Model Predictive Control Strategy to Forecast Employability in Earth Sciences

E Courtial, C Garrouste

To cite this version:

HAL Id: insu-01114917
https://hal-insu.archives-ouvertes.fr/insu-01114917
Submitted on 10 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Model Predictive Control Strategy to Forecast Employability in Earth Sciences

E. Courtial* and C. Garrouste**

* Laboratoire PRISME, University of Orléans, INSA CVL, EA 4229, F45072, Orléans, France
(e-mail: Estelle.Courtial@univ-orleans.fr).

** Laboratory of Economics of Orléans, University of Orléans, CNRS, UMR 7322, Orléans, France
(e-mail: christelle.garrouste@univ-orleans.fr)

Abstract: Energy prices and environmental policies influence more than ever employment trends across the world. The purpose of this paper is to develop a control strategy to enhance the employability of French graduates in a field that is both a key driver and a significant target of these new trends, namely Earth Sciences. The aim is to provide French universities with a predictive tool to adjust efficiently their skills' supply capacity with the demand forecasts at the European level. This task is treated as a tracking problem from the viewpoint of the control theory. The reference trajectory is obtained via a labour market forecasting model. For the first time, an econometric model and a predictive control strategy are combined. Simulations illustrate the feasibility and potentials of the proposed approach.

1. INTRODUCTION

In the past two decades, energy consumption, sustainable development and environmental protection have become priorities of energy policies in most countries (Kyoto Protocol, Grenelle environment in France, Carbon plan in the United Kingdom, 20-20-20 targets in the European Union). The energy transition is leading to many changes in the economy as a whole, in labour markets structure and dynamics, in societal behavior, and in research and education. Among the study and research disciplines the most affected by these changes is Earth Sciences (ES). ES includes the study of atmosphere, hydrosphere, oceans and biosphere as well as solid Earth. Among the many challenges, the management of mineral resources is a real challenge for the society and the environment (green mining) because of its direct impact on consumption, labour and the economy. For instance, rare-earth minerals (REM) are increasingly used for the production of high-tech items such as smart phones, laptops but also in magnets for wind turbines, hybrid-car batteries, etc. After a misleading forecasting of its own consumption needs, the U.S. lost its leading producer position in favour of China and is now constrained to import at a very high price the goods it used to produce domestically. As a consequence of the closure of the mining extractions in the 1990s, the U.S. also decreased its investments in the training of solid Earth scientists. While it now faces the need to re-open its REM mining sites to satisfy an increasing demand for high-tech goods, it suffers from a deficit in qualified Earth System scientists. A similar deficit affects Australia and Canada.

With its technical and scientific competences, its historical assets and its first-class actors in the field of ES, France aims at becoming a worldwide leader in the ES training. In 2011, the French Ministry of Higher Education nominated the project VOLTAIRE as a LABEX (Laboratory of excellence). Among the tasks of this project is the construct of an anticipation tool to ensure the employability of Earth System scientists trained in French universities. The CIPEGE center (International Center for the Prediction of Employment in Earth and Environmental Sciences) was created to handle this task. In the sequel, the anticipation tool will be called the CIPEGE tool. This paper presents the innovative forecasting strategy adopted to develop the CIPEGE tool, which combines economics (an econometric model) and control process strategies (a model predictive control).

The employability of French ES graduates is measured and forecasted using a labour market micro-econometrics model, controlling for European macroeconomic trends. The task of this study is then to track the reference trajectory of the forecasted employability for the French graduates in ES. This objective can be viewed as a tracking problem from a control theory perspective. Among the existing advanced control laws, Model Predictive Control (MPC) is a control strategy well-adapted to deal with tracking problems (Alessio [2009], Camacho [2007]). The success of MPC in several industrial sectors is due to the easy way to formulate the control objective in the time domain and also to the ability to handle constraints (Qin [2003]). A wide variety of applications has been reported in the literature but no application in labor economics exists to our knowledge. MPC is based on the direct use of an explicit model to predict the future behavior of a process. This model plays a crucial role in the MPC strategy. In our case, an econometric-based model is used to forecast the behavior of the process considered (the flow of ES

* This work is supported by the “Investissements d’Avenir” initiative of the French Ministry of Higher Education and Research, through the project VOLTAIRE (ANR-10-LABX-100-01).
graduates in France) over a finite prediction horizon.

