics for the pooled sample to the moment matrix of one of the groups.

constant included in the set of regressors in each equation. Thus, although it is reasonable to argue that the average effect of this term will be small, the unobservable effects in both groups do not have to cancel each other out. Second, OLS will lead to biased estimates when the error term of the equation is correlated with a regressor. Thus, panel data or IV techniques must be used to implement the decompositions.

### 3.2.2 Nonparametric decompositions

The basic feature in a nonparametric model such as

$$f(\mathbf{x}_i) = \sum_{t \in T^*} c_t \, I\{ \mathbf{x}_i \in t \}, \tag{3.15}$$

is that the wage differential between two types of workers will depend on the other set of characteristics and this dependence cannot be captured by independent functions for each group. Thus, the nonparametric case is a natural generalization of the model in which there are different wage premiums for each group.

Suppose that we want to study the average wage differential for a dichotomous variable $x_j$, $x_j$=1,0. Take $\mathbf{x}_j(k)=(\mathbf{x}_{-j},x_j=k)$ where $\mathbf{x}_{-j}$ is a vector of characteristics containing all elements of $\mathbf{x}$ except $x_j$. The expected wage differential for variable $x_j$, $d(\mathbf{x}_{-j}(k))$, is of the form

$$d(\mathbf{x}_j(k)) = f(\mathbf{x}_j(k)) - f^*(\mathbf{x}_{-j}). \tag{3.16}$$

Note that $f^*(\cdot)$ is the expected wage if the market could not distinguish between workers for which $x_j=1$ and workers for which $x_j=0$.

How can we decompose the observed average wage differences between workers with $x_j=1$ and $x_j=0$? Since

$$\overline{w}_{|x_j=1}-\overline{w}_{|x_j=0}=\frac{1}{N_1}\sum_{(i|x_j=1)} f(\mathbf{x}_i)-\frac{1}{N_0}\sum_{(i|x_j=0)} f(\mathbf{x}_i) + \frac{1}{N_1}\sum_{(i|x_j=1)} e_i-\frac{1}{N_0}\sum_{(i|x_j=0)} e_i, \tag{3.17}$$

where $N_k$ is the number of observations for which $x_j=k$, and

$$f(\mathbf{x}_i|x_j=k)=f(\mathbf{x}_j(k))=f^*(\mathbf{x}_{-j})+d(\mathbf{x}_j(k)) \quad \forall\ k=0,1 \tag{3.18}$$

we can therefore decompose the observed average differences in three terms,

$$\begin{aligned}\overline{w}_{|x_j=1}-\overline{w}_{|x_j=0}=&\left(\frac{1}{N_1}\sum_{i|x_j=1} d(\mathbf{x}_j(1))-\frac{1}{N_0}\sum_{i|x_j=0} d(\mathbf{x}_j(0))\right)\\&+\left(\frac{1}{N_1}\sum_{i|x_j=1} f^*(\mathbf{x}_{-j})-\frac{1}{N_0}\sum_{i|x_j=0} f^*(\mathbf{x}_{-j})\right)\\&+\left(\frac{1}{N_1}\sum_{i|x_j=1} e_i-\frac{1}{N_0}\sum_{i|x_j=0} e_i\right).\end{aligned} \tag{3.19}$$

The interpretation of these terms is equivalent to the interpretation in the linear parametric specification. As in **(3.4)**, the first term can be decomposed into "discrimination" and "favoritism".

In order to do so, what is crucial again is the assumption we take on the nondiscriminatory wage structure. For example, if $f^*(\mathbf{x}_{-j})=f(\mathbf{x}_j(1))$, then

$$\overline{w}_{|x_j=1}-\overline{w}_{|x_j=0} = \left(-\frac{1}{N_0}\sum_{i|x_j=0} d(\mathbf{x}_j(0))\right) + \left(\frac{1}{N_1}\sum_{i|x_j=1} f(\mathbf{x}_j(1)) - \frac{1}{N_0}\sum_{i|x_j=0} f(\mathbf{x}_j(0))\right) + \left(\frac{1}{N_1}\sum_{i|x_j=1} e_i - \frac{1}{N_0}\sum_{i|x_j=0} e_i\right). \quad \textbf{(3.20)}$$

The interpretation of each component is now apparent. The third term in the right-hand side of the equation is the average effect of unobserved skills on the wage gap. The second term is the wage gap that would exist if there were no premiums to the variable $x_j$ and the individuals' distribution of unobserved skills were similar across types. It is therefore the effect of the workers' specialization. The first term is the effect of differing premiums for the two different types.

In the last expression, I have assumed that the structure of the 'non-discriminatory' (i.e. the $x_j$-blind) market would be that of the individuals for which $x_j=1$. This is, as noted before, arbitrary. However, a similar approach to that explained in the parametric case is easily implementable here. In particular, we can assume that

$$f^*(\mathbf{x}_{-j}) = a(\mathbf{x}_j)\cdot f(\mathbf{x}_j(1)) + (1-a(\mathbf{x}_j))\cdot f(\mathbf{x}_j(0)), \quad \forall 0<a(\cdot)<1 \quad \textbf{(3.21)}$$

and we could implement decompositions in the same way as in Cotton (1988), Reymers (1983), or Neumark (1988).

We can therefore compute nonparametric decompositions with the same interpretation as in the parametric case. A detailed analysis of the estimated structures may unveil regions where any discriminatory interpretation would differ. Thus, this technique is an interesting alternative to the parametric methods proposed in the literature. In the following section, I will carry out two empirical applications of these decompositions to U.S. micro data on wages.

## 3.3 Empirical findings

Persons of Hispanic origin make up one of the fastest growing worker groups in the United States. Mexican Americans are, by far, the largest single Hispanic group. Some studies have looked at the characteristics of workers of Hispanic origin[3]. Here, I carry out two wage gap decompositions first between male workers of Mexican origin born in the U.S. or Mexico and then between Mexican Americans and white non-Hispanic male workers.

### 3.3.1 Parametric and nonparametric average wage gap decompositions for workers of Mexican origin

---

[3] See, for example, Cattan (1993) for a description of labor statistics for Hispanics. Reimers (1983), Verdugo (1992), and Cotton (1993) all study the earnings differentials between Black, Hispanic, and non-Hispanic workers with a parametric approach.

The data that I will use in this section corresponds to the extracts of the 1994, 1995, and 1996 Annual Earnings File of the Current Population Survey (CPS) prepared by the NBER.

Two variables are recorded in the extracts of the CPS relating to Mexican origin. The first one is country of birth and the second one is ethnicity. This last variable is the answer to the question: "What is your origin or descent?". This is intended to be the national or cultural group a person is descended from and is determined by the nationality or lineage of a person's ancestors. There is no rule on how many generations to consider. A respondent may report origin based on the origin of a parent, or a far-removed ancestor. Origin is not necessarily related to race or country of origin. It shows the respondent´s self-perception in terms of ethnicity. Two ethnic groups are considered in this study: Mexican Americans and Mexicanos.

Country of birth may help us to control for integration in the labor market. There is empirical evidence[4] showing that English proficiency counts. Furthermore, in the sample, the vast majority of workers born in Mexico do not hold full U.S. citizenship. There is also indirect empirical evidence showing

[4] See, for example, Dávila et alia (1993) and Bloom and Grenier(1993).

that this may also be affecting their wages[5].

The sample I study consists of full-time employed male workers of Mexican origin born in the U.S. or Mexico. The data for 1994, 1995, and 1996 are pooled in a single sample to enlarge the number of observations available. In order to avoid duplications in the observations, I choose the first outgoing rotation interview for each household. In the following, I will refer to this data set as the "Mex data"[6].

There is, as said, no perfect relation between ethnicity and country of birth[7]. Thus, we can find Mexican Americans born in Mexico and in the U.S., and Mexicanos again born in Mexico and in the U.S.. More Mexican Americans hold U.S. citizenship than

---

[5] Donato and Massey (1993) show that the Immigration Reform and Control Act (IRCA) of 1986 increased the wage penalties accruing to undocumented status. Pagan and Dávila (1996) find however that IRCA reduced the true wages of male natives most likely to be mistaken as unauthorized.

[6] See also Appendix A for a description of the data set.

[7] The Spearman rank correlation coefficient between the country of birth code and the ethnicity code was 0.73, 0.70, and 0.70 for 1994, 1995, and 1996 respectively.

are born in the U.S.[8].

Wages are the logarithms of earnings per week divided upon hours per week at the job. The variable specification includes education, potential experience, country of birth, and ethnicity. Education refers to whether the worker has completed high school and at least started further education, x1=2, or not, x1=1. Potential experience is divided into 5 categories: from 1 to 10 years of potential experience, from 11 to 20, from 21 to 30, from 31 to 40 and from 40 years of potential experience onwards. Potential experience is computed from age minus education minus six. Values from 1994 are inflated with the index of wage inflation from the entire sample of full time employed workers. Values from 1996 are deflated with the same index for 1995:1996. The data set contains 5265 different cases. The sample is randomly divided into two samples of sizes of 2/3 and 1/3 of the total, respectively.

The result of the splitting process can be seen in Appendix A. The splitting stops if there are less than 50 observations in a node. Figure 3.1 in Appendix B shows the residual sum of squares both for the estimation and the test sample.

The minimum $R^{ts}$ is obtained with tree T21, as it is shown in

---

[8] For example, in 1995, 88% of Mexican Americans had U.S. citizenship whilst only 86% were born in the U.S..

**Table 3.1** Regression results: Goodness-of-fit measures[1] for the Mex data set.

Data: CPS,1994,1995,1996. 1st outgoing rotation group. Born in Mexico or in the USA; full time; male; Mexican American or Mexicano. Estimation Sample:3517 cases. Test Sample: 1748 cases.

Human capital variable specification + Country of birth + Ethnic

| | 1st sample | 2nd sample | Both samples |
|---|---|---|---|
| Non-parametric regression: | | | |
| Tree T21 | 0.2286 | 0.2099 | . |
| Parametric regressions: | | | |
| with Dummy | . | 0.2123 | 0.2154 |
| born in U.S. | . | 0.1826 | 0.2003 |
| born in Mexico | . | 0.0833 | 0.0805 |

Note: Tree T21 refers to the surface estimated with regression trees using the 0-SE rule and 50 as the stop-splitting rule in the algorithms. Parametric regressions refer to "Mincer" quadratic equations extended with ethnic.

[1]For the tree: RE=1-R(T)/R(root) in the second sample, whilst $R^2$ is for the first sample. For the quadratic specification, $R^2$ is computed for the entire sample.

Appendix A. This tree has 13 terminal nodes and the $R^2$ measure based on the estimation sample is 0.2286. For the test sample, it falls to 0.2099. Table 3.1 gives goodness-of-fit indices for the nonparametric regression and for two parametric specifications. The first parametric regression estimates the effect of place of birth and ethnic association on wages by including these variables in the wage equation. The second specification consists of different equations for workers born in the U.S. and workers born in Mexico. Similar results were obtained using different wage equations for workers declaring to be of Mexican American origin and workers considering themselves as Mexicanos. There are very little differences in our goodness-of-fit measures for all these different models, so it seems hard to choose one model on this basis.

How complex are the models? The parametric models incorporate square terms for potential experience and linear terms for all the variables. This implies 40 different expected wages. Therefore, tree T21 is a simpler structure, with only 13 different expected wages.

Figure 3.2 in Appendix B shows tree T21. High average wages will tend to occur on the right branches of the tree due to the variable specification. The influence of experience is clearly nonlinear. Some workers with very low or very high level of experience do have average wages smaller than workers with potential experience of between 10 and 40 years. It is interesting to note that the negative effect of potential experience for old workers only appears in workers born in the U.S. -splits at nodes number 9 and 14. If anything, we observe an increasing earnings experience profile for Mexico-born workers -node 23.

The role of place of birth is clear. Being born in the U.S. usually carries a premium in the wage structure. A simple way of getting this information is by plotting all wage differentials between U.S. and Mexico-born workers. Figure 3.3

in Appendix B shows the predicted wage differentials for the variable country-of-birth in descending order. The graph shows the wage differential between a worker born in the U.S. and a worker born in Mexico in all points $\mathbf{x}_{-j}$. The ordering is however arbitrary. I chose to present wage differentials in descending order to make the range of values more visible to the reader. It can be argued that place-of-birth is a crude indicator of other variables describing the degree of integration of the worker in the country. Therefore we can expect[9] a positive premium for most workers born in the U.S., as it is shown in the graph. Although there are some negative values, most of the differentials are of the expected sign, ranging from around 0.10 to 0.50. The majority are around 0.20. We could expect a value near -0.20 in the estimation of the parameter of the variable country-of-birth in the parametric specification. In Appendix C I present the results of the parametric regressions with dummy variables for country-of-birth and ethnicity. The estimated value for the country-of-birth effect is -0.23, a reasonable value considering what tree T21 tells us about the

[9] As mentioned before, language proficiency is a clear productive factor. However, other arguments have been examined in the literature. Walker (1996), for example, argues that "captivity mechanisms may be present in allocating Mexican immigrants and local women to the lowest paid and most undesirable mobs within production". Addressing this debate is beyond the scope of this empirical analysis.

differentials.

We can also look at the differentials for workers who declared to be of Mexican-American origin and workers who declared to be of Mexicano origin. As mentioned before, this variable shows the respondent's self perception. If this perception corresponds also to the way the market perceives the worker, then a negative effect of this variable on wages could be understood as wage discrimination.

In Figure 3.4 in Appendix B I show ethnic differentials obtained from T21. Clearly ethnicity is not a relevant factor for most workers. Perhaps more interestingly, when it is relevant in explaining differences, it does so with different signs. Among workers born in Mexico, with low education and with more than 10 years of experience, those who considered themselves as Mexican-American had a higher wage than those who considered themselves as Mexicano. However, among workers born in the U.S., with higher education, and with potential experience of between 10 and 40 years, those who considered themselves as Mexicano had higher wages than those who considered themselves as Mexican-American.

The interesting methodological point that ethnicity raises is that the tree structure deals with the strong non-homogeneity already described. The dummy variable approach takes the differential to be constant. The estimated value[10] of the dummy specification is "near" zero, but statistically significant and of negative sign. Looking at the results of the parametric regressions, we could argue that there is overall "discrimination". Using different equations for Mexican-Americans and for Mexicanos does not help: for some Mexican-Americans and for some Mexicanos the distinction matters, for most it does not.

The difference between regression trees and the linear specification can be understood as the result of two factors. It is first a consequence of the algorithm used to estimate the tree structure in regression trees. The recursive algorithm used in the estimation of the nonparametric surface works in practice not only as a nonparametric estimation technique, but also as a variable selection procedure. Secondly, it also reflects that linear regressions use linear covariation to render the estimates.

Thus, although all models score similarly in terms of the fit in and out of sample, the structures are different. Next, I will show how this affects the results of a standard decomposition of average wage differentials.

I first computed the decompositions of average wage gaps

[10] See Appendix C.

**Table 3.2** Average Wage Decompositions. Mex data. Reference variable: Country of Birth. Reference worker: born in Mexico.

| Data | Total | Structure | Sample | Rest |
|---|---|---|---|---|
| Tree T21 | | | | |
| Estimation Sample | 0.35 | 0.28 | 0.068 | 0.00 |
| Test Sample | 0.34 | 0.28 | 0.061 | 0.0006 |
| All | 0.34 | 0.28 | 0.065 | 0.0007 |
| Sees[1] | | | | |
| Estimation Sample | 0.35 | 0.23 | 0.11 | 0.00 |
| Test Sample | 0.34 | 0.23 | 0.11 | 0.0007 |
| All | 0.34 | 0.23 | 0.11 | 0.0007 |
| Sea[2] | | | | |
| Estimation Sample | 0.35 | 0.23 | 0.12 | -0.001 |
| Test Sample | 0.34 | 0.23 | 0.11 | 0.0007 |
| All | 0.34 | 0.23 | 0.11 | 0.00 |
| Tees[3] | | | | |
| Estimation Sample | 0.35 | 0.15 | 0.20 | 0.00 |
| Test Sample | 0.34 | 0.15 | 0.19 | -0.001 |
| All | 0.34 | 0.15 | 0.20 | 0.00005 |
| Tea[4] | | | | |
| Estimation Sample | 0.35 | 0.17 | 0.17 | -0.0004 |
| Test Sample | 0.34 | 0.19 | 0.16 | -0.0009 |
| All | 0.34 | 0.17 | 0.17 | 0.00 |

[1] Single equation estimated with the estimation sample.

[2] Single equation estimated with all observations.

[3] Two equations estimated with the estimation sample.

[4] Two equations estimated with all observations.

The last three columns in all rows do not sum up to the first column due to rounding errors.

assuming that the "blind" wage was similar to that of the worker born in the U.S.. I obtained six decompositions for each parametric model: single linear parametric regression and separate linear regressions. These decompositions depend upon whether all observations or those from the estimation sample are used to get the estimates and on whether the decompositions were carried out on the estimation sample, the test sample, or all observations. For T21 I also carried out three decompositions depending again on the sample used. Table 3.2 gives the results of the decompositions of the average wage differential between workers born in the U.S. and workers born in Mexico. In the first column we can see the positive gap in average wages in favor of workers born in the U.S.. All models suggest the importance of the structure of wages to explain the gap. There is a clear difference, however, between T21 on one side and the parametric models analyzed on the other side regarding the importance of the sample effect. T21 gives very little weight to the sample effect, that is, to the different characteristics of the workers. From T21 we can deduct that what most matters is the fact that the advantaged workers were born in the U.S.. This fact allows us to predict them as more productive holding all other observed characteristics constant.

The parametric models suggest that U.S.-born workers have on average a higher wage in part because they were born in the U.S. and in part because they tend to consider themselves as Mexican-American, a trace that comes with a premium in the labor market in the parametric models.

We can do a similar exercise for the wage gap between workers who considered themselves as Mexican-American and workers who considered themselves as Mexicano.

The results are presented in Table 3.3. The first column of the table shows the average wage premium for workers who considered they were of Mexican-American origin. There is more overall agreement in the models regarding this decomposition. The wage gap is mainly described by the sample differences between the two groups. Crucially, the fact that most Mexican-American were born in the U.S..

From Figure 3.4 in Appendix B we can obtain an explanation of why the overall structural effect is smaller in the parametric models than in T21. The parametric models impose a structure to the data that in practice smooth out large wage differentials between some workers with different ethnic affiliation.

T21 does not smooth this differentials, as seen in figure 3.4. The overall effect in the nonparametric decomposition will depend on the number of observations falling in each category. In this example, the effect of the structural differences is clearly positive.

Note how adding more flexibility to the parametric single equation model by estimating two equations does not solve the problems of parametric regression in this particular example.

**Table 3.3** Average Wage Decompositions. Mex data. Reference variable: Ethnic. Reference worker: Mexicano.

| | Data | Total | Structure | Sample | Rest |
|---|---|---|---|---|---|
| Tree T21 | Estimation Sample | 0.28 | 0.11 | 0.16 | 0.013 |
| | Test Sample | 0.29 | 0.11 | 0.17 | 0.0074 |
| | All | 0.28 | 0.11 | 0.17 | 0.0074 |
| Sees[1] | | | | | |
| | Estimation Sample | 0.28 | 0.05 | 0.23 | 0.00 |
| | Test Sample | 0.29 | 0.05 | 0.24 | -0.005 |
| | All | 0.28 | 0.05 | 0.24 | -0.001 |
| Sea[2] | | | | | |
| | Estimation Sample | 0.28 | 0.04 | 0.24 | 0.001 |
| | Test Sample | 0.29 | 0.04 | 0.25 | -0.003 |
| | All | 0.28 | 0.04 | 0.24 | 0.00 |
| Tees[3] | | | | | |
| | Estimation Sample | 0.28 | -0.004 | 0.29 | 0.00 |
| | Test Sample | 0.29 | 0.002 | 0.30 | -0.03 |
| | All | 0.28 | -0.002 | 0.29 | -0.002 |
| Tea[4] | | | | | |
| | Estimation Sample | 0.28 | 0.005 | 0.27 | 0.002 |
| | Test Sample | 0.29 | 0.01 | 0.28 | -0.006 |
| | All | 0.28 | 0.007 | 0.28 | 0.00 |

[1] Single equation estimated with the estimation sample.

[2] Single equation estimated with all observations.

[3] Two equations estimated with the estimation sample.

[4] Two equations estimated with all observations.

The last three columns in all rows do not sum up to the first column due to rounding errors.

Why do the unobservable factors have a negative effect in the wage gap in all parametric models? A plausible explanation is that the linear parametric models are too restrictive and T21 is nearer the true model. We know from Figure 3.4 that wage differentials for most workers along ethnic origin are null and that the parametric structures nonetheless force a positive wage differential in favor of Mexican-American workers. The adjustment in the wage gap equation comes as an artificial negative effect of the unobservable effect[11].

### 3.3.2 Average non-Hispanic-Mexican wage differentials in the border states

I study the wage gap between Mexican Americans and white non-Hispanic male full-time employed workers in the border states[12] between Mexico and the U.S.. I concentrate on this region to avoid possible regional effects in the wage gap that may arise since almost three quarters of Mexican Americans interviewed were from the four border states.

The data that I will use in this section corresponds to the extracts of the 1995 Annual Earnings File of the Current Population Survey (CPS) prepared by the NBER. Here Mexican Americans are all respondents who stated they were of Mexican

---

[11] To see this, it is enough to consider that the first term in **(3.20)** is overestimated whilst the third term is computed as a residual.

[12] California, Arizona, New Mexico, and Texas.

