Experiences from a socio-economic application of induction trees: The need for relevant validation criteria

Fabio B. Losa[1], Pau Origoni[1], and Gilbert Ritschard[2]

[1] Statistical Office of Ticino Canton, CH-6500 Bellinzona, Switzerland
fabio.losa@ti.ch
[2] Dept of Econometrics, University of Geneva, CH-1211 Geneva 4, Switzerland
gilbert.ritschard@themes.unige.ch

Abstract. This paper presents a full scaled application of induction trees for non-classificatory purposes. The grown trees are used for highlighting regional differences in the women's labor participation, by using data from the Swiss Population Census. Hence, the focus is on their descriptive rather than predictive power. A first tree provides evidence for three separate analyses for non-mothers, married or widowed mothers, and divorced or single mothers. For each group, trees grown by language regions exhibit fundamental cultural differences supporting the hypothesis of cultural models in female participation. From the methodological standpoint, the main difficulties with such a non-classificatory use of trees have to do with their validation, since the classical classification error rate does not make sense in this setting. We comment on this aspect and consider alternatives that are both consistent with our non-classificatory usage and easy to compute.

1 Introduction

Induced decision trees have become, since [1], popular multivariate tools for predicting continuous dependent variables and for classifying categorical ones from a set of predictors. They are called *regression trees* when the outcome is quantitative and *classification trees* when it is categorical. Though their primary purpose is to predict and to classify, trees can be used for many other relevant purposes: as exploratory methods for partitioning and identifying local structures in datasets, as well as alternatives to statistical descriptive methods like linear or logistic regression, discriminant analysis, and other mathematical modeling approaches [2].

This contribution demonstrates such a *non-classificatory* use of classification trees by presenting a full scaled application on female labor market data from the Swiss 2000 Population Census (SPC). The use of trees for our analysis was dictated by our primary interest in discovering the interactions effects of predictors of the women's labor participation. Since the goal is no longer to extract classification rules, but to understand — from a cross-cultural perspective —

the forces that drive women's participation behavior, misclassification rates do not make sense when they are used to validate the trees. We therefore rely on best suited alternative fit criteria like those proposed in [10, 11]. Our experiment brings insight into the limits and practicability of these criteria for large scale applications.

Apart from these methodological aspects, the practical experiment discussed in this paper is original in at least two respects: 1) the use of trees for microeconomic analysis, which does not appear to be a common domain of application; 2) the use of induction trees for a complete population census dataset.

Section 2 briefly recalls the principle of classification trees. In Section 3, we present the socio-economic research objectives and discuss the main findings. Section 4 is devoted to the validation issue. Finally, we conclude in Section 5 with an overall evaluation of the experience and of the application of classification trees for non-classificatory purposes.

2 Classification trees principle

Classification trees are grown by seeking, through successive splits of the learning data set, some optimal partition of the predictor space for predicting the outcome class. Each split is done according to the values of one predictor. The process is greedy. At the first step, it tries all predictors to find the "best" split. Then, the process is repeated at each new node until some stopping rule is reached. This requires a local criterion to determine the "best" split at each node. The choice of the criterion is the main difference between the various tree growing methods that have been proposed in the literature, of which CHAID [5], CART [1] and C4.5 [7] are perhaps the most popular. For our application, we used CART that builds only binary trees by choosing at each step the split that maximizes the gain in purity measured by the Gini index. CART uses relatively loose stopping rules, but proceeds to a pruning round after the preliminary growing phase.

One of the striking features of induction trees is their ability to provide results in a visual form that provides straightforward interpretations. This visual feature, when compared with the outcome of regression models for instance, has exceptional advantages in terms of user-friendliness and in supporting the knowledge discovery process. Furthermore, by their very nature, trees provide a unique description of the predictor interaction effects on the response variable. These advantages remain true as long as the tree does not become too complex. That is why we chose CART for our analysis, despite the gain in purity seems less appropriate for a non-classificatory purpose than, for example, the strength of association criterion used by CHAID. Indeed, the great readability of the binary CART trees was decisive when compared with the n-ary CHAID trees that had, even at the first level, a much too high number of nodes to allow for any useful interpretation.

As with other statistical modeling approaches, it is essential to assess the quality of the obtained tree before drawing any conclusion from it. Our point is to make it clear that the validation criteria are largely dependant of the pursued goal. Especially, it is worth mentioning that the misclassification rate, which is most often the only validation criterion provided by software programs, is of little help in non-classificatory settings.

