Teachers' Preferences for School Attributes: the Case of France

Pierre BUI QUANG

Master Economics and Public Policy
PhD Track in Economics

Master Dissertation

Supervisor: Christian Belzil
Second member of the jury: Robert Gary-Bobo
Academic Defense: the 19th of June 2011

Academic year 2010 - 2011
Introduction

In several developed countries, understanding the allocation of teachers across schools and areas has become a crucial point of focus. Indeed, any education system can be considered to pursue two objectives: the first one is to be efficient, in the sense of good academic achievement, the second is to be fair, in the sense that the system should provide identical opportunities to any students.

In these circumstances, several authors have pointed out the importance of teachers in the achievement of students. Among these studies, one can refer to Rivkin, Hanushek and Kain (2005) or Clotfelter, Ladd and Vigdor (2006).

Therefore, the economic literature has very early been interested in studying the teacher labor market. The various issues associated to the teacher labor market have also been the subject of a large literature review by Guarino, Santibañez and Daley (2006). They notably consider that the literature can be divided into the following categories, defined according to the questions to be dealt with:

1. What are the characteristics of individuals who enter the teaching profession?
2. What are the characteristics of individuals who remain in teaching?
3. What are the characteristics of schools and districts that successfully recruit and retain teachers?
4. What type of policies show evidence of efficacy in recruiting and retaining teachers?

In this framework, the goal of the present dissertation is quite transverse. Specifically, this dissertation is interested in the evaluation of teachers' preferences for school attributes. Indeed, it does not require much more than common sense, given the common knowledge about the heterogeneity in school difficulties, to envisage the fact that working conditions are crucial to the utility of teachers. This fact has been quite largely described – and methodologies have been keenly debated – in the literature, as will be shown later in this dissertation.

The utility of evaluating teachers' preferences has much to do with the issues (2), (3) and (4) of Guarino, Santibañez and Daley (2006). Indeed, teachers' preferences for school attributes allow us to describe the supply side of the teacher labor market. In turn, this may have an important impact on the repartition of teachers across schools.

This repartition of teachers can be harmful for the normative objective of providing students with equal opportunities. In many countries for example, teachers’ preference toward the most able students tends to increase the difficulties faced by disadvantaged schools in recruiting and retaining teachers. Therefore, these schools will tend to face an higher turnover of their pedagogical staff, and will be likely to concentrate the less experienced and the less skilled teachers.

Nonetheless, the impact of such preferences on the allocation of teachers is likely to crucially depend on the structure of the national labor market for teachers. On this topic, we are not aware of any study that has attempted to formally model the mobility decisions of French
teachers. This dissertation will maybe provide a first insight in this matter, and in particular, it will be shown that the structure of the French teacher labor market does not allow to evaluate teachers' preferences using the methodologies that the literature has used until now. This model for example, will make clear the reasons of the following paradox: it can be the case that a teacher who enjoys a lot teaching advantaged students is found to consume more disadvantaged schools than other teachers.

The explanation for this paradox is to be found in the particularities of the French system of assigning teachers. Overall, this system consists in a centralized mechanisms that does not allow schools and teachers to agree on a match, such that the “classical” rules of decentralized allocation do not hold. However, it is known that this system induces an unequal repartition of teachers. In fact, this was one of the points made by a recent report from the French Cour des Comptes.\(^1\)

Last but not least, it must be said that in this dissertation, the question of the evaluation of teachers' preferences will mainly be dealt with theoretically. Unfortunately, we were not allowed to have access to any source of data relevant for our purpose.

The rest of the dissertation will be organized as follows. Section 1 will present in a non-formal way the important features of the educational system in France, it will also review the literature that has studied the mobility of French teachers. Section 2 is a literature review of the methodologies and results that have been used and obtained on the topic of teachers' preferences for school attributes. Section 3 provides a theoretical model of teachers' mobility decisions in the case of the French labor market for teachers, and discusses how it could be used in order to obtain teachers' preferences from duration data. Section 4 proposes an econometric specification to account for heterogeneity of preferences in a plausible way. Section 5 discusses alternative perspectives and further topics that we have met in this dissertation.

---

\(^1\) “L'éducation nationale face à l'objectif de la réussite de tous les élèves”, report from the Cour des Comptes, May 2010
Section 1 – Teacher mobility in France

In this section, we present in a non-formal way the main features of education in France, with particular attention for the assignment of teachers. Our description corresponds to the educational system in secondary education in the public sector (that accounts for around 80% of secondary students).

This market is characterized by a strong administrative centralization. Not only – as this is the case for many countries – are wages set nation-wide, but also, the mobility schedule of teachers across school is planned centrally. The first paragraph will be devoted to the description of the French educational system.

As far as we know, no studies have attempted to evaluate preferences of French teachers for school attributes. However, several studies have been looking at teachers mobility patterns and composition of pedagogical staff depending on school characteristics. The second paragraph will review this literature.

Unfortunately, for the present dissertation, we did not obtain any relevant data that would have allowed us to perform an empirical estimation of teachers' preferences. We made contact with the administration within the ministry of Education that is responsible of managing statistical data, but we were not allowed to access them. However, our experience allowed us to have some idea concerning the data that one can expect from a collaboration with the ministry of Education. In a third paragraph, we will provide a description of these datasets, that may be useful for researchers.

**The French educational system**

General remarks on French teachers and the diversity of treatments

The French educational system does not allow much diversity in the way teachers are treated. Unlike countries like the US, wages are fixed and set nationally. And so are working hours. Nevertheless, three sources of diversity remain.

First, everything else being equal, there is a hierarchy between teachers according to their initial certificate. More specifically, students who are willing to become secondary teachers must pass a competitive examination that evaluates mostly their academic skills. But several options are available to candidates, that induce different types of certificate, and there exists a clear hierarchy of difficulty between them. This hierarchy also ranks teachers according to their certificate. For example, the two main certificates for the “enseignement général” are the “agrégation” and the “CAPES”. And it is quite clear that the “agrégation” is considered as more difficult than the “CAPES”. Therefore, the “agrégés” (teachers who have the certificate of “agrégation”) are offered better working conditions compared to the “certifiés” (teachers who hold the “CAPES”), among which higher wages and the possibility to teach in higher education.

Secondly, this hierarchy according to the certificate is doubled with a hierarchy of grades that

---

2 Or, to be more accurate, it is partially planned by local administration, but according to national rules
reflects the promotion of a teacher within her class of certificate. The speed with which higher grades are attained depends partially on the evaluation of the teacher's quality by the administrative staff. However, seniority is the main driving force.

Thirdly, since the end of the eighties, there exists in France a class of schools, called priority schools (“Zones d’Education Prioritaire”, abbreviated ZEP) that gathers the most disadvantaged schools that have for example low academic results and suffer from violence issues. Teachers who have tenure in such schools are offered some compensation, such as increased wages.

As an illustration of this diversity, one may consider the following table.

### Table 1: The diversity of treatments of French teachers

<table>
<thead>
<tr>
<th>Seniority in teaching</th>
<th>Teachers “certifiés”</th>
<th>Teachers “agrégés”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“lower bound” net wage</td>
<td>“upper bound” net wage</td>
</tr>
<tr>
<td>Internship</td>
<td>1 584 €</td>
<td>1 584 €</td>
</tr>
<tr>
<td>&gt; 2 years</td>
<td>1 666 €</td>
<td>1 666 €</td>
</tr>
<tr>
<td>&gt; 10 years</td>
<td>1 805 €</td>
<td>1 913 €</td>
</tr>
<tr>
<td>&gt; 20 years</td>
<td>2 365 €</td>
<td>2 543 €</td>
</tr>
<tr>
<td>&gt; 30 years</td>
<td>2 543 €</td>
<td>3 026 €</td>
</tr>
</tbody>
</table>


In this table, the “lower bound” net wage corresponds to the minimum wage that a teacher would receive depending on her seniority, ie it is the wage that a teacher would receive if during her entire career, her evaluations by the administrative staff was making her promotion the slowest possible. Identically, the “upper bound” net wage is the maximum wage depending on seniority.

For each cell in this table, one can add the gratification associated with teaching in a ZEP school. This takes the form of a lump sum gratification of 96 €, that represents between 6% and 2,5% of the baseline net wage.

### Teacher mobility rules

Every year, teachers who already hold a position can request a transfer. To this purpose, they must declare a hierarchical list of preferred positions. New teachers, who are not assigned yet, also have to go through this procedure. The administration then recovers all the wishes and allocates teachers to their preferred school with an available position. In other words, for each teachers, the administration reads the list of wishes from top to bottom and stops as soon as it finds a wish that can be satisfied.

What happens if several teachers are competing for the same position? In the French educational system, teachers are holding *bonuses* that are the result of their career. In the case of competition for a position, the teacher with higher bonuses usually gets priority. We
precise “usually”, because rules on this matter specify that the administration can always go over it. However, it is broadly respected and monitored by unions.

Obviously, the way bonuses are attached to teachers is crucial to understand the resulting allocation. Indeed, it is straightforward that higher bonuses are associated with an higher probability to obtain a “good” position, and therefore is associated with higher expected utility. Broadly, the level of bonuses is determined by (i) the civil situation, (ii) the initial certificate and teachers evaluations by the administrative staff, (iii) the seniority in teaching, and (iv) the seniority in the current school.

Two remarks can be done. First, (ii) and (iii) are likely to be correlated with teacher's quality. Therefore, quality teachers (in the sense of (ii) and (iii)) would tend to have less constraint when choosing their position and thus, good schools (in the sense of generally valuable to teachers) will tend to concentrate the best teachers.

Second, (iv) allows us to consider a strategic component in the decision of transferring or not, because a successful transfer will induce a loss of the bonuses associated to (iv). Specifically the question for a teacher is: “should I ask for a transfer now, and risk to have only second-best wish satisfied and lose bonuses, or should I stay and capitalize bonuses hoping to have first-best wish satisfied in the future?”.

Another important point to mention is the fact that teachers may be forced to accept a position even if they did not express the desire to fill this position. For example, this happens as soon as a teacher requests a tenure in another district. In this case, the teacher is assumed to accept any position in this district, if the procedure were to fail in assigning her to one of the schools she applied for. We will often refer to this possibility as the “bottom line wish”, because it must appear physically at the bottom line of the list of preferred positions. Thus, as soon as a teacher is asking for a position which does not belong to her current district (the administration refers to this part of the allocation as “mouvements inter-académiques”), she must accept, as a bottom line wish, to be assigned to any schools in the district. This rule also applies to new teachers, who do not have a current district yet. Hence, the worst schools are likely to welcome new teachers who are more likely to touch the bottom in the administrative procedure, because they did not accumulate enough bonuses.

Broadly speaking, what is interesting from an econometric point of view is that because of this system, the selection of teachers is almost free of “non-objective” criteria. School principals have almost (it may depend on the status of the school) no power to choose their pedagogical staff or to bargain so as to attract particular teachers. Administrators have some arbitrary power, but it is almost unused due to the pressure of unions. Thus, in the observed school-teacher matching, we have removed all the issues related to the impact of unobserved characteristics on the selection of the teacher by the school.

**Inter-district mobility: an illustration**

As an illustration, one can consider Table 1 displayed below. In this table are represented the “barres d'entrée” per districts from the years 2005 to 2010, as recorded by the SNES (one of the principal teacher union in France) for mathematics teachers. Loosely speaking, these “barres d'entrée” represent the minimum level of bonuses required to be accepted in the associated district. More precisely, it corresponds to the level of bonuses of the last mathematics teachers accepted in the district for the associated year.

As the distribution of these thresholds is very difficult to obtain from the general game
implemented by the administrative framework, it is not easy to interpret them. Even for “normal” districts (not overseas and that gathers a lot of school such that convergence properties can be expected), like Paris, the thresholds do not appear stable. However, we can draw some qualitative lessons from the most extreme cases.

Consider for example the cases of the districts CRETEIL and VERSAILLES. One can see from the table that the level of bonuses required to enter these districts are very low (21 for every years). This can be understood by the fact that these districts are considered among the worst in France, as far as teaching conditions are concerned.

Table 2: Thresholds for assignments in inter-district transfers

<table>
<thead>
<tr>
<th>District</th>
<th>Barre d'entrée 2010</th>
<th>Barre d'entrée 2009</th>
<th>Barre d'entrée 2008</th>
<th>Barre d'entrée 2007</th>
<th>Barre d'entrée 2006</th>
<th>Barre d'entrée 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIX-MARSEILLE</td>
<td>186</td>
<td>58</td>
<td>71</td>
<td>28</td>
<td>71.1</td>
<td>31</td>
</tr>
<tr>
<td>AMIENS</td>
<td>21</td>
<td>71</td>
<td>21,1</td>
<td>21</td>
<td>71</td>
<td>38</td>
</tr>
<tr>
<td>BESANÇON</td>
<td>71</td>
<td>35.1</td>
<td>128</td>
<td>298.2</td>
<td>231.2</td>
<td>261.2</td>
</tr>
<tr>
<td>BORDEAUX</td>
<td>111,1</td>
<td>186</td>
<td>180</td>
<td>200</td>
<td>221.3</td>
<td>244</td>
</tr>
<tr>
<td>CAEN</td>
<td>78</td>
<td>127</td>
<td>100</td>
<td>171.3</td>
<td>55</td>
<td>221,2</td>
</tr>
<tr>
<td>CLERMONT-FERRAND</td>
<td>71,1</td>
<td>71.1</td>
<td>91</td>
<td>118</td>
<td>137</td>
<td>221,3</td>
</tr>
<tr>
<td>CORSE</td>
<td>1231,2</td>
<td>1106.2</td>
<td>896.3</td>
<td>631</td>
<td>681</td>
<td>771,3</td>
</tr>
<tr>
<td>CRETEIL</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>DION</td>
<td>21</td>
<td>35</td>
<td>35</td>
<td>71.1</td>
<td>51</td>
<td>171.2</td>
</tr>
<tr>
<td>GRENOBLE</td>
<td>51</td>
<td>110</td>
<td>61</td>
<td>111</td>
<td>98</td>
<td>141</td>
</tr>
<tr>
<td>GUADÉLOUPE</td>
<td>206</td>
<td>107</td>
<td>21</td>
<td>163</td>
<td>21</td>
<td>81</td>
</tr>
<tr>
<td>GUYANE</td>
<td>71</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>LILLE</td>
<td>71,1</td>
<td>88</td>
<td>221.3</td>
<td>221.3</td>
<td>71</td>
<td>118</td>
</tr>
<tr>
<td>LIMOGES</td>
<td>71</td>
<td>38</td>
<td>111</td>
<td>93</td>
<td>81</td>
<td>171,2</td>
</tr>
<tr>
<td>LYON</td>
<td>71</td>
<td>71</td>
<td>51</td>
<td>71.1</td>
<td>41</td>
<td>71</td>
</tr>
<tr>
<td>MARTINIQUE</td>
<td>1071,1</td>
<td>180</td>
<td>78</td>
<td>71</td>
<td>21</td>
<td>91</td>
</tr>
<tr>
<td>MAYOTTE</td>
<td>21</td>
<td>71</td>
<td>21</td>
<td>111</td>
<td>121</td>
<td>137</td>
</tr>
<tr>
<td>MONTPELLIER</td>
<td>71,1</td>
<td>71.1</td>
<td>71.1</td>
<td>100</td>
<td>58</td>
<td>137</td>
</tr>
<tr>
<td>NANCY-METZ</td>
<td>221,3</td>
<td>100</td>
<td>71.1</td>
<td>28</td>
<td>173</td>
<td>221,3</td>
</tr>
<tr>
<td>NANTES</td>
<td>111,1</td>
<td>221.3</td>
<td>128</td>
<td>171.2</td>
<td>171.2</td>
<td>271,3</td>
</tr>
<tr>
<td>NICE</td>
<td>28</td>
<td>71</td>
<td>48</td>
<td>21</td>
<td>91</td>
<td>31</td>
</tr>
<tr>
<td>ORLEANS-TOURS</td>
<td>71</td>
<td>71</td>
<td>51</td>
<td>71</td>
<td>21.1</td>
<td>81</td>
</tr>
<tr>
<td>PARIS</td>
<td>71</td>
<td>114</td>
<td>68</td>
<td>41</td>
<td>141</td>
<td>221,3</td>
</tr>
<tr>
<td>POITIERS</td>
<td>187</td>
<td>198.2</td>
<td>185</td>
<td>187</td>
<td>138</td>
<td>221,3</td>
</tr>
<tr>
<td>REIMS</td>
<td>71.1</td>
<td>71.1</td>
<td>31</td>
<td>71</td>
<td>21.1</td>
<td>171.2</td>
</tr>
<tr>
<td>RENNES</td>
<td>256,2</td>
<td>303.3</td>
<td>381.2</td>
<td>296.3</td>
<td>240</td>
<td>373,2</td>
</tr>
<tr>
<td>LA RÉUNION</td>
<td>181,2</td>
<td>414.2</td>
<td>121</td>
<td>135</td>
<td>124</td>
<td>121</td>
</tr>
<tr>
<td>ROUEN</td>
<td>71</td>
<td>31</td>
<td>21</td>
<td>71</td>
<td>81.1</td>
<td>71</td>
</tr>
<tr>
<td>STRASBOURG</td>
<td>120</td>
<td>71.1</td>
<td>71.1</td>
<td>138.1</td>
<td>71.1</td>
<td>82</td>
</tr>
<tr>
<td>TOULOUSE</td>
<td>65</td>
<td>81</td>
<td>110</td>
<td>142</td>
<td>173</td>
<td>216</td>
</tr>
<tr>
<td>VERSAILLES</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

Source: SNES website [http://www.snes.edu/Barres-des-mouvements-inter.html](http://www.snes.edu/Barres-des-mouvements-inter.html)
The district of CRETEIL is characterized by an important share of minorities, low academic results (the proportion of students who passed the “brevet” – a national examination for ninth grade students – was 79% against 83% at the national level), and an important share of priority schools (in 2001, 32.5% of lower secondary schools in CRETEIL were labeled as ZEP, whereas the average for metropolitan France was 16.7%).