The main advantage of this strategy lies in the continuous
and systematic nature of its prediction, which enables us
to define different time horizons (at the contrary of classic econometrics) and to correct the trajectory continuously
thanks to the feedback mechanism of the MPC approach. Another advantage is the fact that this strategy can deal
with estimation errors and modeling errors and take them into account to correct the reference value at each run.
This two-step error correction procedure constitutes a valuable tool towards more robust estimates. It improves
the econometric model and, consequently, the MPC’s efficiency. Moreover, the MPC approach determines the con-
trol inputs exogenously in such a way that it is applied simultaneously to the process and the model. In an econo-
metric model, the effect of the exogenous control input is usually estimated inside the model and is considered as
endogenous to the process. For the first time, the definition of the control inputs as completely exogenous inputs pro-
vides policy makers with full flexibility to test an unlimited range of possible interventions.

The remainder of this paper is organized as follows. The section II presents the definition of the employability re-
tained for the CIPEGE tool and the model applied to measure and forecast the French ES graduates’ employability. Section III briefly recalls the principle of Model Predictive Control and details the control structure used. Section IV addresses the way of combining an econometric model of employability and a predictive control approach. Then, in section V, different simulations serve at testing the feasibility of the control strategy. In the last section, we synthesize our preliminary results and draw the perspectives of the proposed approach.

2. EMPLOYABILITY

Measuring employability of graduates is a controversial issue due to the difficulty to apply a straight-forward definition (Gazier [1998], McQuaid [2005], Arjona Perez [2010]). Employability is a complex and multi-faceted concept. Therefore, either because of a lack of compatibility between dimensions or a lack of data, a holistic measure of employability has so far been recognized to be impossible. Employability measures are instead reduced to the most pertinent dimensions for the study at hand.

2.1 Definitions

In McQuaid [2005], the authors highlight the existence of two alternative perspectives in the employability debate. One focuses only on the individual’s characteristics and skills, referring to the individual’s potential to obtain a job. The other perspective takes into account external factors (e.g. labor market institutions, socio-economic status) that influence a person’s probability of getting into a job, of moving between jobs or of improving his or her job. In De Grip [2004], these factors are called “effectuation conditions”, i.e. the conditions under which workers can effectuate their employability. In addition, the literature also considers the aspects of the time lag between leaving

education and employment (Boateng [2011]), the degree of
skills matched between one’s educational background and
his or her occupation, and the type of contractual arrange-
ment (full-time vs. part-time; permanent vs. temporary) (Arjona Perez [2010]).

Employability is about having the capability to gain ini-
tial employment, maintain employment and obtain new employment if required (Cedefop [2008]). In other words, the employability of a graduate is the predisposition of the graduate to exhibit attributes that employers anticipate will be necessary for the future effective functioning of their organization (Harvey [1998]). Hence, employability is a combination of capacity and willingness to be and to remain attractive for the labor market, for instance, by anticipating changes in tasks and work environment and reacting on them (De Grip [2004]). For the individual, em-
ployability depends on the knowledge, skills and attitudes they possess, the way they use those assets and present them to employers (Hillage [1998]).

In the context of the CIPEGE tool, we define employability as the capacity of a French Earth Sciences graduate to be employed at a fulfilling job that enables him or her to make use of the skills acquired during the training, given the demand trends of the relevant sectors of activity at the European level.

2.2 Modeling

Let \( y^F_j(t) \) denote the number of employed individuals in France who are graduated in Earth Sciences at degree level \( j \), with \( j \in \{3, 5, 8\} \), such that \( j = 3 \) to a 3-year degree (i.e. Bachelor degree); \( j = 5 \) to a 5-year degree (i.e. Master degree) and \( j = 8 \) to an 8-year degree (i.e. Ph.D. degree). The number of employed ES graduates with a level-\( j \) degree is the total number of working aged \( j \)-level ES graduates minus the number of unemployed \( j \)-level ES graduates, at the sampling time \( t \):

\[
y^F_j(t) = S^F_j(t) - un^F_j(t) \tag{1}
\]

where \( S^F_j \) is the stock of ES skills on the French market and \( un^F_j \) is the stock of unemployed ES graduates in France. We model employment as a matching function as suggested by Mortensen et al. (Mortensen [1994]) to describe the formation of new relationships (also called 'matches') from unmatched individuals (or agents) of the appropriate types. In our case, we are interested in the formation of matches between the suppliers and the demanders of ES skills. We assume our matching function to have the following Cobb-Douglas form:

\[
m^F_j(t) = M(un_j(t), v_j(t)) = \mu(un_j(t))^a(v_j(t))^b \tag{2}
\]

where \( m^F_j(t) \) is the number of new matches created at current time \( t \) on the European market and \( \mu, a, b \) are positive constants. While \( un_j \) is now the stock of unemployed ES graduates in Europe, \( v_j(t) \) is the number of job vacancies in ES field at degree level \( j \). The matching function is increasing, concave and homogeneous of degree 1. As reviewed by Petrongolo [2001], the Cobb-Douglas form of the matching function can be justified by empirical evidence of constant returns to scale, i.e. \( a + b \approx 1 \). If the fraction of workers separating from a firm per period
of time (due to firing, quits, and so forth) is \( \delta \), then the change in employment from one period to the next is calculated by adding the formation of new matches and subtracting the separation of old matches. Combining equations (1) and (2) yields the following representation of the dynamics of employment over time in France:

\[
y_p^F(k + 1) = m_j^{EU}(t) + (1 - \delta) y_p^F(t) \\
y_p^F(k + 1) = \mu(u_j(t))^a(v_j(t))^b + (1 - \delta) y_p^F(t)
\]  

(3)

where \( \delta \) is a parameter to be estimated.

The influence of international environmental and economic shocks on the demand for skills in ES is captured by \( m_j^{EU} \), which is estimated using an extended version of the energy-environment-economy model of Europe (E3ME), developed by Cambridge Econometrics, to forecast skills supply and demand in Europe (Wilson [2010]). All other parameters of the model are estimated using the 1990-2011 microdata collected by INSEE (French National Institute of Statistics and Economic Studies) through the Employment Survey.

3. MODEL PREDICTIVE CONTROL (MPC)

3.1 Principle

MPC is a mature control strategy. Initially developed for linear systems in the 70s, MPC had extensively been studied for nonlinear systems with constraints and successfully applied in numerous industrial domains (Alessio [2009], Qin [2003]). The MPC strategy is based on the receding horizon principle and is formulated as solving online a nonlinear optimization problem (Camacho [2007]). The basic concepts of MPC are the explicit use of a model to predict the process behavior over a finite prediction horizon \( N_p \) and the minimization of a cost function with respect to a sequence of \( N_c \) controls where \( N_c \) is the control horizon. At the current instant \( t \) (see Fig. 1), the process output is measured and the MPC algorithm computes a sequence of \( N_c \) control inputs by minimizing the tracking error (difference between the reference trajectory and the predicted model output) over \( N_p \). Only the first element of the obtained optimal control sequence is really applied to the process. At the next sampling time (see Fig. 2), the finite prediction horizon moves a step forward, the measurements are updated and the whole procedure is repeated. Given its formulation in an optimization problem, MPC is well suited to take into account constraints. It is the most effective way to satisfy all kinds of constraints (on states, inputs or outputs) by adding them explicitly to the optimization problem.

3.2 Internal Model Control (IMC) Structure

Predictions based on data are inevitably subject to disturbances and modeling errors. To gain in robustness, the well-known Internal Model Control (IMC) structure (see Fig. 3) is considered in this approach.

The real process is described by its mathematical model. The control inputs \( u \) are simultaneously applied to the process and the model. The difference between the process output \( y_p \) and the predicted model output \( y_m \) provides an error signal \( e \). This signal embeds disturbances and modeling errors and constitutes the feedback information impacting on the reference trajectory. The feedback information is taken into account in an original way rarely used in economics. Due to the nature of sample, a discrete-time formulation is considered where \( t \) is the sampling time.

According to Fig. 3, we can write:

\[
y_d(t) = y_{ref}(t) - e(t) \\
y_d(t) = y_{ref}(t) - (y_p(t) - y_m(t)) \\
y_d(t) - y_m(t) = y_{ref}(t) - y_p(t).
\]  

(4)

The tracking of the reference trajectory \( y_{ref} \) by the process output \( y_p \) is equivalent to the tracking of the desired trajectory \( y_d \) by the model output \( y_m \).
4. ECONOMETRIC MODEL-BASED PREDICTIVE CONTROL

This section addresses the way of combining an econometric model and the MPC approach. The common points to all predictive strategies are discussed according to the control objective: the improvement of French graduates’ employability in the field of ES.