3 The applied study

We begin by setting the applied research framework, then we sketch our global analysis procedure and, finally, we present selected findings.

3.1 The topic: Female labor market participation in Switzerland

Female labor market participation reveals significant differences across countries. In Europe, scholars often identify at least two general models: a Mediterranean one (Italy, Greece, Portugal, etc.) versus a model typical for Central and Northern Europe [9]. The first is represented by an inverse L-shaped curve of the *activity* or *participation rate* by age, where after a short period of high rate (at entry in the labor market) the proportion of women working or seeking work begins to steadily decline up to retirement. The same graph depicts a M-shaped curve in Central and Northern European countries, characterized by high participation at entry, followed by a temporary decline during the period of motherhood and childbearing, and a subsequent comeback to work, up to a certain age where the process of definite exit starts.

In this respect, Switzerland is an interesting case. Firstly, Switzerland is a country placed in a nutshell across the Alps, which are considered as one of the cleavages dividing Southern Europe from Central and Northern Europe. Secondly, there are three main languages, spoken by people living in three geographically distinct regions: French in the western part on the border with France, German in the northern and eastern parts on the border with Germany and Austria, and Italian south of the Alps in a region leading to Italy.

The existence of three regions, with highly distinctive historical, social and cultural backgrounds and characters, and the fact that the Italian-speaking part is divided from the other two by the Alps highlight the very specific particularity of this country for a cross-cultural analysis of the female participation in the labor market. Moreover, the fact that the comparative analysis is performed amongst regions of the same country guarantees, despite differences stemming from the Swiss federal system, a higher degree of comparability on a large series of institutional, political and other factors than one would get with cross-country studies.

The idea of the research project was to verify the existence of differing cultural models of female labor market participation, by analysing activity rates and hours worked per week — in terms of proportions of full-timers and part-timers — across the three linguistic regions in Switzerland, by using the SPC 2000 data.

To shortly describe the data, we can say that the Federal Statistical Office made us available a clean census dataset covering the about 7 millions inhabitants of Switzerland. For our study, only the about 3.5 millions women were indeed of interest. In the preprocessing step we disregarded young ($< 20, 23\%$) and elderly ($> 61, 18\%$) women, as well as non Swiss women not born in Switzerland (1.6%), i.e. about 43% of the women. This left us with about 2 millions cases. Finally, we dropped about 350000 cases with missing values, and hence included 1667494 cases into the analysis.

3.2 The empirical research design

The research procedure used classification trees at two different stages, with differing but complementary purposes.

A tree was first grown in what we refer to as the *preliminary step*. Its main goal was, in the spirit of structured induction [13] and local pattern detection [4, 12], to find a sound partition of the analysed population into a limited number of homogeneous groups — homogeneous female labor supply behavior in terms of activity and choice between full-time and part-time employment — over which a tailored analysis could be performed. In other words, in order to avoid an "average" analysis at global level, classification trees have been used to structure the research and to identify those groups of the population which could be used to guide subsequent analysis.

This first step was run on the whole Swiss female population of age 20 to 61, using their *labor market status*[3] as outcome variable, and general socio-demographic characteristics (civil status, mother/non mother, ...) as predictive attributes. From this, a robust partition in three groups was chosen, as the best compromise between level of details for the subsequent analysis and population size of each group. The three groups are the *non-mothers*, the *married or widowed mothers*, and the *divorced or single mothers*. The first group is composed by 609,861 women (36.6%), the second one by 903,527 (54.2%) and the third one by 154,106 (9.2%).

The second application of classification trees took place in the analysis of cross-cultural female labor supply behavior for each selected group. Here again the outcome variable was the *labor market status* of the women. A much broader series of predictive variables was retained however: age, profession, educational level, mother/non mother, number of kids, age of last-born kid, type of household, etc. Before growing trees, we carried out a series of simple bivariate analyses between the labor market status and each selected predictive attribute. This helped to identify the most relevant attributes for the retained cross-cultural perspective. The analysis of their raw impact on the labor status provided useful indications on how important each one is when it comes to explaining the female labor supply behavior.

Classification trees have been produced separately for each region and then compared, as described in the next section, in order to analyse cultural patterns

[3] Labor market status is a categorical variable with four values: full-time active (at least 90% of standard hours worked per week), long part-time active (50% to 90%), short part-time active (less than 50%) and non active, where active means working or seeking for a job.