For the district of VERSAILLES, things are interestingly not so clear. For example, the results at the “brevet” are on average better than in metropolitan France (86% passed it in 2010). On the other hand, the share of priority schools among lower secondary schools was slightly higher than the average in metropolitan France (18.4% in 2001), and the share of priority schools among upper secondary schools was very important compared to the metropolitan France average (18.6% versus 5.2% in 2001). This could be a good illustration of the fact that during inter-district transfers, teachers may be arbitrarily assigned. Indeed, the weak demand for VERSAILLES does not seem to be driven by a low “average” school quality, since average results are quite good. But it can be the case that this weak demand is due to the fact that since this district gathers many ZEP schools, the “bottom-line” wish has a very low value, such that teachers would prefer not to apply to this district.

TEACHER MOBILITY IN FRANCE

In France, few studies have examined the mobility of teachers. This can be explained by the fact that French schools do not usually have to face an important shortage of teachers, and recruitment is not generally speaking an issue. Overall, the value associated with being a teacher is quite high, a value that is driven by an important social consideration, the safety of the position and a good income. Identically, exits from the public school sector are not frequent: in 2002 for example, Cros and Obin (2003) say that 3.7% of teachers in secondary schools stopped teaching, among which less than 3% because of resignation or layoffs (approximately equivalent to the share explained by deceases) and more than 80% because of retirement.

For authors who have studied French teachers’ mobility, the main focus concerns the unequal provision in quality teachers between disadvantaged schools labeled as ZEPs, and the other schools. Even if it is not the core of their article, Bénabou, Kramarz and Prost (2005) are among the first to assess the impact of the ZEP policy on the composition of the pedagogical staff.

For example, they find that despite the increased wage, and everything else being equal, the ZEP status was associated with a deterioration of the statistical qualification of the teaching staff. They interpret this result as a negative signaling effect of the ZEP status that the increased gratifications were unable to compensate. For example, during the period from 1987 to 1992, the share of young teachers – defined as teachers aged below 30 – increased on average by 1.3 percentage points per year after 1989 for schools that were labeled ZEP in 1989.

Prost (2010) pursues this reflection with a greater focus on teachers mobility. Using a logit
model, Prost (2010) estimates the probability of switching schools (versus staying) conditional on teacher and school characteristics. Eventually, she finds that the less advantaged is the school, the higher is the probability that a teacher would switch. For example, an increase by one percent in the share of North African students leads to a marginal increase by 0.018 points of the probability that a teacher would switch (measured at average characteristics), and an increase by one percent in the share of Central or South African students leads to a marginal increase of 0.063 points of the same probability. Identically, an increase by one percent of the share of students not having lunch at school (which is a proxy for the income level of child's family) increases by 0.006 points this probability.

One of the surprising result from Prost (2010) is that when controlling for the social composition of schools, the ZEP status tends to decrease the probability of switching schools. She explains this by the fact that teaching staffs in ZEP school were volunteer for teaching in difficult areas, for example, because of some particular pedagogical program. This hypothesis is supported by the fact that this negative effect is stronger for older teachers (who have accumulated enough bonuses such that their choice is not constrained) than for younger teachers for which it is not significant.

That a positive coefficient on the ZEP variable is consistent with a negative signaling effect of the ZEP status.

On this topic, we would like to provide another explanation based on the negative signaling effect of Bénabou, Kramarz and Prost (2005). Assume for example, that the control variables included by Prost (2010) entirely capture the effect of disadvantaged students on teachers probability of switching such that, the coefficient on ZEP would really reflect the pure effect of the status. One can expect that for teachers who are already in a ZEP school, the status, or label, or reputation of the school should not matter : the decision of switching school will only be based on real variables measuring the difficulty of the school, so that, everything else being equal, teachers in ZEPs should not behave differently that teachers that are not in ZEP. In other words, in the absence of signaling effect, the coefficient on ZEP should be zero.

Now, a positive coefficient could be explained by the ex ante selection that is implemented by a negative signaling effect... Consider the signaling effect of Bénabou, Kramarz and Prost (2005), and imagine that this effect is heterogeneous : for a majority of teachers, it is negative but for some teachers who are interested in teaching in difficult areas, it is positive. For example, let us say that Pierre has a negative a priori on ZEP status because he does not enjoy teaching in difficult schools, whereas it is positive for Jean because he enjoys teaching in difficult areas. Suppose now that we offer two perfectly identically disadvantaged schools to Pierre and Jean, but one has the ZEP status whereas the other does not. According to their preferences, Jean should locate in the ZEP, and Pierre in the non-ZEP. But since Pierre does not enjoy teaching in difficult areas, the probability that he will switch is higher than the probability for Jean, since both schools are really identical. This in turn implies a positive coefficient on the ZEP variable in Prost (2010).

To summarize, if we assume that switching decisions are unaffected by the ZEP status, but that application decisions are, and in a consistent way (the label effect is more negative for individuals who dislike more teaching in difficult schools), then, for identical schools, the label effect will induce an ex-ante selection such
that the ZEP staff will on average be composed of teachers that tend to suffer less dis-utility from teaching difficult students. This in turn could explain the positive coefficient of Prost (2010), assuming that the effect of “school difficulty” is essentially captured by other variables.

Conceptually, this is because the signaling effect magnifies the selection for entrance, but does not affect the probability of exits.

Moreover, Prost (2010) looks at mobility patterns of teachers. She shows that when teachers become older, they tend to move to more advantaged schools, with less students not having lunch at school, less students with North-African nationality and less sixth grade students older than usual. Also, the proportion of teachers who teach in ZEP schools decreases with age: ZEP schools gather 22% of teachers under 25, but 16% of teachers between 31 and 40 and only 12.3% of teachers above 51. This is consistent with a dis-utility for teaching disadvantaged students: when teachers get older, they accumulate bonuses that make them able to widen their choice of schools, and they move to more advantaged schools. Identically, Paris gathers only 1% of teachers under 25 but 4.5% of teachers above 51, consistently with an average preference for the capital.

Also, and interestingly, she computes the average school characteristics among different classes of teacher certificate. It appears that the “agrégés”, on average, are teaching in schools with higher proportion of students not having lunch at school, and higher proportion of students with North-African nationality. They also teach more frequently in ZEP schools (15.1%) compared to the teacher that are “certifiés” (14.7%), despite the fact that the “agrégés” are privileged in their choice of assignment. Prost explains that these figure may be induced by a composition bias: since the “agrégés” have the possibility to require tenures in upper secondary schools, those we observe in lower secondary schools may have specific preferences for disadvantaged students.

On this topic, we will provide an alternative explanation (Section 3) based on the fact that the “agrégés” may be strategically willing to stay longer in ZEP schools in order to be able to go to the best schools later. For example, they may be willing to attain the top Parisian schools. This interpretation would be supported by the fact that Paris gathers an important share of “agrégés” (5.3%) compared to the “certifiés” (2.8%).

Last but not least, Prost shows that between the year 1989 and 2000, the situation in ZEP school has overall deteriorated. For example, in 2000, the share of teachers under 25 in ZEP schools was 35%, whereas it was 22% in 1989.

**IDEAS FOR A DATABASE**

For the purpose of studying the teacher labor market in France, most statistics are not publicly available and should be request to the *Direction de l'Evaluation, de la Prospective et de la Performance* (DEPP). Information regarding the management of teachers should be
asked to this department, or potentially to local administrations that manage infra-district transfers.

School-level data can be obtained from the *Fichiers Standards Enrichis* (FSE), the *Indicateurs pour le Pilotage des Etablissements scolaires* (IPES) and the *Base Centrale Scolarité* (BSC). In particular, the BSC is at the disposition of researchers *via* the Maurice Halbwachs center.

Panel datasets concerning students can also be obtained from the same center (*Panels d'élèves du second degré* year 1980, 1989, 1995). Overall, the Maurice Halbwachs center provide several sources of data on education, one may look at the following website: [http://www.cmh.ens.fr/greco/enquetes.php#FORMATION](http://www.cmh.ens.fr/greco/enquetes.php#FORMATION). The panel dataset *Génération 98*, by the Céreq, can also prove useful regarding the achievement of students.

Also, we believe that a dataset describing the list of preferred positions of teachers could be obtained, since it is used at some point to allocate teachers, and since all teachers are declaring their list of wishes on a single website called I-prof. Moreover, matching teachers to their list of preferred positions should be possible, since the NUMEN number that identifies teachers in the management system of the administration is the same that the one used for expressing mobility wishes.

Moreover, in the previous paragraph, we used the thresholds published by the SNES concerning the district assignment. We know that these thresholds also exist at a school-level and are published (and very likely, are recorded) by the SNES, but access to these data is restricted to the members of the union. To our knowledge, these data have never been used, despite the fact that they may prove very informative. Therefore we believe that further researchers should try to obtain them.
Section 2 - Literature review: measuring preferences for job attributes

In this section, we review the literature – both methods and results – concerning the estimation of teachers' preferences for schools' attributes, and about teachers' mobility across schools. This literature is based on the theoretical work of Rosen (1974) about individual preferences for multi-dimensional products. Three main approaches are illustrated: estimations based on wage differentials, estimations based on duration data, and estimations based on game-theoretic two-sided models. These three approaches will provide the four paragraphs of this section, as we will make a difference between competing risks models and duration models, even if they have much in common.

For each approach, we provide comments on selected articles that are particularly illustrative from our point of view. It is important to notice however, that for some of these articles, the primary focus is not to determine the teacher's valuation for schools' attributes. This is true in particular for duration models, where several authors are interested in exit probabilities. Nonetheless, these exit probabilities are clearly linked with teachers' utilities – in a way that will be illustrated by Gronberg and Reed (1994) – such that we felt authorized to comment these articles in the light of our problematic.

Throughout this section, we advise the reader to draw particular attention to the assumptions underlying each model, more precisely about the optimality of the allocations, the exogeneity of frictions on the labor market, and the representation of the space of opportunities (continuum).

Estimations based on wage differentials: the theory of hedonic wages

The estimation of individual preferences using wage differentials is based on the theory of hedonic wages. This theory has been used in various context: Smith (1979) provides a review of the early literature on this matter. Usually, the literature refers to Rosen and Thaler (1975) as a starting point for a theoretical presentation of the use of wage differentials. However, Rosen and Thaler (1975) is based on a particular job attributes, namely, the life-risks embedded in job positions, such that it harms the generality on this subject. Therefore, we find it more instructive, given the topic of the present dissertation, to comment on Antos and Rosen (1975) who focus on the teacher labor market.

Antos and Rosen (1975)
In Antos and Rosen (AR) (1975), the question is the following: “how much is required to induce white teachers to teach in black schools?” In order to provide an answer to this question, AR use an estimation of the wage equation (the “wage gradient”) in the teacher market. Why does it help?
The reason why the wage gradient may provide us with the marginal willingness to pay for an amenity is based on the fact that wage rates are the results of transactions on the labor market, in which teachers provide a service to a school – service that will be valued according to their characteristics – and where schools provide utility to teachers not only through wage, but also
through working conditions or utility bearing school characteristics. In other words, the decrease in the level of wage that one teacher is willing to accept in exchange of an increase in one dimension of the vector of school characteristics (from a given level) is the equivalent of the hedonic pricing of the teacher for this characteristic (at this level). Now, considering that teachers behave optimally and consider offered wages as exogenous, one can compute this hedonic price.

Suppose that we know the wage function, defined for a given value of the teacher's and the school's characteristics. This function is known to any teacher who considers it as parametric. Thus, the teacher will choose to locate herself in the \((wage, school\ characteristics)\) space so as to maximize her utility. Assuming that this space can be represented as a continuum, it is quite natural to think that this allocation would equalize, for a given dimension, the marginal cost of an increase in this dimension, with the marginal rate of substitution between this dimension and wage.

More formally, denote by \(t\) (fixed) the vector of characteristics of some teacher, and by \(s\) the vector of school characteristics. The wage function of this teacher \(w(t, s)\) is known by the teacher – and considered exogenous – for all \(s\), and \(s\) takes value on a continuous support. Let \(c\) be consumption, taken as a numeraire (dollars) such that \(c = w(t, s) + y\) where \(y\) represents the non-wage income of the teacher. Utility is given by a function \(U(c, s)\) that satisfies usual properties.

The teacher's problem is then to maximize her utility on the set of possible location \((w(t, s), s)\) subject to the constraint \(c = w(t, s) + y\).

Then first-order conditions for this problem yield quite straightforwardly (for any dimension \(s_i\)):

\[
\frac{\partial w(t, s^*)}{\partial s_i} = -\frac{\partial U(c^*, s^*)}{\partial c} \frac{\partial}{\partial s_i} \frac{\partial U(c^*, s^*)}{\partial c}
\]

Or, in other words, the indifference curves in the \((w, s)\)-plane will be tangent to the set of opportunities \((w(t, s), s)\), as displayed in Figure 1.

It is important to point out that these partial derivatives allow us to identify the marginal rate of substitution at a given level of characteristics, for the individuals who locate at this level. For example, in the previous figure, individual \(X\) and individual \(Y\) (identical characteristics \(t\)) are locating at different points in the plane, because they have different valuation for the attribute \(s_i\). More specifically, the teacher \(Y\) has a more important dis-taste for \(s_i\).
Figure 1: $w_i(s_i)$ represents the opportunity set of teachers with characteristics $t$, where all other amenities $s_j$ are fixed. The amenity $s_i$ is here a "bad", in the sense that individuals require to be compensated for an increase in $s_i$. $XX$ and $YY$ represent the indifference curve of two individuals $X$ and $Y$.

In their empirical analysis, AR focus on the marginal willingness to exchange one point of percentage of black students with wages. They assume a linear function of wages with school and teacher characteristics: $w(T, S) = T^\alpha + S^\beta$. In a first specification, they find that between 5$ and 7$ (depending if male or female) should be paid to white teachers for them to accept an increase of 1% in the proportion of black students. For black teachers, estimates suggest a preference for black students: they are willing to accept a decrease of 3$ to 1$ in their wages for an increase of 1% in the proportion of black students.

However, estimates are not significant, and a second specification challenges the first results: the sign of the estimates are reversed (but still not significant)! In this second specification, AR in particular are controlling for student ability and they find strong valuation of white teachers for students with higher abilities (stronger for male than for female), whereas it is the opposite for black teacher: these results are significant.

Hwang, Mortensen and Reed (1998)
The theory of hedonic wages is based on the assumption that individuals maximize their
utility in a static world, with a perfect knowledge of the constraints they are facing. Can the estimation method still holds if we consider a new paradigm?

Hwang, Mortensen and Reed (1998) provide a theoretical answer to this question, in the sense of a criticism of the use of the method of hedonic wages. More specifically, they show that with labor market frictions, as represented under the paradigm of job-search models, the method of hedonic wages is likely to lead to bias estimates.

**DURATION MODELS**

As an alternative way to measure workers’ valuation for job attributes in the case of labor market frictions, Gronberg and Reed (1994) propose to look at duration data.