American, Chicano, or Mexicano ethnic origin and their parents and themselves were born in the U.S.. Non-Hispanic white workers were all those white Non-Hispanic workers born in the U.S. whose parents were also born in the U.S..

Wages are the logarithms of earnings per week divided upon hours per week at the job. The variable specification consists of education and potential experience. Education refers to whether the worker has not completed high school, has just high school, or has higher education. Potential experience is divided into five categories: from 1 to 10 years of potential experience, from 11 to 20, from 21 to 30, from 31 to 40 and from 40 years of potential experience onwards. Potential experience is computed from age minus years of education minus six.

The sample consists of 5,527 cases of which 661, 11%, were Mexican American. To obtain an honest size tree, I randomly split the Non-Hispanic sample into an estimation and a test sample. All Mexican Americans are included in both samples[13]. I will call this data set[14] the "Texmex data".

---

[13] This is a procedure somewhere between v-fold cross validation and spliting the sample into two. The reason to do this here is that I want to minimize the possibility of one group of Mexican Americans not being sufficiently represented in either sample.

[14] See also Appendix A.

The result of the splitting process can be seen in Appendix A. The splitting stops if there are less than 10 observations in a node. Figure 3.5 in Appendix B shows the residual sum of squares for the test sample. The minimum $R^{ts}$ is obtained with tree T1. Nevertheless, the fundamental structure in the data can be analyzed by looking at simpler trees, such as T8. This tree has 23 terminal nodes and the $R^2$ measure based on the estimation sample is 0.25. For the test sample, it rises to 0.26. Tree T8 is shown in Figure 3.6 in Appendix B.

Ethnicity is a split criterion in all regions of the independent variables' space. Casual inspection of the figure shows that Mexican origin comes with a penalty, that is, all ethnic left-branch projections are higher than right-branch projections. Figure 3.7 in Appendix B shows all predicted wage differentials for ethnic in descending order. But for a few exceptions[15], wage differentials are around 0.20.

The interesting methodological point that this analysis shows is that regression trees may unveil also linear structures. The dummy variable approach takes the differential to be constant. In this example, this is almost right, so we might expect small

[15] Consider, for example, workers with more than 40 years of experience who went into higher education. The projected log wage for non-Hispanic white workers was 2.82, whilst the projection for Mexican origin workers was only 1.54.

**Table 3.4** Average Wage Decompositions. Texmex data. Reference variable: Ethnicity. Reference worker: Mexican American, Chicano or Mexicano.

| Data | Total | Structure | Sample | Rest |
|---|---|---|---|---|
| Tree T21 | | | | |
| Estimation Sample | 0.36358 | 0.23052 | 0.13306 | 0.00 |
| Test Sample | 0.37291 | 0.23052 | 0.13063 | 0.011764 |
| All | 0.36825 | 0.23052 | 0.13184 | 0.0058821 |
| Sea[1] | | | | |
| All | 0.36825 | 0.23146 | 0.13679 | 0.00 |

[1]Single equation estimated with all observations.

The last three columns in all rows do not sum up to the first column due to rounding errors.

differences in the results of the decompositions.

Table 3.4 gives the results of the decompositions of the average wage differential between white Non-Hispanic and Mexican ethnicity workers. As suspected, the decompositions are almost equivalent, reflecting that the parametric decomposition is meaningful because almost all wages follow a linear structure with respect to ethnicity[16].

[16] I implemented other variable specifications to see whether the result was robust. In particular, an aggregate of occupation and also union membership were introduced in the analysis. Although the estimated tree was sensitive to these

## 3.4 Conclusions

In this chapter I show that average wage gap decompositions between any two groups of workers can be carried out without any compromise in their interpretation using a nonparametric wage structure. Oaxaca type decompositions are simply generalized to decompositions when the differentials do not have a simple parametric structure.

I proceed by studying wage gap decompositions for two groups of Mexican workers. I choose a human capital specification augmented with ethnic/origin variables. Of course, there may be some relevant variables not considered in the analysis. It is thus of interest to obtain realistic estimates of the effect on the wage gap of these unobservable factors.

The nonparametric approach differs in this example from parametric specifications in that it gives different country-of-birth and ethnic differentials depending on the worker's characteristics. In particular, ethnicity has very little effect on expected wages, even affecting in different ways U.S. born workers and Mexico-born workers and lower- and higher-education workers.

---

changes, the decompositions of the wage gap between mexican and white non-Hispanic were similar.

In the nonparametric model, the structural effect -the part of the wage gap due to the pricing of the worker's characteristics in the labor market- is larger than the sample effect -the effect of differing characteristics in the two populations- after decomposing the average wage gap between workers born in Mexico and workers born in the U.S.. For the parametric models the difference between the structural and the sample component is very small.

It is reasonable to take birth-of-place as a proxy of integration in and accessibility to the U.S. labor market. The nonparametric approach suggests that this factor is a predictor of overall productivity. The wage gap between Mexican workers born in the U.S. and those born in Mexico is primarily due to their productivity differentials. Their characteristics' differentials play a more limited role. The parametric decompositions fail in recognizing this fact.

In the second empirical analysis, I study the average wage gap between Mexican Americans and white non-Hispanic male workers in the border states between Mexico and the U.S.. Regression Trees unveils a linear relation in the wage structure with respect to ethnicity so that most ethnic differentials are around 20%. Due to this linear behavior, the nonparametric decomposition is very similar to a simple decomposition with a dummy variable.

## Appendix A

**Mex Data:**
**CPS,1994,1995,1996.1ST OUTGOING ROTATION GROUP.**
**BORN IN MEXICO-BORN IN THE USA**
**MALE, EMPLOYED, WORKS MORE THAN 35 HOURS**
**ETHNIC: MEXICAN AMERICAN, MEXICANO**
**5265 OBSERVATIONS: Estimation Sample:3517 cases. Test Sample: 1748 cases.**

Variables:
x1: Potential experience 1:1-10 2:11-20 3:21-30 4:31-40 5:+40
x2: Education 1: At most, High School 2: more than High School
x3: Country of Birth 1: USA 2: Mexico
x4: Ethnic 1: Mexican American 2: Mexicano

| Tree: | (Yes) | (No) | Average |
|---|---|---|---|
| Nodo 1 : x2<=1.5 | 2682 | 833 | 2.1422908 |
| Nodo 2 : x3<=1.5 | 867 | 1815 | 2.0412585 |
| Nodo 3 : x1<=1.5 | 344 | 489 | 2.467583 |
| Nodo 4 : x1<=1.5 | 283 | 584 | 2.2087824 |
| Nodo 5 : x1<=1.5 | 433 | 1382 | 1.9612348 |
| Nodo 6 : x3<=1.5 | 226 | 118 | 2.2783496 |
| Nodo 7 : x3<=1.5 | 356 | 133 | 2.6007041 |
| Nodo 8 : x4<=1.5 | 211 | 72 | 2.0157391 |
| Nodo 9 : x1<=4.5 | 538 | 46 | 2.302329 |
| Nodo 10 : x4<=1.5 | 35 | 398 | 1.8321234 |
| Nodo 11 : x4<=1.5 | 114 | 1268 | 2.0016872 |
| Nodo 12 : x4<=1.5 | 168 | 58 | 2.3425075 |
| Nodo 13 : x4<=1.5 | 17 | 101 | 2.1554711 |
| Nodo 14 : x1<=4.5 | 348 | 8 | 2.6799407 |
| Nodo 15 : x1<=2.5 | 89 | 44 | 2.3886124 |
| Nodo 18 : x1<=2.5 | 276 | 262 | 2.3184431 |
| Nodo 22 : x1<=2.5 | 39 | 75 | 2.1784505 |
| Nodo 23 : x1<=2.5 | 569 | 699 | 1.9857952 |
| Nodo 28 : x4<=1.5 | 280 | 68 | 2.6937847 |
| Nodo 30 : x4<=1.5 | 11 | 78 | 2.3332398 |
| Nodo 32 : x4<=1.5 | 213 | 63 | 2.2804515 |
| Nodo 33 : x1<=3.5 | 170 | 92 | 2.3584648 |
| Nodo 35 : x1<=4.5 | 63 | 12 | 2.2429412 |
| Nodo 37 : x1<=3.5 | 402 | 297 | 2.017662 |
| Nodo 38 : x1<=2.5 | 158 | 122 | 2.6556589 |
| Nodo 39 : x1<=3.5 | 56 | 12 | 2.8507733 |
| Nodo 44 : x4<=1.5 | 134 | 36 | 2.3958163 |
| Nodo 45 : x4<=1.5 | 80 | 12 | 2.2894457 |
| Nodo 46 : x1<=3.5 | 38 | 25 | 2.2566656 |
| Nodo 49 : x1<=4.5 | 203 | 94 | 2.0288882 |
| Nodo 51 : x1<=3.5 | 92 | 30 | 2.7272823 |
| Nodo 52 : x1<=2.5 | 33 | 23 | 2.8284754 |

Terminal Nodes' Averages:

| | | | |
|---|---|---|---|
| 16 | 2.0183977 | 17 | 2.007948 |
| 19 | 2.1138643 | 20 | 1.941488 |
| 21 | 1.8225059 | 24 | 2.3488662 |
| 25 | 2.3240891 | 26 | 2.3875722 |
| 27 | 2.1164045 | 29 | 2.0777244 |
| 31 | 2.5006162 | 34 | 2.05443 |
| 36 | 1.9466478 | 40 | 2.3438962 |
| 41 | 2.3317369 | 42 | 2.2746859 |
| 43 | 2.2999447 | 47 | 2.170888 |
| 48 | 2.009368 | 50 | 2.6003548 |
| 53 | 2.9548298 | 54 | 2.3935341 |
| 55 | 2.4043114 | 56 | 2.2671271 |
| 57 | 2.4382362 | 58 | 2.2696462 |
| 59 | 2.2369352 | 60 | 2.0351454 |
| 61 | 2.0153752 | 62 | 2.7075008 |
| 63 | 2.7879454 | 64 | 2.8280592 |
| 65 | 2.8290726 | | |

Residual Sum of Squares 737.94732
Complexity: |T| = 33

Tree Pruning:

| Tree | Node | Critical Value | S(T) | \|T\| |
|---|---|---|---|---|
| T1 | 0 | 0 | 737.9 | 33 |
| T2 | 52 | 1.392e-05 | 737.9 | 32 |
| T3 | 30 | 0.001425 | 737.9 | 31 |
| T4 | 44 | 0.003296 | 738 | 30 |
| T5 | 8 | 0.005862 | 738 | 29 |
| T6 | 46 | 0.01614 | 738 | 28 |
| T7 | 49 | 0.02511 | 738 | 27 |
| T8 | 12 | 0.02647 | 738 | 26 |
| T9 | 32 | 0.03102 | 738.1 | 25 |
| T10 | 37 | 0.06508 | 738.1 | 24 |
| T11 | 35 | 0.07417 | 738.2 | 23 |
| T12 | 51 | 0.1464 | 738.3 | 22 |
| T13 | 39 | 0.1578 | 738.5 | 21 |
| T14 | 45 | 0.3055 | 738.8 | 20 |
| T15 | 10 | 0.4554 | 739.3 | 19 |
| T16 | 33 | 0.6754 | 739.9 | 18 |
| T17 | 18 | 0.818 | 740.8 | 17 |
| T18 | 15 | 0.8249 | 741.6 | 16 |
| T19 | 22 | 0.9118 | 742.5 | 15 |
| T20 | 13 | 1.07 | 743.6 | 14 |
| T21 | 38 | 1.109 | 744.7 | 13 |
| T22 | 23 | 1.582 | 746.3 | 12 |
| T23 | 9 | 1.774 | 748 | 11 |
| T24 | 28 | 2.083 | 750.1 | 10 |
| T25 | 6 | 2.712 | 752.8 | 9 |
| T26 | 14 | 2.968 | 755.8 | 8 |
| T27 | 11 | 3.882 | 759.7 | 7 |
| T28 | 7 | 8.218 | 767.9 | 6 |
| T29 | 5 | 9.479 | 777.4 | 5 |
| T30 | 4 | 15.66 | 793 | 4 |
| T31 | 3 | 20.98 | 814 | 3 |
| T32 | 2 | 35.95 | 850 | 2 |
| T33 | 1 | 115.5 | 965.5 | 1 |

T1 is the largest tree.

Test sample Impurity:

| Tree | S(T) | StDev(S(T)/N) |
|---|---|---|
| T1 | 411.5 | 0.02274 |
| T2 | 411.5 | 0.02274 |
| T3 | 411.5 | 0.02274 |
| T4 | 411.5 | 0.02274 |
| T5 | 411.5 | 0.02274 |
| T6 | 411.4 | 0.02274 |
| T7 | 411.4 | 0.02271 |
| T8 | 411.3 | 0.02271 |
| T9 | 411.2 | 0.02271 |
| T10 | 411.3 | 0.0227 |
| T11 | 411.2 | 0.02269 |
| T12 | 411.2 | 0.0227 |
| T13 | 411.5 | 0.0227 |
| T14 | 411.7 | 0.0227 |
| T15 | 411.8 | 0.02271 |
| T16 | 411.3 | 0.02271 |
| T17 | 410.8 | 0.0227 |
| T18 | 412 | 0.0227 |
| T19 | 410.9 | 0.02269 |
| T20 | 409.9 | 0.02268 |
| T21 | 409.9 | 0.02268 |
| T22 | 411 | 0.02243 |
| T23 | 411.3 | 0.02243 |
| T24 | 410 | 0.02242 |
| T25 | 411.7 | 0.02241 |
| T26 | 411.5 | 0.02241 |
| T27 | 410.9 | 0.02252 |
| T28 | 418.9 | 0.02257 |
| T29 | 420.9 | 0.02223 |
| T30 | 428.5 | 0.02224 |
| T31 | 439.8 | 0.0223 |
| T32 | 454.2 | 0.02297 |
| T33 | 517.8 | 0.0241 |

Honest Tree: Prune until node 38 (SE rule: 0 )

**Texmex Data:**
**CPS,1995. 1ST OUTGOING ROTATION GROUP.**
**BORN IN US. PARENTS BORN IN US.**
**MALE, EMPLOYED, WORKS MORE THAN 35 HOURS. CALIFORNIA, NEW MEXICO, ARIZONA, TEXAS.**
**ETHNIC: MEXICAN, WHITE NONHISPANIC.**
**5527 OBSERVATIONS: Estimation Sample: 3094 cases. Test Sample: 3094 cases.**

Variables

| | | |
|---|---|---|
| x1: | Education: | 1.Less than High School 2.High School 3.More than High School |
| x2: | Experience: | 1.1-10 2.11-20 3.21-30 4.31-40 5.+40 |
| x3: | Ethnic: | 1.White nonhispanic 2.Mexican American, Chicano or Mexicano |

Tree:

| | (Yes) | (No) | Average |
|---|---|---|---|
| Nodo 1 : x1<=2.5 | 1155 | 1939 | 2.6049518 |
| Nodo 2 : x1<=1.5 | 278 | 877 | 2.3823716 |
| Nodo 3 : x2<=1.5 | 520 | 1419 | 2.7375357 |
| Nodo 4 : x3<=1.5 | 127 | 151 | 2.1187014 |
| Nodo 5 : x2<=1.5 | 200 | 677 | 2.4659523 |
| Nodo 6 : x3<=1.5 | 423 | 97 | 2.4821661 |
| Nodo 7 : x3<=1.5 | 1249 | 170 | 2.8311173 |
| Nodo 8 : x2<=1.5 | 35 | 92 | 2.2931483 |
| Nodo 9 : x2<=1.5 | 31 | 120 | 1.9719811 |
| Nodo 10 : x3<=1.5 | 133 | 67 | 2.1999375 |
| Nodo 11 : x3<=1.5 | 501 | 176 | 2.5445387 |
| Nodo 14 : x2<=2.5 | 562 | 687 | 2.8581375 |
| Nodo 15 : x2<=4.5 | 167 | 3 | 2.6325979 |
| Nodo 17 : x2<=3.5 | 52 | 40 | 2.3935835 |
| Nodo 19 : x2<=2.5 | 42 | 78 | 2.0194799 |
| Nodo 22 : x2<=2.5 | 222 | 279 | 2.5973261 |
| Nodo 23 : x2<=3.5 | 156 | 20 | 2.3942744 |
| Nodo 25 : x2<=4.5 | 659 | 28 | 2.8811077 |
| Nodo 26 : x2<=2.5 | 92 | 75 | 2.6521559 |
| Nodo 28 : x2<=2.5 | 32 | 20 | 2.3388681 |
| Nodo 29 : x2<=4.5 | 21 | 19 | 2.4647137 |
| Nodo 31 : x2<=4.5 | 54 | 24 | 2.0611008 |
| Nodo 33 : x2<=4.5 | 247 | 32 | 2.6313505 |
| Nodo 34 : x2<=2.5 | 98 | 58 | 2.3770661 |
| Nodo 35 : x2<=4.5 | 17 | 3 | 2.5284999 |
| Nodo 36 : x2<=3.5 | 471 | 188 | 2.8834933 |
| Nodo 39 : x2<=3.5 | 61 | 14 | 2.7060273 |
| Nodo 44 : x2<=3.5 | 31 | 23 | 2.1161947 |
| Nodo 46 : x2<=3.5 | 140 | 107 | 2.6358814 |

Proyecciones de los Nodos Terminales:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12 | 2.5132 | 13 | 2.3468 | 16 | 2.0291 | 18 | 1.7881 |
| 20 | 2.2665 | 21 | 2.0679 | 24 | 2.8301 | 27 | 1.5439 |
| 30 | 1.9422 | 32 | 2.5546 | 37 | 2.825 | 38 | 2.6082 |
| 40 | 2.3693 | 41 | 2.2902 | 42 | 2.6208 | 43 | 2.2922 |
| 45 | 1.9371 | 47 | 2.5964 | 48 | 2.3729 | 49 | 2.3841 |
| 50 | 2.528 | 51 | 2.5315 | 52 | 2.8848 | 53 | 2.8802 |
| 54 | 2.6857 | 55 | 2.7947 | 56 | 2.1416 | 57 | 2.0819 |
| 58 | 2.6277 | 59 | 2.6466 | | | | |

Residual Sum of Squares: = 648.24877
Number of Nodes: 30

Pruning

| Tree | Node | Critical Value | R(T) | \|T\| |
|---|---|---|---|---|
| T1 | 0 | 0 | 0.2095 | 30 |
| T2 | 35 | 3.271e-05 | 0.2095 | 29 |
| T3 | 36 | 0.002775 | 0.2095 | 28 |
| T4 | 34 | 0.00462 | 0.2095 | 27 |
| T5 | 46 | 0.02172 | 0.2095 | 26 |
| T6 | 33 | 0.04421 | 0.2095 | 25 |
| T7 | 44 | 0.04699 | 0.2096 | 24 |
| T8 | 28 | 0.07699 | 0.2096 | 23 |
| T9 | 25 | 0.09201 | 0.2096 | 22 |
| T10 | 39 | 0.1354 | 0.2097 | 21 |
| T11 | 26 | 0.3951 | 0.2098 | 20 |

| | | | | |
|---|---|---|---|---|
| T12 | 23 | 0.4065 | 0.2099 | 19 |
| T13 | 19 | 0.4594 | 0.2102 | 17 |
| T14 | 17 | 0.7176 | 0.2107 | 15 |
| T15 | 22 | 0.7289 | 0.2109 | 14 |
| T16 | 14 | 0.8056 | 0.2112 | 13 |
| T17 | 9 | 1.319 | 0.2116 | 12 |
| T18 | 10 | 1.757 | 0.2122 | 11 |
| T19 | 6 | 2.186 | 0.2129 | 10 |
| T20 | 8 | 3.367 | 0.214 | 9 |
| T21 | 15 | 3.62 | 0.2151 | 8 |
| T22 | 11 | 5.37 | 0.2169 | 7 |
| T23 | 4 | 7.115 | 0.2192 | 6 |
| T24 | 7 | 7.612 | 0.2216 | 5 |
| T25 | 5 | 18.33 | 0.2276 | 4 |
| T26 | 2 | 25.45 | 0.2358 | 3 |
| T27 | 3 | 46.34 | 0.2508 | 2 |
| T28 | 1 | 91.31 | 0.2803 | 1 |

T1 is the largest tree.