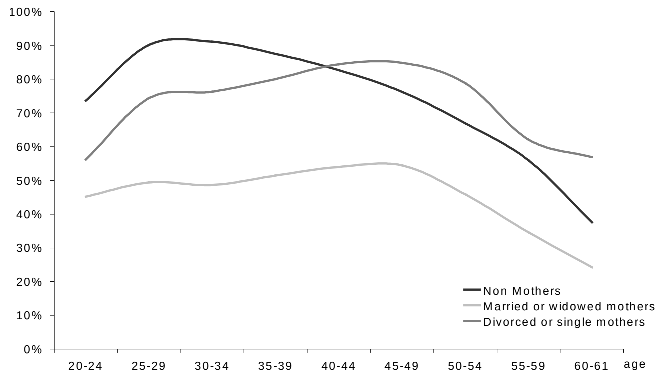

Fig. 1. Activity rates by age of the three groups selected

in the participation behavior of the main language regions in Switzerland. At this stage, classification trees and traditional analyses have been used in a complementary way allowing for interplay between them. This proved to be highly productive in stimulating the knowledge discovery process as well as in analysis and understanding of relevant phenomena.

It is worth mentioning here that the final trees retained are simplified versions of those that resulted from the stopping and pruning criteria. They were selected on the basis of comprehensibility and stability factors. We checked for instance that the splits retained stayed the same when removing randomly 5% of the cases from the learning data set.

3.3 Results

Definition of the groups of analysis. The three groups identified in the preliminary local pattern detection step appear to exhibit a high degree of inter-group diversity combined with a significant intra-group homogeneity. Inter-group diversity is highlighted by the very specific participation rates by age depicted in Figure 1.[4]

Comparison of part-time versus full-time employment reinforces the picture by highlighting the very different choices made by working women of the three groups: a majority of the non-mothers choose full-time employment all along their professional life, divorced and single mothers switch from part-time jobs during motherhood and early childbearing to full-time (or long part-time) jobs, and the married and widowed mothers prefer short part-time employment in the majority of cases.

[4] Figure 1 demonstrates that the M- or L-shaped curves encountered in cross-country studies may result from the superposition of group specific curves.

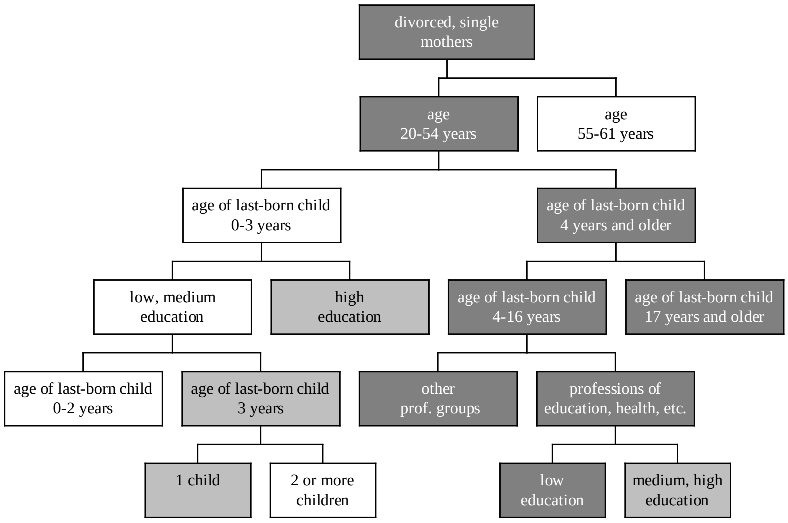

Fig. 2. Tree for participation of divorced or single mothers, Italian speaking region

The determinants of labor supply behavior of divorced and single mothers. In order to identify cultural models of female labor supply, three trees (one per region) were generated for each group. These — in combination with the results of the traditional bivariate analyses — were compared and thoroughly analysed in terms of structure and results. We give hereafter a very brief overview of the main results for the third group, i.e. divorced or single mothers.[5] For details interested readers may consult the research report [6]. In Figure 2 and 3, white background is used for nodes with a majority of non active women, light grey for a majority of part-timers and dark grey for a majority of full-timers. We see that opting for inactivity seems to be much more frequent in the Italian speaking region.