The proof goes as follows (it is a adaptation of the model used by GR). Let \( s \lambda \) represents the arrival rate of new job offers (\( s \) represents the effort of the individual, which is not necessary infinite since effort is conveniently costly). Let \( X \) describe multi-dimensional job offers characteristics (including for example wage and probability of death), that is valued by the worker according to \( v(X) \). Let \( F(v) \) be the cumulative distribution function describing the distribution of the values associated to a job offer. in other words, if an individual receives a job offer, \( F(v) \) is the probability that the value of this offer is less than \( v \). Finally, let \( \delta \) represents the exogenous separation rate.

In this case, the duration \( Y \) of a job spell is exponential-distributed, such that :

\[
P[Y > y] = e^{-h(w)y}
\]

Where \( h(w) \) is the hazard rate given by :

\[
h(v(X)) = \frac{\delta}{\text{exogeneous rate of separation}} + \frac{s*(v(X))\lambda}{\text{arrival rate of opportunities}} \left[1 - F(v(X))\right] + \frac{\text{probability for an opportunity to be valued above } v(X)}{1 - F(v(X))}
\]

This job hazard rate can be used to compute the marginal willingness to trade amenities \( i \) and \( j \) : \( MWT_{ij} \). Indeed, small computations lead to :

\[
\frac{\partial v}{\partial X_i} = \frac{\partial h}{\partial X_i}
\]

In this case, it becomes quite easy to obtain the hazard rate function – and thus the marginal willingness to trade amenities – from duration data.

Among the important assumptions to be remarked, one may notice that the distribution \( F(v) \) of the values of job opportunities does not depend on the characteristics of the current job position. Nor do the exogenous separation rate or the arrival rate of opportunities.

**Dolton and van der Klaauw (1995)**

The duration approach has been applied by Dolton and van der Klaauw (1995) for teachers in the UK. Their econometric assumption (which is close to the one proposed by Gronberg and Reed (1994)) is to suppose the following function for the hazard rate :

\[
h_i(t) = h(t)e^{X_i(t)' \beta}
\]
Two other specifications allow for unobserved heterogeneity in the hazard rate function, the first one is adding a error-term $\nu_i$ in the linear argument of the exponential with the hypothesis that $e^\nu$ is Gamma distributed with mean 1; in the second one, the error-term takes discrete values according to a multinomial distribution. Overall, including this heterogeneity does not change the results, such that we decide to focus on the results associated with the first specification.

The basic aim of Dolton and van der Klaauw (1995) is not to compute the marginal willingness to trade amenities, but merely to study the effect of job/teacher characteristics on the hazard rate and the probability of leaving the public sector. However, following Gronberg and Reed (1994), one can infer this MWT from their results. For example, they find that an increase in the relative teacher earnings decrease the exit rate, with an estimated elasticity of $-1.5$. Identically, teaching in London decreases the exit rate by $-0.108$. Thus, we can deduce that an individual located in London would be compensated for teaching out of London by an increase of 7.2 percentage points in their relative earnings.

**Competing Risks Models**

Competing risks models are to be seen as an extension of the duration models of previous paragraph, as the similarities between Dolton and van der Klaauw (1999) (“A Competing Risks Explanation”) and Dolton and van der Klaauw (1995) (“A Duration Analysis”) would reveal. The specificity of competing risks model is to consider that the individual is subject to various, mutually exclusive, exit possibilities (various “risks” that may end a duration spell).

**Boyd, Lankford, Loeb and Wyckoff (2005)**

Boyd, Lankford, Loeb and Wyckoff (2005) (BLLW) offer a good illustration of competing risks model on the teacher labor market: notably because of the way they model heterogeneity.

More specifically, BLLW assume that every year, teachers are given the following options: stay in the same school, transfer to another school (in the same district or in another district) or quit teaching in the public sector. Every period, the teacher must chose between these options, and decides according to the one that brings the maximum utility.

More formally, BLLW assume that the utility associated with each option $k$ is given by:

$$ U_{it}^k = \nu_{it}^k + \varepsilon_{it}^k $$

where the $\varepsilon$'s are extreme-value type 1 distributed, such that the probability to choose the option $k$ is given by:

$$ p_{it}^k = \frac{e^{\nu_{it}^k}}{\sum_{j=1}^{4} e^{\nu_{it}^j}} = \frac{1}{1 + \sum_{j=2}^{4} e^{\tilde{\nu}_{it}^j}} $$

where $\tilde{\nu}^k = \nu^k - \nu^1$.

BLLW assume the following econometric specification:
\[ \hat{V}_{it}^k = \beta_0 + \beta_1^k z_{it} + \beta_2^k x_{it} + z_{it}^T \theta^k x_{it} + \tau_i + \delta_i x_{it} + \eta_i \]

where \( z_{it} \) are individual characteristics, \( x_{it} \) are school characteristics, and \( \tau \) is a time specific error-term, and \( \delta \) and \( \eta \) are individual error-terms.

Two things are important to point out. First, the model allows for an important heterogeneity of individual tastes for school attributes. Indeed, not only is the taste for a single attribute given by a linear function of the teacher's characteristics \((z_{it}^T \theta^k)\) but also, this linear function allows for an individual random effect \( \delta_i \) for each school attribute. To our point of view, this is a great improvement of the description of tastes for school attributes\(^3\).

Second, the interpretation of this specification is to be done with caution, in particular, we believe it is not supported by the mechanisms described by BLLW, and that it does not allow to interpret parameters as preference for school attributes. For example, parameters associated to \( z_{it} \) are a reduced-form specification in which both the supply and the demand side of the teacher market are accounted for. Considering the case \( k = 2 \) (transferring within the district), these parameters may reflect both the preference of the teacher for alternative schools, and the preference of alternative schools for the teacher (the composition of this set of alternative schools does not need to be modeled since the authors are considering teachers in the same district).

Moreover, how should the parameter associated to \( x_{it} \) be interpreted? Potentially, we would like to interpret \( \beta_2^k x_{it} + z_{it}^T \theta^k x_{it} + \delta_i x_{it} \) as the utility that the individual enjoys from being in the school with attributes \( x_{it} \). And this could be done, indeed, if we assume that job offers do not depend on the attributes of the current school:

“If a teacher has the option of remaining in his current schools and the hiring authorities in other schools base their hiring decisions on the attributes of the candidate – not the attributes of the school where they previously taught – the attributes of the initial school would affect the probability of transferring only by changing the relative attractiveness of the school from the perspective of the teacher.”

With this assumption, the values \( V_{it}^2, \ V_{it}^3 \) and \( V_{it}^4 \) are distributed independently of \( x_{it} \), such that parameters on \( x_{it} \) could be directly interpreted as participation to the value \( V_{it}^1 \) which is the valuation of the teacher for her school. This, in turn, would allow us to identify the marginal willingness to trade amenities. But in this case, why are the parameters associated to \( x_{it} \) dependent on \( k \)? Not restricting the parameters \( \beta_2, \theta \) and \( \delta \) to be identical across \( k \) is, from our point of view, a major limitation in their interpretation.

Nonetheless, their results (obtained for teachers in New-York) are consistent with an overall preference for more performing students. But this preference is heterogeneous depending on the academic level of the teachers: it is more accentuated for teachers with the highest academic level (measured by the results at the certificate exam). Also, white and hispanic teachers tend to move away for schools gathering an high share of black students, whereas the racial composition does not seem to impact the behavior of black teachers. Moreover, teachers are shown to prefer schools that are close to where they live, and BLLW states that this can participate to an unequal provision of teachers since teachers tend to live far from disadvantaged schools.

\(^3\) See Section 4
Bonhomme, Jolivet and Leuven (2011)

Bonhomme, Jolivet and Leuven (2011) criticize the duration approach for computing teachers' valuation for job characteristics. More specifically, they note that the assumptions underlying such approaches – notably, the job destruction rate, the job offer rate, and the distribution of the attributes of job offers do not depend on current job characteristics – are unlikely to be met.

They assert that in this case, estimates of teachers' preferences are likely to be biased. In order to correct for this bias, their strategy is to explicitly account for the distribution of job offers characteristics in the computation of the marginal willingness to trade amenities.

Formally, if we denote $V(A, X)$ the valuation of teacher with characteristics $X$ of a position with characteristics $A$, and $c(X, Z)$ the moving cost, then the probability to observe $Q = 1$, the teacher changing for a school with characteristics $A^*$ is given by:

$$P[Q=1|A, A^*, X, Z] = P[c(X, Z) < V(A^*, X) - V(A, X)]$$

Now, $A^*$ is not observed for job offers that are not accepted. However, using Baye's rule, one has:


and the left-hand side of this equality is likely to be observable from data. In this case, one can compute the value of the following ratio for any value $z_1$ and $z_2$ of $Z$ (assuming it is well-defined):

$$R(A^*, A, X, z_1, z_2) = \frac{P[Q=1|A, X, Z] f(A^*|A, X, z_1, Q=1)}{P[Q=1|A, X, Z] f(A^*|A, X, z_2, Q=1)}$$

And, if we assume further that the distribution of $A^*$ conditionally on $X$ and $A$ is independent of $Z$, this ratio would be equivalent to:

$$R(A^*, A, X, z_1, z_2) = \frac{P[c(X, z_1) < V(A^*, X) - V(A, X)]}{P[c(X, z_2) < V(A^*, X) - V(A, X)]}$$

Now, we would like to interpret the MWT amenities $a_i$ and $a_j$ as the compensating variation that allows $R(.)$ to remain constant. This interpretation is made possible by assuming that variation in amenities does not affect mobility costs (conditionally on $A^*, V(A), X, Z$). In this case one can re-write the ratio as:

$$R(A^*, A, X, z_1, z_2) = H(A^*, V(A), X, z_1, z_2)$$

Such that, eventually, one has under few technical assumptions (mostly about the differentiability of the ratio) and with small computations:

$$\frac{\partial R}{\partial a_i} = \frac{\partial V}{\partial a_i}$$

Performing this analysis on the Dutch market for teachers, BJL find that teachers dislike disadvantaged minority pupils and large classes. More importantly, they find that the
corresponding estimates using the classical duration models of Gronberg and Reed (1994) are different, which supports the idea of the bias they were describing.

THE GAME-THEORETIC TWO-SIDED MATCHING MODEL


Boyd, Lankford, Loeb and Wyckoff (2003) come up with another model designed to obtain teachers’ valuation for school characteristics. Basically, they account for the fact that matching teacher-school is a two sided decision, in a game-theoretical framework, and estimate the parameters of the model using simulated moments.

Formally, they assume that the value of a teacher \( j \) for a school \( k \) is given by:

\[
 u_{jk} = u(z_k^1, d_{jk} | q_j^2, \beta) + \delta_{jk}
\]

where \( z \) are school characteristics, \( q \) are and \( d \) are teacher-school characteristics (for example, distance) and \( \beta \) is the parameter to be estimated. Identically, schools value teachers according to:

\[
 v_{jk} = v(q_k^1 | z_k^2, \alpha) + \omega_{jk}
\]

These valuations of teachers for schools, and schools for teachers, induce an allocation of teachers across available positions that can be reproduce using Gale-Shapley algorithm. This algorithm relies on the assumption of the stability of the resulting matching. Namely, it must not be the case that in the resulting allocation both a teacher and a school could increase their utility with a match. Formally, this stability condition states that for all candidate \( g \) in position \( g' \) and candidate \( h \) in \( h' \), one has \( u_{gg'} > u_{gh'} \) or \( v_{hh'} > v_{gh'} \) or both.

In this case the expected value of \( z_{jk} \) – the attributes that teacher \( j \) will face after allocation – is “somehow” given by the parameter \( \alpha \) and \( \beta \) and the variable of characteristics \( q_j \), even if we do not know the its analytical distribution. Therefore, the moment restrictions will be given by (NB: for simplicity we report only one of the moment conditions, but BLLW also use the value \( d_{jk} \)):

\[
 E[z_j - E[z_j | q_j, \alpha, \beta] | q_j] = 0
\]

where \( z_j \) is the observed school attributes faced by the newly hired individual \( j \). This leads to:

\[
 E[q_j(z_j - E[z_j | q_j, \alpha, \beta])] = 0
\]

BLLW point out that we do not know the distribution of the conditional expected value of \( z_j \). Therefore, in order to estimate the parameters, we need to use simulated moments. The principle is quite simple: take some value for \( \alpha \) and \( \beta \), then simulate the error-terms and obtain from simulated utilities an allocation using Gale-Shapley algorithm – \( ie \), find the allocation that satisfies stability conditions. Record the simulated value of \( z_j \) and simulate again a certain number of times for the same value of \( \alpha \) and \( \beta \).

These simulations allow us to have an approximation of \( E[z_i | q_i, \alpha, \beta] \), such that one can compute the sample moment conditions.
Performing this procedure for various values of $\alpha$ and $\beta$, one obtains the estimation of the parameters by selecting those for which the sample moment conditions “fit” the theoretical restrictions the best.

It is interesting to remark that to a certain extent, the assumptions underlying this approach have some similarities with Antos and Rosen (1975)! From our point of view, they could be seen as an adaptation in a discrete case of the static optimization of Antos and Rosen (1975). Indeed, the use the Gale-Shapley algorithm supposes implicitly that the applicants have information about all available positions, or that there are no search costs.

Potentially, one may be willing to add frictions on this market to see what happens. Two options can be envisage. First, one can make applications costly such that teachers will be willing to limit their iterations. Second, we could imagine remove the “maybe” option of schools and place ourselves in a “take-it-or-leave-it” framework. In this second case, finding the optimal stopping strategy of schools would be quite challenging. For example, we know that for a school to maximize its probability to obtain the best teacher when there are $N$ candidates, the optimal strategy is the following: “leave all applicants before $\alpha_1$, and take the first applicant after $\alpha_1$ that has a value greater than the best of the $\alpha_1$ first applicants” (for example, for 100 candidates, the optimal $\alpha_1$ is around 36). Therefore, we guess that the optimal strategy if the school is willing to maximize its expected utility will be of the following form: “leave all applicants before $\alpha_1$; between $\alpha_1$ and $\alpha_2$, take the first applicant that has a value greater than the $\alpha_1$ first applicants; between $\alpha_2$ and $\alpha_3$, take the first applicant that has a value greater than the second best of the $\alpha_1$ first applicants etc.”

Using this method on New-York teacher labor market, they find that white teachers prefer schools who gather an higher proportion of white students. Also, they underline the importance of distance: teachers preferring schools that are close to where they live.

---

4 BLLW use several moment conditions even if we report only one. They use an identity matrix for weighting this various conditions.
Section 3: A Discrete-Choice Programming Model for Teachers' Mobility Decisions

The French system of teachers' promotion does not allow to use wage differentials to measure teachers' preferences for school attributes, as in Antos and Rosen (1975). Indeed, in France, wages are set nationally with variations being almost entirely due to differences in seniority and in certificates\(^5\). Therefore, as the literature we reviewed in the previous section suggests to, we propose to move to the analysis of mobility decisions in order to measure these preferences. In other words, the question in this section is: what can the time before first application to transfers teach us about teachers' preferences?

These kinds of models have already been seen in the literature (Boyd, Lankford, Loeb and Wyckoff (2005)). However, there is an important particularity in the French system because the rules of promotion are set nationally, and such that the availability of opportunities depends on the seniority in the current school. Therefore one cannot in general infer that an high duration in a school is associated with an high valuation for this school. For example, we know from Prost (2010) that the “agrégés” are less likely to leave disadvantaged schools: but this is not necessarily because they value more this kind of schools, potentially, they may behave so because they are targeting the top-level Parisian high schools for which one must capitalize an important seniority if she is willing to be assigned there.

In fact, this issue is quite similar to the one addressed by Bonhomme, Jolivet and Leuven (2010) saying that the opportunities faced by an individual may depend on the attributes of the school she is currently assigned to. The key difference is that this relationship between current school and opportunities is institutionalized in France, such that one can expect to be able to disentangled the part of the school value that is to be attributed to school characteristics, and the part of the school value that is to be attributed to its effect on the availability of other opportunities.

Overall, the present section has two purposes. The first one is to propose a framework to analyze mobility decisions of French teachers, as it has never been done as far as we know. The second is to provide an insight on how mobility decisions could be used in order to recover teachers' valuation for school amenities.

The section will be organized as follows. In a first paragraph, we will briefly recall the administrative framework of teachers' mobility in France, and we will make a general presentation of the teacher's problem. We will then move to a presentation of discrete-time programming models that will allow us in a third paragraph to obtain a new representation of the theoretical model of the teacher's problem. Finally, we will propose an econometric model that accounts for the French specificities mentioned above in order to obtain the preferences of teachers for school attributes.