Test Sample Impurity:

| Tree | R(T) | Standard Deviations |
|---|---|---|
| T1 | 0.2077 | 0.005154 |
| T2 | 0.2077 | 0.005153 |
| T3 | 0.2077 | 0.005155 |
| T4 | 0.2077 | 0.005155 |
| T5 | 0.2077 | 0.005154 |
| T6 | 0.2079 | 0.005159 |
| T7 | 0.2079 | 0.005163 |
| T8 | 0.2078 | 0.00516 |
| T9 | 0.2081 | 0.005181 |
| T10 | 0.2081 | 0.005181 |
| T11 | 0.2083 | 0.005188 |
| T12 | 0.2084 | 0.005197 |
| T13 | 0.2087 | 0.005212 |
| T14 | 0.2093 | 0.005211 |
| T15 | 0.21 | 0.005211 |
| T16 | 0.2102 | 0.005207 |
| T17 | 0.2106 | 0.005216 |
| T18 | 0.2114 | 0.005222 |
| T19 | 0.2122 | 0.005273 |
| T20 | 0.2136 | 0.005282 |
| T21 | 0.2148 | 0.005533 |
| T22 | 0.2168 | 0.005538 |
| T23 | 0.2194 | 0.005549 |
| T24 | 0.222 | 0.005682 |
| T25 | 0.2277 | 0.005726 |
| T26 | 0.236 | 0.005799 |
| T27 | 0.2505 | 0.005921 |
| T28 | 0.2798 | 0.00626 |

Honest tree: Prune from T1 until node 28 (SE rule: 0.05 )

## Appendix B

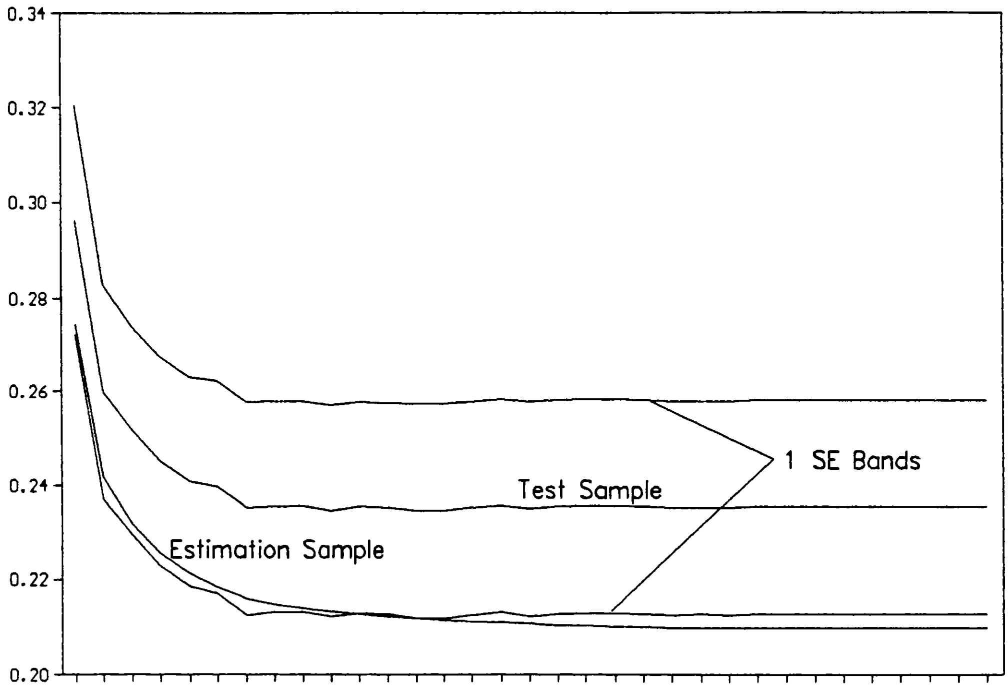


**Figure 3.1** Average residual sum of squares for the sequence of optimal subtrees. Estimation and Test Samples. Mex data.

Standard Errors are computed assuming independence in the observations

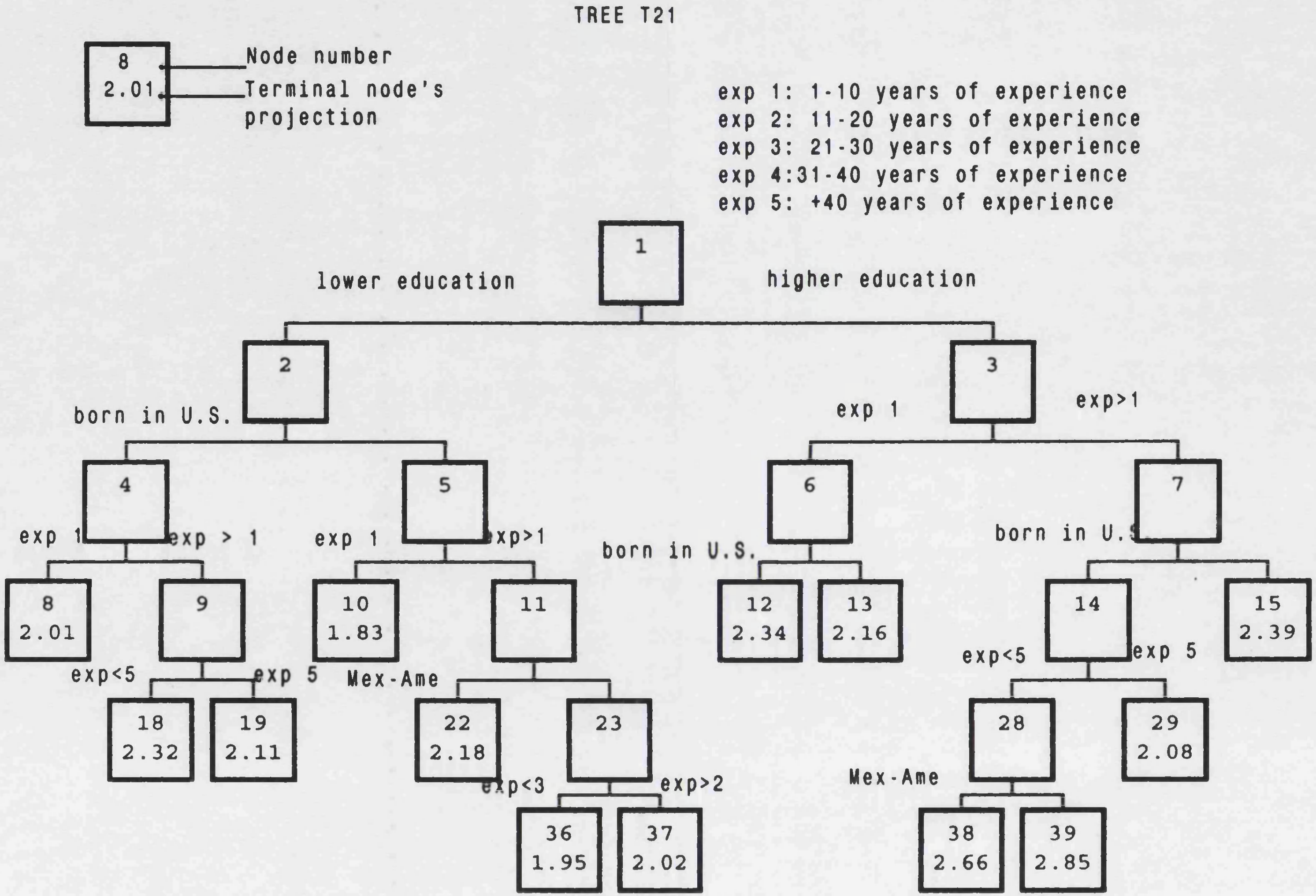


Figure 3.2 Tree T21. Mex data

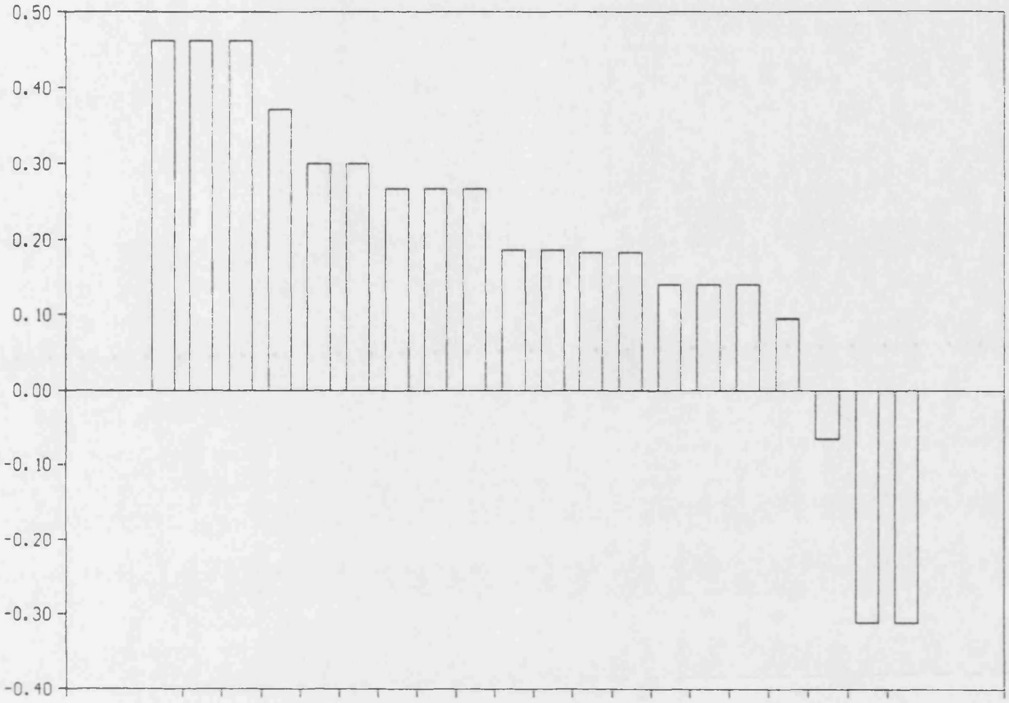


**Figure 3.3** Wage differentials projected by T21 between workers born in the U.S. and workers born in Mexico.

The horizontal axis shows types of workers ordered by the size of the wage differentials according to country of birth. There are 20 different possible comparisons: 5 experience levels x 2 education levels x 2 ethnic. The reference worker was born in the US.
Types of worker as they appear ordered in the horizontal axis:

1 Higher education, Mexicano, 11-20 years of experience
2 Higher education, Mexicano, 21-30 years of experience
3 Higher education, Mexicano, 31-40 years of experience
4 Lower education, Mexicano, 11-20 years of experience
5 Lower education, Mexicano, 21-30 years of experience
6 Lower education, Mexicano, 31-40 years of experience
7 Higher education, Mexican-American, 11-20 years of experience
8 Higher education, Mexican-American, 21-30 years of experience
9 Higher education, Mexican-American, 31-40 years of experience
10 Higher education, Mexicano, 1-10 years of experience
11 Lower education, Mexicano, 1-10 years of experience
12 Higher education, Mexican-American, 1-10 years of experience
13 Lower education, Mexican-American, 1-10 years of experience
14 Lower education, Mexican-American, 11-20 years of experience
15 Lower education, Mexican-American, 21-30 years of experience
16 Lower education, Mexican-American, 31-40 years of experience
17 Lower education, Mexicano, +40 years of experience
18 Lower education, Mexican-American, +40 years of experience
19 Higher education, Mexican-American, +40 years of experience
20 Higher education, Mexicano, +40 years of experience

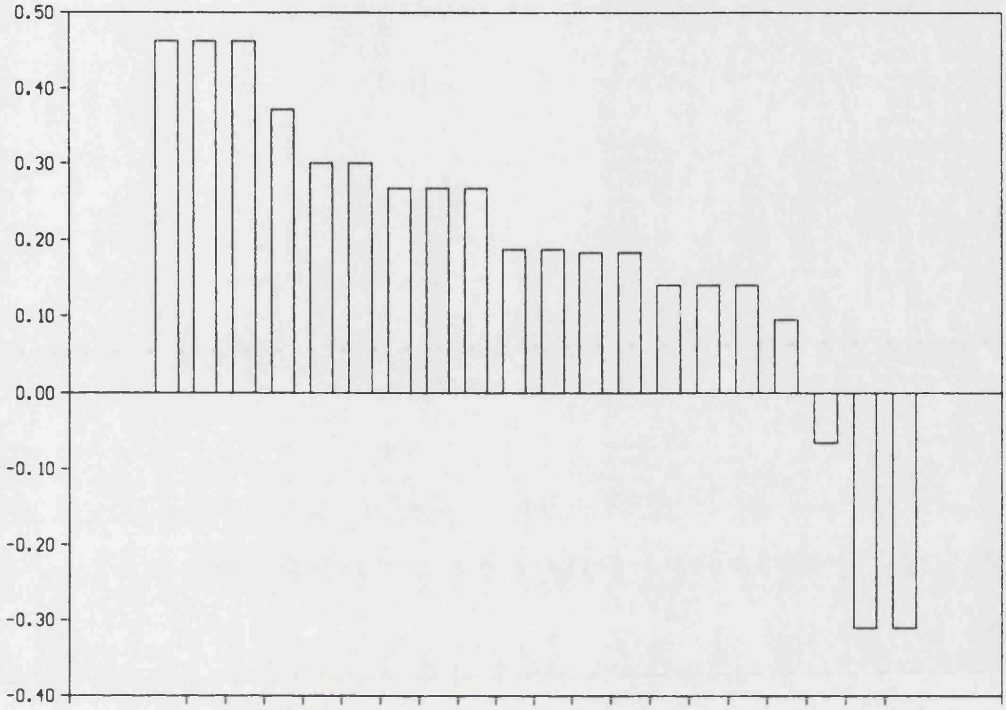


**Figure 3.3** Wage differentials projected by T21 between workers born in the U.S. and workers born in Mexico.

The horizontal axis shows types of workers ordered by the size of the wage differentials according to country of birth. There are 20 different possible comparisons: 5 experience levels x 2 education levels x 2 ethnic. The reference worker was born in the US.
Types of worker as they appear ordered in the horizontal axis:

1 Higher education, Mexicano, 11-20 years of experience
2 Higher education, Mexicano, 21-30 years of experience
3 Higher education, Mexicano, 31-40 years of experience
4 Lower education, Mexicano, 11-20 years of experience
5 Lower education, Mexicano, 21-30 years of experience
6 Lower education, Mexicano, 31-40 years of experience
7 Higher education, Mexican-American, 11-20 years of experience
8 Higher education, Mexican-American, 21-30 years of experience
9 Higher education, Mexican-American, 31-40 years of experience
10 Higher education, Mexicano, 1-10 years of experience
11 Lower education, Mexicano, 1-10 years of experience
12 Higher education, Mexican-American, 1-10 years of experience
13 Lower education, Mexican-American, 1-10 years of experience
14 Lower education, Mexican-American, 11-20 years of experience
15 Lower education, Mexican-American, 21-30 years of experience
16 Lower education, Mexican-American, 31-40 years of experience
17 Lower education, Mexicano, +40 years of experience
18 Lower education, Mexican-American, +40 years of experience
19 Higher education, Mexican-American, +40 years of experience
20 Higher education, Mexicano, +40 years of experience

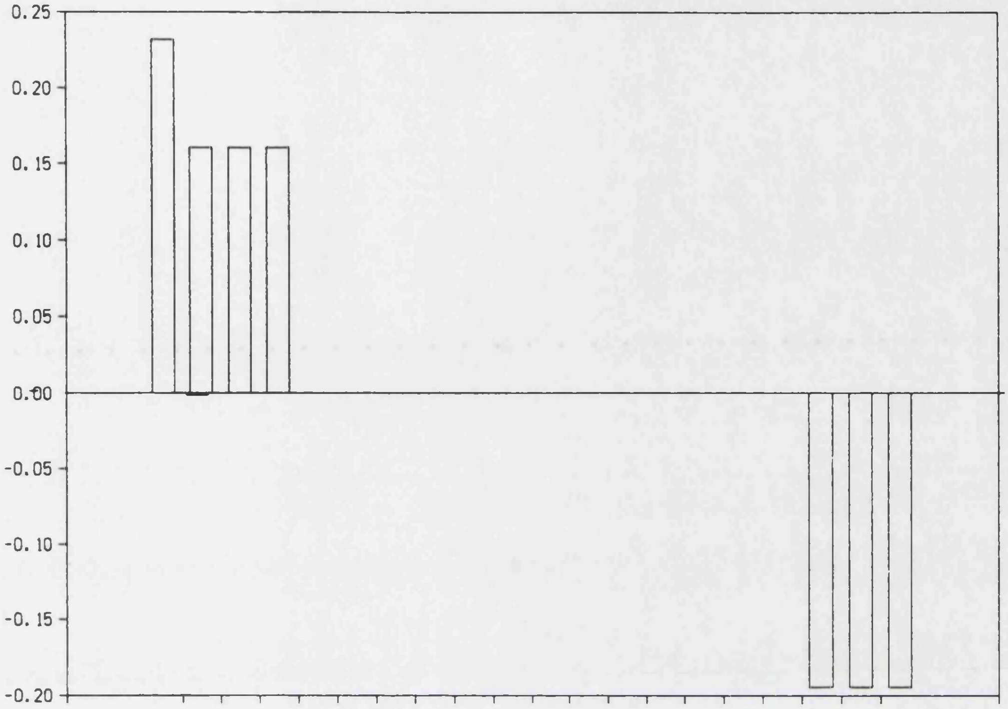


**Figure 3.4** Wage differentials projected by T21 between Mexican-Americans and Mexicanos.

The horizontal axis shows types of workers ordered by the size of the wage differentials according to ethnicity. There are 20 different possible comparisons: 5 experience levels x 2 education levels x 2 countries of birth. The reference worker declares to be Mexican American.

Types of worker with non-zero differential as they appear in the graph:

- Positive differentials:
  - Lower Education, born in Mexico, 11-20 years of experience
  - Lower Education, born in Mexico, 21-30 years of experience
  - Lower Education, born in Mexico, 31-40 years of experience
  - Lower Education, born in Mexico, +40 years of experience
- Negative differentials:
  - Higher Education, born in US, 11-20 years of experience
  - Higher Education, born in US, 21-30 years of experience
  - Higher Education, born in US, 31-40 years of experience