Profession and *age of the mother* point out specific groups with particular distinct behaviors. The former puts apart a group of professions — in the fields of health, education, sciences, etc. — which are known to be characterized by high proportions of part-time jobs. Age plays a central role in the Swiss Italian (Figure 2) and in the (not shown) Swiss German tree by clearly splitting the period of active life (up to age 54-55), from that of the definite withdrawal from the labor market.

The *age of the last-born child* appears as the most discriminative factor in all the regions demonstrating the very central role within the family-work conflict of being mother for the women of this group, who live mainly in single-parent households. The most significant differences across the Swiss language regions

[5] For space reasons, only the (slightly simplified) trees of the Italian and French speaking regions are presented.

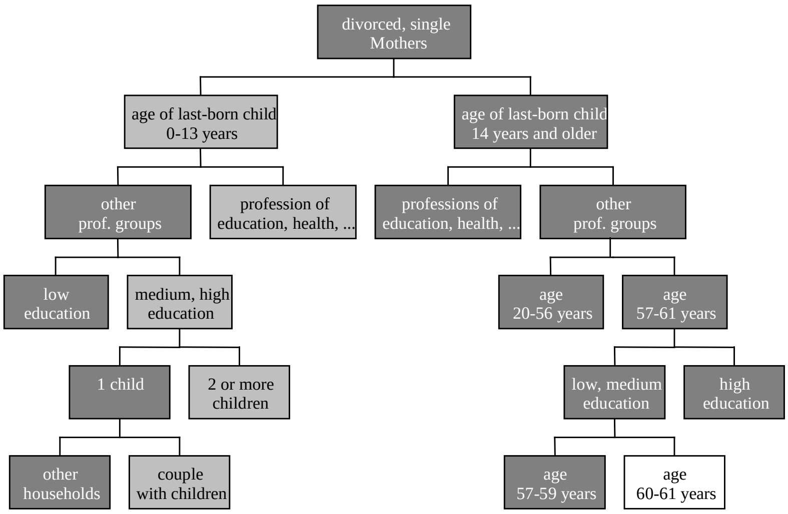

Fig. 3. Tree for participation of divorced or single mothers, French speaking region

appear in this variable, namely on its position in the tree, its split values and the distribution in the classes of the resulting partition. There is a high proportion of inactivity among Swiss Italian women living in single-parent household when last-born child is 2 years old or younger. This proportion decreases for the first time when the child is 3 (access to public kindergarten) and for the second time when the child reaches 6 (access to primary school). Swiss German women also quit the labor market, but re-enter sooner, while Swiss French are almost indifferent to this factor, showing constant activity rates per age of last-born child.

In all the three regions, *educational level* has a strong influence on female labor supply. The higher the educational level, the higher the proportion of active mothers and the lower the proportion of full-timers. This double effect is particularly evident in the Italian speaking and German speaking regions, when last-born child is very young (less than 4 respectively 6 years). Mothers with elementary or intermediate level education decide in the majority of cases to quit their jobs and to stay at home during this period, while mothers of higher education work on a part-time basis.

The *presence of the partner* and the *number of children*, which strongly influence the behavior of married women have only limited effect on divorced women.

4 Validating non-classificatory trees

As mentioned earlier, the classification error rate is not satisfactory when applied to trees for non classification purposes. For instance, if the majority outcome class is the same in all leaves, the reduction in the classification error provided

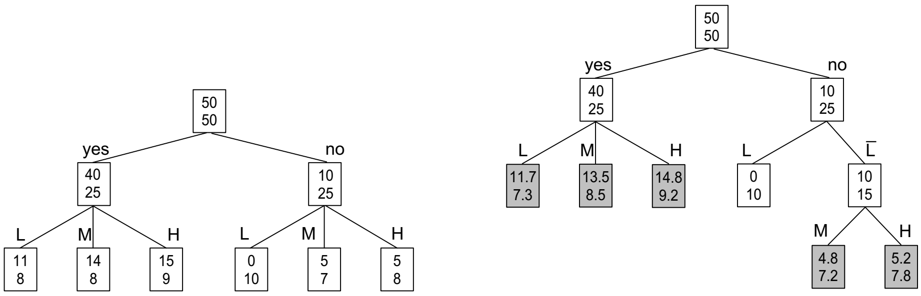

Fig. 4. [Left] Saturated tree for two predictors: partner (yes, no) and education (L=low, M=medium, H=high). Response is activity rate (part time, full time). [Right] Induced tree (white node) and expected counts for the leaves of the saturated tree.

by the tree when compared with the root node is null. Despite this zero gain in terms of classification error, the tree may nevertheless exhibit significant differences between leaves regarding their probability distributions. These differences are valuable knowledge from the descriptive standpoint and should hence be accounted for.