---

\(^5\) See Section 1
To begin with, let us recall the basic features of the promotion of teachers in the French system:

During her career, a teacher is allowed every year to apply for transfers. This application consists in an administrative procedure, in which she declares a list of preferred positions: the wishes. It is made on an Internet website such that the administration can record the wishes of all teachers applying to transfers. Also, we can consider that this declaration is made simultaneously by all teachers (in the sense that when applying, teachers do not know what other teachers have done). The list is hierarchical, such that if two wishes are accepted by the administration, the teacher will be assigned to the wish that has the highest rank in the list.

Wishes can be schools or districts. When all wishes of the teacher are schools that belong to her current district, the procedure applicable are “mouvements intra-académiques” that we will refer to as infra-district transfers. If at least one wish is a school that belongs to another district, then the list of wishes must also contain this district, the procedure applicable is then the “mouvements inter-académiques” and we will refer to it as inter-district transfers.

In the case of infra-district transfers, when no assignment is found, the teacher stays in her current position.

In the case of inter-district transfers, when no assignment is found, for both districts and schools, the teacher stays in her current position. If a district is accepted, but no schools within this district is, the teacher is arbitrarily assigned by the local administration to a school in this district, even if it does not belong to the list of wishes.

In case of competition for an assignment, priority is given to the teacher with higher bonuses, such that overall, for all wishes, the probability that it is accepted increases with the bonuses that the teacher is entitled for it.

There are four parts in the bonuses. The first part depends on the certificate of the teacher; the second part depends on the seniority in the public sector; the third part depends on the seniority of the teacher in the current school; the fourth part is school-specific, and depends on the civil situation of the teacher.

We assume that at date \( t \), a teacher \( i \)'s utility from deciding \( f_t \) (a mobility decision) is given by the expectation of the sum (over a finite number of periods) of the discounted instantaneous utilities that she will get in any period from the school she will be assigned to.

\[
U_{it}^{j(i,t)}(f_t) = E_{t, f_t} \left[ \sum_{k=t}^{T} \beta^{k-t} u_{ik}^{j(i,k)} \right] \tag{1}
\]

where \( \beta \) is the discount factor and \( j(i,t) \) stands for the school attended by individual \( i \) at time \( t \). \( j(i,k)_t \) is a career-path for the individual \( i \). The feasibility of a career path is described by a probability \( P_{0, f_t}[j(i,k)] = 1 \) such that even if we assume that anticipated instantaneous utilities are perfectly known, the expectation operator makes sense.

The teacher's problem in mobility decision is to find the decision \( f_t \) that will maximize her
expected utility in (1). However, the representation of the teacher's problem in (1) is not practical, therefore we are willing to move to another equivalent representation using Bellman's equations.

**DISCRETE-CHOICE PROGRAMMING MODEL : INTRODUCING DEFINITIONS AND NOTATIONS**

Here we are introducing the various notations and definitions that we will be using in this section:

**Notations:**

- \( \Omega_i(t) \) is the state-space of the individual \( i \) at time \( t \). It consists in all arguments, known to the individual, that affect both instantaneous utilities and the probability distribution of future utilities.

- \( R^a(\Omega_i(t), t) \) is the instantaneous rewards enjoyed by the individual \( i \) for opting for the alternative \( a \) at time \( t \).

- \( v^a(\Omega_i(t), t) \) is the alternative-specific value function for alternative \( a \) at some date \( t \) for some teacher \( i \). This value function is composed of the instantaneous rewards and the discounted value of all future opportunities associated with opting for the alternative \( a \).

- \( v(\Omega_i(t), t) = \max_a \{ v^a(\Omega(t), t) \} \) is the value function of the individual \( i \) at time \( t \).

- \( d_i^a(t) \) is an indicator equal to 1 if the alternative \( f_i = a \) is chosen by individual \( i \) at time \( t \). Because alternatives are mutually exclusive in this setting and that at least one alternative is chosen at each period, we have that \( \sum_{a \in A} d_i^a(t) = 1 \).

Having this notation, the Bellman's equations provide us with a general formula for expressing the teacher's problem:

\[
v^a(\Omega_i(t), t) = R^a(\Omega_i(t), t) + \beta E \left[ v(\Omega_i(t+1), t+1 \mid \Omega_i(t), d_i^a(t) = 1) \right] \quad (2)
\]

\[
v(\Omega_i(t), t) = \max_{f_i} U_{it}^{j(i,t)}(f_i) \quad (3)
\]

**THE TEACHER'S PROBLEM IN DISCRETE-CHOICE PROGRAMMING SETTING :**

**A THEORETICAL APPROACH**

---

6 A general description of discrete-choice programming models can be found in Keane and Wolpin (1994)
At the end of each school year, a teacher is given four options: (0) she can leave teaching in the public sector, (1) she can stay in her current position, (2) she can apply for infra-district transfers, or (3) she can apply for inter-district transfers.

The state-space will be described by:

\[
\Omega_i(t) = \{\Gamma; \ D; \ \{\delta^d_i\}_{d \in D}; \ \{p^j_{it}\}_{j \in \Gamma}; \ \{p^d_{it}\}_{d \in D}; \ \{u^j_{it}\}_{j \in \Gamma}; \ j(i,t); \ d(j); \ c_{it}\}
\]  

where \( \Gamma = \{0, 1, \ldots, J\} \) is the set of schools (with 0 representing the position out of the teaching public sector); \( j(i,t) \) associates an individual to her school at \( t \); \( D \) is the set of districts and \( d(j) \) associates schools to districts; \( p_{it}^j \) represents the probability that the teacher \( i \) has to be accepted in school \( j \) if applying to it at time \( t \); \( p_{it}^d \) is the probability that the teacher \( i \) has to be accepted in district \( d \) if applying at date \( t \); \( \delta^d_i \) is the value that the teacher associates with being arbitrarily assigned within the district \( d \); \( u_{it}^j \) is the instantaneous utility associated with being in school \( j \) and \( c_{it} \) is a cost associated to undertaking administrative procedures for transferring.

We are aware that this state-space is associated with a very theoretical model: this is because we were willing to make as few assumptions as possible at this stage. But moving to an empirical perspective would allow us to greatly simplify this state-space.

We have particularly emphasized this theoretical model, because it turns out that this representation is not obvious at all. If we compare our problem with the classical problem of occupational choice in Keane and Wolpin (1994) for example, there is a key difference, because in our framework, a choice does not correspond to an actual position of the teacher. For instance, when a teacher chooses to apply for transfer, she does not know certainly what will be her future position. Therefore at some point, we will need to distinguish the concept of instantaneous reward from the concept of instantaneous utilities. This point is also the reason why we did not achieve to find a procedure to estimate \( u_{it}^j \).

The alternative 0: leaving teaching in the public sector

If a teacher decides to quit, she will get an instantaneous reservation utility denoted by \( u_{it}^0 \) which is assumed known to the teacher. Thus, the reward function is simply:

\[
R^0(\Omega_i(t), t) = u_{it}^0 \quad (5)
\]

And the value of this alternative is given by:

\[
v^0(\Omega_i(t), t) = u_{it}^0 + \beta E[v(\Omega_i(t+1), t) \mid \Omega_i(t), d_{it}^0(t) = 1] \quad (6)
\]
We assume that any decision to quit teaching in the public sector is permanent. Notice that this is not an obvious assumption.

First, there exists several procedures that allow teachers to spend a transitory year out of teaching, for example, teachers can ask a sabbatical year, or a “congé formation” which is a year out of teaching devoted to the preparation of a competitive exam that would enable them to obtain a better certificate. However, enjoying this kind of advantages is never granted and depends on the decision of the administration : therefore, we do not model them as alternatives.

Nevertheless, we will account for these possibilities in the alternative 1, considering that an individual who spends a year out of teaching with the agreement of the administration still holds his position, with her particularity being embedded in the utility associated to her current position.

Second, there may exist ways for a teacher who spent time out of the public teaching sector to go back in. The administration usually refers to this possibility as “mise à disposition”. However, we will consider that this possibility is embedded in the distribution of $u^0_{i+k|t}$.

Eventually, one could even write the value function of this alternative :

$$v^0_j(O_j(t), t) = u^0_{j(t)} + \beta E \left[ \sum_{T=t+1} \beta^{T-t} u^0_T \mid O_j(t), d^0_i(t) = 1 \right]$$ (7)

The alternative 1 : staying in the same school

If the teacher decides to stay in her current position $j(i,t)$, she receives an instantaneous utility $u^0_{j(i,t)}$. Thus the reward function is given by :

$$R^1_j(O_j(t), t) = u^0_{j(i,t)}$$ (8)

and the value function corresponding to this alternative is given by :

$$v^1_j(O_j(t), t) = u^1_{j(i,t)} + \beta E \left[ v(\Omega_j(t+1), t+1) \mid \Omega_j(t), d^1_j(t) = 1 \right]$$ (9)

This specification allows us to point out the fact that the value of staying in a school for a teacher does not only depend on the stream of utilities that she expects to enjoy at different dates by teaching in this school (ie the law of motion of $(u^j_{i+k})_{k \geq 1}$ for $j = j(i,t+1)$) : it is also driven by the influence of the decision of staying on the state-space.

In particular, in the French system, the decision of staying is associated with an increase in $p^j_{i+1}$ for all $j$’s and $p^d_{i+1}$ for all $d$’s. Therefore, one can decide to stay because the value she associates with an increase in these probabilities induces her to do so.

For example, the instantaneous utility associated with staying can be very low for every $t$ (for
instance : \( u^i_0 > u^j_0 \ \forall \ t \), and still, the individual can decide to stay in the school because this increases the probability to obtain a school that provides to the teacher high levels of instantaneous utilities.

Or again, one can imagine two teachers in the same schools such that at all date, the instantaneous utility associated with teaching in this school is lower for the first one than for the second. But still, the first teacher can decide to stay when the second decides otherwise, for example, if the first one values a lot schools that are very difficult to obtain, such that she values a lot the increase in \( p^j_{t+1} \) associated with the decision to stay\(^7\).

**Alternative 2 : applying to infra-district transfers**

The teacher can apply for a transfer within her district. In this case, her future position is uncertain : it will be chosen over a list of wishes that she would have declared. If we denote by \( W_\alpha \) this list of wishes, and if we write \( L \mid W_\alpha, \Omega_\alpha(t) \) the random assignment resulting from the procedure, her *instantaneous utility* will be given by :

\[
u^L_{it} \mid \Omega_\alpha(t), W_\alpha \quad (10)\]

Now, it is important to point out that \( W_\alpha \) does not appear in the state-space. This decision is supported by the fact that assuming optimal behavior of the agent, we have \( W_\alpha = W_\alpha^* \) and that \( W_\alpha^* \) is almost determined by the state-space as it is. Indeed, it is determined by (i) the set of alternatives schools, (ii) the *value* that the teacher associates to each of these schools, and (iii) probability she has to be accepted if the teacher applies to this school. Now, (i) , and (iii) already belong to the state-space, but (ii) does not *a priori* : indeed, the *value* of a school is not to be confused with the instantaneous *utility* associated to this school. However, we can make assumptions such that the value of a school is a function of the state-space at date \( t \) : we will provide a proof of this point below. Nevertheless for the moment, we consider that

\[
L \mid W_\alpha, \Omega_\alpha(t) = L \mid \Omega_\alpha(t)
\]

Here is where we stop matching instantaneou**s rewards** and **instantaneous utilities**. Conceptually, we consider that the instantaneous reward from applying to an infra-district transfer depends only on the “statistical” flow of utility that one enjoys from the different positions available to her, and not on the particular realization of this flow. Indeed, we believe that this particular realization is to be considered as a component of the (stochastic) motion of the state-space if we want the model to be consistent.

Hence, we write the reward function as :

\[
R^2(\Omega_i(t)) = E[u^L_{it} \mid \Omega_i(t)] - c_{it} \quad (11)
\]

where the parameter \( c \) can be interpreted as moving cost, cost of applying to the procedure (in terms of time consumption and screening costs), dis-utility from the uncertainty of the assigned position, or dis-utility from the uncertainty about the value of other schools.

---

\(^7\) An example is provided in the next paragraph
And the value associated with this alternative is given by:

\[ v^2(\Omega_i(t), t) = E[u^L_{it} | \Omega_i(t)] - c_{it} + \beta E[v(\Omega_i(t+1), t+1) | \Omega_i(t), d^2_i(t) = 1] \]  

(12)

The decision to apply to a transfer has an ambiguous effect on \( v(\Omega_i(t+1), t) \), indeed, if the application is successful (meaning that the teacher is assigned to one of the schools she applied for), this will decrease \( p_{it+1} \) and \( p'_{it+1} \) compared to what she would have obtained if she had decided to stay. Therefore, we have in general that

\[ v^2(\Omega(t+1), t+1)|d^2(t) = 1 \geq v^2(\Omega(t+1), t+1)|d^1(t) = 1 \]

Also, it is important to point out the fact that when the procedure fails to assign a school to the teacher, everything goes as if she never applied to infra-district transfers.

**Proof:** the state-space is well-defined (\( W_i^* \) is determined by the state-space)

We are willing to provide a proof that under few assumptions, the state-space is well-defined, in the sense that it determines \( W_i^* \). However, in order to make this proof most understandable, we simplify the problem by considering that there exists no inter-districts transfers. Hopefully, it will appear obvious that adding the possiblility of such transfers does not change much the issue.

In general, this optimal list of wishes is not determined by \( \Omega_i(t) \). For example, one can imagine that a particular value of \( j(i,t+1) \) is associated with a particular law of motion of \( p_i^j \). Then, this particular law of motion will be accounted for in the computation of the value of a school, but it is not included in the state-space as we defined it, and we will therefore have inconsistency in our model.

Conceptually, the issue arises because when applying to a transfer, an individual selects a particular school for the next year ( \( j(i,t+1) \) ), and this selection may affect the realization of all variables of the state-space in year \( t+1 \). In order to prevent this from happening, we need the first following assumption:

**Assumption 1:** In the state-space \( \Omega_i(t-1) \), everything except \( j(i,t-1) \) evolves independently of \( j(i,t) \), conditionally on \( \Omega_i(t-1) - j(i, t-1) \) and \( d^k(t-1) \):

\[
\Omega_i(t) | \{\Omega_i(t-1), d^k(t-1) = 1, j(i,t)\} = \Omega_i(t) | \{\Omega_i(t-1), d^k(t-1) = 1\} \forall k
\]

where \( \Omega_i(t) = \{\Gamma ; \{p_i^j\}_{j \in \Gamma}; \{u_{it}^j\}_{j \in \Gamma}; c_{it}\} \)

Basically, this assumption states that the evolution of the set of probabilities and of the set of utilities between period \( t-1 \) and period \( t \) does not depend on the school attended between these two dates. Even if it does not appear unrealistic at first sight, the assumption is not verified in the French system, because some schools allow for a faster promotion of teacher. Then, one could imagine to add
to the state-space a set of functions \( F_j'(.) \) describing the particular effect of being in school \( j \) on the law of motion of \( p_{it}^{j'} \). We prefer not to do so, in order to keep things simple. However, in a econometric perspective, the particularities of the French system will make us able to consider this set of functions in a simple way.

We also need the following assumption:

**Assumption 2**: Removing one element in the set of maximization of the expected utility associated to a transfer does not change this maximization

\[
\argmax_{\{W_1, \ldots, W_N\} \in \Gamma - j} \sum_{j=1}^{N} V_{it}^j(2) p_{it}^j \prod_{i=1}^{j-1} (1 - p_{it}^i) + V_{it}^j(1) \prod_{j=1}^{N} (1 - p_{it}^j)
\]

\[
= \argmax_{\{W_1, \ldots, W_N\} \in \Gamma} \sum_{j=1}^{N} V_{it}^j(2) p_{it}^j \prod_{i=1}^{j-1} (1 - p_{it}^i) + V_{it}^j(1) \prod_{j=1}^{N} (1 - p_{it}^j)
\]

\( \forall j \in \Gamma \)

where \( V_{it}^j(2) \) refers to the value of school \( j \) in a transfer procedure, and \( V_{it}^j(1) \) refers to the value of school \( j \) when deciding to stay: we will provide an expression for each of these value later.

This assumption states that the optimization behavior of the agent is unaffected by the fact of being in school \( j \), conditionally on knowing the value \( V_{it}^j(1) \).

**First step : writing the value of schools**

In order to prove our statement, we need to change our perspective and consider, the value of schools, instead of the value of the decisions.