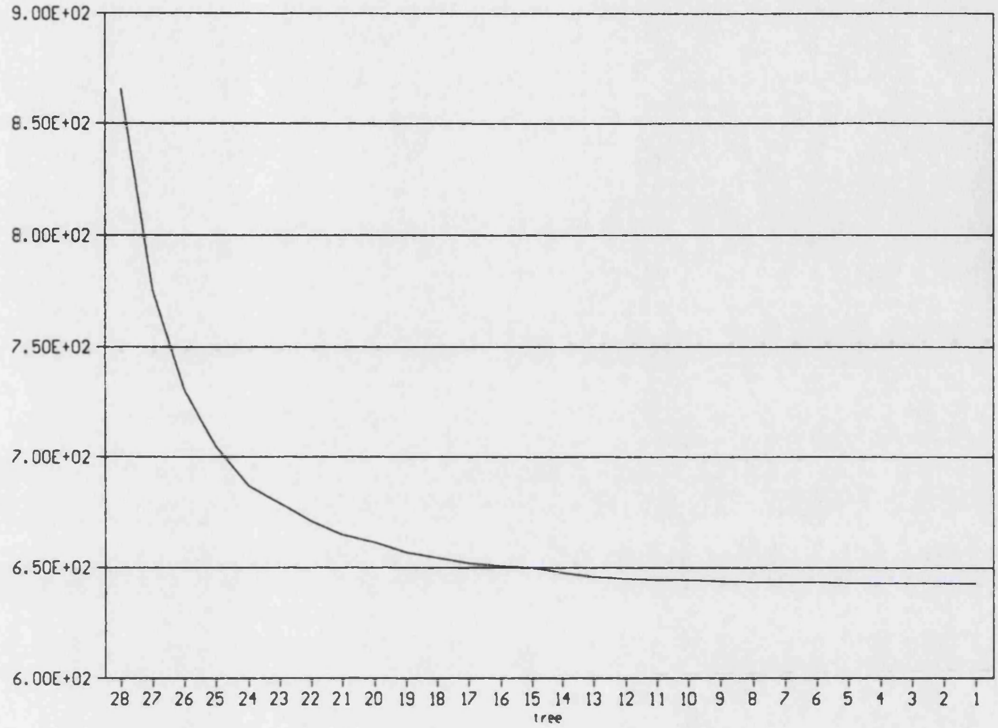


**Figure 3.5** Residual sum of squares for the sequence of optimal subtrees. Test Sample. Texmex data

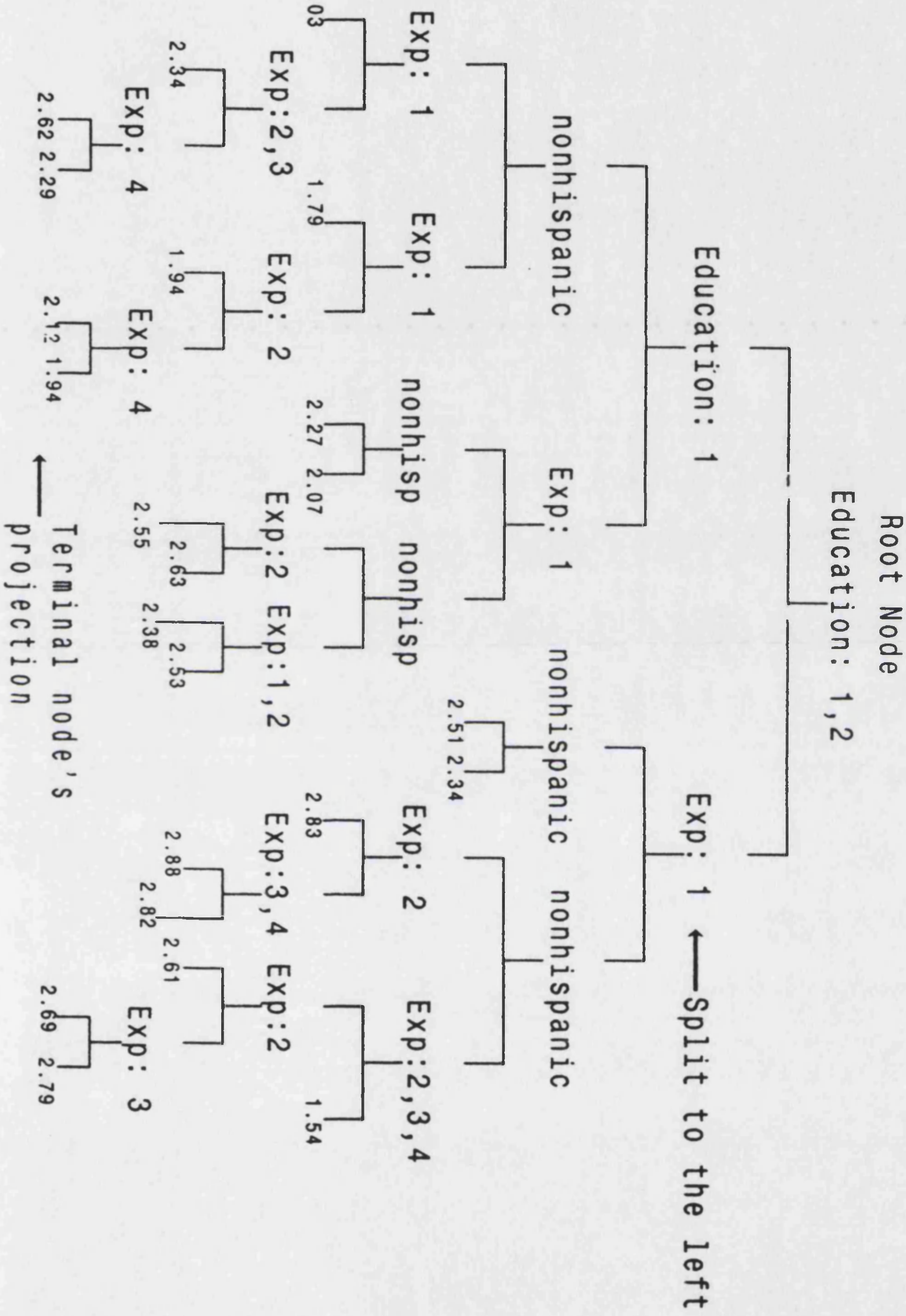


Figure 3.6 Tree T8. Texmex data

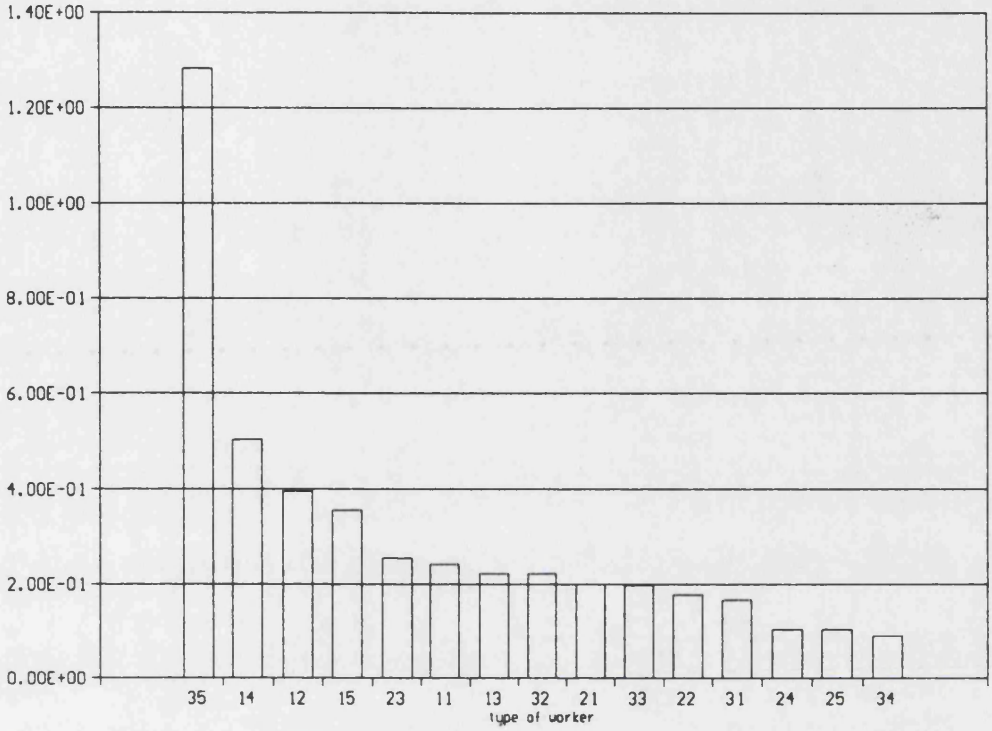


**Figure 3.7** Wage differentials projected by T8 between Mexican and White Nonhispanic

First digit in *type of worker* refers to Education, second digit refers to Experience.

The horizontal axis shows types of workers ordered by the size of the wage differentials according to ethnicity. There are 15 different possible comparisons: 5 experience levels x 3 education levels. The reference worker is white non-hispanic.

Type of worker:
35: More than High School; +40 Years of experience
14: Less than High School; 31-40 years of experience
12: Less than High School; 11-20 years of experience
15: Less than High School; +40 years of experience
23: High School; 21-30 years of experience
11: Less than High School; 1-10 years of experience
13: Less than High School; 21-30 years of experience
32: More than High School; 11-20 years of experience
21: High School; 1-10 years of experience
33: More than High School; 21-30 years of experience
22: High School; 11-20 years of experience
31: More than High School; 1-10 years of experience
24: High School; 31-40 years of experience
25: High School; +40 years of experience
34: More than High School; 31-40 years of experience

# Appendix C

Single Parametric Equation Results

**Mex Data:**
**CPS,1994,1995,1996.1ST OUTGOING ROTATION GROUP.**
**BORN IN MEXICO-BORN IN THE USA**
**MALE, EMPLOYED, WORKS THAN 35 HOURS**
**ETHNIC: MEXICAN AMERICAN, CHICANO, MEXICANO**
**5265 OBSERVATIONS: Estimation Sample:3517 cases. Test Sample: 1748 cases.**

Variables:
x1: Potential experience 1:1-10 2:11-20 3:21-30 4:31-40 5:+40
x2: Education 1: At most, High School 2: more than High School
x3: Country of Birth 1: USA 2: Mexico
x4: Ethnic 1: Mexican American 2: Mexicano or Chicano

SINGLE EQUATION OLS estimation sample

| | | | |
|---|---|---|---|
| Valid cases: | 3515 | Dependent variable: | logwages |
| Missing cases: | 0 | Deletion method: | None |
| Total SS: | 965.484 | Degrees of freedom: | 3509 |
| R-squared: | 0.219 | Rbar-squared: | 0.218 |
| Residual SS: | 753.951 | Std error of est: | 0.464 |
| F(5,3509): | 196.901 | Probability of F: | 0.000 |

| Variable | Estimate | Standard Error | t-value | Prob >\|t\| | Standardized Estimate | Cor with Dep Var |
|---|---|---|---|---|---|---|
| CONSTANT | 1.673534 | 0.056325 | 29.712364 | 0.000 | --- | --- |
| EXP | 0.351444 | 0.030059 | 11.691782 | 0.000 | 0.760256 | 0.100941 |
| EXP2 | -0.051551 | 0.005530 | -9.322746 | 0.000 | -0.605373 | 0.068913 |
| EDUCA | 0.363401 | 0.019655 | 18.488641 | 0.000 | 0.294851 | 0.345905 |
| COUNTRY | -0.230939 | 0.023108 | -9.993865 | 0.000 | -0.216901 | -0.324550 |
| ETHNIC | -0.048701 | 0.022902 | -2.126544 | 0.034 | -0.044958 | -0.259945 |

SINGLE EQUATION OLS ALL OBSERVATIONS

| | | | |
|---|---|---|---|
| Valid cases: | 5262 | Dependent variable: | logwages |
| Missing cases: | 0 | Deletion method: | None |
| Total SS: | 1482.842 | Degrees of freedom: | 5256 |
| R-squared: | 0.214 | Rbar-squared: | 0.213 |
| Residual SS: | 1165.341 | Std error of est: | 0.471 |
| F(5,5256): | 286.402 | Probability of F: | 0.000 |

| Variable | Estimate | Standard Error | t-value | Prob >\|t\| | Standardized Estimate | Cor with Dep Var |
|---|---|---|---|---|---|---|
| CONSTANT | 1.666944 | 0.046957 | 35.499059 | 0.000 | --- | --- |
| EXP | 0.326815 | 0.024770 | 13.194033 | 0.000 | 0.701631 | 0.098411 |
| EXP2 | -0.046764 | 0.004522 | -10.341912 | 0.000 | -0.549416 | 0.068020 |
| EDUCA | 0.378114 | 0.016380 | 23.083655 | 0.000 | 0.300420 | 0.347497 |
| COUNTRY | -0.230654 | 0.019121 | -12.062930 | 0.000 | -0.213030 | -0.317615 |
| ETHNIC | -0.044363 | 0.019103 | -2.322338 | 0.020 | -0.040283 | -0.257735 |

Chapter 4

---

# An application of regression trees to the analysis of the evolution of the U.S. wage structure since 1980

## 4.1 Introduction: Relative growth performance and dispersion changes in wages

In this chapter I aim to apply regression trees to study two related empirical problems in wage structures. First, I will look at the relative growth performance of male workers' real wages since 1980 in the U.S. labor market and study the role that several economic and social variables had on it. Second, I will analyze wage dispersion and its evolution also from 1980 onwards.

The analysis of relative growth performance may seem a rather simple task. Consider the problem of comparing the evolution of the wages of two workers in two consecutive periods. Growth rates will give us all we need to know since by looking at their relative growth rates, we will be obviously able to ascertain whose wage increased the most. Also, we will be able to say whether their wages are converging or diverging in the sense that their absolute difference decreased or increased during the period. If the higher wage experienced higher growth, then there was increased dispersion. If the lower wage increased the most, then either convergence or divergence with overtaking took place.

However, the problem of relative growth performance becomes more subtle when more than two agents or more than two periods are considered.

For example, with three periods and two agents, suppose that the growth in the variable for the first agent was positive in the first period and stagnant thereafter while the reverse was true for the second agent. Then, it is not clear to what extent the overall growth rate is a good descriptor of both growth patterns. The basic idea is that an average growth rate does not entirely describe what happened since only information at the beginning and at the end of the time span is considered.

Relative performance and dispersion are also more complicated to study when more than two agents are included in the analysis.

What matters is the evolution of the distribution of the variable across agents and an overall measure of dispersion may again destroy too much relevant information.

A simple strategy that may succeed in overcoming these potential problems is to study the nature of the data and propose general reasonable patterns that these data should satisfy. Sometimes, this will be enough to allow us to judge whether indicators such as average growth rates and standard deviations are meaningful. In other words, it may be possible to restrict the data-generating process to a statistical model that gives structure both in the cross-section and the time series dimensions of the problem and also allows us to judge the properties of one-dimensional indicators of multidimensional phenomena.

Following this introduction, I will present an econometric model that I understand is well suited to study these dynamic issues for large cross-sections of relative real wages. Namely, the dynamic index model with one latent trend. It will be further assumed that the parameters of the model have a nonparametric tree structure. The advantage of this approach lies in that it allows us to have clear-cut definitions of what we understand by growth performance and dispersion in real wages. I will argue that the model is very general indeed and reflects well our understanding of wage structures.

In the following section, I will describe alternative indicators of growth and inequality and comment on their statistical properties when they are applied to a sample generated by a dynamic index model with one latent variable. In the fourth section of the chapter, empirical results on the evolution of wages are presented and compared to results in the literature. Next, empirical results of wage dispersion are given. The chapter ends with a reminder of the main conclusions.

## 4.2 Random fields and dynamic wage structures

### 4.2.1 Dynamic wage structures

I have in previous chapters assumed that the labor market is fragmented and consists of many sectors, each with an equilibrium wage. For any worker of type **x**, her log wage is:

$$w_i = f(\mathbf{x}) + v_i \quad . \tag{4.1}$$

In this chapter, I will assume that $v_i$ is independent of any of the regressors although the techniques developed could be generalized to more realistic assumptions. This model is clearly static and we need to enlarge it to study dynamic phenomena in the labor market. We assume that both the structural component and the idiosyncratic component may evolve through time. Then, the log wage of any worker, i, of group **x** at any period, t, will have the form:

$$w_{it} = f_t(\mathbf{x}) + v_{it} \quad . \qquad (4.2)$$

We can further assume that at any point in time the wage structural component $f_t$ can be described by means of a tree plot.

If the vector of characteristics, **x**, were constant in time for any worker, equation **(4.2)** would also describe the evolution of wages for individuals. In this chapter, **x** may include time dependent variables for individuals, such as experience, so that **(4.2)** is only the evolution of the representative worker's wage.

I will use this general formulation to study wage dispersion trends within groups since only the idiosyncratic term effect differs across individuals of the same group. As suggested in the introduction, an overall analysis of wage dispersion may be misleading. Even if we control for variables that affect expected wages, a parametric approach may not be appropriate. The reason is that changes in the structural component in **(4.2)** will also affect the distribution of log wages. In section 4, I will use this nonparametric framework to study the role of structural change in the evolution of wage dispersion.

An important property of **(4.2)** is that wages are double-indexed. In other words, if we conceptualize the observed wage as an observation from a stochastic process, then it is neither a cross-section point, nor a time series realization, but a random field observation. In other words, the data may show structure not only along the cross-section dimension but also along the time dimension. The model that I consider in this chapter is a random field with a nonparametric specification for the cross-section and a parametric specification for the time series. The main methodological strategy consists of choosing among all reasonable models for wages the family that incorporates the least structure so that no relevant features of growth and wage inequality are lost in the process.

4.2.2 A dynamic index model for dynamic wage structures

The model proposed here is a restricted version of the dynamic index model[1]. Three features of this model make it interesting in the analysis of the evolution of the wage structure:

(a) Changes in the individual specific components only

---

[1] See Quah and Sargent (1993). Their work is an extension of the common factors model to nonstationary models where the statistical techniques to estimate the model are robust orthogonality conditions that qualify as quasi-maximum likelihood estimators. See Lawley and Maxwell (1963) for an introduction to factor analysis. For a dynamic stationary factor analysis model, see Geweke (1977). Engle and Kozicki (1993) set up a general framework in which cointegrated models and common factor models appear as special cases.

have temporary effects on wages. This is a self-evident desirable property for the synthetic worker model.

(b) Trends in wages will be the result of the interaction of two types of factors:

(b1) Global factors: They affect all wages. A typical example is the technological level of the economy. By its own nature, this factor is dynamic and cumulative or integrating. All new inventions are added to the stock of knowledge for ever. This is a powerful source of trending behavior.

(b2) Local characteristics: The global factors will affect each type of worker in a different way. Local characteristics may allow some workers to take more advantage of new changes in the global conditions of the labor market. These effects are by definition constant within each class.

(c) Wages tend to be log-proportional between groups: In the absence of new permanent and temporary disturbances, relative growth of observed wages will be proportional. In other words, workers who are doing better in terms of growth in wages will keep on doing better to the same extent in the steady-state equilibrium.

Feature (c) is very restrictive. To see why, consider two types of workers, $\mathbf{x}_1$ and $\mathbf{x}_2$. For simplicity, suppose that their real log wages follow deterministic trends:

$$w_{1t} = a_1 t + a_2 t^2 \qquad (4.3)$$
$$w_{2t} = b_1 t,$$

where $w_{it}$ denotes the real log wage of worker i. Only if $a_2=0$ will relative log wages be constant.

Why should we disregard $a_2>0$ for our analysis? If $a_2>0$ growth in wages in the second group will be increasing while growth in the first group will be constant. Relative growth performance will therefore not be constant along the period. If $a_1<b_1$, then it will make sense to split the time span of our sample in two periods, the splitting point in time being $t^*=(b_1-a_1)/2a_2$. The fundamental problem in **(4.3)** is of course the fact that growth in the first group has two independent sources. By assuming just one source of long-term growth, we impose long-term constant relative growth performance in observed wages. This will allow us to talk of a single relative growth performance indicator for the entire period.

As already stated, the model is a restricted version of the dynamic index model. As in the original model, the data results from the interaction of both a common source of shocks and idiosyncratic shocks. In contrast to the original model, the common source of shocks is constrained to be a single factor.

The data set is a realization of a random field. Workers belong to groups within which the structural component in the wage is equal. The number of groups in the sample is unknown. Following Quah and Sargent(1993), assume that $\{w_{it}\}$ is an observed segment of a random field that satisfies the relation

$$w_{it} = a_x \star u_t + v_{it}, \qquad \textbf{(4.4)}$$

where $\star$ denotes convolution. The source of the permanent component on $w_{it}$, $u_t$, is a scalar unobserved nonstationary process, common to all elements. The idiosyncratic component, $v_{it}$, is the idiosyncratic term with zero mean and possibly nonstationary variance. It is further assumed independent of any stationary transformation of the stochastic component in $u_t$.

The transference filter operator, $a_x$, is assumed to have all roots outside the unit circle and to be such that log wage variance conditional on the initial value $u_0$ is finite.

Expression **(4.4)** is a general model that incorporates as special cases well-known examples of structural time series models. The simplest case is the so-called deterministic polynomial trend model in which the trend takes a polynomial functional form in time.

When this form is linear, we have the linear deterministic trend model:

$$w_{it} = d_x \cdot (1 + b \cdot t) + v_{it}, \qquad \textbf{(4.5)}$$

with $d_x$ and b real numbers.

In another family of models stemming from **(4.4)** the common factor, $u_t$, has a stochastic trend component. For example, if $u_t$ is an unobserved random walk and $v_{it}$ is a noise term that is uncorrelated with any stationary transformation of $u_t$ at all leads and lags, the trend will have a stochastic component and the common variance in the cross-sample will originate in the trend. The model for the trend would be:

$$\begin{aligned} u_t &= u_{t-1} + d + e_t \\ \operatorname{cov}(v_{it}, e_s) &= 0, \ \forall\, t,s, \end{aligned} \qquad \textbf{(4.6)}$$

with $e_t$ random noise with variance one. If the transference operator $a_x$ is a real number, then we have the one common factor version of the common factor model (Harvey, 1989, page 450).

A fully dynamic version of the common factor model is obtained by relaxing some assumptions. We can call this the dynamic factor model, and it is a nonstationary version of common factor models as in Geweke (1977).

The model takes the form

$$w_{it} = a_x(B)\, u_t + v_{it}\,, \qquad \textbf{(4.7)}$$

with real nonnegative $a_x(1)$ and with $a_x(B)$ a polynomial in the lag operator. For the trend, the model is:

$$(1-B)\cdot g(B)u_t = d + e_t \qquad (4.8)$$
$$\text{cov}(v_{it}, e_s) = 0, \; \forall\, t,s,$$

where g(B) is a lag polynomial with g(0)=1 and $|g(z)|\neq 0 \;\forall |z|\leq 1$.

Model **(4.4)** encompasses these and other model specifications. They all share an important feature. Relative long-term growth performance for each element of the cross-section of the random field depends on $a_{\mathbf{x}}(1)$. Heuristically, if $a_{\mathbf{x}}(1)>a_{\mathbf{y}}(1)$, for any two groups of workers, **x** and **y**, then the effect of the sources of growth are larger for group **x** than for group **y**.

A second interesting property of model **(4.4)** is that it admits pairwise projections between the elements such that relative growth performance can be studied from estimates of the filters. More precisely, since

$$E[w_{it}|\, w_{jt}] = d_{\mathbf{xy}}(B)w_{jt} \qquad (4.9)$$

where

$$d_{\mathbf{xy}}(B) = \frac{a_{\mathbf{x}}(B)}{a_{\mathbf{y}}(B)}, \qquad (4.10)$$

then the coefficients in $d_{xy}$ can be consistently estimated by a least squares estimation. For the dynamic factor model **(4.7)** and **(4.8)**, the variables are pairwise cointegrated[2] so that pairwise

[2] See Appendix A.

cointegrating regressions will also give consistent estimates.

### 4.3 A review of trend estimators

In this section I present a brief account of the properties of simple estimators of trends in dynamic index models. Since the objective is to obtain estimates of relative growth performance, I will only review estimators for $d_{xy}$. I will consider estimators for the special cases introduced in the previous section for the brevity of exposition, but the arguments can be extended to more general models, as pointed out at the end of the previous section.

If $w_{it}$ follows the linear deterministic trend model we can carry out OLS regression of the wage on the linear trend and on any linear combination of the other wages. The estimators converge $O(T^{-1})$ to the parameter of the structural model under very general assumptions for $v_{it}$. The estimators are superconsistent since OLS minimizes the error sample variance and only the correct parameter will give sample errors that do not explode in variance.

Although the model is more complex, this is also the basic reason for the consistency result in the cointegrating regression. Assume that $w_{it}$ and $w_{jt}$ are cointegrated with a cointegrating vector $(1,-d_{ij})$. Under general regularity conditions[3] we have that $T^{1-\delta}(d_{OLS}-d_{ij})\rightarrow 0$ for $\delta>0$, where $a_{ij}$ is the OLS slope estimator of regressing $w_{it}$ onto $w_{jt}$ plus a constant.

Thus, pairwise cointegrating regressions give consistent estimates for the normalized cointegrating vectors in the dynamic factor model.

Maximum Likelihood estimation in the dynamic factor model is accomplished via the EM algorithm[4]. Consider the reduced MA form of the dynamic factor model, $z_t=C(B)\xi_t$, where $z_t=\nabla w_t-E[\nabla w_t]$, $C(0)=I_N$, $\Sigma_\xi$ is p.d. and $C(1)$ is of rank 1. The maximum likelihood estimators of the MA parameters are consistent as T goes to infinity[5]. Consistency of the structural parameters just

---

[3] See Stock (1987).

[4] See Watson and Engle (1983), Watson and Kraft (1983), Engle and Watson (1981) and Quah and Sargent (1993) for applications of the EM algorithm to the dynamic factors model. At every step, this method involves the maximization of the expected likelihood of the latent data. Under general conditions, this iterative procedure will lead to a fixed point which is also a local maximum of the likelihood. See Ruud (1991) for an introduction to the algorithm and Wu (1983) for convergence results.