In [10], among other alternatives, a deviance measure is proposed to evaluate how far the fitted tree m is from the target table m_T associated to the saturated (finest) tree that can be drawn from the data. This saturated (or finest) tree is the one that results from all possible splits that can be defined from the values of the predictors. For example, considering only the two predictors partner (yes, no) and education (low, medium, high) we get a saturated tree with $2 \times 3 = 6$ leaves, each of them corresponding to a possible profile in terms of predictors (see Figure 4, left).

We call target table the cross tabulation of the response variable with the set of possible different profiles (leaves of the saturated tree), i.e. the table which columns are the outcome counts in the leaves. Let n_{ij} be the count of cases that belong to the ith outcome value in the jth leaf of the saturated tree m_T, and let $\hat{n}_{ij}$ be the count expected from the induced tree m. The latter is obtained by distributing the total $n_{\cdot j}$ of cases in the saturated leaf j, according to the distribution of the leaf of the fitted tree that covers those cases (see the example in Figure 4, right). The deviance reads then

$$D(m|m_T) = -2 \sum_{i=1}^{\ell} \sum_{j=1}^{c} n_{ij} \ln(\hat{n}_{ij}/n_{ij}) \ ,$$

considering the expression under the summation sign as zero when $n_{ij} = 0$.

This deviance permits to build useful statistical tests. Especially, the difference in deviance between two nested trees can be compared with a Chi-square distribution to test the significance of the difference between them (See [10]). We could thus provide statistical arguments for simplifying or complexifying some descriptive tree. We can also derive from it a penalized deviance like the BIC that could help to determine the best trade-off between fit and complexity.

Though the idea is appealing, we must first define the target table in order to compute the deviance. This is quite easy as long as only a limited number of attributes with each a limited number of values are used. In our real full-scale application, it happened, nevertheless, to be a virtually unmanageable task. Indeed, the combination of the observed values of the attributes considered gives rise to more than a million different profiles, i.e. columns for the target table.

We therefore considered only a partial deviance $D(m|m_{T^*})$ that measures the departure from the partition m_{T^*} defined by the mere split values used in the tree. In other words, we compare the partition defined by the tree with the finest partition that can be achieved by combining the groups of values defined by the splits. For example, the induced tree of our illustrative example (Figure 4) distinguishes nowhere between medium (M) and high (H) education level. The target m_{T^*} for the partial deviance would then be defined by the mere 4 profiles (yes,L), (yes,$\bar{L}$), (no,L), (no,$\bar{L}$) instead of the 6 defined by the saturated tree. The resulting target table is clearly somewhat arbitrary. The consequence is that the partial deviance has no real meaning by itself. However, letting m_T be the true target, we have, thanks to the additivity property of the deviance, $D(m|m_T) = D(m|m_{T^*}) + D(m_{T^*}|m_T)$. Hence, $D(m_{T^*}|m_T)$ being independent of the fitted tree, the difference in the partial deviance of two nested trees m_1 and m_2 remains unchanged, whatever target m_{T^*} is used. Thus, comparing the deviance of the fitted tree with that of any of its subtree permits us to test the statistical significance of the gain over this subtree. In particular, comparing with the deviance of the root node we get a Likelihood Ratio Chi-square test similar to those used in logistic regression.

For our application, we obtained the partial deviances with SPSS. Two deviances were computed, namely $D(m_0|m_{T^*})$ and $D(m_0|m)$, where m_0 is the root node and m the fitted tree. We first recoded the attributes so as to group the values that remain together all over the tree. It was then easy to build a profile variable taking a different value for each observed combination of the recoded values. The target table m_{T^*} results from the cross tabulation of this profile variable with the outcome variable, i.e. the type of participation in the labor market. The deviance $D(m_0|m_{T^*})$ is finally simply the independence Log Likelihood Ratio Chi-square statistic (LR) for this target table. Likewise, the deviance $D(m_0|m)$ between the root node and the fitted tree is the LR statistic for the table that cross tabulates the leave number with the response variable. Since the trees were grown with Answer Tree [14], we readily obtained the leave number of each case with the SPSS code generated by this software. The deviance $D(m|m_{T^*})$ that measures how far the tree is from the target, is obtained as the difference between the two computed deviances:

$$D(m|m_{T^*}) = D(m_0|m_{T^*}) - D(m_0|m) \ .$$

Similar relations hold for the degrees of freedom. Recall, however, that the partial deviance has no real meaning by itself. Its interest lies in that it permits to testing statistically differences between nested trees.