For example, consider a teacher who is in \( j' \) at \( t \) and decides to stay, what is the value \( V_{it}^j(1) \) of the school \( j' \)? It will be given by the instantaneous utility from teaching in \( j' \) plus the value of future opportunities that will be drawn at \( t+1 \), which will be made of the three alternatives (we are in a world without inter-district transfers) : leaving teaching in the public sector (\( v^0 \)) ; apply to a transfer (\( v^2 \)) ; stay in the same school \( j' \). Hence, we can write (ignoring individual subscripts):

\[
V_{i}^{j'}(1) = u_{i}^{j'} + \beta E[\max(v_{i+1}^{0}, v_{i+1}^{2}, V_{i+1}^{j'}(1)) | \Omega(t) , j(t+1) = j', d^{1}(t) = 1] \quad (13)
\]

Identically, suppose a teacher decides to apply for a transfer. In this procedure, the value of the school \( j' \) will be given by:
\[ V_t^{j'}(2) = u_t^{j'} + \beta E\left[ \max(v_t^0, v_t^2, V_{t+1}^{j'}(1)) \mid \Omega(t), j(t+1) = j', d^2(t) = 1 \right] \quad (14) \]

These two expressions are making clearer the relevance of our assumption: when at date \( t \), the teacher is assessing the value of a school, she considers the fact that she must choose a value for \( j(t+1) \), and thus, she is concerned with the impact of \( j(t+1) \) on \( v^0 \) and \( v^2 \).

Second step: \( W_t^* \) is determined by (i) the set of alternatives schools, (ii) the value that the teacher associates to each of these schools, and (iii) probability she has to be accepted if the teacher applies to this school.

Straightforwardly, we have, when a teacher decides to apply to a transfer and when she is initially in \( j' \) (supposing that the restricted number of wishes is \( N \)):

\[ W_t^* = \text{argmax}_{(W_1, ..., W_N) \in \Gamma} \sum_{j=1}^{N} V_t^{W_j}(2) p_t^{W_j} \prod_{i=1}^{j-1} (1 - p_t^{W_i}) + V_t^{j'}(1) \prod_{j=1}^{N} (1 - p_t^{W_j}) \quad (15) \]

that shows our proposition, and we used the assumption 2°.

Since (i) and (iii) already belong to the state-space, \( W_t^* \) will be determined by the state-space if (ii) is determined by the state-space.

Third step: \( V_t^{j}(1) \) and \( V_t^{j}(2) \) are determined by the state-space

We want to prove that \( V_t^{j}(1) \) and \( V_t^{j}(2) \) are functions of the state-space \( \Omega(t) \), a proposition that we will write:

\[ V_t^{j}(1) = V_t^{j}(1)(\Omega(t)) \]
\[ V_t^{j}(2) = V_t^{j}(2)(\Omega(t)) \]

To demonstrate this point, we proceed recursively.

→ At time \( T \)
We have:
\[ V_T^{j'}(2) = V_T^{j'}(1) = u_T^{j'} \in \Omega(T) \]

→ At time T-1

Remark that we considered as granted the fact that \( W_t^* \) is unique. In fact, it is not necessarily the case, but this is unimportant. Potentially, one may consider that * is a two stage operator, that first provides the set of optimal wishes, and second, selects one element of this set using some arbitrary rule (for example, minimize the concatenation \( W_1W_2...W_N \)).
By definition, we have:

\[ V_{T-1}^{(j')} (1) = u_{T-1}^{(j')} + \beta E \left[ \max \left\{ v_{T,1}^0, v_{T,2}^0, V_{T-1}^{(j')} (1) \right\} \mid \Omega (T-1), j(T) = j', d^2(T-1) = 1 \right] \]

\[ V_{T-1}^{(j')} (2) = u_{T-1}^{(j')} + \beta E \left[ \max \left\{ v_{T,1}^0, v_{T,2}^0, V_{T-1}^{(j')} (1) \right\} \mid \Omega (T-1), j(T) = j', d^2(T-1) = 1 \right] \]

From the definitions of the value of alternative 2 (equation 15) and 0 (equation 7):

\[ v_T^2 = \max_{(w_1, \ldots, w_L) \in \Gamma} \sum_{j=1}^N u_T^W p_T^W \prod_{i=1}^{j-1} (1 - p_T^W) + u_T^{(j')} \prod_{j=1}^N (1 - p_T^W) - c_T \]

and

\[ v_T^0 = u_T^0 \]

Therefore we can write:

\[ E \left[ \max \left\{ v_T^0, v_T^2, V_T^{(j')} (1) \right\} \mid \Omega (T-1), j(T) = j', d^k(T-1) = 1 \right] \]

\[ = \int \max \left\{ u_T^{(j')}, u_T^0, \max_{(w_1, \ldots, w_L) \in \Gamma} \sum_{j=1}^N u_T^W p_T^W \prod_{i=1}^{j-1} (1 - p_T^W) + u_T^{(j')} \prod_{j=1}^N (1 - p_T^W) - c_T \right\} \times d \left[ \left\{ \left( p_T^j \right)_{j \in \Gamma}; \left( u_T^j \right)_{j \in \Gamma}; \ c_T \right\} \mid \Omega (T-1), j(T) = j', d^k(T-1) = 1 \right] \]

\[ = \int \max \left\{ u_T^{(j')}, u_T^0, \max_{(w_1, \ldots, w_L) \in \Gamma} \sum_{j=1}^N u_T^W p_T^W \prod_{i=1}^{j-1} (1 - p_T^W) + u_T^{(j')} \prod_{j=1}^N (1 - p_T^W) - c_T \right\} \times d \left[ \left\{ \left( p_T^j \right)_{j \in \Gamma}; \left( u_T^j \right)_{j \in \Gamma}; \ c_T \right\} \mid \Omega (T-1), d^k(T-1) = 1 \right] \]

\[ = E \left[ \max \left\{ v_T^0, v_T^2, V_T^{(j')} (1) \right\} \mid \Omega (T-1), d^k(T-1) = 1 \right] \]

Therefore we used the assumption 1 to go from the second to the third equality.

Eventually, we can write:

\[ V_{T-1}^{(j')} (2) = u_{T-1}^{(j')} + \beta E \left[ \max \left\{ v_T^0, v_T^2, V_T^{(j')} (1) \right\} \mid \Omega (T-1), d^2(T-1) \right] \]

\[ = V_{T-1}^{(j')} (2) (\Omega (T-1)) \]

and identically,

\[ V_{T-1}^{(j')} (1) = V_{T-1}^{(j')} (1) (\Omega (T-1)) \]

→ Recursive step

Assume that for all \( j \) and \( k=1,2 \),

\[ V_i^{(j)} (k) = V_i^{(j)} (k) (\Omega (i+1)) \]

Then we have as before:
\[ V_i^{j'}(1) = u_i^{j'} + \beta E \left[ \max \left( v_{i,t+1}^0, v_{i,t+1}^2, V_{i,t+1}^{j'}(1) \right) \mid \Omega(t), j(t+1) = j', d_i^1(t) = 1 \right] \]

\[ V_i^{j'}(2) = u_i^{j'} + \beta E \left[ \max \left( v_{i,t+1}^0, v_{i,t+1}^2, V_{i,t+1}^{j'}(1) \right) \mid \Omega(t), j(t+1) = j', d_i^2(t) = 1 \right] \]

Using the fact (equation 7) that \( v^0 = v^0(\Omega(t)) \), Assumption 1 and Assumption 2, and the expression for \( v^1 \) (equation 15), we can express the expectations as:

\[ E \left[ \max \left( v_{i,t+1}^0, v_{i,t+1}^2, V_{i,t+1}^{j'}(1) \right) \mid \Omega(t), j(t+1) = j', d_i^k(t) = 1 \right] \]

\[ = \int \max \{ v^0(\Omega(t+1), t+1); V^{j'}(1)(\Omega(t+1), t+1); \}
\[ \max_{(w_1...w_N) \in R} \sum_{j=1}^{N} V^{W_j}(2)(\Omega(t+1), t+1) p_{i,t+1}^{W_j} \prod_{i=1}^{j-1} (1 - p_{i,t+1}^{W_i}) \]
\[ + V^{j'}(1)(\Omega(t+1), t+1) \prod_{j=1}^{N} (1 - p_{i,t+1}^{W_j}) - c_{t+1} \}
\[ \times d P[\Omega(t+1); \Omega(t), j(t+1) = j', d_i^k(t) = 1] \]

\[ = \int \max \{ v^0(\Omega(t+1), t+1); V^{j'}(\Omega(t+1), t+1); \}
\[ \max_{(w_1...w_N) \in R} \sum_{j=1}^{N} V^{W_j}(\Omega(t+1), t+1) p_{i,t+1}^{W_j} \prod_{i=1}^{j-1} (1 - p_{i,t+1}^{W_i}) \]
\[ + V^{j'}(\Omega(t+1), t+1) \prod_{j=1}^{N} (1 - p_{i,t+1}^{W_j}) - c_{t+1} \}
\[ \times d P[\Omega(t+1); \Omega(t), d_i^k(t) = 1] \]
\[ = E \left[ \max \left( v_{i,t+1}^0, v_{i,t+1}^2, V_{i,t+1}^{j'}(1) \right) \mid \Omega(t), d_i^k(t) = 1 \right] \]

Therefore, we have for all \( t \) and for all \( j : \)

\[ V_i^{j'}(1) = V_i^{j'}(1)(\Omega(t)) = V_i^{j'}(1)(\Omega(t)) \]
\[ V_i^{j'}(2) = V_i^{j'}(2)(\Omega(t)) = V_i^{j'}(2)(\Omega(t)) \]

**END OF THE PROOF**

**Alternative 3 : applying to inter-district transfers**

When the list of wishes of a teacher applying to a transfer contains at least one wish that belongs to another district, she goes through a particular procedure, called “mouvements inter-académiques”.

This procedure could be described in two steps. First, the teacher is allocated to a district according to her favorite district. Second, she is assigned to a school according to her favorite
school in this district. It is important to notice that in the case the second step fails to assign a school belonging to the list of wishes of the teacher, whereas she has been assigned to a new district in the first step, she will be assigned to an arbitrary school designated by the local administration.

Overall, this alternative is not much different than the alternative (2). If we denote $M$ the random variable that assigns the teacher to a school, we can write:

$$R^2(\Omega_i(t)) = E[u_{it}^M | \Omega_i(t)] - c_{it} \tag{16}$$

However, it is important to see that the random variable $M$ does not have the same distribution than the previous variable $L$. Indeed, between districts transfers are associated with a particular risk to be worse-off after the assignment, if a district but no schools are assigned.

In our specification, this risk is represented by the parameter $\delta^d_i$ : it is the only parameter in the model that is not “structural” in the sense that it is to be considered as a reduced-form expression of the value of a district when the teacher is arbitrarily assigned to a school in this district. Finding a more structural expression for this parameter would be quite challenging : intuitively, it will be equal to the sum of the values that the teacher associates to each school in the district, weighted by the probability to be assigned in the school when the assignment is arbitrary.

More details on this alternative will be found in the equation (24).

**ESTIMATING THE PREFERENCES OF TEACHERS FOR SCHOOL ATTRIBUTES USING TIME BEFORE FIRST APPLICATION TO TRANSFERS**

In the case of France, the basic purpose of an econometric model will be to disentangle, in the mobility decisions of a teacher, the pure effect of the school attributes and the effect of the seniority in school on the opportunities available to the teacher.

This fact has been pointed out in several occasions, let us now illustrate it in a very simple framework. Assume for example a perfectly deterministic world, such that for every level of bonuses, and every available school, one has either a probability 1 to be accepted in the school, or a probability 0. In this case, the value of asking a transfer is simply the value of the best school available to the individual at her given level of bonuses. Bonuses increase by one if the teacher decides to stay, and it goes back to 0 when the teacher decides to move to a new school. The utility out of teaching is 0 and there are no districts. The discount rate is equal to 1.

Consider the representation of the teacher's problem below. The teacher is working four periods (T) and there are four schools available (A, B, C and D). There are two individuals “Blue” and Red”. For “Blue” (resp. “Red”) the instantaneous utility provided by being in school X at time Y is given by the blue (resp. red) figure in the cell associated to the row of school X and the column of period Y.

The “price” in the last column refers to the level of bonuses required to be accepted in the
school on the same row (for example, one requires 4 bonuses at the end of period \( T - 1 \) in order to be accepted in school D at \( T \)).

Individuals begin the game in school A with 0 bonus. At the end of period 1, they have one bonus (according to the law of motion we described above), and they are able to move to school B. If they do move to school B, they lose their bonus, such that at the end of period 2, they will have 1 bonus. If they stay in school A, they will have 2 bonuses at the end of period 2.

<table>
<thead>
<tr>
<th>School</th>
<th>“price”</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>6/4</td>
</tr>
<tr>
<td>C</td>
<td>5/3</td>
</tr>
<tr>
<td>B</td>
<td>2/3</td>
</tr>
<tr>
<td>A</td>
<td>1/2</td>
</tr>
<tr>
<td>T</td>
<td>1</td>
</tr>
</tbody>
</table>

Since the accumulation of bonuses does not depend on the school here, it is quite obvious that for these profiles of utilities, both individuals will always “go up” (in the alphabetical ordering) in this universe, such that one can focus on these trajectories.

In this case, possible trajectories are: AAAA; AABB and AAAC. Since the discount rate is 1, “Blue” will prefer AAAC, since the utility associated to this path is 8, when it is 6 for AABB and 4 for AAAA. “Red” will choose ABBB, since the associated utility is 10, when it is 9 for AAAC and 8 for AAAA.

Therefore, “Blue” will stay a longer period in school A than “Red” (three periods versus four), even if “Red” enjoys a greater utility from being in A for all dates (2 versus 1). This is due to the fact that “Blue” enjoys a greater utility than “Red” for expensive schools.

In order to separate the effect of the increased opportunities in the explanation of duration, our idea is to use the law of motion of bonuses in the teacher's career. To be more formal, we assume that the bonuses of a teacher is composed of two parts: a part \( g \) that is given by the seniority in teaching and that cannot be lost as long as the teacher remains in teaching, and a part \( s \) that is given by the seniority in the current school and that can be lost if the teacher moves to a new school. The total bonuses is given by \( k = g + s \).

We assume that bonuses accumulate in the following way:

\[ g = \text{seniority in teaching} \]
\[ s = \text{seniority in current school} \]

\[ k = g + s \]

Recall that there was a third part in the description of bonuses: the specific bonuses that one teacher may be entitled to because of her civil situation. We were not able to introduce them in the model because these bonuses are school-specific and therefore quite incompatible with our objective of “reducing” the teacher's problem. However, one is allowed to hope that this simplification does not change much the issue: indeed, the bonuses are allocated according to the civil situation of the individual independently of mobility decisions such that (i) it does not impact the strategy of the teacher and (ii) it may be captured by individual characteristics. There were also a fourth part that was depending on the certificate of the teacher: we assume that this part only affects the initial level of bonuses (observed) not the accumulation. A more general model could make the law of motion of bonuses depends on teachers' characteristics.
\[ g_{it+1} = g_{it} + y \text{ if } j(i,t+1) \neq 0 \text{ and } 0 \text{ otherwise} \]
\[ s_{it+1} = s_{it} + \sigma_{j(i,t)} \text{ if } j(i,t+1) = j(i,t) \text{ and } 0 \text{ otherwise} \] (17)

About this law of motion, two things are to be said. First, we assume that bonuses accumulate additively with a constant parameter, and it is not necessarily the case. It can be the case that the accumulation also depends on seniority, for example one can imagine that young teachers are offered faster promotion than more experienced teachers. This does not matter: the law of motion of the bonuses can be changed if necessary. Second, the accumulation of \( s \) depends on \( j(i,t) \). This is because we know that some schools allow for faster promotion of teachers (for example, priority schools). Recall that in the previous paragraph, we had to assume otherwise (assumption 1) but we announced that we would be able to remove this assumption for the law of motion of probabilities quite simply: this subscript \( j(i,t) \) on \( \sigma \) is the assumption being removed.

Now, we assume that the value to a teacher \( i \) at date \( t \) for a school \( j \) is given by:

\[ V^j = G(X_{it}, A_j) + W(X_{it}, k_{it+1}) \] (18)

This specification assumes that the value of school is made of two additive components. The first component \( G(.) \) is the intrinsic valuation of the teacher with characteristics \( X_{it} \) (column vector of size \( k_X \)) for the amenities \( A_j \) of school \( j \) (column vector of size \( k_A \)). It is the function we are willing to estimate. The second component \( W(.) \) is the valuation of the teacher with characteristics \( X_{it} \) for the set of opportunities associated to the level of bonuses she will have in \( t+1 \).