[5] See Dunsmuir and Hannan (1976). No central limit theorems are available as the spectrum density matrix of the process is

follows identification[6]. ML estimation is even more relevant when the interest lies on the estimation of the unobserved common index, as in Quah and Sargent (1993). When all that is needed is an indicator of growth performance, then simple robust estimators may be worth studying. Notably, ratios of average growth give consistent estimates of $d_{xy}$ when the common trend is stationary after differencing and when the trend is deterministic.

---

not of full rank at frequency zero. Indeed, we cannot use Dunsmuir and Hannan's law of large numbers for general processes to check consistency of the Quasi-maximum likelihood estimator of the structural parameters since the conditions are not satisfied. Check condition B4 in Dunsmuir and Hannan (1976).

[6] The identification problem for the dynamic factor model in stationary variables has been discussed by Geweke and Singleton (1981). As in the conventional factor model, necessary and suficient conditions for the identification of the model are unknown. As stressed by Geweke (1993), additional identification problems are introduced if the common trends are non-stationary, since there are several ways to represent the stationary transforms of the common trends as a linear combination of serially uncorrelated processes. This situation, however, does not necessarily arise in the model of a single non-stationary common factor. For example, Hotta (1983) studies the univariate case and finds neccesary and sufficient order conditions.

As pointed out at the end of the previous section, pairwise OLS regressions also lead to consistent estimators. For small samples, however, estimates will depend on the normalization chosen. Thus, in practice, we will have several estimates of relative growth performance.

A solution to this problem consists of adding a further step to the estimation procedure. Let there be N different workers. Note that we can always carry N-1 pairwise OLS regressions of any $w_{it}$ on any other group's wage. This will give us N-1 parameter estimates for each $w_{it}$. Define $a_i=(a_{1i},\ldots,a_{ii},\ldots,a_{Ni})'$ where $a_{ji}$ is the consistent estimate resulting from regressing $w_{jt}$ on $w_{it}$, $\forall j\neq i$ and $a_{ii}=1$. Also let $d_i$ be $a_i(1)$.

Since plim $a_{ji}=d_j/d_i$, $d_i\cdot(\text{plim } a_i)=d$ where $d=(d_1,\ldots,d_N)'$. If we define $M_a= \sum_i(a_i \; a_i')$ and normalize so that, without loss of generality, $\sum_i(d_i)^{-2}=1$, we get plim $M_a=d\cdot d'$. Thus, as T goes to infinity, the moment matrix $M_a$ is of rank one, its eigenvector being the parameter vector d.

The finite sample eigenvector associated with the largest eigenvalue in $M_a$ is the least squares estimator of d in the latent variable model:

$$\text{plim } a_i = d\, u_i, \qquad \textbf{(4.11)}$$

where $1=d_iu_i$. Least squares is defined as:

$$\min_{d,u} S_N = \sum_{i=1}^{N} \left(a_i - du_i\right)'\left(a_i - du_i\right). \qquad \textbf{(4.12)}$$

First order conditions are:

$$\begin{aligned} \frac{\partial s_N}{\partial d} &= -2 \sum_{i=1}^{N} u_i (a_i - du_i) = 0 \\ \frac{\partial s_N}{\partial u_i} &= -2d'(a_i - du_i) = 0, \quad i=1,\dots,N. \end{aligned} \qquad \textbf{(4.13)}$$

Then $u_i=(d'a_i)$ and $d=(\sum u_i^2)^{-1}(\sum u_i a_i)$. Combining these two conditions, we have

$$\left(d'M_a d\right)d = M_a d. \qquad \textbf{(4.14)}$$

If $\lambda_v \neq 0$ is the largest eigenvalue of $M_a$ and v its corresponding normalized eigenvector, $\lambda_v v = M_a v$, so that $\lambda_v = v'M_a v$. Therefore, $d_{LS}=v$ and $u_i=1/v_i$ is a solution to the first order conditions. Let us call v the latent estimator.

To sum up, we can transform the parameter estimates of all possible pairwise regressions so that all vectors resulting from different normalizations are on the same one-dimensional space. For samples with unbounded T, matrix $M_a$ is of rank one and its unique, up to normalization, nonnegative eigenvector is the growth performance indicator.

In the following section I apply the latent indicator to the study of growth performance in wage structures, therefore departing from the common practice of directly using average growth rates.

As average growth rates, the latent estimator is a simple robust estimator. The fundamental advantage of this estimator of growth performance results from its exploiting cross-section information as much as time series variation[7].

**4.4 Trends in U.S. real wages**

It is generally agreed that, during the eighties, there was a slow down of growth in real wages. There were also increases in the wage differentials between workers with college and high school education. Finally, experience differentials continued a long-term increasing trend, while race differentials remained stable[8].

---

[7] Studying consistency results as N goes to infinity is however beyond the scope of this empirical study. To obtain a taste of this advantage in small samples, I carried out a very simple simulation exercise for two models. Results are presented in Appendix B.

[8] The literature on the evolution of wages in the U.S. is indeed vast. See, amongst others, Levy and Murname (1992), Bound and Johnson (1992), Katz and Murphy (1992), and Murphy and Welch (1992) for overall evaluations. See Ashraf (1994) for a study of trends in white-black earnings differentials and Buschinsky (1994) for a study of the changes in the wage structure using

Were there nonlinear features in the trend behavior of U.S. wages since 1980? The objective in this section is to judge to what extent nonlinear effects of workers' characteristics were prevalent among the basic features of wage trends since 1980.

### 4.4.1 Description of the data

I use the extracts of the Annual Earnings Files of the Current Population Survey (CPS) prepared by the NBER for the years 1980, 1985, 1990, and 1995. Each individual can only be interviewed at most twice and in two consecutive years, so that no individual observation is repeated in this data set.

The universe is reduced to male adults working more than 35 hours every week, and employed in any industry group but agriculture who live in either metropolitan or non-metropolitan areas.

Wages here are the logarithm of earnings per week divided upon hours per week at the job and deflated by wage inflation. To distinguish among different representative workers, I consider six characteristics:

---

quantile regression. On inequality, see, for example, Juhn, Murphy and Pierce (1993), Borjas and Ramey (1994), Topel (1994), and Bound and Johnson (1995).

1. Education: I classify education into three categories. lower education, some college, and higher education. Lower education includes all those workers who at most completed High School. Some college refers to those workers who started some form of higher education but did not finish it. Finally, higher education refers to workers with a higher education degree.

2. Experience: This is an index of potential experience, since the sample does not give a direct measure of the workers' experience. The usual procedure is used here: age-education-six. Then, individuals are divided into five categories of experience. The first category includes those individuals with less than ten years of experience. The second are those with no less than 10 and less than 20, and so on. The fifth group includes individuals with more than forty years of experience.

3. Region: This variable segments the labor market into four geographical regions. South includes the East South Central division, the South Atlantic division except Delaware, Maryland and D.C., and Arkansas and Louisiana. East includes the New England division, Middle Atlantic, and D.C., Delaware, and Maryland. Middle West includes all remaining states from central divisions, and finally the west includes Mountain and Pacific. This grouping was chosen to obtain an ordered variable with geographical and historical intuition, starting from South and ending in West.

4. Type of worker: This variable refers to whether the worker is a blue collar or a white collar worker.

5. Industry: This is another binary variable that describes whether the job is in a goods-producing industry or in a services-producing industry.

6. Race: workers are either white or black workers.

This variable specification segments the labor market in at most 480 markets. However, due to the sample size and the fact that the sample design did not contemplate surveying all these groups, only 334 of all possible cells are represented in all years so that two samples, estimation and test, can be extracted.

The overall sample is then randomly divided into two samples of sizes of 1/2 each. Within each subsample, individuals' observations are transformed into 'groups' observations by averaging across individuals with the same demographic characteristics. This is done for each single year. I will refer to this data as the "trends" data.

The need for grouping observations arise from the fact that individuals are not interviewed in different years. If groups are not weighted for the number of observations within each group, the results from regression trees may depend on the variable specification chosen. However, an advantage of carrying unweighted regression trees is that poorly represented groups will come soon out in the splitting algorithm.

### 4.4.2 Empirical results

Growth indicators were computed for both samples. Figures 4.1 and 4.2 in Appendix B plot latent growth indicators against average growth rates for the entire period. Although the plots indicate a positive correlation between both indicators, the relation is clearly less than perfect.

Figures 4.3 and 4.4 in Appendix C show histograms for latent growth indicators both in the estimation and the test sample. They can be thought of as the effect of the overlapping of several sample distributions of the growth estimators. Groups with the same long-term growth performance will have their growth sample indicators around their true value.

A nonparametric tree structure for growth patterns is assumed. A tree will partition the input space so that the overlapping of distributions will be unveiled. Figure 4.5 in Appendix C plots within-nodes sum of squares for the sequence of optimal trees both for the estimation and the test sample. The minimum value for the test sample residual sum of squares is achieved at tree T182, with only seven different groups. Figure 4.6 in Appendix C shows T182.

The estimation sample residual sum of squares for tree T182 was 0.4171. This contrasts with the value for the simpler one single group model, 0.6528. The test sample validates these results. The residual sum of squares takes now the value 0.4349. These values imply that the coefficient of determination for the estimation sample is approximately 36% while the test sample value drops to 30%.

We can now study the structure of tree T182 with the help of Figure 4.6 in Appendix C. Nonlinearities exist concerning the interaction of variables such as experience, race, and education. For higher education workers with more than 40 years of experience, growth shows poor performance. However, performance was good for more experienced black workers with lower education whose type of occupation fell into the category of blue collar workers.

Other interactions are worth mentioning. In general, blue collar workers are associated with better performance than white collar workers. The exception is black workers with less than 40 years of experience.

Education splits the sample into higher and lower performance workers. In general, workers with lower education, including those with some college, had worse growth in real wages than those with higher education. For higher education workers, nodes 12 and 13 are the more numerically important. Education seems the most important factor in explaining relative growth performance in the labor market in the last 15 years. However, experience, type of worker, and race also help in explaining local growth performances.

Neither geographical nor industrial factors seemed to have influenced growth performance. This result corroborates the idea that whatever the wage differentials between workers from different large regions and different aggregate sectors were, they have, on average, remained the same.

In order to have a better understanding of the growth experiences of the seven different groups, Table 4.1 gives basic statistics for real wages of the workers in each of the seven groups for each of the four years.

The groups are presented in ascending order according to their average latent growth indicators. This gives the opportunity to understand the relation between Figure 4.1 in Appendix C and tree T182. Terminal nodes are of two general types. First there are the marginal small groups formed, for example by workers with more than forty years of experience. The number of individual cases of these types of workers is very low indeed. Their trend behavior is nonetheless well isolated by the growth indicators.

The second type of node is obviously more interesting in the sense that it involves bigger segments of the labor market. Two nodes, terminal nodes 4 and 10, are particularly important, since they account for around two thirds of the entire sample in each year.

Node number 4 represents lower education, white collar workers. Table 4.1 suggests a nearly linear fall in their real wages at an annual rate of 1.1%. Node number 10 represents lower education, white, blue collar workers. Table 4.1 shows how differently their average real wage evolved from that of the equivalent white collar group.

They first experienced a steady increase in their average real wage. During the eighties, it grew at an annual rate of 1.1%. However, this gain was lost between 1990 and 1995. Overall, average growth in real wages was near zero.

Black workers with the same characteristics also experienced some real gains from 1980 to 1990, but their losses from 1990 to 1995 were greater, so their growth performance was among the worst ones. Their average annual real growth was -0.6%. It is this important fall from 1990 to 1995 that made their indicator of growth performance to be worse on average than that of lower education, white collar workers.

Blue collar workers with higher education also suffered some real loss in their wages for the entire period. A very small one

**Table 4.1** Main growth experiences from regression trees. Latent growth indicator. Trends data.1980-1995.

Data: CPS 1980-85-90-95. Male, full-time,non-farm workers. Individual observations.

| | | Nodes in tree T182[1] | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Year | | 7 | 22 | 4 | 12 | 10 | 13 | 23 |
| 1980 | No.Obs.[2] | 36 | 635 | 40317 | 1191 | 6646 | 1672 | 32 |
| | Real wage | 1.84 | 1.77 | 1.87 | 1.96 | 1.88 | 2.03 | 1.89 |
| | Std.Dev. | 0.48 | 0.40 | 0.35 | 0.43 | 0.41 | 0.44 | 0.34 |
| 1985 | No.Obs. | 24 | 627 | 33084 | 1176 | 6312 | 1762 | 29 |
| | Real wage | 2.23 | 1.81 | 1.86 | 1.97 | 1.89 | 2.09 | 1.84 |
| | Std.Dev. | 0.52 | 0.43 | 0.45 | 0.45 | 0.44 | 0.42 | 0.50 |
| 1990 | No.Obs. | 183 | 1655 | 43212 | 5495 | 26251 | 7632 | 63 |
| | Real wage | 2.11 | 1.79 | 1.72 | 1.98 | 2.02 | 2.32 | 1.67 |
| | Std.Dev. | 0.74 | 0.48 | 0.46 | 0.52 | 0.51 | 0.49 | 0.53 |
| 1995 | No.Obs. | 203 | 1059 | 30111 | 4657 | 11407 | 18023 | 44 |
| | Real wage | 2.04 | 1.68 | 1.70 | 1.92 | 1.88 | 2.20 | 1.94 |
| | Std.Dev. | 0.73 | 0.49 | 0.51 | 0.49 | 0.53 | 0.50 | 0.53 |

Note:
[1] Tree 182 refers to the tree with 0-SE rule and 1 as stop-splitting rule obtained with dependent variable the latent growth indicators after pooling individual observations according to: potential experience -five levels-, education -three levels-, region -four large areas-, type of worker -binary-, industry -binary-, and race. Each node represents a region in the independent variables space:

7: Higher Education; Experience>40 years.
22: Lower Education; Experience<40 years; Black; Blue collar.
4: Lower Education; White collar.
12: Higher Education; Experience <40 years; White collar.
10: Lower Education; White; Blue collar.
13: Higher Education; Experience <40 years; Blue collar.
23: Lower Education; Experience>40 years; Black; Blue collar.

[2]No.Obs. is the number of individual observations in each erminal node. Real wage is the average logwage within each terminal node. Std.Dev. is the standard deviation in each node.

indeed, at an annual rate of 0.3%. This decline concentrated in the last five years, since there had been gains, although very modest, during the first 10 years of the period.

The best performing large group, node 13, was the only one that experienced average gains in their real wages. Again, there were losses in the last five years, but these were small compared to the gains that had taken place from 1980 to 1990. Overall, real wages for higher education, blue collar workers with less than 40 years of experience increased at an annual rate of 1.1%.

Tree T182 suggests a linear structure for the growth experience of white workers with less than forty years of experience. Table 4.2 gives the results of linear parametric regression for these workers. The results show that for this group of workers, blue collar and higher education workers had always better performance than white collar and lower education workers.

Note, however, that linearity does not extend to a wider segment of the labor market, although the number of groups in tree T182 is only seven.

For each year, I have also computed the standard error within each group to study whether these trend differentials were the only source of changes in the dispersion of wages.

The results can be found in Table 4.1. For all groups the standard deviation at the end of the period was higher than in 1980. For some groups, this increase represented around 50% of the original level of within groups dispersion.

**Table 4.2** Growth structure for white workers with less than 40 years of experience. Trends data.

(A) Tree T182 Projections of latent growth indicators:

| | Lower education[1] | Higher education | Differential |
|---|---|---|---|
| White Collar | -0.034 | -0.0092 | 0.0248 |
| Blue Collar | 0.0051 | 0.0356 | 0.0305 |
| Differential | 0.0391 | 0.0448 | |

(B) Linear regression results:

Data: Latent growth indicators for white workers with less than 40 years of experience
Independent Variables: region -four large areas-, experience -five groups-, education -binary-, type of worker -binary-, and industry -binary.

Valid cases: 189
R-squared: 0.424

| Variable | Estimate | Standard Error | t-value |
|---|---|---|---|
| CONSTANT | -0.1086 | 0.0153 | -7.0919 |
| region | -0.0019 | 0.0023 | -0.8215 |
| experience | 0.0038 | 0.0023 | 1.6120 |
| education | 0.0248 | 0.0032 | 7.7998 |
| type | 0.0426 | 0.0052 | 8.2360 |
| industry | -0.0084 | 0.0052 | -1.621 |

Note:
[1]Lower education includes College dropouts.

If standard deviations had not shown any trend, then we could argue that all the sources of the changes in the dispersion of wages came from the different growth behavior of the seven groups already studied. Since there has been an increase in wage dispersion, as shown in the next section, some of these increases must come from changes in the dispersion of wages within groups.

In the following section I will assess the importance of each source of increased dispersion in the wage distribution.

## 4.5 Wage dispersion and nonparametric dynamic wage structures

In this section I will study changes in the dispersion of wages due to three possible factors:

a. Changes in the dispersion of unobservable factors.

b. Changes in the value of observable factors.

c. Changes in the structure of wages.

The last element can only be studied if we do not impose a constant wage structure. Therefore, it is natural to compute the wage structure in these demographic and economic groups of workers using regression trees. If the structure is nonparametric and can evolve, then this evolution will likely have an effect on the dispersion of wages.

At the beginning of section 2 we assumed that the log wage of any worker, i, of group $\mathbf{x}$ at any period, t, would have the form:

$$w_{it} = f_t(\mathbf{x}) + v_{it} \quad . \tag{4.15}$$

This model is very general. For each year, not only may wage differentials between different groups change, but also the relevant groups may also change. In terms of tree structures, not only can the projections in the tree change with time, but the tree itself may change.

Taking 1980 as the reference year, we can easily compute for the following years the wage any worker would have had if no change had occurred in the tree structure. In order to do so, we add her estimated idiosyncratic residual to the projected wage she would have had if the tree structure had not changed from 1980.

More precisely, if $f_t(\mathbf{x},\mathbf{b}_t)$ is the expected wage of any worker from group $\mathbf{x}$ at time t, then we can define, for any synthetic worker at time t+s, her wage with no structural change since t as:

$$w_i(t+s,t) = f_t(\mathbf{x},\mathbf{b}_{t+s}) + v_{i(t+s)} \quad . \qquad \textbf{(4.16)}$$

To evaluate the importance of changes in the wage structure, I set up an algorithm in three steps:

(a) Estimation of the tree structure for each year. This is the standard output in regression trees. From this, residuals for each observation can be computed: $v_{i(t+s)}$.

(b) Using the structure obtained for 1980, we can compute projections for all the other years. These projections are sample averages of wages within terminal nodes in the tree structure for 1980: $f_t(\mathbf{x},\mathbf{b}_{t+s})$.

(c) Creation of an artificial sample: this is simply done by adding to each residual obtained in (a) the projection for that worker obtained in (b): $w_i(t+s|t)$.

**Table 4.3** Regression Trees: Validation Results. Dispersion data. 1980-1995.

Data: CPS 1980-85-90-95. Male, full-time,non-farm workers. Individual observations.

| | Residual Sum of Squares | | | | Complexity | |
|---|---|---|---|---|---|---|
| | Estimation Sample | | Test Sample | | | |
| | Root | Best[1] | Root | Best | No. Obs. | No. Nodes |
| 1980 | 6043 | 4955 | 2987 | 2453 | 33671 | 115 |
| 1985 | 5801 | 4630 | 2851 | 2304 | 28734 | 110 |
| 1990 | 14245 | 10023 | 7132 | 5088 | 53100 | 219 |
| 1995 | 13146 | 9774 | 6111 | 4552 | 43757 | 206 |

Note:
Regression trees was carried out for each year with a 0-SE rule and a 50 stop-splitting rule. The independent variables were education, experience, region, type of worker, type of industry, and race.
[1]Root refers to the entire sample and Best is the honest tree with a 0-SE rule. An $R^2$-alike measure of goodness of fit can be computed as 1-RSS(root)/RSS(best), where RSS denotes the residual sum of squares.

Heuristically, if trees become ever more complex, then projections will tend to introduce more variance to the distribution of wages. However, if changes in the complexity of the trees do not lead to substantial differences in the projections, then changes in the structure cannot be regarded as an important source of increasing wage dispersion.

I use the extracts of the Annual Earnings Files of the Current Population Survey (CPS) prepared by the NBER for the years 1980, 1985, 1990, and 1995. I use the same variable specification as in the previous section but must implement the analysis obviously on individual observations. I will refer to this data set as the "dispersion" data.

Results of step (a) are shown in Tables 4.3 and 4.4. Table 4.3 gives some statistics on the tree structures obtained. Table 4.4 presents basic statistics for real log wages and residuals in the trees. Figures 4.7, 4.8, 4.9, and 4.10 in Appendix C show histograms of test sample residuals for each year.

Several comments are worth making. Wage dispersion for each year within each terminal node is a fairly large part of total wage dispersion. This follows from the statistics in Table 4.3 and from the residual and log wage basic statistics in Table 4.4.

The residual sum of squares estimates in Table 4.3 imply coefficients of determination never greater than 29.64% for the estimation sample and 28.66% for the test sample.

Residual variance increased during the period less than total variance. Changes in the wage structure account for 53.25% of the increase in wages dispersion measured with variance. Thus, although residual variance was around 82% of total variance in 1980, it fell to 73% in 1995.