In the context of this study, the term "prediction" actually refers to an "anticipation" or "forecast".

4.1 The reference trajectory

The reference trajectory corresponds to the expected behavior of the process. In our case, the reference trajectory \(Y_{ref}\) to be tracked corresponds to the employability of French ES graduates. This reference has been determined off-line by estimating equations (1), (2) and (3) using E3ME outputs and French microdata (INSEE Employment Survey). The following figure (see Fig. 4) describes the future general trend of employability in ES in France at level \(j = 3\), i.e. Bachelor degree. The predicted values of employability correspond to the fitted values of employability, i.e. the points where a particular \(x\)-value fits the line of best fit. They are found by substituting a given value of \(x\) into the regression equation \(\hat{y} = b_0 + b_1x\). In economics, because of the high degree of unpredictability of individual behaviors, fitted values are likely to variate within a 90 percent confidence interval.

![Reference trajectory for the employability in ES at Bachelor degree level.](image)

Concerning the random component of (eq. 5), we assume that the vector \(\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iJ})\) has a multivariate normal distribution with mean vector 0 and arbitrary correlation matrix \(R\). The main advantage of this model is that it allows correlation between the utilities that an individual assigns to the various alternatives. Finally, the number of students estimated to choose alternative \(i\) is given by:

\[
\hat{N}(t) = \sum_{s=1}^{S} w_s P_{si}
\]

where \(P_{si}\) is the probability (result of eq. 6) that a student in study field \(s\) chooses alternative \(i\) and \(w_s\) is the number of students enrolled in study field \(s\).

For control purpose, the model is written and identified under a state-space representation. For example, the employability model at Bachelor degree level \(j = 3\) can be given by:

\[
\begin{bmatrix}
\hat{N}_1 \\
\hat{N}_2 \\
\hat{N}_3 \\
\end{bmatrix}
(t)_i = \begin{bmatrix}
\hat{N}_1 \\
\hat{N}_2 \\
\hat{N}_3 \\
\end{bmatrix}
(t-1) + B \begin{bmatrix}
u_1(t) \\
u_1(t-1) \\
u_1(t-2) \\
\end{bmatrix}
\]

\[
A = \begin{bmatrix}
r_3 & pr_2 & 0 \\
0 & r_2 & pr_1 \\
0 & 0 & r_1 \\
\end{bmatrix}, B = \begin{bmatrix}
0 & 0 & en_3 \\
0 & en_2 & 0 \\
en_1 & 0 & 0 \\
\end{bmatrix}.
\]

\(\hat{N}_j\) represents the potential (or estimated) number of graduated students respectively at level \(j = 1, 2, 3\). The control input is the number of students enrolled at time \(t\), \(t - 1, t-2\), at level \(j = 1\). The diagonal coefficients \(r_j\) are the repetition rate at level \(j\). All the other parameters are constants.
4.3 The cost function (optimization criterium)

It is usually a quadratic function of the tracking error. The error signal $e$ over the prediction horizon is computed thanks to a linear interpolation over the past measured errors, and it is updated at each measurement. Since the reference $y_{ref}$ is known over the whole working horizon, the desired trajectory can be computed:

$$y_d(k) = y_{ref}(k) - e(k), \quad k \in [t + 1, t + N_p],$$

(12)

and the cost function can be written in discrete-time as:

$$J(u) = \sum_{k=t+1}^{t+N_p} e_{tra}(k)^T Q e_{tra}(k) + \Delta u(k-1)^T R \Delta u(k-1)$$

(13)

where $e_{tra} = y_d - y_{mu}$. $Q$ and $R$ are symmetric definite positive matrices and $\Delta u(k-1) = u(k-1) - u(k-2)$.

4.4 The solving optimization method

The cost function $J$ is to be minimized with respect to a sequence of $N_c$ different controls noted $\hat{u} = \{u(t), u(t+1),..., u(t+N_c),..., u(t+N_p-1)\}$ where $N_c$ is the control horizon ($N_c < N_p$). From $u(t + N_c + 1)$ to $u(t + N_p - 1)$, the inputs are constant and equal to $u(t + N_c)$. The mathematical formulation of MPC is then given by the following optimization problem:

$$\min_{\hat{u}} J(u).$$

(14)

Although the prediction and optimization steps are performed over the prediction horizon, only the value of the input for the current time $u(t)$ is really applied to the process.