This “additive” specification is justified by the fact that a variation in the bonuses of a teacher should not impact the way this teacher values the attributes of a school.

Also, it is important to notice that in this specification, we implicitly assumed that all teachers were belonging to the same district (call it district 1), such that for equal levels of \( k \) the set of opportunities are identical across individuals, only the valuation for this set of opportunities varies, according to \( X_{it} \).

Using the notation we have introduced above (equations (13) and (14)) the equation (18) together with the equations (17) allow us to re-write (extending to the case \( k = 3 \)) \( V_{it}^j(k) \):

\[ V_{it}^j(1) = G(X_{it}, A_j) + W(X_{it}, k_{it} + y + \sigma_j) \]
\[ V_{it}^j(2) = G(X_{it}, A_j) + W(X_{it}, k_{it} + y + \sigma_j - s_{it}) \] (19)
\[ V_{it}^j(3) = G(X_{it}, A_j) + W_{d(j)}(X_{it}, k_{it} + y + \sigma_j - s_{it}) \]

Notice that in the case of equation \( V_{it}^j(3) \) (for which we did not provide a expression before) we needed to add a subscript \( d(j) \) to the equation \( W(.) \) in order to account for the fact that different districts are associated with different perspectives about the set of opportunities.
Using these equations, we can write the values of alternative 1, 2 and 3.

For alternative 1, we have:

\[
v^1 = G(X_{it}, A_{j(i,t)}) + W(X_{it}, k_{it} + y + \sigma_{j(i,t)}) \quad (20)
\]

For the alternative 2, things are more complicated, we have:

\[
v^2 = \max_{(w_1, \ldots, w_N) \in \Gamma^1} \sum_{j=1}^{N} V^{W}_{it}(2) p^W_{it} \prod_{i=1}^{j-1} (1-p^W_{it}) + V^{j'}_{it}(1) \prod_{j=1}^{N} (1-p^W_{it}) - c_{it} \quad (21)
\]

Where \( \Gamma^1 \) stands for the set of available schools restricted to those who belong to district 1.

This expression is quite complicated, and, which is worse, it cannot be available from the data. Indeed, in no circumstances could it be recovered, since one cannot hope to observe the list of wishes when teachers do not apply to transfers.

However, we can find a reduced expression for (21). To this purpose, we need the following assumptions:

**Assumption 3**: there exists a function \( p(.) \) such that for any school \( j \), \( p^j_{it} = p(k_{it}, A_j) \). This function is continuous, and increasing in \( k_{it} \).

**Assumption 4**: \( c_{it} \) is a function of \( X_{it} \).

Such that the following proposition holds quite obviously:

**Proposition (reduced-form)**: Under the previous assumptions, \( v^2 \) is continuous and increasing in \( k_{it} \). Moreover, for individuals in the same district, with same characteristics, and identical \( V^{j(i,t)(1)}_{it} \), \( v^2 \) is only a function of \( k_{it} \):

\[
v^2 = f^2_{2}(X_{it}, k_{it}, V^{j(i,t)}_{it}(1)) \quad (22)
\]

**Proof**: see the proof below, which is quite similar in more complicated way.

For alternative 3, we have the equivalent, providing that:

**Assumption 5**: there exists a function \( p^d(.) \) such that for any district \( d \), \( p^d_{it} = p(k_{it}, d) \). This
function is continuous, and increasing in $k$. 

**Assumption 6**: $\delta^d_{it}$ is a function of $X_{it}$ for all $d$'s

As before, the following holds:

**Proposition**: Under the previous assumptions, $v^3$ is continuous and increasing in $k_{it}$. Moreover, for individuals in the same district, with same characteristics, and identical $V_{it}^{j(i,t)}$ (1), $v^3$ is only a function of $k_{it}$:

$$v^3 = f_3(X_{it}, k_{it}, V_{it}^{j(i,t)}(1)) \quad (23)$$

**Proof**: The value associated to the alternative 3 is given by the maximum expected value among all possible list of wishes. The procedure applicable to the individual is such that first she is assigned to a district and then she is assigned to a school within this district according to the wishes of the individual. If the procedure fails to assign the individual to a wish, but succeed in assigning her a district, then she is arbitrarily assigned in the district such that her expected value is $\delta^d_{it}$.

We denote $N(d)$ the number of districts included in the list of preferred positions, and $N(w)$ the number of schools, such that $N(d) + N(w) = N$, the limited number of wishes. We denote $D_1, \ldots, D_{N(d)}$ the hierarchical list of district wishes. We denote $Q_j$ the number of school wishes belonging to the district $D_j$. $W_1(j), \ldots, W_{Q_j}(j)$ is the hierarchical list of schools belonging to the district $D_j$. Finally, let $W_1, \ldots, W_{Q_0}$ be the hierarchical list of wishes that belong to the current district.

All these variables constitute the set on which the individual maximizes her expected utility:

$$v^3 = \max \sum_{d=1}^{N(d)} \left\{ \left[ \sum_{j=1}^{Q_d} p(k_{it}, A_{it}^{W_{it}^{j(i,t)}}) V_{it}^{W_{it}^{j(i,t)}}(3) \prod_{j=1}^{Q_d-1} (1-p(k_{it}, A_{it}^{W_{it}^{j(i,t)}})) + \delta_{it}^{D_d} \prod_{j=1}^{Q_d} (1-p(k_{it}, A_{it}^{W_{it}^{j(i,t)}})) \right] \right\}$$

$$+ \left[ \sum_{j=1}^{Q_0} p(k_{it}, A_{it}^{W_{it}^{j(i,t)}}) V_{it}^{W_{it}^{j(i,t)}}(2) \prod_{j=1}^{Q_0-1} (1-p(k_{it}, A_{it}^{W_{it}^{j(i,t)}})) + V_{it}^{W_{it}^{j(i,t)}}(1) \prod_{j=1}^{Q_0} (1-p(k_{it}, A_{it}^{W_{it}^{j(i,t)}})) \right]$$

$$\times \prod_{i=1}^{N(d)} (1-p(k_{it}, D_i)) \quad (24)$$

$v^3(\Omega(i))$ is increasing in $k_{it}$. Fix everything in the state-space but $k_{it}$. Suppose that at $k_{it}$, the individual behave optimally concerning her list of preferred positions and call $Z(k_{it})$ the random variable that describes the valuation of the
realized assignment such that all \( E[Z(k_1)] = v'(k_1) \). Let \( Z(k_j) \) be random variable that describes the realized assignment when the list of preferred positions is identical to the one in the case \( k_1 \), but \( k_2 > k_1 \) (such that the list is not optimal for \( k_2 \)). Since the list is hierarchical (the individual will be assigned to her best wishes), and the probability to be assigned to any position is increasing in bonuses, we have for any \( x \), \( P[Z(k_2) > x] > P[Z(k_1) > x] \). Therefore, \( v'(k_1) = E[Z(k_1)] \leq E[Z(k_2)] \leq v'(k_2) \).

\( v'(Ω_1(t)) \) is continuous in \( k_{it} \). Fix everything in the state-space and consider the value \( v'(k^*) \). Consider a sequence \( (k_{it}) \) that converges to \( k^* \) from below. Let \( g(k_{it}) \) be the expected value that one has when the list of preferred position is the same as for the level \( k^* \), but the actual bonuses is \( k_{it} \). Then \( g(k_{it}) \rightarrow v'(k^*) \) (in equation 24, fix everything but \( k \)). Since \( v' \) is increasing in \( k \), we have \( g(k_{it}) \leq v'(k_{it}) \leq v'(k^*) \), such that \( v'(k_{it}) \rightarrow v'(k^*) \).

For teachers in the same district, the value \( v^3 \) is a function of individual characteristics \( X_{it} \), bonuses \( k_{it} \), and the value of “fallback” position \( V_{it}^{(1)} \). If the individuals have identical characteristics, identical bonuses, and identical bottom-line position, then, for a given list of wishes, the value given by the expression under the max operator in (23) will be identical. Since the set of maximization is identical as well, the value of the alternative 3 will be identical across these individuals.

At this stage, a very important remark needs to be done. Recall that our objective is to estimate the parameters of the function \( G(.) \), once a functional form would have been agreed on. In order to do so, we need to identify \( G(.) \) in a single alternative-specific value function, more precisely, in the value \( v' \). In this case, if we were to observe two perfectly identical individuals A and B with identical bonuses and in perfectly identical schools except for one amenity \( a \), we would be able to say:

“the only reason why individual A waits longer before applying to transfer than B is because A enjoys an higher value of \( a \). But \( a \) affects the probability of transferring through the value \( v' \), and the value \( v' \) is affected by \( a \) only through the function \( G(.) \). Therefore, the marginal willingness to pay for the amenity \( a \) of individuals identical to A and B would be the decrease in the wage of A that is required to equalized between A and B the duration before applying to transfers.”

Unfortunately, our analysis (equations 21 and 24) reveals that \( G(.) \) is also an argument of \( v^2 \) and \( v^3 \), and in a way that cannot be untied by the econometrician. For example, considering equation 21, the value \( v' \) was depending on \( G(A, X) \) through the following term:

\[
G(A, X) \prod_{j=1}^{N} (1 - p(k_j, A^W)) \tag{25}
\]

10 Plug equations (19) in (24) and use assumptions 4 and 6
Therefore, we require the following identification assumption:

**Assumption 7**: the effect of \( A_{j(i,t)} \) on \( v^2 \) and on \( v^3 \) cannot be separated from \( k_{it} \).

This assumption is justified by the fact that the effect of \( v^1 \) (and hence, the effect of \( G(.) \), and hence the effect of \( A_{j(i,t)} \)) should be asymmetrical depending on \( k \). For example, the effect of \( G(.) \) on \( v^2 \) is merely due to the value associated to the “fallback” position of one who applies to transfers. But the probability that one has to be left with no wishes fulfilled, i.e. the probability that the value of the fallback position matters, is decreasing in \( k \).

In this case, we can eventually have the following econometric model:

\[
\begin{align*}
v^0 &= f_0(X_{it} | \beta^0) + \epsilon^0_{it} \\
v^1 &= g(X_{it}, A_{j(i,t)} | \theta) + w_1(X_{it}, k_{it} + \gamma + \sigma_{j(i,t)} | \beta^1) + \epsilon^1_{it} \\
v^2 &= f_2(X_{it}, k_{it}, A_{j(i,t)}, k_{it} + \gamma + \sigma_{j(i,t)} | \beta^2) + \epsilon^2_{it} \\
v^3 &= f_3(X_{it}, k_{it}, A_{j(i,t)}, k_{it} + \gamma + \sigma_{j(i,t)} | \beta^3) + \epsilon^3_{it}
\end{align*}
\]  

(26)

Assuming for example that the \( \epsilon \)'s are distributed as extreme-value type 1 would allow us to estimate and interpret the parameters \( \theta \) using the assumption 7. Functional forms for \( G(.) \) are discussed in details in the next section, but for the sake of illustration, one may for example consider:

\[
\begin{align*}
v^0 &= X^T_{it} \beta^0 + \epsilon^0_{it} \\
v^1 &= X^T_{it} \theta A_{j(i,t)} + X^T_{it} \beta^1 \ln(k_{it} + \gamma + \sigma_{j(i,t)}) + \epsilon^1_{it} \\
v^2 &= X^T_{it} \beta^2 A_{j(i,t)} \ln(k_{it}) + \epsilon^2_{it} \\
v^3 &= X^T_{it} \beta^3 A_{j(i,t)} \ln(k_{it}) + \epsilon^3_{it}
\end{align*}
\]  

(27)

where \( \beta^0, \beta^1, \beta^2 \) and \( \beta^3 \) are \((k \times 1)\), and \( \theta \) is \((k \times k_{ij})\).

In this case, the probability that a teacher chooses the alternative \( k \) will be given by:

\[
P(k) = \frac{e^{v^k}}{\sum_{i=0}^{3} e^{v^i}}
\]  

(28)

Here, the function \( G(.) \) will be identified by \( X^T_{it} \theta A \) and the marginal willingness to trade amenities will be given by (for a given individual \( X \)):
$$MWT_{ij} = \frac{\partial G_X / \partial a_i}{\partial G_X / \partial a_j} \quad (29)$$

**Database**

Last but not least, it is quite obvious that performing this kind of estimation relies on the following database hypothesis:

**Database hypothesis 1**: in addition to teachers' characteristics, matched teacher-school data, and schools' characteristics, we observe the duration before the teacher first applies to transfers or leaves the school public sector.

Notice that this database hypothesis is quite strong. However, for consistency, we found no way out of this hypothesis. Potentially however, if we assume that teachers *almost always* find a new assignment when they ask for transfers, we may base the estimation on the following database:

**Database hypothesis 2**: in addition to teachers' characteristics, matched teacher-school data, and schools' characteristics, we observe the duration before the teacher transfers or leaves the school public sector.

If we were to use this hypothesis, we would in fact assimilate the duration before applying to transfers, to the duration before actually switching schools. Therefore we would tend to overestimate the duration before applying to transfers. Then, it all comes down to the question “how much biased would be our observations?” For this bias to be small, we require two qualitative assumptions. First, we need the probability that one finds no assignment after her application to be low. And second, we also need some perseverance from the teacher, saying that if her application was to fail, it is very likely that she tries again the following year.
Section 4: The heterogeneity of tastes and mixture models

In the previous section, we provided a function $G(A, X)$ to describe the valuation of teachers for school attributes, without discussing it. In this section, we are willing to provide a plausible functional form for this function. More specifically, we will explore the issue of individual heterogeneity in a way that has not been yet been considered in the literature, as far as we know it. We will first present our model of heterogeneity, based on mixture models, and compare it with the existing literature. Then, we will derive the estimation procedure, based on maximum likelihood. Finally, we will provide a simulation of the estimation, that will illustrate important “unexpected” issues justifying several precautions in the use of mixture models.

It is important to notice that in this section, we assume that we are able to directly observe some ordinal preferences of teachers for schools. This assumption allows us to simplify the presentation of this section, and it was necessary for the purpose of providing a simulation. However, it is quite obvious that the econometric specification we are about to present can be adapted to the analysis of mobility decisions as in the previous section.

“Soft” Heterogeneity: Attitudes toward the Teaching Profession

There are strong reasons to believe that tastes for school attributes are heterogeneous across individuals. First, because it makes sense: tastes are psychological factors for which there exists no reason to prejudice homogeneous distributions. As an example (NB: we will often use this example), teachers may consider their social contribution in various ways: some may consider to have a greater contribution to social welfare by helping the neediest students (“altruistic-oriented”), while other may consider more important to support the most talented students (“elitist-oriented”), while other may consider more important to support the most talented students so that they get the highest achievements (“elitist-oriented”).

Second, because we have some statistical clues of such heterogeneity. For example, Prost (2010) finds that the “agrégés” (teachers with the highest teaching certificate in France) tend to have a lower probability to quit priority schools (who gather a lot of disadvantaged students) than other groups. Boyd et al. (2003) showed that teachers had heterogeneous preferences over school types: teachers coming from rural (resp. urban) regions would tend to prefer rural (resp. urban) schools. Or again, Antos and Rosen (1975), found that black teachers would tend to prefer teaching to disadvantaged students, whereas it is the opposite for white teachers.

To discuss this topic, let us now introduce the notations we are going to use in this section. As before, $A$ and $X$ represent respectively the school's and the teacher's characteristics. In this

---

11 We discuss this topic already: this does not mean that the agrégés are more “altruistic”: one can also explain this behavior by high preferences for elitist schools.
section, we put aside the question of bonuses: not only does it allow us to simplify notations, but also, this adds generality to our point, since the issue of bonuses is quite specific to France. If we were to do parallels with the previous section, the following function $V_i(.)$ that describes the value of a school $j$ by individual $i$ should be seen as the equivalent of the function $G(.)$:

$$V_{ij} = \alpha_i^T A_j + \eta_{ij} \quad (1)$$

Here, $\alpha_i$ is $(k_A \times 1)$. Each element of this vector is the general formulation that represent the taste of individual $i$ for the corresponding amenity. The econometric problem is to find plausible way to describe the heterogeneity of this parameter.

However, it is worth noticing that for most of the literature, the heterogeneity of $\alpha_{it}$ does not seem to be a problem. The heterogeneity is often accounted for using various hypothesis for the error-term $\eta_{ij}$: for example individual or school fixed effects... Sometimes, heterogeneity is included in $\alpha_i$ using interactive dummies of running separate regressions.