Figures 4.7 to 4.10 in Appendix C give a fair idea of the way that regression trees work with the data. Since the splitting rule consists of minimizing within-node sum of squares, it is natural that the estimation sample residuals show a symmetric

**Table 4.4** Wages and tree residuals. Dispersion data. Basic statistics.

Data: CPS 1980-85-90-95. Male, full-time,non-farm workers. Individual observations.

| Log wages (all observations) | mean | sta.dev. | min | max | no.obs. |
|---|---|---|---|---|---|
| 1980 | 1.88 | 0.42 | 0.23 | 3.24 | 50406 |
| 1985 | 2.10 | 0.44 | 0.47 | 3.34 | 42418 |
| 1990 | 2.43 | 0.50 | 0.42 | 3.99 | 78547 |
| 1995 | 2.56 | 0.49 | 0.70 | 4.00 | 63814 |
| Residuals (test sample) | | | | | |
| 1980 | -0.005 | 0.38 | -1.69 | 1.58 | 16735 |
| 1985 | -0.001 | 0.39 | -1.66 | 1.40 | 14083 |
| 1990 | -0.003 | 0.42 | -1.95 | 1.72 | 26077 |
| 1995 | 0.004 | 0.42 | -2.24 | 1.59 | 21186 |

Note: Residuals are obtained with the test sample and the projections of the trees computed each year with a 0-SE rule and 50 stop-splitting rule. See Table 4.3 for the variable specification.

normal distribution resemblance. This feature is thus fabricated in the residuals of the estimation sample in the same way that linear regression with a constant will give zero mean residuals. This property has passed nicely onto the residuals plotted in the figures, which come from the test sample observations.

A summary of statistics for the artificial data computed in steps (b) and (c) can be found in Table 4.5 and 4.6. In Table 4.5 I present univariate descriptive statistics for all variables: original wages, constructed wages, residuals, and the structures under the assumption of no change in the functional form and with change.

**Table 4.5** The effect of structural change in observed wages dispersion. Dispersion data.1980-1995.

Data: CPS 1980-85-90-95. Male, full-time,non-farm workers. Individual observations.Test Sample. Reference Year: 1980.

1985

| Variable | Mean | Std Dev | Minimum | Maximum | No.Obs. |
|---|---|---|---|---|---|
| $w_i(t+s\|t)$ | 2.0946 | 0.4356 | 0.722 | 3.288 | 14083 |
| $f_t(\mathbf{b}_{t+s})$ | 2.0959 | 0.1872 | 1.641 | 2.418 | 14083 |
| $v_{i(t+s)}$ | -0.0013 | 0.3939 | -1.657 | 1.403 | 14083 |
| $f_{t+s}(\mathbf{b}_{t+s})$ | 2.0965 | 0.1921 | 1.609 | 2.784 | 14083 |
| $w_{i(t+s)}$ | 2.0951 | 0.4357 | 0.693 | 3.301 | 14083 |

1990

| Variable | Mean | Std Dev | Minimum | Maximum | No.Obs. |
|---|---|---|---|---|---|
| $w_i(t+s\|t)$ | 2.4231 | 0.4924 | 0.782 | 4.196 | 26077 |
| $f_t(\mathbf{b}_{t+s})$ | 2.4263 | 0.2560 | 1.858 | 2.936 | 26077 |
| $v_{i(t+s)}$ | -0.0032 | 0.4207 | -1.954 | 1.724 | 26077 |
| $f_{t+s}(\mathbf{b}_{t+s})$ | 2.4271 | 0.2748 | 1.757 | 3.211 | 26077 |
| $w_{i(t+s)}$ | 2.4240 | 0.5006 | 0.734 | 3.994 | 26077 |

1995

| Variable | Mean | Std Dev | Minimum | Maximum | No.Obs. |
|---|---|---|---|---|---|
| $w_i(t+s\|t)$ | 2.5697 | 0.4815 | 0.682 | 4.196 | 21186 |
| $f_t(\mathbf{b}_{t+s})$ | 2.5656 | 0.2379 | 2.060 | 2.921 | 21186 |
| $v_{i(t+s)}$ | 0.0041 | 0.4164 | -2.239 | 1.548 | 21186 |
| $f_{t+s}(\mathbf{b}_{t+s})$ | 2.5646 | 0.2541 | 2.045 | 3.190 | 21186 |
| $w_{i(t+s)}$ | 2.5688 | 0.4884 | 0.738 | 3.948 | 21186 |

Note:
$w_i(t+s|t)$: computed wage at t+s under no structural change since 1980.
$f_t(\mathbf{b}_{t+s})$: projection at t+s under no structural change since 1980.
$v_{i(t+s)}$: residual of observation i,t+s.
$f_{t+s}(\mathbf{b}_{t+s})$: projection at t+s with honest tree.
$w_{i(t+s)}$: logwage of observation i,t+s.

For all years, the variance of the constructed wages is smaller than the variance of wages. This suggests that part of the

**Table 4.6** The effect of structural change. Correlation matrix. Dispersion data.1980-1995.

Data: CPS 1980-85-90-95. Male, full-time,non-farm workers. Individual observations.Test Sample. Reference Year: 1980.

| 1985 | $w_i(t+s\|t)$ | $f_t(\mathbf{b}_{t+s})$ | $v_{i(t+s)}$ | $f_{t+s}(\mathbf{b}_{t+s})$ | $w_{i(t+s)}$ |
|---|---|---|---|---|---|
| $w_i(t+s\|t)$ | 1.0000 | 0.4269 | 0.9030 | 0.4019 | 0.9936 |
| $f_t(\mathbf{b}_{t+s})$ | 0.4269 | 1.0000 | -0.0031 | 0.9663 | 0.4233 |
| $v_{i(t+s)}$ | 0.9030 | -0.0031 | 1.0000 | -0.0147 | 0.8976 |
| $f_{t+s}(\mathbf{b}_{t+s})$ | 0.4019 | 0.9663 | -0.0147 | 1.0000 | 0.4276 |
| $w_{i(t+s)}$ | 0.9936 | 0.4233 | 0.8976 | 0.4276 | 1.0000 |

| 1990 | $w_i(t+s\|t)$ | $f_t(\mathbf{b}_{t+s})$ | $v_{i(t+s)}$ | $f_{t+s}(\mathbf{b}_{t+s})$ | $w_{i(t+s)}$ |
|---|---|---|---|---|---|
| $w_i(t+s\|t)$ | 1.0000 | 0.5196 | 0.8542 | 0.4779 | 0.9803 |
| $f_t(\mathbf{b}_{t+s})$ | 0.5196 | 1.0000 | -0.0003 | 0.9329 | 0.5119 |
| $v_{i(t+s)}$ | 0.8542 | -0.0003 | 1.0000 | -0.0083 | 0.8358 |
| $f_{t+s}(\mathbf{b}_{t+s})$ | 0.4779 | 0.9329 | -0.0083 | 1.0000 | 0.5420 |
| $w_{i(t+s)}$ | 0.9803 | 0.5119 | 0.8358 | 0.5420 | 1.0000 |

| 1995 | $w_i(t+s\|t)$ | $f_t(\mathbf{b}_{t+s})$ | $v_{i(t+s)}$ | $f_{t+s}(\mathbf{b}_{t+s})$ | $w_{i(t+s)}$ |
|---|---|---|---|---|---|
| $w_i(t+s\|t)$ | 1.0000 | 0.5021 | 0.8695 | 0.4624 | 0.9818 |
| $f_t(\mathbf{b}_{t+s})$ | 0.5021 | 1.0000 | 0.0094 | 0.9311 | 0.4924 |
| $v_{i(t+s)}$ | 0.8695 | 0.0094 | 1.0000 | 0.0029 | 0.8540 |
| $f_{t+s}(\mathbf{b}_{t+s})$ | 0.4624 | 0.9311 | 0.0029 | 1.0000 | 0.5227 |
| $w_{i(t+s)}$ | 0.9818 | 0.4924 | 0.8540 | 0.5227 | 1.0000 |

Note:
$w_i(t+s|t)$: computed wage at t+s under no structural change since 1980.
$f_t(\mathbf{b}_{t+s})$: projection at t+s under no structural change since 1980.
$v_{i(t+s)}$: residual of observation i,t+s.
$f_{t+s}(\mathbf{b}_{t+s})$: projection at t+s with honest tree.
$w_{i(t+s)}$: logwage of observation i,t+s.

increased dispersion in wages may have been due to changes in the structure of wages.

This is also corroborated by looking at the correlation matrices in Table 4.6. The correlation results show that the structures and the residuals are almost orthogonal decompositions of both types of wages and thus these variance measures approximate variance decompositions. The results depend however on the year of study. In 1985, observed variance[9] was 8.4% higher than in 1980.

Residual variance was only 7.7%. The combined effect on the variance of changes in prices and structure must have increased by almost 11.5%, or 0.0034. However, only 0.0001 remains unexplained after accounting for the change in prices.

In 1990, nonresidual variance increased by 0.0429 units, more than 100%. Around 0.0081 can be attributed to the effect of structural change on wages dispersion. This is near 18.98% of all the effect from variability in the structure.

In 1995, the amount unexplained by changes in prices is 0.0069. Total nonresidual variance increased by 0.03064, so the effect of a changing structure contributed in 22.52% of all increases in nonresidual variance.

To sum up, overall dispersion of wages increased from 1980 to 1985 by 0.0148. Increase in the dispersion within groups, or

[9] These computations are carried out with test sample data. The variance of observed wages was 0.1746 in 1980.

increase in inequality, accounts for 75.7% of this increase. The rest must be mainly attributed to price changes in the labor market, but not to changes in the interaction of the workers' characteristics. For 1990 and 1995 I find similar results. Less than half of the increase in the variance of log wages was due to increased wage inequality. At least 10% of the increase in variance came from changes in the interaction of different variables. In other words, due to changes in what the labor market would understand as homogeneous workers with a single expected wage. If these changes had not taken place, the increased variance in wages since 1980 would have been 10% lower.

## 4.6 Conclusions

In this chapter I applied regression trees to study the relative growth performance of workers' real wages and the sources of wage dispersion and its evolution in the U.S. from 1980 onwards. In order to study these problems I assumed that real wages follow a random field with a nonparametric specification for the cross-sample and a parametric specification for the time series. By doing so I can unveil nonlinear features in the trend behavior of real wages and compute the extent to which dispersion in wages has a structural source different from changes in prices.

The main results on trends agree with the results in previous studies. Education was an important factor of growth. Industry and region were not. Here I add two remarks.

First, the estimated tree suggests a linear structure for the growth experience of white workers with less than forty years of experience. The results show that for this class of workers, blue collar and higher education workers had always better performance than white collar and lower education workers.

Second, nonlinearities were present in other segments of the labor market. For example, the "effect" of race and experience on growth was not uniform and even differentials were of different sign.

On wage dispersion I started by noting how dispersion increased within groups with the same growth experience, therefore vindicating previous work on growing inequality within groups defined by socioeconomic variables. I then assessed the importance of each source of increased dispersion in the wage distribution. There are three potential sources of increased wage dispersion in observed wages: growing within groups inequality, changes in the premiums, and changes in the segmentation of the labor market. The last element can only be studied if the structural form of the wage structure is estimated.

Increase in within groups inequality accounts for most of the increase in wage dispersion from 1980 to 1985. The rest must be mainly attributed to price changes in the labor market, but not to changes in the interaction of the workers' characteristics. For 1990 and 1995, however, at least 10% of the increase in variance came from changes in the interaction of different variables.

## Appendix A

The implication of cointegration in integrated dynamic common factor models is easy to obtain. The other direction in the relationship cannot be established when the elements of the cointegrated vector have no idiosyncratic disturbances. This highlights the important difference between common factor models on one side and common trend extraction in cointegrating systems as in Stock and Watson (1988) on the other.

I will show that a dynamic factor model entails cointegrating relations amongst the elements of the vector. Assume that $x_{jt}$ satisfies the relation

$$x_{jt} = a_j(B)\ u_t + y_{jt}\ ,\ j=1,\ldots,N.$$

Here $u_t$ is a Kx1 vector of unobserved orthogonal random walks; $y_{jt}$ is a noise term with the usual stationary and orthogonal properties that is uncorrelated with the first differences in $u_t$ at all leads and lags. The common factors have a multivariate representation of the form

$$\Gamma(B)(1-B)u_t = b + e_t,$$

where $\Gamma(B)$ is a diagonal lag polynomial matrix and $e_t$ is a vector white noise with variance matrix the identity. Let $a(B)$ be $(a_1(B),\ldots,a_N(B))'$. Assume that $a_j(1)\neq 0$ and that $a(1)$ is of full rank K. Further assume that a(B) and $\Gamma(B)$ are finite matrix polynomials with $\Gamma(0)=I_K$. Finally, suppose that $|\Gamma(z)|\neq 0 \; \forall |z|\leq 1$, so that $(1-B)u_t$ follows a multivariate stationary invertible process.

The common factors structure originates cointegration and the number of factors determines the rank of the cointegrating space, N-K.

To see that $x_t$ is C(1,1), first note that it is I(1) since[1] it is the linear combination of I(1) and I(0) processes and the I(1) processes are not cointegrated. The vector $(1-B)x_t$ is stationary since it is the sum of I(0) elements. Since the spectrum of $(1-B)x_t$ is not of full rank at frequency zero, it is strictly noninvertible, as one should expect if cointegration is to hold. I will now show that there are N-K independent cointegrating relationships. We can always decompose a(B) as

$$a(B) = \sum_{i=0}^{M} a_i B^i = a(1) + a^{o}(B)(1-B),$$

where

$$a_i^{o} = -\sum_{s=i+1}^{M-1} a_s.$$

[1]See Engle and Granger (1987).

Therefore, we have that $x_t=a(1)u_t+w_t$ where $w_t=a^o(B)\Gamma^{-1}(B)e_t+d+y_t$. Note that $w_t$ is always stationary.

On the other hand, since its power spectrum is $f_w(\lambda)=(2\pi)^{-1}a^o(e^{-i\lambda})\Gamma^{-1}(e^{-i\lambda})\Gamma^{-1}(e^{i\lambda})'a^o(e^{i\lambda})'+f_y(\lambda)$, the invertibility condition for $w_t$ is simply that the process $y_t$ be invertible.

Since a(1) is of full rank K, there exists a base H, matrix Nx(N-K) of rank N-K, of the subspace orthogonal to that spanned by the columns of a(1) and so $H'a(1)=0_{(N-K)xK}$. Thus $H'x_t=H'w_t$ must be stationary. Invertibility is assured by the invertibility of $y_t$ and the fact that H is a full rank matrix. When $H'f_y(\lambda)$ is of full rank, $H'x_t$ is jointly invertible and, thus, $x_t$ is C(1,1).

When K=1 then for any two elements of $x_t$, $x_{it}$ and $x_{jt}$, i<j, the vector $(1,-a_i(1)/a_j(1))'$ is a (normalized) cointegrating vector.

Cointegration does not imply a common factor structure with idiosyncratic disturbances. Simply consider the previous model without idiosyncratic shocks. The elements in $x_t$ are cointegrated although the model is not one with idiosyncratic shocks.

Common trend decompositions do not build a bridge between cointegration and dynamic factor models where each element is influenced by disturbances that are orthogonal to the other disturbances. A very simple example is that of the structural random walk with noise model. The model can be postulated as:

$$y_{it} = x_t + e_{it},\ i=1,2\ ;$$

where $x_t$ is unobservable and follows a random walk process:

$$x_t = x_{t-1} + v_t.$$

We can think of $e_{it}$ as purely idiosyncratic to $y_{it}$ whilst $v_t$ would be common to $y_{it}$ and $x_t$.

Clearly, all nonstarionary variables are cointegrated. The common trend extraction methods usually proposed do not identify the "true" trend $x_t$. What is more important, the decompositions into a nonstationary and a stationary component give a single stationary element that cannot be idiosyncratic to each element. Usually, this element takes the form:

$$u_t = k\cdot(e_{1t} - e_{2t}).$$

Escribano and Peña (1994) propose $k=1/\sqrt{2}$, Stock and Watson (1988) propose k=1 and Kasa (1992) proposes k=1/2.

**Appendix B**

Since the cross-section information is not used when computing average growth rates for each variable, there will not be gains in accuracy when these rates are computed for large cross-sections. The latent growth indicator described in section 3 of the chapter does use cross-section information by computing OLS pairwise regression between all elements of the vector and extracting the principal component of the moment matrix of those estimates.

I carried out a simulation to assess the importance of these gains in small samples for two simple examples of dynamic structures.

The first model corresponds to equation **(5)**: the linear deterministic trend model. The simulation was implemented for parameter values $d_x$=1, b=1, and var($v_{it}$)=1. The number of periods was always 5. Six different sizes for the cross section were evaluated: N=25,50,75,100,200, and 300.

The second model corresponds to the common factor model in Harvey (1989, p.450) with the common factor equal to equation **(6)** and the transference matrix a column vector of ones. The simulation was implemented for parameter values d=1, and var($v_{it}$)=1. The number of periods was again 5, and also six different sizes for the cross-section were evaluated:

**Table 4.7** Simulation results.

Number of periods: 5. Number of iterations: 200.

Model 1: Deterministic Linear Trend Model.

Parameter Value=1.

| | Average Growth Rates | | Latent Indicator | |
|---|---|---|---|---|
| N | mean | sta.dev. | mean | sta.dev. |
| 25 | 0.7484 | 0.3465 | 0.9541 | 0.0605 |
| 50 | 0.7563 | 0.3527 | 0.9544 | 0.0422 |
| 75 | 0.7471 | 0.3532 | 0.9450 | 0.0344 |
| 100 | 0.7484 | 0.3520 | 0.9547 | 0.0298 |
| 200 | 0.7480 | 0.3536 | 0.9542 | 0.0212 |
| 300 | 0.7492 | 0.3540 | 0.9537 | 0.0174 |

Model 2. Common Factor Model. Parameter Value=1.

| | Average Growth Rates | | Latent Indicator | |
|---|---|---|---|---|
| N | mean | sta.dev. | mean | sta.dev. |
| 25 | 0.7436 | 0.3462 | 0.8765 | 0.0660 |
| 50 | 0.7231 | 0.3524 | 0.9220 | 0.0453 |
| 75 | 0.7324 | 0.3533 | 0.8933 | 0.0374 |
| 100 | 0.7606 | 0.3522 | 0.9009 | 0.0325 |
| 200 | 0.7636 | 0.3533 | 0.8989 | 0.0228 |
| 300 | 0.7247 | 0.3541 | 0.8727 | 0.0191 |

N=25,50,75,100,200, and 300.

The results of the simulation are given in Table 4.7. They show the expected good behavior of the latent estimator in terms of accuracy. This gain for small samples when the number of periods is very small may be a very important advantage when we try to partition the cross section according to growth performance.

## Appendix C

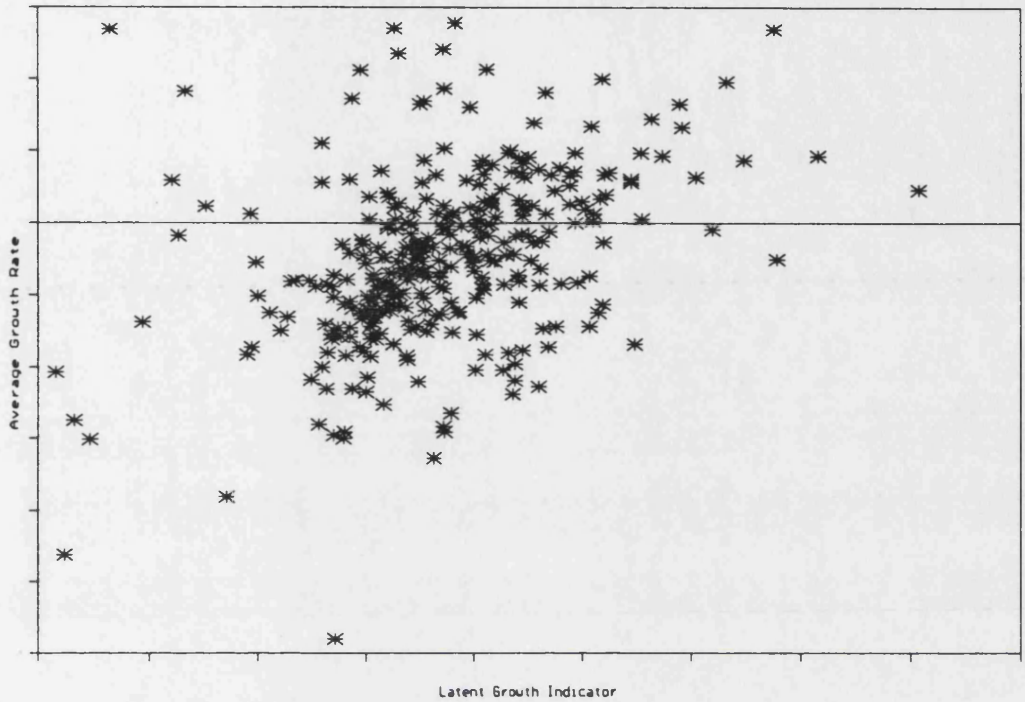


**Figure 4.1** Latent Growth Indicator vs Average Rates.

Trends data. Estimation Sample. 1980-1995.

Note:
Data: Pooled observations from CPS 1980-85-90-95, male, full time, nonfarm workers according to education, experience, race, type of worker, type of industry and region. On the vertical axis, average growth rates in real wages for each of the groups are plotted. On the horizontal axis, the laten growth estimators are plotted. The estimators have been normalized so that the sum of squares equals to one. The correlation between the average growth rates and the latent growth estimators is 0.74.