The partial deviance can also be used for defining AIC and BIC criteria, since only differences in the values of the latter matter. We thus compute the BIC value for a tree m as

$$\mathrm{BIC}(m) = D(m|m_{T^*}) - \ln(n)(c^* - q)(\ell - 1) \ ,$$

where n is the number of cases, c^* is the number of different profiles in the target table m_{T^*}, q the number of leaves of the tree and ℓ the number of outcome classes, i.e., in our application, the 4 types of participation in the labor market. The product $(c^* - q)(\ell - 1)$ gives the degrees of freedom associated with the partial deviance. Recall, that according to Raftery [8], a difference in BIC values greater than 10 provides strong evidence for the superiority of the model with the smaller BIC, in terms of trade-off between fit and complexity.

It is also very convenient to measure the gain in information in relative terms. Pseudo R^2's are not very informative when computed from partial deviances, due to the arbitrariness of the target table. It is preferable to consider the percent reduction in uncertainty about the outcome distribution achieved with the tree when compared to the root node. The association measure τ of Goodman-Kruskal [3] and the uncertainty coefficient u of Theil [15], both provided by SPSS, are two such measures. The first is the proportion of reduction in the quadratic entropy and the second in Shannon's entropy. These two indices produce always very close values. They evolve almost in a quadratic way from no association to perfect association. Their square root is therefore more representative of the position between these two extreme situations.

Table 1 reports some of the quality figures we have computed for each of the three regional trees: CHI for the Italian speaking, CHF for the French speaking and CHG for the German speaking region. The deviances $D(m_0|m)$ are all very large for their degrees of freedom. This tells us that the grown trees clearly improve the description as compared to the root node. The deviances $D(m|m_{T^*})$, not shown here, are also very large indicating that there remains room for improving the fit. The difference ΔBIC in the BIC values between the root node and the grown trees lead to a similar conclusion. They are largely superior to 10, providing evidence of the superiority of the grown trees over the root node. For CHI and CHF, the BIC values of the grown trees are also much smaller than those of the associated saturated trees. This is not the case, however, for CHG. There is thus definite room for improvement in this last case. Remember, however, that we are interested in pointing out the main forces that drive the female participation in the labor market. Hence, we have a comprehension purpose, for which increasing complexity would undoubtedly be counter productive. This is

Table 1. *Trees quality measures*

	q	c^*	p	n	$D(m_0\|m)$	d	sig.	ΔBIC	u	$\sqrt{u}$
CHI	12	263	299	5770	822.2	33	.00	536.4	.056	.237
CHF	10	644	674	35239	4293.3	27	.00	4010.7	.052	.227
CHG	11	684	717	99641	16258.6	30	.00	15913.3	.064	.253

typical in socio-economic modelling, where we cannot let the modelling process be entirely driven by purely statistical criteria. Indeed, the trees need to make sense.

The Theil uncertainty coefficient u seems to exhibit a low proportion of gain in uncertainty. However, looking at its square root, we see that we have covered about 25% of the distance to perfect association. Furthermore, the values obtained should be compared with the maximal values that can be achieved with the attributes considered. For the target table, which retains a partition into c^* classes, the u is, respectively, .28, .24 and .23. The square root of these values is about .5, i.e. only about twice the values obtained for the trees. Thus, with the grown trees that define a partition into q classes only, we are about half the way from the target table.

To illustrate how these measures can be used for tree comparison, consider the simplified tree in Figure 2 obtained from CHI by pruning a branch grown from the node "age 55-61 years". The original tree CHI has $q = 12$ leaves, while the simplified tree has only 9 terminal nodes. For the latter, we get $D(m_0|m) = 799.4$ with $d = 24$, which leads to a significant difference in deviances of 22.8 for 9 degrees of freedom. The $\Delta(BIC)$ between the two trees is however 55.1 in favor of the simplified tree, the loss in fit being more than compensated by the complexity reduction. This statistically grounds the retained simplification.