From our point of view, one of the best specification about the heterogeneity of tastes is given by Boyd et al. (2005). Indeed, they assume that the tastes of the teacher for school attributes is given by a linear function of the teacher's characteristics. For example, if $k$ represents the $k^{th}$ element of $\alpha_i$, we have according to Boyd et al. (2005)\footnote{In fact, this is a re-interpretation of Boyd et al. (2005), using the concept of “tastes” that we have introduced. In the article, the authors directly give a formulation for the value of a school. It is from this formulation that we have derived the implicit expression of the individual tastes.}:

$$\alpha_i^k = X_i^T \beta^k + \pi_i^k \quad (2)$$

where $\beta^k$ is $(k_A \times k_{D})$. Here $\pi_i$ represents a vector of random effects, that allows an important source of heterogeneity. These error-terms are normally distributed, such that overall, $\alpha_{it}$ is normally distributed with (conditional on the observations of the $X$'s) mean $(X^T \beta)^T$ (where $\beta = (\beta^1 \ldots \beta^{k_{D}})$) and covariance matrix equal to the covariance of $\pi$.

We believe however that assuming that tastes are normally distributed could be very misleading, as far as the description of tastes is concerned. Indeed, the “regularity” (symmetric and uni-modal distribution) implied by the normal distribution is unrealistic. The story we have told above, for example, (with “altruistic” and “elitist”-oriented teachers) can justify a very different type of distribution. Consider the school characteristic “share of disadvantaged students”, in this framework, one can easily imagine a distribution of tastes of the kind drawn below (Figure 2):
This kind of distribution cannot be accounted for by individual random effects. In order to allow them, we propose the use of mixture models for the distribution of $\alpha_i$ conditional on $X_i$.

The general idea behind this, is to consider that the values a teacher gives to school characteristics depend on her general orientation, or her attitude, toward the teaching profession. We think that it is important to account for this “soft” heterogeneity if one is willing to interpret the data. How to explain for example that the “agrégés” are valuing schools differently than the “certifiés”? One story could be that the type of certificate has a psychological impact on tastes... another story, more convincing from our point of view, is to say that attitudes toward teaching are composing the two groups unequally. And identically for the Blacks and the Whites of Antos and Rosen (1975).

More formally, we assume $\alpha_i$ (the parameter of interest) follows a mixture distribution. This mixture distribution is in fact a particular type of hierarchical model: we assume that there exists a set of profiles, or categories or strata $\{C_1, \ldots, C_p\}$ across which individual allocates according to a categorical distribution defined conditionally on the set of observables, and within each individual allocates according to classical (for example, Gaussian) continuous distributions. Hence, if the individual belongs to the stratum $cat$, her tastes for the vector of attributes will be given by:

$$\alpha_i^{cat} = X_i^T \beta^{cat} + \pi_i^{cat} \quad (3)$$
where the error-term $\pi^\text{cat}$ is a continuous multivariate distribution with density $\varphi^\text{cat}(\cdot)$.

The probability of belonging to a given stratum will be given by a function (for example, logistic) of $X$:

$$p_{\text{cat}} = L[ X^T y^\text{cat} ] \quad (4)$$

that will satisfies

$$\sum_{k=1}^{p} p_k = \sum_{k=1}^{p} L[ X^T y^k ] = 1 \quad .$$

**THE LIKELIHOOD FUNCTION WHEN ORDINAL PREFERENCES ARE OBSERVED**

In the case of the model described above, the density of $\alpha$ will be given by:

$$f_{\text{mix}}( \alpha | X_j ) = \sum_{k=1}^{p} p_k ( X_j ) \varphi( \alpha | X_j, \beta^k, S_k ) \quad (5)$$

Let us assume now that we observe ordinal preferences of teachers across some schools, that is, we assume that for each individual, we can observe $V_{ij} > V_{ij'} > V_{ij''}$ for some schools $j$ and $j'$ and $j''$.

This case arises in the French system when teachers are applying to transfers, as they are asked to provide a hierarchical list of wishes. In this case however, teachers have the possibility to make district wishes\(^{13}\), which may imply that the list is not completely ordered. Yet, one could then focus on intra-district transfers, or in general, consider that only the wishes that belong to the same district are ordered.

Moreover, let us assume that the error-terms in the value function, $\eta$, are distributed as extreme-value type 1 with scale parameter equals to 1, that the error-term in the taste function $\pi^k$ is distributed as a multivariate normal with covariance matrix $S^k$, and that the categorical distribution is such that the function $L(\cdot)$ is logistic. In this case, the probability to observe $V_j > V_{j'}$ is given by:

---

\(^{13}\) See section 3
\[
P(V_j > V_{j'}) = \sum_{k=1}^{n} \left[ L(X^T \gamma^k) P(V_j > V_{j'} | C_k = 1) \right] \\
= \sum_{k=1}^{n} \left\{ \frac{e^{X \gamma_k}}{\sum_{j=1}^{n} e^{X \gamma_j}} P(\tilde{\eta}_j > (X \beta^k + \pi^k)[A_j - A]) \right\} \\
= \sum_{k=1}^{n} \left\{ \frac{e^{X \gamma_k}}{\sum_{j=1}^{n} e^{X \gamma_j}} \int \frac{1}{1 + e^{(X \beta^k + \pi^k)[A_j - A]}} \varphi(\pi^k | S^k) d \pi^k \right\} 
\] (6)

where we used the fact that the difference between two independent extreme-value random variables with identical distribution is distributed as a logistic random variable.

Already, when we consider that for each individual, we only observe one ordinal relationship, computational difficulties can be expected, since the log-likelihood requires to take the log function over a sum.

Moving to the case where the hierarchy includes more than two schools makes the log-likelihood function even more complicated. For example, what is the probability to have \( V_j > V_{j'} > V_{j''} \)? Notice that in this case, assuming that the \( \eta \)'s are distributed as extreme-value does not help much, therefore, one may want to assume that these terms are distributed as standard Gaussian (that we can expect to be easier to deal with in the case of simulated likelihood).

Hence, the probability will write (where \( A_{ij} = A_i - A_j \)):

\[
P[ V_j > V_{j'} > V_{j''} ] = \sum_{k=1}^{n} \left[ L(X^T \gamma^k) P(V_j > V_{j'} > V_{j''} | C_k = 1) \right] \\
= \sum_{k=1}^{n} \left\{ \frac{e^{X \gamma_k}}{\sum_{j=1}^{n} e^{X \gamma_j}} P(\eta_j > \alpha^k [\tilde{A}_{j,j'}] + \eta_{j'}, \eta_{j''} > \alpha^k [\tilde{A}_{j'',j}] + \eta_{j''}) \right\} \\
= \sum_{k=1}^{n} \left\{ \frac{e^{X \gamma_k}}{\sum_{j=1}^{n} e^{X \gamma_j}} \right\} \int \int \int \int \varphi_0(\eta_j) \varphi_0(\eta_{j'}) \varphi_0(\eta_{j''}) \varphi(\alpha | X, \beta^k, S^k) d \eta_j d \eta_{j'} d \eta_{j''} d \alpha 
\] (7)

Computing this kind of expression is not practicable in general. One way to overcome this difficulty is to drop from the observations the bilateral orderings that are not independent. For example, if one observes \( V_1 > V_2 > V_3 > V_4 \) then the observation associated to \( V_2 > V_3 \) could be dropped such that we would end up the two independent observations \( V_1 > V_2 \) and \( V_3 > V_4 \).

Also, one could make the assumption that \( \pi^k \tilde{A} + \eta_j = \epsilon_j \) is distributed as extreme-value type
1: in this case, the likelihood function greatly simplifies as there is no more integrals in the likelihood function. Of course, this yields endogeneity issues, such that one would in fact be assuming that $\pi_k$ is equal to zero. We will make this assumption in our simulation, it will be quite clear then how this can simplify the log-likelihood.

These methods however involve a loss of information and important simplifications. If one is willing to use the “true” likelihood, there exists some ways, based on simulation and Monte-Carlo estimator to approximate the integral in (7).

**Simulated Data and Estimation**

In order to assess the performance of mixture models, we perform an estimation on simulated data. To this purpose, we use the toolbox associated with the software MATLAB, and we provide the code of this simulation in Appendix.

**Set-up of the simulation**

In the simulation, we set a number $N$ of individuals. We consider that each of these individual is characterized by a single number $X$ that takes value between 0 and 1, and such that it is uniformly distributed on the interval $[0;1]$. One may think of this parameter as representing age for example (up to an affine transformation, obviously).

Identically, schools are characterized by a single attribute $A$ that takes value between 0 and 1. One may think of this parameter as representing the proportion of disadvantaged students.

The taste for this attribute depends on the “type” of the individual, that defines a general attitude toward the teaching profession. We model two types of individuals: the type 0, that corresponds to individuals who tend to dislike the attribute $A$; and type 1, that corresponds to individuals who tend to like $A$. If for example $A$ is the proportion of disadvantaged students, type 0 can be seen as corresponding to “elitist-oriented” teachers and type 1 to “altruistic-oriented” teachers.

Formally in the model, the value associated to teaching in a school with attribute $A$ will be:

$$V(A, X) = \beta_0 X + \epsilon_0$$ if the individual is type 0 and,

$$V(A, X) = \beta_1 A + \epsilon_1$$ if the individual is type 1.

and where the $\epsilon$'s are distributed as extreme-value type 1 with scale parameter equals to 1. Notice that this is equivalent to the simplification assumptions we were making before, which is to consider that $\pi_k A + \eta_k = \epsilon_j$ is distributed as extreme-value. For the sake of illustration, one may also consider that $\beta_0$ is negative and $\beta_1$ is positive.

To use again the notation previously used, if $\alpha$ represents the taste for the attribute, we have:

---

14 But this would still be an “acceptable” assumption compared to many other models of heterogeneity in the literature.
\[ \alpha = \beta_0 X \] if type 0 and,  
\[ \alpha = \beta_1 \] if type 1.

Notice that we allow \( \alpha \) to be a function of \( X \) in the case of individuals of type 0. In the framework of the metaphor we are using, this means that the dis-utility that suffer "elitist-oriented" teachers from teaching in a school with high proportion of disadvantaged increases with age. On the contrary the age of "altruistic-oriented" teachers does not have any effect on the utility they enjoyed from teaching in schools with an high proportion of disadvantaged children.

Also, we know that the difference of two extreme-value random variables is distributed as a logistic random variable, such that for two values \( A_0 \) and \( A_1 \), we have:

\[ P[V_1 > V_0] = P[\alpha A_1 + \varepsilon_1 > \alpha A_0 + \varepsilon_0] = P[\varepsilon_1 - \varepsilon_0 > \alpha(A_0 - A_1)] = \frac{1}{1 + e^{\alpha(A_0 - A_1)}} \]

Having generated the vector of size \( N \) of individual characteristics, we generate the type of each individual, such that the probability for an individual with characteristics \( X \) to be of type 1 is:

\[ P[\text{type} = 1 \mid X] = \frac{e^{\gamma X}}{1 + e^{\gamma X}} \]

where \( \gamma \) is negative, such that the probability of being type 1 decreases when \( X \) increases (the probability to be "altruistic-oriented" decreases with age).

Knowing the type and the characteristic of the individual, we know her taste parameter. The simulation then offers two schools to the individual, with attributes being respectively \( A_0 \) and \( A_1 \) drawn from \( \text{Uniform}[0;1] \). Two random error-terms are then drawn from extreme-value distribution in order to obtain the associated \( V_0 \) and \( V_1 \). The individual then "chooses" the option that brings the highest value and we record its choice: \( id=1 \) if the individual chooses the option 1, and \( id=0 \) if she chooses the option 0.

In this set-up, the econometrician observes \( X, A_0, A_1 \) and \( id \) and wants to estimate \( \beta_0, \beta_1 \) and \( \gamma \). To do so, we maximize the log-likelihood function:
\[ l(b_0, b_1, g \mid X, A_0, A_1, id) = \sum_{i=1}^{N} \{ P[\text{type}_i = 1 \mid X_i] P[V_{i1} > V_{0i} \mid X_i, \text{type}_i = 1, A_{0i}, A_{1i}] + P[\text{type}_i = 0 \mid X_i] P[V_{i1} > V_{0i} \mid X_i, \text{type}_i = 0, A_{0i}, A_{1i}] \} \]

\[ + (1-id_i) \times \{ P[\text{type}_i = 1 \mid X_i] P[V_{i1} < V_{0i} \mid X_i, \text{type}_i = 1, A_{0i}, A_{1i}] + P[\text{type}_i = 0 \mid X_i] P[V_{i1} < V_{0i} \mid X_i, \text{type}_i = 0, A_{0i}, A_{1i}] \} \]

Using the formulas above, this yields\(^{15}\):

\[ = \sum_{i=1}^{N} \{ \frac{I(b_0, b_1, g \mid X, A_0, A_1, id)}{1 + e^{b_1(A_{0i} - A_{1i})}} + \frac{1}{1 + e^{b_0 X}} \frac{1}{1 + e^{b_1 X}} \} \]

\[ + (1-id_i) \times \{ \frac{e^{g X}}{1 + e^{b_1 X}} + \frac{1}{1 + e^{b_0 X}} \frac{1}{1 + e^{b_1 X}} \} \]

**Calibration and Results**

We set the parameters to be \(\beta_0 = -1, \beta_1 = 1\) and \(\gamma = -1\). We consider three cases: in the first one 1 000 individuals are simulated, in the second one they are 5 000, and in the third one, they are 10 000. For each case, we proceed to 1 000 simulations/estimations and we record the vector of the 1 000 estimated parameters.

With these calibrations, the probability to be “type 1” goes from 0,26 (individuals with characteristic \(X = 1\)) to 0.5 (individual with \(X = 0\))

The code is provided in Appendix. The file *mixture.m* is the central file of the simulation. In this file, one sets the number of drawn in the variable \(MC\), and the number of individuals in the variable \(N\). This file uses the opposite of the log-likelihood function provided by the file *loglik.m*. The minimization of the opposite of the log-likelihood function (which is the same as the maximization of the log-likelihood) is performed using the algorithm implemented in MATLAB called *fminsearch*.

A set of statistics and graphics are provided below, for simulations with 1 000, 5 000 and 10 000 individuals. Several lessons are to be drawn from them.

First, and unsurprisingly since it is the basic property of maximum likelihood, the estimator seems consistent. On average, increasing the number of observations allows to have better estimates: for example considering the parameter \(\beta_0\) (recall its true value is - 1), we can see that the average estimation improves from –2,18 to –1,21 when the number of individuals goes from 1 000 to 10 000. Identically, for this parameter, the variance of the estimator decreases from 14,97 to 0,78.

\(^{15}\) Notice how the assumption have simplified the log-likelihood: we have no integrals.
However, it seems that the estimator is biased. From the three simulations, it appears that the direction of this bias depends on the sign of the parameter: it is upward (estimates are greater than true parameter) when the true parameter is positive, and downward (estimates are lower than the actual value) when the true parameter is negative. At this stage, it is worth noticing that the medians of the distribution of the estimates are very close to the true value of these parameters, even for low values of N. For example, with only 1 000 observations, the median of the distribution of $\beta_0$ estimates is $-0.98$ when its empirical mean is $-2.18$.

This bias can potentially be very important. For example, for $N = 1 000$, some estimations were giving a value less than $-50$ for the parameter $\beta_0$. Even for $N = 10 000$, the probability that the estimates for $\beta_0$ is less than $-2$ is close to 20%.