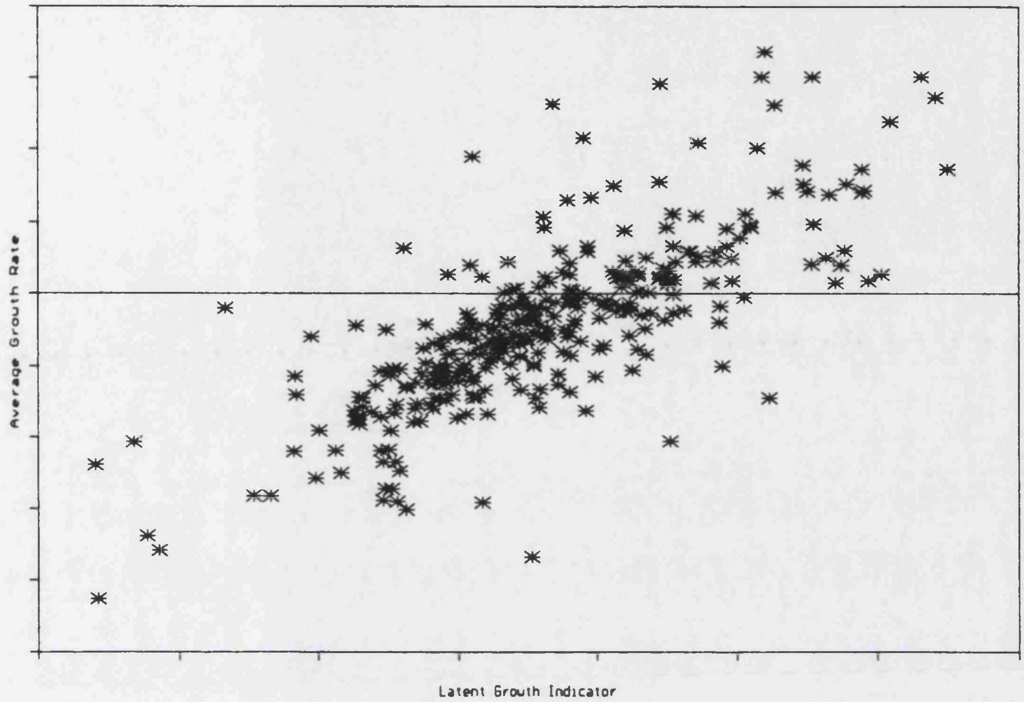


**Figure 4.2** Latent Growth Indicators vs Average Rates.

Trends data. Test Sample.1980-1995.

Note:
Data: Pooled observations from CPS 1980-85-90-95, male, full time, nonfarm workers according to education, experience, race, type of worker, type of industry and region. On the vertical axis, average growth rates in real wages for each of the groups are plotted. On the horizontal axis, the laten growth estimators are plotted. The estimators have been normalized so that the sum of squares equals to one. The correlation between the average growth rates and the latent growth estimators is 0.79.

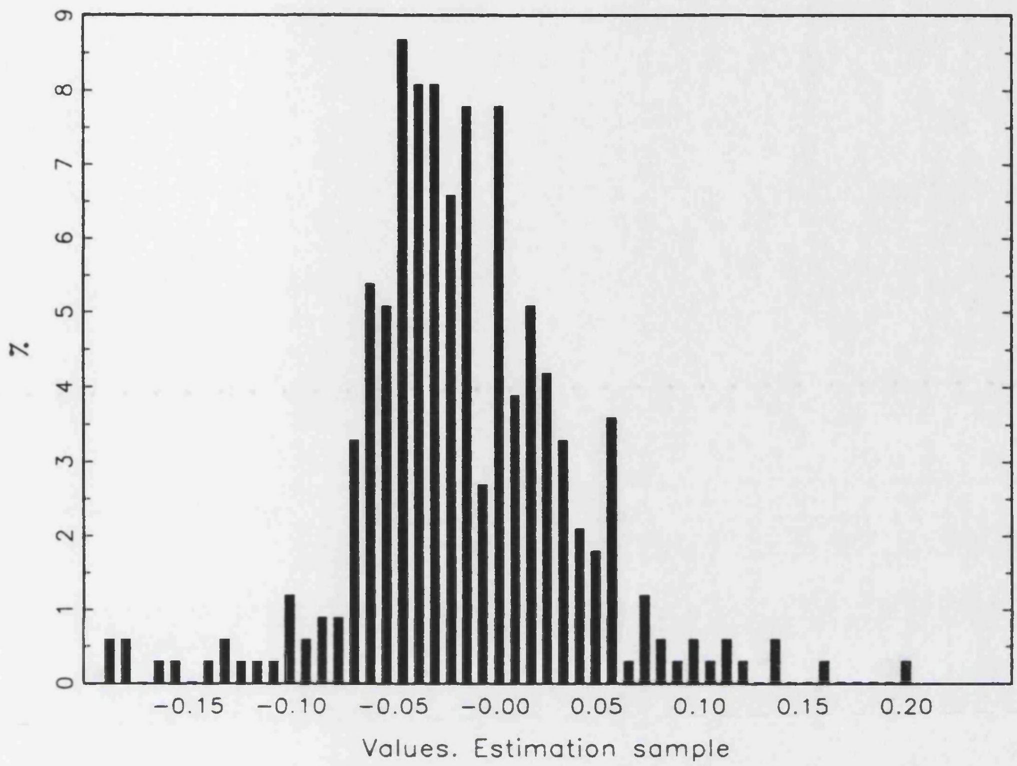


**Figure 4.3** Histogram for Latent Growth Indicator. Trends data. Estimation Sample. 1980-1995.

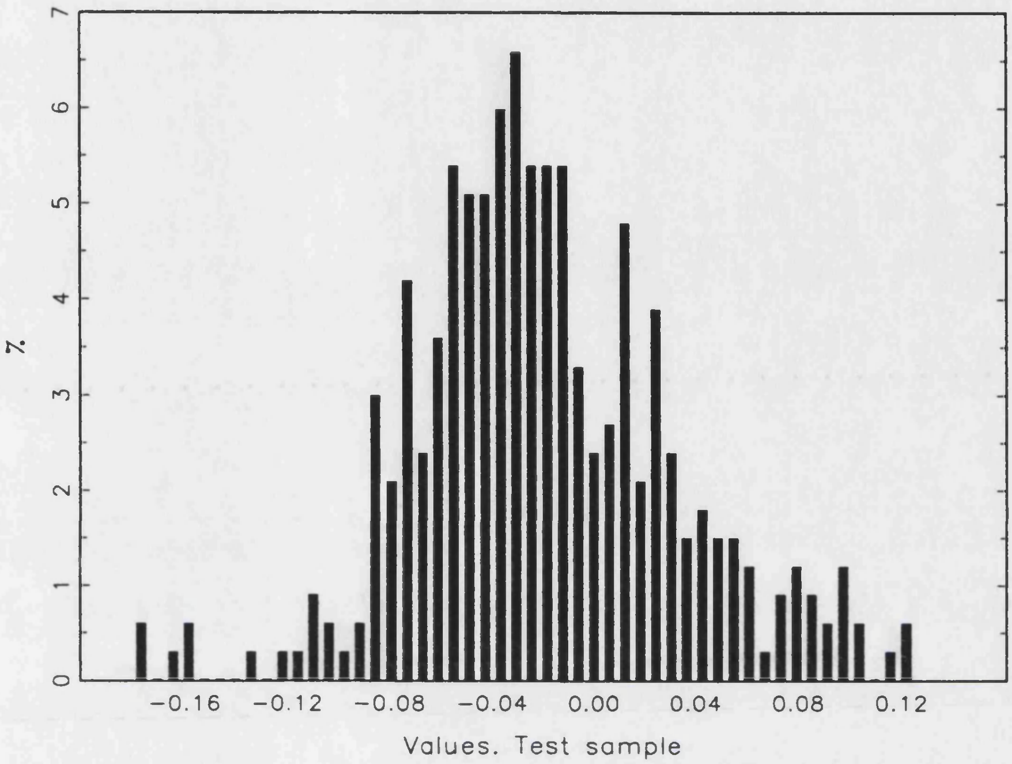


**Figure 4.4** Histogram for Latent Growth Indicator. Trends data. Test sample. 1980-1995.

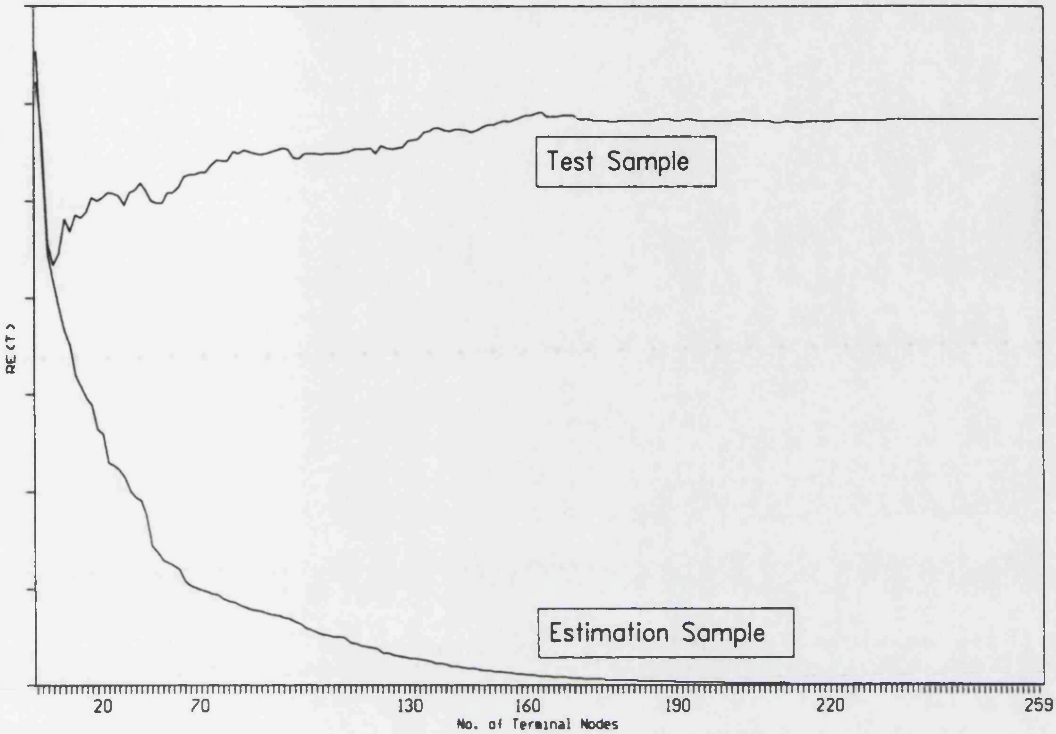


**Figure 4.5** Residual sum of squares for the sequence of optimal trees. Trends data. Latent growth estimator

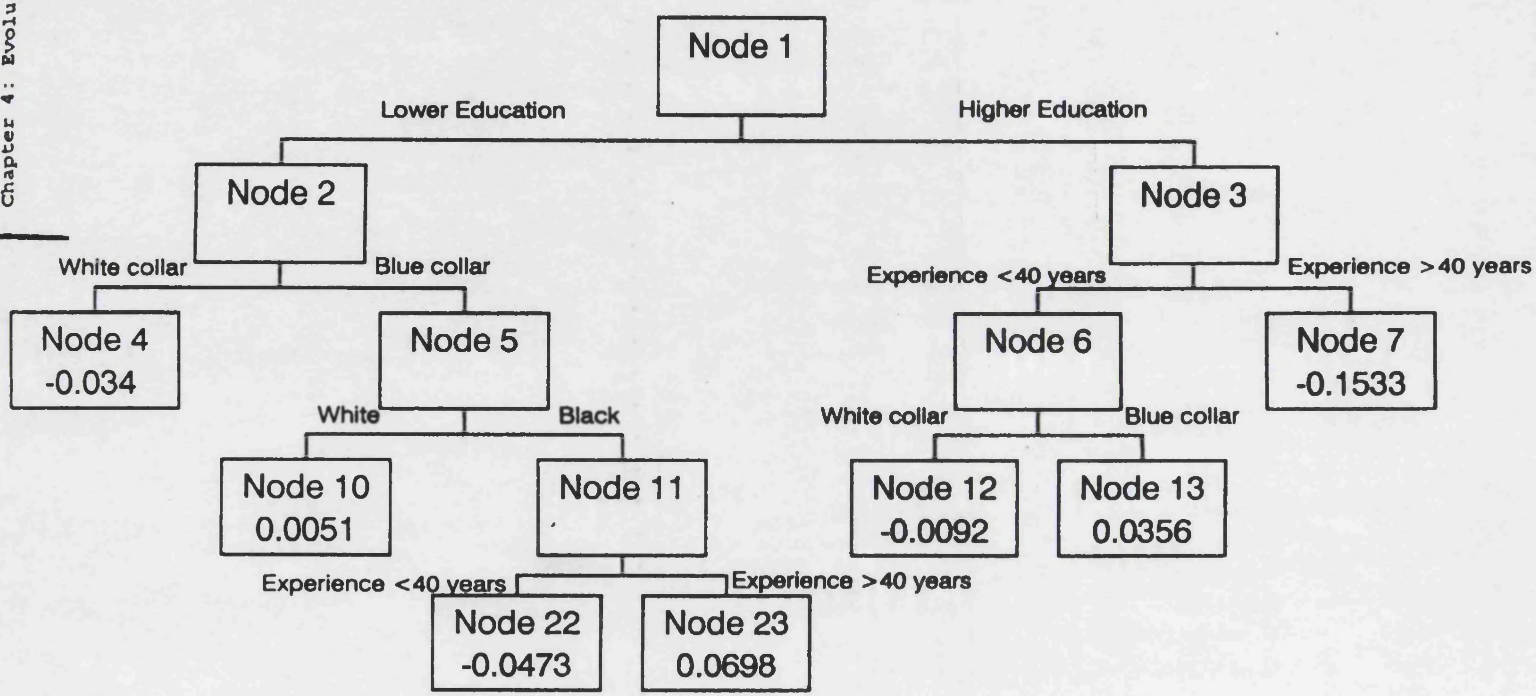


**Figure 4.6** Tree T182. Trends data

Lower Education includes College dropouts

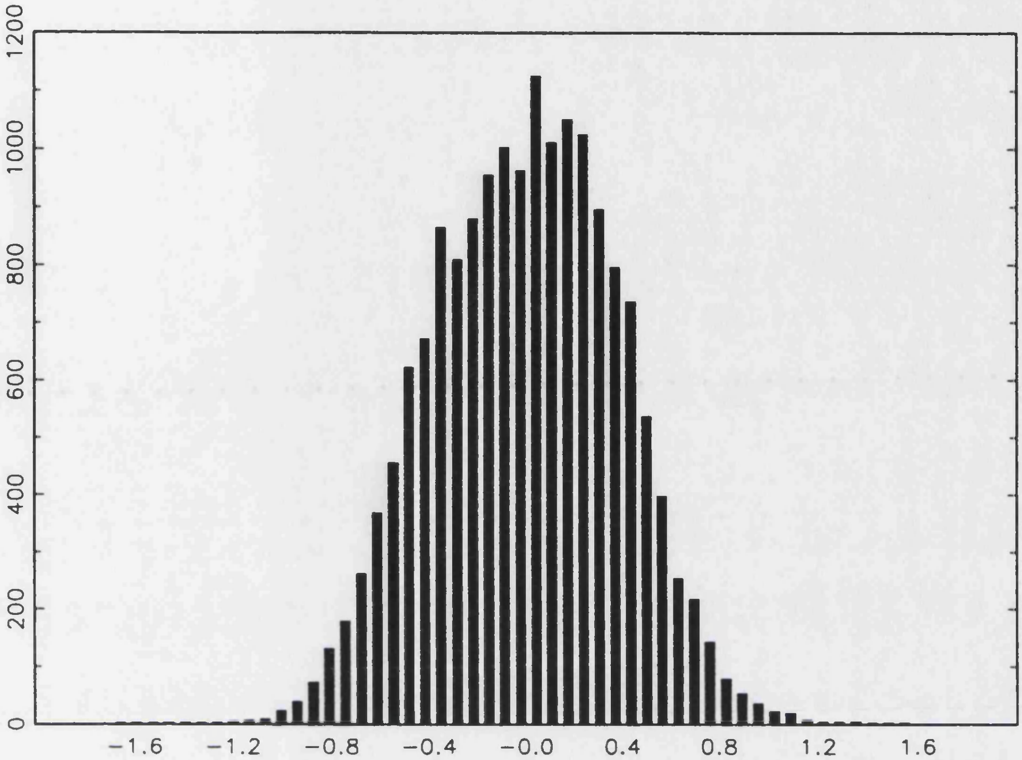


**Figure 4.7** Histogram for regression tree test sample residuals. Dispersion data. 1980.

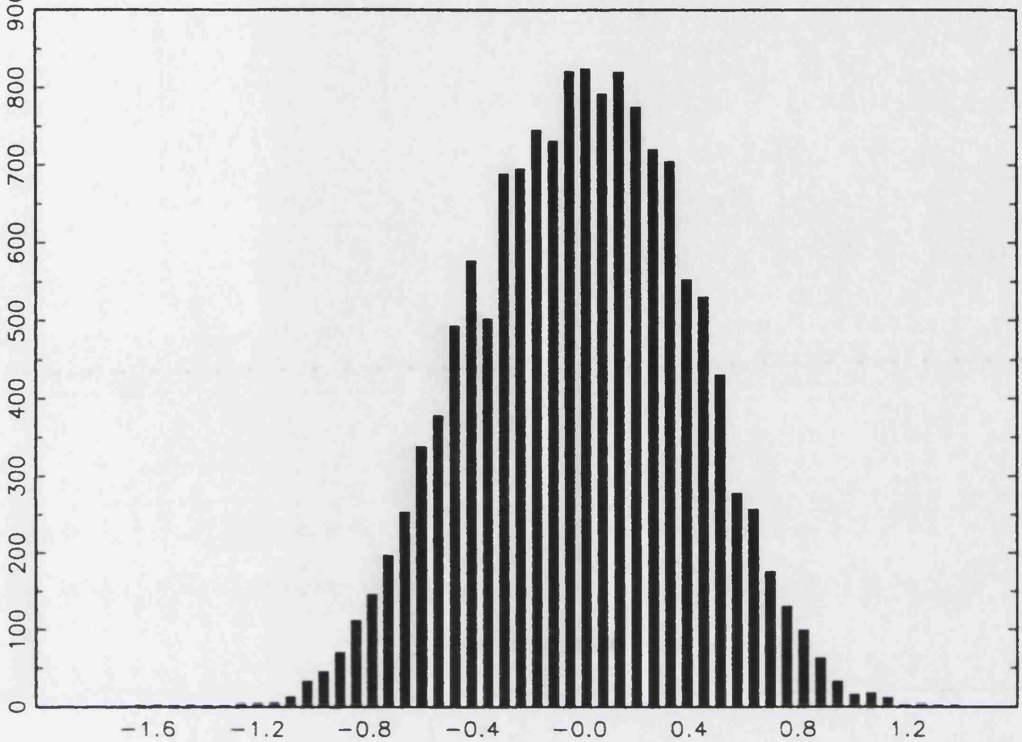


**Figure 4.8** Histogram for regression tree test sample residuals. Dispersion data. 1985.

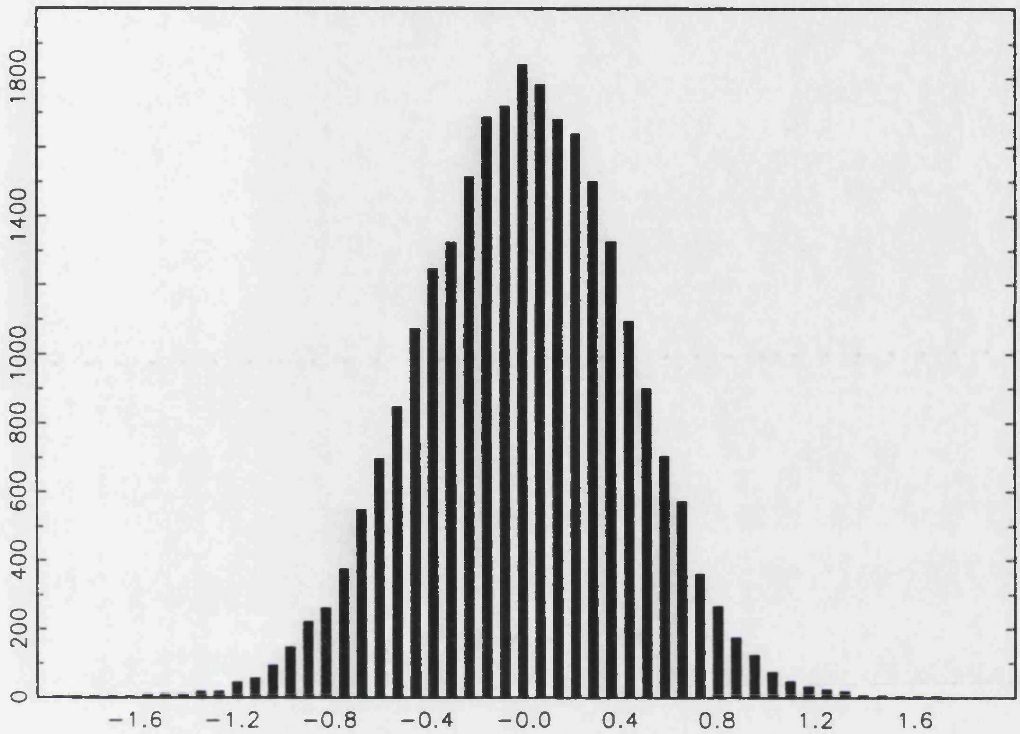


**Figure 4.9** Histogram for regression tree test sample residuals. Dispersion data. 1990.

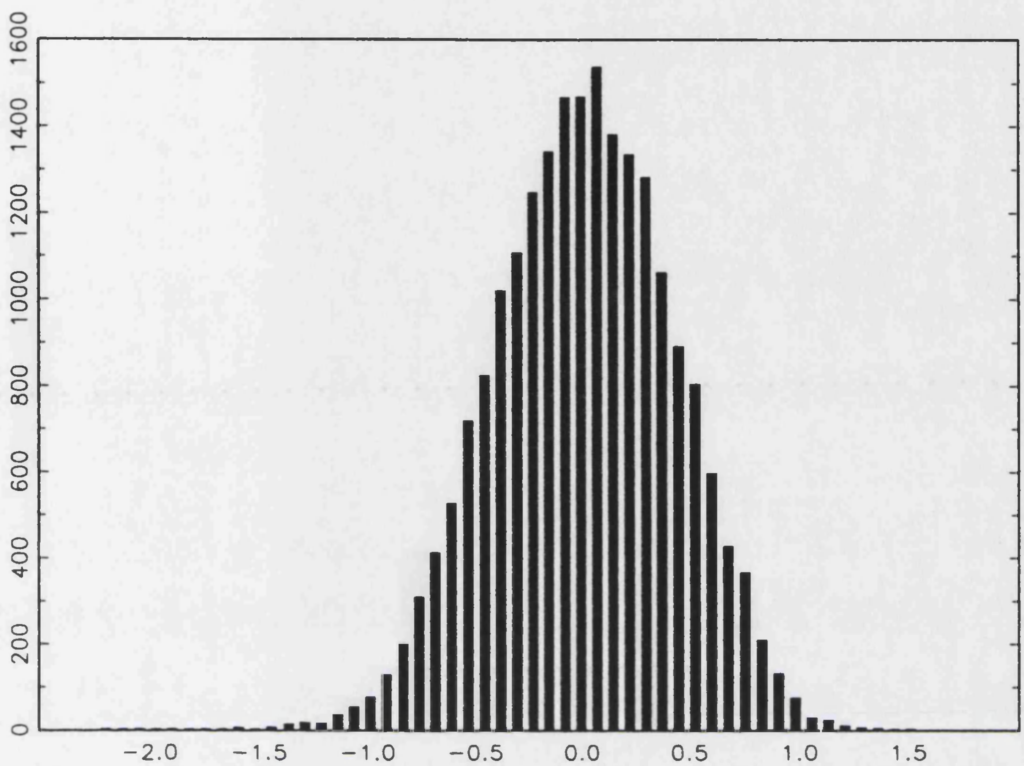


**Figure 4.10** Histogram for regression tree test sample residuals. Dispersion data. 1995.

Chapter 6

---

# Conclusions

This study was concerned with the application of multivariate nonparametric models known as regression trees to the analysis of the U.S. wage structure.