5 Conclusion

The experiment reported demonstrates the great potential of classification trees as an analytical tool for investigating socio-economic issues. Especially interesting is the visual tree outcome. For our study, this synthetic view of the relatively complex mechanisms that steer the way women decide about their participation in the labor market provided valuable insight into the studied issue. It allowed us to highlight regional cultural differences in the interaction effects of attributes like age of last-born child, number of children, profession and education level that would have been hard to uncover through regression analysis, for example.

It is worth mentioning that generating reasonably sized trees is essential when the purpose is to describe and understand underlying phenomenon. This is not the case with classification settings. Indeed, complex trees with many levels and hundred of leaves, even with excellent classification performance in generalization, would be too confusing to be helpful. Furthermore, in a socio-economic framework, like that considered here, the tree should make sense from the social and economic standpoint. The tree outcomes should therefore be confronted with other bivariate analyses and modeling approaches. Our experience benefited a great deal from this interplay.

Now, as end users, we had to face the lack of suitable validation measures provided by the tree growing software programs for our non-classificatory purpose. The main novelty proposed here is the partial deviance and the efficient way to compute it. The relevance of the partial deviance is based on the additivity property of the deviance. Alternative chi-square divergence measures

(Pearson for example) could be considered. However, since they do not share the additivity property, we could not as easily derive partial forms of them.

Although we were able to obtain relevant indicators and statistics afterwards by means of classical cross tabulation outcomes, we would urge software developers to include such validation measures in their software output. Even more, we are convinced that better descriptive trees can be generated when maximal change in overall deviance or BIC values is used as a criterion for growing trees.

References

[1] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: *Classification And Regression Trees.* Chapman and Hall, New York (1984)
[2] Fabbris, L.: *Statistica multivariata: analisi esplorativa dei dati.* McGraw Hill, Milano (1997)
[3] Goodman, L.A., Kruskal, W.H.: Measures of association for cross classifications. *Journal of the American Statistical Association* **49** (1954) 732–764
[4] Hand, D.J.: The framing of decisions as distinct from the making of decisions. In Herzberg, A.M., Oldford, R.W., eds.: *Statistics, Science, and Public Policy VI: Science and Responsibility*, Kingston, Ont., Queen's University (2002) 157–161
[5] Kass, G.V.: An exploratory technique for investigating large quantities of categorical data. *Applied Statistics* **29** (1980) 119–127
[6] Losa, F.B., Origoni, P.: Partecipazione e non partecipazione femminile al mercato del lavoro. Modelli socioculturali a confronto. Il caso della Svizzera italiana nel contesto nazionale. Aspetti statistici, Ufficio cantonale di statistica, Bellinzona (2004)
[7] Quinlan, J.R.: *C4.5: Programs for Machine Learning.* Morgan Kaufmann, San Mateo (1993)
[8] Raftery, A.E.: Bayesian model selection in social research. In Marsden, P., ed.: *Sociological Methodology.* The American Sociological Association, Washington, DC (1995) 111–163
[9] Reyneri, E.: *Sociologia del mercato del lavoro.* Il Mulino, Bologna (1996)
[10] Ritschard, G., Zighed, D.A.: Goodness-of-fit measures for induction trees. In Zhong, N., Ras, Z., Tsumo, S., Suzuki, E., eds.: *Foundations of Intelligent Systems, ISMIS03.* Volume LNAI 2871. Springer, Berlin (2003) 57–64
[11] Ritschard, G., Zighed, D.A.: Qualité d'ajustement d'arbres d'induction. *Revue des nouvelles technologies de l'information RNTI* **E-1** (2004) 45–67
[12] Rüping, S.: Learning with local models. In Morik, K., Boulicaut, J.F., Siebes, A., eds.: *Local Pattern Detection.* Volume 3539 of *LNCS.*, Berlin, Springer (2005) 153–170
[13] Shapiro, A.D.: *Structured Induction in Expert System.* Adison-Wesley, Wokingham (1987)
[14] SPSS, ed.: *Answer Tree 3.0 User's Guide.* SPSS Inc., Chicago (2001)
[15] Theil, H.: On the estimation of relationships involving qualitative variables. *American Journal of Sociology* **76** (1970) 103–154