The reason for these results is likely to be linked with the asymmetrical distribution of the parameters. This asymmetry is clearly illustrated in the graphics provided below, where we have plotted the kernel density of the estimators for the $N = 10 000$ case. It was also illustrated by the shift between the median and the mean of the distributions, and by the difference between the probability that the estimation is less than the true value minus one, and the probability that it is more than the true value plus one.
### Table 3: Statistics on the estimated parameters \((N=1\,000)\)

<table>
<thead>
<tr>
<th></th>
<th>Mean of estimates</th>
<th>Variance of estimates</th>
<th>Min of estimates</th>
<th>Max of estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0)</td>
<td>-2,18</td>
<td>14,97</td>
<td>-66,04</td>
<td>5,97</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>2,1</td>
<td>86,37</td>
<td>-60,08</td>
<td>250,34</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>-3,26</td>
<td>77,76</td>
<td>-148,77</td>
<td>19,39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Probability Estimates &lt; True Value - 1</th>
<th>Probability Estimates &lt; True Value + 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0)</td>
<td>-0,98</td>
<td>0,23</td>
<td>0,07</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>1,38</td>
<td>0,04</td>
<td>0,27</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>-1,06</td>
<td>0,3</td>
<td>0,22</td>
</tr>
</tbody>
</table>

### Table 4: Statistics on the estimated parameters \((N=5\,000)\)

<table>
<thead>
<tr>
<th></th>
<th>Mean of estimates</th>
<th>Variance of estimates</th>
<th>Min of estimates</th>
<th>Max of estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0)</td>
<td>-1,33</td>
<td>1,46</td>
<td>-8,05</td>
<td>0,19</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>1,2</td>
<td>0,21</td>
<td>-0,21</td>
<td>5,14</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>-1,92</td>
<td>11,34</td>
<td>-54,35</td>
<td>0,96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Probability Estimates &lt; True Value - 1</th>
<th>Probability Estimates &lt; True Value + 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0)</td>
<td>-0,93</td>
<td>0,22</td>
<td>0,01</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>1,14</td>
<td>0</td>
<td>0,04</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>-1,07</td>
<td>0,27</td>
<td>0,18</td>
</tr>
<tr>
<td></td>
<td>Mean of estimates</td>
<td>Variance of estimates</td>
<td>Min of estimates</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------</td>
<td>-----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-1,21</td>
<td>0,78</td>
<td>-5,2</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1,12</td>
<td>0,07</td>
<td>0,5</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-1,47</td>
<td>2,37</td>
<td>-10,26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Probability Estimates $&lt;$ True Value - 1</th>
<th>Probability Estimates $&lt;$ True Value + 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-0,94</td>
<td>0,19</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1,11</td>
<td>0</td>
<td>0,0040</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-1,06</td>
<td>0,26</td>
<td>0,14</td>
</tr>
</tbody>
</table>
Figure 3: Kernel Densities of the Estimates for simulations with 10 000 individuals
Lessons for using mixture model

Overall, this experiment shows that mixture model can be consistently used in order to represent the “soft” heterogeneity that one may want to introduce in order to represent teachers' preferences. For example, we also tried the simulation for very high value of N (more than 500 000), and obtained very accurate estimates: even if we still had the bias that we were noticing before, its order was less than 0.1\textsuperscript{16}.

However, we have seen that in finite sample framework (ie in estimations such that the number of observations in not so high compared to the number of parameters), the estimates could be very wrong, with an important bias.

Therefore, we believe that this kind of estimations should be done with precaution. In particular, we strongly recommend bootstrapping procedures that will allow the econometrician to have a clear idea about the distribution of the estimator. As we noticed earlier, the median of the obtained distribution is quite a good estimator of the true parameter, even for relatively low values of $N$. As a consequence, noticing that the value of the median is close to the value of the estimated parameters should be a good test to see if we have achieved proper estimates of the true parameter. In general, this bootstrapping should be used to construct confidence intervals for the estimations.

\textsuperscript{16} We did not report these results because it was not possible to simulate a number of draws important enough such that we would have a clear idea of the distribution (mean, variance...) of the estimators in the case $N = 100\ 000$. Therefore, these results should be seen as qualitative.
Section 5: Further Research

In this short section, we propose two other models, based on various assumptions, that may be interesting for further research.

A Conceptual Equivalent to the Theory of Wage-Differentials

It has been said that in the case of the French teacher labor market, the theory of compensating wage differentials could not be used for the estimation of teachers' willingness to pay for school attributes, since it does not allow a decentralized setting of wages. However, the framework and the assumptions proposed by Rosen (1974) or Antos and Rosen (1975) can provide us with a very simple way to obtain the Marginal Willingness to Trade amenities, that does not require much data.

Conceptually, our approach is similar to the wage differentials approach, as it is based on the tangency of indifference curves with the “budget constraint gradient” represented by the level of bonuses. Suppose for example that we observe a “new hired” teacher at time $t$ in some school $j$. Then, in certain circumstances (for example, the new assignment was obtained after infra-district transfers), we can deduce that the school $j$ was the best choice for individual $i$ at her given level of probability in time $t-1$. In order to make such a conclusion, several assumptions are required, that are very similar to Antos and Rosen (1975).

To begin with, let us rewrite the simplified value function as in Section 4:

$$V(a_1, a_2, A, X)$$

Where $A = A - \{ a_1, a_2 \}$.

Assume further that there exists a function that associates to a level of bonuses $k$ and a vector of school characteristics $A$ a probability that the teacher with $k$ bonuses is accepted in the school with characteristics $A$. Call this function: $p(k, a_1, a_2, A)$.

Recall that this probability function could be recover from the data. Indeed (see Section 1), unions record every year the level of bonuses required to be accepted in a school. And this information is available to teachers that are members of this union. Looking at the series of these thresholds would allows us to derive the probability distribution of the thresholds conditionally on school characteristics.

Hence, if we denote $M | A$ the random variable (with cumulative distribution function $F_{M|A}(.)$, known from estimation) that describes the future level of bonuses required to be accepted in a school with characteristics $A$, we have $p(k, A) = F_{M|A}(k)$.

We can make assumptions on $a_1$ and $a_2$ such that the marginal willingness to trade $a_1$ and $a_2$ can be recovered from the probability function. In particular, we may require that $V(.)$ is
continuous, increasing concave and differentiable in \( a_1 \) and \( a_2 \). Identically, we may require that \( p(.) \) is differentiable in \( a_1 \) and \( a_2 \).

Also, assume that the boundaries of \((a_1, a_2)\) are known and finite, such that we are able to identify interior solutions.

In this case, consider that we observe a new hired teacher in \( j' \) in \( t \) (knowing that her previous assignment was in the same district as \( j' \)). Then, if the vector \( A_{j'} \) belong to the interior of the space of available \((a_1, a_2)\), we have that:

\[
\frac{\partial V(A_{j'}, X_{t-1})}{\partial a_1} = \frac{\partial p(k_{it-1}, A_{j'})}{\partial a_1} \quad \frac{\partial V(A_{j'}, X_{t-1})}{\partial a_2} = \frac{\partial p(k_{it-1}, A_{j'})}{\partial a_2}
\]

This point is quite clearly illustrated by the following graphic:

---

**Figure 4**: In this figure, the grey area represents the continuous space of available amenities. The curves \( p_1 \) and \( p_2 \) are two elements of the "gradient of budget constraint" we were mentioning before: they represent the sets of loci for which, every other amenities being fixed, the

---

17 This assumption is partly "illustrative", where we conventionally assume that \( a_1 \) and \( a_2 \) can be described as "goods". It can be adapted for "bads".
probability to be accepted are identical. In other words, they are iso-
probability curves. Identically, $V_i$ and $V_j$ represent indifference curve in
the $(a_1, a_2)$-space. The superscript R or B stands for two different
individuals with different valuation of the amenities $a_1$ and $a_2$.

The curves $C(.)$ represents the set of optimal wishes at each level of iso-
probability.

AN APPROACH USING THE METHOD OF SIMULATED MOMENTS

Another approach can be envisaged from Boyd et al (2003). Indeed, the French case
explicitly involves a matching game, from which it would be difficult to obtain theoretical
moments or distributions, but that can be simulated such that one could find the parameters
that that allows the empirical conditions to be close to the theoretical ones.

For example, if we consider infra-district transfer, one could run exactly the same analysis as
in Boyd et al (2003) only by setting (using the same notation as in Section 2 paragraph 4) $v_{ji} = k_i$,
i.e, the preference of any school for any teacher depends only on the level of bonuses of
this teacher. This setting – which is truly implemented in France – would reduce greatly the
dimension of the estimation problem, such that this analysis could be more accurate in France
than in New-York.

However, recall that we criticized the fact that the stability conditions required to perform
Gale-Shapley algorithm were not necessarily met in reality. Is it still the case when we move
to the French case? In fact, we are at the same time in a better and in a worse position.
Suppose for example that the list of wishes in infra-district transfers were unconstrained
(illimited number of wishes), then the mechanism that allocates teachers would precisely be
the Gale-Shapley algorithm and stability conditions would be met. But since the number of
wishes is constrained, we cannot ensure that stability conditions will be satisfied.

For example, suppose individual 1 has higher bonuses than individual 2 and that they have
identical preferences over schools. There are three schools A, B and C such that A is preferred
to B is preferred to C by the two individuals. Imagine the constraint is such that individuals
can make only two wishes, then it can be the case that individual 1 (who holds more bonuses)
ask for schools A and C and individual 2 ask for schools B and C. Then, there is a positive
probability that individual 1 is not accepted in A but is in C, whereas individual 2 is accepted
in B. In this case, both school B and individual 1 would prefer to match at the expense of
individual 2, which contradicts the stability condition.

Nonetheless, it may be the case that we find an adaptation of the Gale-Shapley algorithm that
would allow us to use the method of Boyd et al (2003), but with better relevance since the
adapted stability conditions would consist in those conditions that are precisely monitored by
unions. However, our preliminary results indicate that this requires to be able to identify all
applicants to infra-district transfers.

Be that as it may, suppose we can only identify movers (i.e, successful applicants). Another
possibility would be to run the following estimation by simulated moments :
1 – as an econometrician, we locate in a single district (and within an academic discipline). The analysis requires that we observe all schools in this district, and all newly hired teachers in the schools of the district. It also obviously requires information on the teachers' characteristics and more particularly, the level of bonuses (potentially approximated using data on the teacher's career). We also need the series of the observed thresholds of bonuses by school.

2 – estimate the function \( p(k,A) \) using the observed thresholds by schools (like in previous paragraph).

3 – Assume that in the previous period, all teachers were valuating schools according to a function \( V(A, X | \theta ) + \varepsilon \), where \( \theta \) are the parameters to be estimated and \( \varepsilon \) is some error-term.

4 – Pick up a \( \theta \)

5 – For all movers, simulate values for \( \varepsilon \), this provides us with a value \( V(A, X | \theta ) + \varepsilon \) for all individuals and all schools.

6 – Knowing the values from step 5, the probabilities from step 2, and assuming that we can observe or consistently estimate the level of bonuses of the movers, infer for each individual the optimal list of wishes when the constrained list contains \( N \) wishes (this constraint can be easily found : it is 20 for example for CRETEIL, which is the general case as far as we know). Notice that in order to do this, we need to observe the position of the individual before she moved (see Section 3). Also, we do not know if there is a fast algorithm to do this. Brutally, one can nevertheless do it by computing the expected value for all hierarchical (and non-redundant) combinations of \( N \) schools, and record the maximizing argument for each individual.

7 – Simulate the accepted wish for all movers. It is very important to notice that simulating this acceptance cannot be done by simulating acceptance for each wish and taking the best accepted wish. This could be done if we were also observing those individuals who were applicants but were not assigned by the procedure. But since we assume that we only observe applicants because they are movers, the simulation should be conditional on the fact that at least one wish is accepted. To sum up, what we must simulate is a categorical variable that indicates which of the wish is validated.

One way to do this if there are two wishes. First wish has probability \( p_1 \) to be accepted, second wish has probability \( p_2 \). Then, knowing that one wish has been accepted, the probability that it is the first wish is given by:

\[
\frac{p_1}{p_1 + p_2 - p_1 p_2}
\]

And identically, the probability that it is the second wish that has been accepted.
is:

\[ \frac{p_2(1-p_1)}{p_1 + p_2 - p_1 p_2} \]

The simulation can then be obtained by drawing a random number distributed as uniform on the support \([0 ; p_1 + p_2 - p_1 p_2]\): if the obtained number is less than \(p_1\), the first wish is accepted, otherwise it is the second wish.

8 – Repeat several time from step 5 to 7, such as to obtain simulated moment condition:

\[ E_{\text{emp}}[X_i(A_{i,j(i)}^{\text{obs}} - E_{\text{sim}}[A_{i,j(i)} \mid X_i, k_i])] \]

where \(E_{\text{emp}}\) is the empirical mean, and \(E_{\text{sim}}\) is the simulated mean.


10 – the estimated \(\theta\) is the one for which the simulated moment condition fits the best the theoretical moment condition:

\[ E[X_i(A_{i,j(i)}^{\text{obs}} - E[A_{i,j(i)} \mid X_i, k_i])] = 0 \]
Conclusion

Understanding teachers' preferences for school attributes is crucial if one is willing to implement optimal policies of equal opportunities.

Indeed, the economic literature on this point is quite clear: the preferences of teachers tend to drive the most qualified away from disadvantaged schools. In France, the policy of priority education zones did not allow to improve the situation, and from Bénabou, Kramarz and Prost (2005), we may even suspect that it made it worse.

The very low wage compensations associated with teaching in disadvantaged schools in France are likely to be an explanation for this inefficiency. Increasing the wages of teachers in disadvantaged areas could be a solution. However, several studies have pointed out the fact that pecuniary factors were not determinant in mobility decisions of teachers, even if it obviously depends on the magnitude of the compensation.

Be that as it may, financial compensations are unlikely to be the least costly way of improving the repartition of teachers. For example, the literature has shown the importance of the social composition in the teacher's assessment of a school's value. Quite ironically, it seems that the issue of teachers allocation has much to do with the issue of students allocation. On this topic, it is an important and debated question in France to know how should be implemented student repartition: to what extent is the State allowed to restrict student mobility so as to ensure social diversity? As an illustration of these debates, the recent reforms that have permitted a greater mobility of students have been criticized, notably by the Cour des Comptes, for their negative effects on the most disadvantaged schools.

The present dissertation was willing to provide a contribution, by proposing a theoretical framework in order to analyze the mobility decisions of teachers in the French teacher labor market. Indeed, the classical analysis we can find in the literature concerning the teacher labor market cannot directly be used in the French case. This fact has to do with the specificity of the French system of teacher promotion, that may generate paradoxical behaviors. For instance, what would be the effect of increasing the speed of bonus accumulation for teachers in disadvantaged areas? For some teachers, it offers the possibility to leave the school sooner, whereas other teachers may consider it as a good reason to stay longer...

However, we believe that these complications may be an opportunity to undertake original estimations of teachers' preferences, and we have proposed some ideas going in this direction. In particular, we want to stress the importance of looking at the bonuses thresholds that have been recorded by unions: these data can indeed prove very fruitful in understanding the budget constraint and the strategic arbitrage faced by French teachers.

Also, we have tried to put forward mixture models that we believe will be more accurate in describing the heterogeneity of tastes, by comparison with the ones that we have seen in the literature. Indeed, they allow to consider that teachers have different attitudes toward their profession, which we think is likely to be true.

As a last word, it is important to say – even if one may find it obvious – that the dimension along which we have been looking at the teacher labor market is only a small part of teacher-related issues. Indeed, it can be the case that the unequal repartition is in fact a good thing, for example, if one considers that young teachers are more efficient in teaching disadvantaged
schools. Therefore, if the analysis of teachers' preferences is necessary, it is certainly not sufficient to achieve optimal policy.
Bibliography


“L'éducation nationale face à l'objectif de la réussite de tous les élèves”, rapport public thématique, Cour des Comptes, Mai 2010.

PROST, Corinne (2010), Teacher mobility: can financial incentives help disadvantaged schools to retain their teachers?, mimeo


Appendix

1. The file mixture.m

MC=1000;
z = zeros(3,MC);
for j = 1:MC
    N=100000;
    X=rand(N,1);
    A0=rand(N,1);
    A1=rand(N,1);
    gamma = -1;
    beta0 = -1;
    beta1 = 1;
    p = exp(gamma*X)./(1 + exp(gamma*X));
    C = binornd(ones(N,1),p);
    alpha = beta1*C + beta0*X.*(1-C);
    V0 = alpha.*A0 + evrnd(zeros(N,1), 1);
    V1 = alpha.*A1 + evrnd(zeros(N,1), 1);
    id = V1 >= V0;
    [x, fval, exitflag]=fminsearch(@(x)liklh(X,A0,A1,id,x(1),x(2),x(3)),[-1,1,-1]);
    z(1,j) = x(1);
    z(2,j) = x(2);
    z(3,j) = x(3);
end
2. The file liklh.m

function [L]=liklh(X, A0, A1, id, b0, b1, gamma)

m= log((1./(1+exp(gamma*X))).* (1./(1+exp(b0*X.*(A0-A1)))) + ((exp(gamma*X))./(1+exp(gamma*X))).* (1./(1+exp(b1*(A0-A1)))));

n= log((1./(1+exp(gamma*X))).* (1./(1+exp(-b0*X.*(A0-A1)))) + ((exp(gamma*X))./(1+exp(gamma*X))).* (1./(1+exp(-b1*(A0-A1)))));

g= m.*id + n.*(1-id);

L=-sum(g);