5. SIMULATIONS

All the presented simulations are performed with Matlab software. The constrained optimization problem is solved by using the Matlab function `fmincon`. Due to a lack of space, only the results of the employability at Master degree level (i.e. $j = 5$) are presented.

5.1 Data and modeling

The internal model uses data from a student tracking survey collected by the Students’ Life Observatory (OVE), which describes the transition trajectories of Master graduates 3 years after graduation, by degree field; university administrative records of the number of intakes and graduates, per year; data from ES Master programmes’ curricula; and complementary data from the report by Varet [2008]. The data collected are represented on Fig. 5. As explained in section 4.1, the reference trajectory is calculated using INSEE and E3ME data. The inputs are obtained from the OVE data and the administrative data. As can be seen, the trajectory of the measured employability is very nonlinear.

According to the same modeling procedure described in section 4.2, the state-space representation of the employability model at Master degree level can be written as follows:

$$\begin{bmatrix} N_e(t) \\ N_4(t) \end{bmatrix} = \begin{bmatrix} r_5 & p r_4 \\ 0 & r_4 \end{bmatrix} \begin{bmatrix} N_e(t-1) \\ N_4(t-1) \end{bmatrix} + \begin{bmatrix} e n_5 & 0 \\ 0 & e n_4 \end{bmatrix} \begin{bmatrix} u_5(t) \\ u_4(t) \end{bmatrix}$$

(15)

Fig. 5. Reference trajectory for the employability in ES at Master degree level.

$N_j$ represents the potential (or estimated) number of graduated students respectively at level $j = 4, 5$. Due to the fact that parallel admissions are possible in the second year of the Master, the number of graduated students at level $j = 5$ is impacted by the number of students enrolled at level $j = 4$ and $j = 5$, at time $t$.

Thanks to an identification procedure, we obtained a model which matches the process with a relative error of 9.33% (see Fig. 6).

5.2 Predictive Control

The econometric model-based predictive control described in section 4 is implemented. The simulation was performed under the following conditions: $N_p = 5$, $N_c = 4$, $R = [10; 00.1]$ and $Q(j) = Q(1)^j$ with $Q(1) = 2$. The future tracking errors are more and more weighted in order to give importance to the final objective, i.e. the desired employability at the end of the prediction horizon, and this, at each sampling time. The reference employability is obtained applying the model described in section 4.1, using INSEE and E3ME data. Several simulations were carried out according to different horizons of control and prediction. The prediction horizon $N_p = 5$ seems to be the best compromise between the tracking accuracy, the intrinsic dynamic (the current control will impact the employability of Master degree graduates in at least five years) and the stability of the controlled system.

We can see (see Fig. 7) that the process output tracks the reference trajectory by remaining within the range of...
6. CONCLUSIONS

This paper presented an innovative tool to predict the employability of French graduates in Earth Sciences. Rather than formulating this control objective as a classic computational general equilibrium (CGE) problem, this control objective has been formulated into an optimization problem. For the first time, Model Predictive Control was combined with an econometric model of employability. The MPC enables to take into account disturbances and modeling errors through an internal model control, which complements efficiently the error correction model (ECM) implemented in the econometric model used to measure the reference trajectory. Moreover, combining the econometrics approach and the MPC yields a predictive tool where the control inputs are held exogenous to the optimization process, which makes it possible to test an unlimited range of possible interventions. The calculated control inputs can then serve as potential action-tools for policy makers. Furthermore, because this approach is very flexible, it can easily be adapted to other disciplines (chemistry, medicine, ...) but also to other countries. Hence, the statistical capacity (in terms of error control), the economic relevance (in terms of control inputs formulation) and the unlimited potentialities for application expansions of the CIPEGE tool, makes it an attractive and valuable decision tool for universities and policy makers.

As with any approach of predictive control, the model is the cornerstone of the strategy and needs to be clearly identified from consistent data.

At this early stage of the project, the results obtained from different simulations are very encouraging. Additional data (which are expected shortly) should improve the model and, thus, the tracking accuracy of the reference employability.

ACKNOWLEDGEMENTS

The authors would like to thank E. Vergès and Y. Touré, initiators of the project, for having given them the opportunity to work on this challenging project and M. Fodha for fruitful discussions.

REFERENCES