In the first application I have estimated a nonparametric experience-wage profile in a multivariate environment to search for local workers who suffered a sudden loss in wages in their last stage of their working life. This approach to experience-wage profiles in cross-sections mimics the work done on growth paths by Durlauf and Johnson (1995). The main methodological difference is that in my study I am only interested in the possible effects of sudden losses of human capital in the profile of a cross-section. The regression trees approach to the estimation of the experience profile fitted the sample better than a simple quadratic specification. When the projections of the tree were carefully studied, linear splines appeared as reasonable alternatives to the quadratic function for many workers. Further, simple parametric specifications are for some types of workers clearly inappropriate when an extended human capital specification is chosen. The algorithm used to identify sudden losses of human capital at the end of the working life is a reasonable strategy when overall human capital depreciation is small. It is totally justified if we assume that no overall human capital depreciation takes place. The results show that for 1980, 1985, and 1990, average wages of the last experience group would be 1% higher if no breaks had occurred. For 1980, the sudden losses help reducing the negative experience differential by 50%. For 1985, they do so by 33%. The value for 1990 is only 16.6%. Finally, the reduction in the negative differential is null for the last year of the analysis.

In the second application, I propose average wage gap decompositions between any two groups of workers for nonparametric structures. Even when it is not possible to talk of simple race, ethnic, or gender differential, Oaxaca-type decompositions are still useful decompositions of the observed average wage gap. These decompositions do take account of reasonable features such as context sensitivity. I think that the methodology is well suited to unveil local differentials and therefore it can help in understanding the complexity of the labor market. I carry out the proposed decompositions for two data sets. In the first empirical application, I study average wage gaps between male workers of Mexican origin born in Mexico and in the U.S. and between male workers of Mexican origin who labelled themselves as Mexicano and Mexican American. Although a very simple human capital specification is implemented, I obtain interesting differences with respect to the parametric analysis. Somewhat intuitively, the structural differential is bigger in the nonparametric case when I decompose average wage gaps between workers born in the U.S. and workers born in Mexico. However, the nonparametric decompositions for average wage gaps by ethnic origin show a very interesting and contrasting feature of regression trees. This technique not only performs nonparametric multivariate analysis, but implicitly carries out variable selection in its search for a simple structure. In this example, ethnicity appears as an important factor in only very specific local cases, rendering a smaller structural effect than the sample effect. In the second empirical analysis, I study the average wage gap between Mexican Americans and white non-Hispanic male workers in the border states between Mexico and the U.S.. Regression trees unveils a linear relation in the wage structure with respect to ethnicity so that most ethnic differentials are around 20%. Due to this linear behavior, the nonparametric decomposition is very similar to a simple decomposition with a dummy variable. In other words, regression trees can encompass parametric approaches.

Finally, in my third application of regression trees to the study of the wage structure I look at the evolution of wages and the sources of wages dispersion in the U.S.. In order to study these problems I assumed that real wages follow a random field with a nonparametric specification for the cross-sample and a parametric specification for the time series. By doing so I can unveil nonlinear features in the trend behavior of real wages and compute the extent to which dispersion in wages has a structural source different from changes in prices. A robust estimator of the trend is implemented to exploit the large cross-section information.

As in previous studies, I find that education was an important factor for growth performance. Two interesting features are unveiled by regression trees. First, the estimated tree suggests a linear structure for the growth experience of white workers with less than forty years of experience. Second, nonlinearities were present in other segments of the labor market. For example, the effect of race and experience on growth was not uniform and even differentials were of different sign.

On wage dispersion, I found that the increase in inequality accounts for most of the increases in wage dispersion from 1980 to 1985. The rest must be mainly attributed to price changes in the labor market, but not to changes in the interaction of the workers' characteristics. For 1990 and 1995, however, at least 10% of the increase in variance came from changes in the interaction of different variables.

**References**

Adamson,D.W. (1993) Differences in Union Relative Wage Effects across Gender and Race: A Longitudinal Analysis, *Journal of Economics* 19(2), Fall 1993, pages 79-91.

Allen, S.G. (1995) Updated Notes on the Interindustry Wage Structure, 1890-1990, *Industrial and Labor Relations Review* 48(2), January 1995, pages 305-21.

Ashraf, J. (1994) Recent Trends in White-Black Earnings Differentials, *Economics Letters* 45(4), August 1994, pages 487-93.

Becker, G. S. (1993) *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Chicago and London: The University of Chicago Press, 3rd.ed.

Ben-Porath, Y. (1967) The production of human capital and the life cycle of earnings, *Journal of Political Economy*, 75:352-365.

Bishop, J.H. (1991) Achievement, Test Scores, and Relative Wages, in Kosters, M.H., ed. *Workers and their wages*, AEI Press, pages 146-86.

Blackburn, M. (1990) What Can Explain the Increase in Earnings Inequality Among Males, *Industrial Relations*, Fall 1990, 29(3), pages 441-456.

Blackburn, M. and S. Korenman (1994) The Declining Marital-Status Earnings Differential, *Journal of Population Economics* 7(3), July 1994, pages 247-70.

Blau, F.D. and A.H. Beller (1992) Black-White Earnings over the 1970s and 1980s: Gender Differences in Trends, *Review of Economics and Statistics* 74(2), May 1992, pages 276-86.

Bloom, D.E. and G. Grenier (1993) Language, Employment, and Earnings in the United States: Spanish-English Differentials From 1970 to 1990, *National Bureau of Economic Research* Working Paper: 4584, December 1993, pages 21.

Borjas, G.J. and V.A. Ramey (1994) Time-Series Evidence on the Sources of Trends in Wage Inequality, *American Economic Review* 84(2), May 1994, pages 10-16.

Bound, J. and H.J. Holzer (1993) Industrial Shifts, Skills Levels, and the Labor Market for White and Black Males, *Review of Economics and Statistics* 75(3), August 1993, pages 387-96.

Bound, J. and G. Johnson (1992) Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations, *American Economic Review* 82(3), June 1992, pages 371-92.

Bound, J. and G. Johnson (1995) What Are the Causes of Rising Wage Inequality in the United States? *Federal Reserve Bank of New York Economic Policy Review* 1(1), January 1995, pages 9-17.

Breiman, L., J.L. Friedman, R.A. Olshen, and C.J. Stone (1984), *Classification and Regression Trees*, Wadsworth, Belmont, CA.

Bryant, P. and J.A. Williamson (1978) Asymptotic Behaviour of Classification Maximum Likelihood Estimates, *Biometrika*, Vol. 65:273-281.

Buchinsky, M. (1994) Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression, *Econometrica* 62(2), March 1994, pages 405-58.

Cameron, S.V. and J.J. Heckman (1991) The Nonequivalence of High School Equivalents, *Journal of Labor Economics*, 11(1), Part 1, pages 1-47.

Cattan, P. (1993) The Diversity of Hispanics in the U.S. Work Force, *Monthly Labor Review* 116(8), August 1993, pages 3-15.

Cotterman, R. and F. Peracchi (1992) Classification and Aggregation: An Application to Industrial Classification in CPS Data, *Journal of Applied Econometrics*; 7(1), Jan.-March 1992, pages 31-51.

Cotton, J. (1988) On the decomposition of wage differentials, *Review of Economics and Statistics* 70, pages 236-243.

Cotton, J. (1993) Color or Culture?: Wage Differences among Non-Hispanic Black Males, Hispanic Black Males and Hispanic White Males, *Review of Black Political Economy* 21(4), 53-67.

Davies, S.J. and Haltiwanger, J.(1991) Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963-86. *Brookings Papers on Economic Activity*, Special Issue, 1991, pages

115-180.

Dávila, A., Bohara, A.K., Saenz, R.(1993) Accent Penalties and the Earnings of Mexican Americans. *Social Science Quarterly* 74(4), December 1993, pages 902-16.

Donato, K.M. and D.S. Massey (1993) Effect of the Immigration Reform and Control Act on the Wages of Mexican Migrants, *Social Science Quarterly*, 74(3), pages 523-41.

Dunsmuir, W. and E.J. Hannan (1976), "Vector Linear Time Series Models", *Adv. Appl. Prob.* 8, 339-364.

Durlauf, S.N. and P.A. Johnson (1995) Multiple Regimes and Cross-Country Growth Behaviour, *Journal of Applied Econometrics*, VOL. 10, 365-384 (1995).

Eberts, R.W. and Schweitzer, M.E.(1994) Regional Wage Convergence and Divergence: Adjusting Wages for Cost-of-Living Differences. *Federal Reserve Bank of Cleveland Economic Review* 30(2), 2nd Quarter 1994, pages 26-37.

Engle, R.F. and C.W.J. Granger (1987) Co-integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*; 55(2), March 1987, pages 251-76.

Engle, R.F. and S. Kozicki (1993) Testing for Common Features, *Journal of Business & Economic Statistics*, October, Vol. 11, No. 4.

Engle, R. and M. Watson (1981) A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates, *Journal of the American Statistical Association*; 76(376), December 1981, pages 774-81.

Escribano, A. and D. Peña (1994) Cointegration and Common Factors, *Journal of Time Series Analysis*, Vol. 15, No 6:577-586.

Fair, R.C. (1994) How Fast Do Old Men Slow Down?, *The Review of Economics and Statistics*, 76(1), pages 103-118.

Fan, J. and I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Friedman, J.H. (1994) An Overview of Predictive Learning and Function Approximation, in Cherkassky, Friedman, and Wechsler, (eds.), *From Statistics to Neural Networks*, NATO ASI Series.

Geweke, J.F.(1977) The Dynamic Factor Analysis of Economic Time-Series Models, in Aigner, Dennis J. and Arthur S. Goldberger, (eds.), *Latent Variables in Socio-Economic Models*. North Holland.

Geweke, J.F. (1993) Comment to "A Dynamic Index for Large Cross Sections", in Stock, J. and M. Watson (eds), *New Research on Business Cycles, Indicators, and Forecasting*,University of Chicago Press.

Geweke, J.F. and K.J. Singleton (1981) Maximum Likelihood "Confirmatory" Factor Analysis of Economic Time Series, *International Economic Review*, Vol. 22:37-54.

Gordon, A.D.(1993), *Classification: Methods for the Exploratory Analysis of Multivariate Data*, Chapman and Hall.

Greene, M. and J.E. Rogers (1994) Education and the Earnings Disparities between Black and White Men: A Comparison of Professionals in the Public and Private Sectors, *Journal of Socio Economics* 23(1-2), Spring-Summer 1994, pages 113-30.

Griliches, Z. (1977) Estimating returns to schooling: some econometric problems, *Econometrica*, 45, 1-22.

Gustman, A.L. and T.L. Steinmeier (1984) The Effect of Partial Retirement on the Wage Profiles of Older Workers, *Industrial Relations* 24, pages 257-265.

Gyimah B., and R. Fichtenbaum(1994) Regional Differences in Labor Market Discrimination, *Review of Regional Studies* 24(1), Summer 1994, pages 13-36.

Hartigan, J.A.(1975) *Clustering Algorithms*, John Wiley and Sons.

Harvey, A.C. (1989) *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press.

Heckman, J.J. (1976) A life-cycle model of earnings, learning, and consumption, *Journal of Political Economy* (Suplement), 84:S11-S44.

Heckman, J.J. (1979) Sample selection bias as a specification error, *Econometrica*, 47:153-162.

Hotta, L.K. (1983) Estimation and Testing of Hipothesis in Unobservable Components Models, PhD (Econ), London School of Economics.

Huber, P.J. (1985) Projection Pursuit, *Annals of Statistics*, 13,435-475.

Johnson, R.W. and D. Neumark (1996) Wage Declines among Older Men, *The Review of Economics and Statistics*, 78(4), pages 740-748.

Juhn, C., Murphy, K. M., and B. Pierce (1993) Wage Inequality and the Rise in Returns to Skills, *Journal of Political Economy*, 101(June 1993): 410-42.

Kasa, K. (1992) Common stochastic trends in international stock markets, *J. Monetary Econ*. 29, 95-194.

Katz, L.F. and A. Revenga (1989) Changes in the Structure of Wages: the U.S. vs. Japan, *Journal of Japanise International Economics*, Dec 1989, 3(4), pages 522-553

Katz, L.F. and K.M. Murphy (1992) Changes in Relative Wages, 1963-1987: Supply and Demand Factors, *Quarterly Journal of Economics* 107(1), February 1992, pages 35-78.

Kim, M.K. and S.W. Polachek (1994) Panel Estimates of Male-Female Earnings Functions, *Journal of Human Resources* 29(2), Spring, pages 406-28.

Kuan, C.M. and H. White (1994) Artificial Neural Networks: An Econometric Perspective, *Econometric Reviews*, 13(1),1-91.

Lalonde, R.J. and R.H. Topel (1991) Immigrants in the American Labor Market: Quality, Assimilation, and Distributional Effects, *American Economic Review* 81(2), May 1991, pages 297-302.

Lawley, D.N. and A.E. Maxwell (1963) *Factor Analysis as a Statistical Method*. Butterworths, London.

Levy, F. and R.J. Murnane (1992) U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations, *Journal of Economic Literature*, 30(3), September 1992, pages 1333-81.

Mincer, J. (1974) *Schooling experience and earnings*. New York: National Bureau of Economic Research.

Mincer, J. and H. Ofek (1982) Interrupted Work Careers: Depreciation and Restoration of Human Capital, *Journal of Human Resources*, 17(1), pages 3-24.

Murphy, K.M. and Welch, F.(1990) Empirical Age-Earnings Profiles, *Journal of Labour Economics*, April 1990, pages 202-229.

Murphy, K.M. and Welch, F.(1992) Inequality and Relative Wages. *American Economic Review Papers and Proceedings*, Vol 83, No. 2.

Neuman , S. and A. Weiss (1995) On the Effects of Schooling Vintage on Experience-Earnings Profiles: Theory and Evidence, *European Economic Review* 39(5), pages 943-955.

Neumark, D.(1988) Employers' Discriminatory Behavior and the Estimation of Wage Discrimination, *Journal of Human Resources* 23: 279-295.

Neumark, D. and S. Korenman (1994) Sources of Bias in Women's Wage Equations: Results Using Sibling Data, *Journal of Human Resources* 29(2), Spring, pages 379-405.

Oaxaca, R.(1973) Male-female Wage Differentials in Urban Labor Markets, *International Economic Review* 9, 693-709.

Oaxaca, R.L. and M.R. Ransom (1994) On Discrimination and the Decomposition of Wage Differentials, *Journal of Econometrics* 61(1), March 1994, pages 5-21.

Pagán, J.A. and A. Dávila (1996) On-the-Job Training, Immigration Reform, and the True Wages of Native Male Workers, *Industrial Relations*, Vol.35, No.1

Piccolo, D. (1990) A Distance Measure for Classifying ARIMA models, *Journal of Time Series Analysis*, Vol. 11, No. 2:153-164.

Poirier, D.J. (1976) *The Economics of Structural Change with Special Emphasis on Spline Functions*, Amsterdam: North-Holland.

Polachek, S. and M.K. Kim (1994) Panel Estimates of the Gender Earnings Gap: Individual-Specific Intercept and Individual-Specific Slope Models, *Journal of Econometrics* 61(1), March 1994, pages 23-42.

Quah, D. and T. Sargent (1993) A Dynamic Index for Large Cross Sections, in Stock and Watson (eds), *New Research on Business Cycles, Indicators, and Forecasting*, University of Chicago Press.

Reimers, C. (1983) Labor Market Discrimination Against Hispanic and Black Men, *Review of Economics and Statistics* 65, pages 570-579.

Ruud, P.A. (1991) Extensions of Estimation Methods Using the EM Algorithm, *Journal of Econometrics*; 49(3), September 1991, pages 305-41.

Scott, A.J. and M.J. Symons (1971) Clustering Methods Based on Likelihood Ratio Criteria, *Biometrics*, Vol. 27:387-397.

Sheshinski, E. (1968) On the individual's life time allocation between education and work, *Metroeconomica*, 20:42-49.

Stock, J. (1987) Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors, *Econometrica*, Vol. 55, No. 5, 1035-1056.

Stock, J.H. and M.W. Watson (1988) Testing for Common Trends, *Journal of the American Statistical Association*; 83(404), December 1988, pages 1097-1107.

Stokey, N.L. and R.E. Lucas (1989) *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge, Massachusetts, and London, England.

Summers, R. and A. Heston (1988) A New Set of International Comparisons of Real Product and Price Levels Estimates for 130 Countries, 1950-1985, *Review of Income and Wealth*, 34, pages 1-25.

Topel, R.H. (1994) Regional Land Markets and the Determinants of Wage Inequality, *American Economic Review* 84(2), May 1994, pages 17-22.

Tronstad, R. (1995) Importance of Melon Type, Size, Grade, Container, and Season in Determining Melon Prices, *Journal of Agricultural and Resource Economics*, 20(1):32-48.

Verdugo, R.R. (1992) Earnings Differentials between Black, Mexican American, and Non-Hispanic White Male Workers: On the Cost of Being a Minority Worker, 1972-1987, *Social Science Quarterly* 73(3), September 1992, pages 663-73.

Walker, J.R. (1996) Earnings Differentials of Mexican Immigrants: The Case of Turkey Processing, *Journal of Socio Economics*, 25 (3), pages 271-283.

Watson, M.W. and R.F. Engle (1983) Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models, *Journal of Econometrics*; 23(3), December 1983, pages 385-400.

Watson, Mark W. and Kraft, Dennis F.(1984) Testing the Interpretation of Indices in a Macroeconomic Index Model, *Journal of Monetary Economics*; 13(2), March 1984, pages 165-81.

Weiss, Y. (1986) The Determination of Life Cycle Earnings: A Survey, in Ashenfelter and Layard, (eds.) *Handbook of Labor Economics*, Elsevier Science Publishers BV.

Welch, F. (1979) Effects of Cohort Size on Earnings: The Baby Boom Babies, *Journal of Political Economy*, Oct. 1979, pages 565-597.

Williams, J.T. (1978) Risk, Human Capital, and the Investor's Portfolio, *Journal of Business*; 51(1), pages 65-89.

Williams, J.T. (1979) Uncertainty and the Accumulation of Human Capital over the Life Cycle, *Journal of Business*; 52(4), pages 521-48.

Willis, R.J. (1986) Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions, in Ashenfelter and Layard, (eds.) *Handbook of Labor Economics*, Elsevier Science Publishers BV.

Wu, C.F.Jeff (1983) On the Convergence Properties of the EM algorithm, *Annals of Statistics*, Vol. 11, 95-103.

Yuengert, A.M. (1994) Immigrant Earnings, Relative to What? Importance of Earnings Function Specification and Comparison Points, *Journal of Applied Econometrics*, Vol. 9, pages 71-90.