Évaluation comparative de modèles de prédiction : application à la masse salariale publique

Benoît Sainct [1] & Philippe Besse [2]

[1] *Adelyce SAS, benoit.sainct@adelyce.fr*
[2] *INSA Toulouse, philippe.besse@insa-toulouse.fr*

Résumé. Dans cet article, on s'intéresse à la prédiction à long terme de la masse salariale des collectivités, et plus précisément à l'évolution du coût mensuelle par individu. Les dépenses de la fonction publique territoriale présentent à la fois de nombreuses disparités et de grands volumes de données, ce qui nécessite une prise en charge appropriée si l'on souhaite réussir à prédire correctement les valeurs futures. Deux modèles rapides et automatiques sont présentées, et l'efficacité de leur application est comparée à l'aide d'outils standards. Les résultats montrent une coïncidence significative entre la prévision et la réalité.

Mots-clés. Prévision, Séries temporelles, ARIMA, ETS, Application

Abstract. In this paper, we deal with the long term forecast of payroll in territorial communities, and more specifically the evolution of the monthly cost per individual. Expenditure in the public service presents both large disparities and high volumes of data, which requires appropriate care in order to be able to correctly predict future values. Two quick and automatic models are presented, and the effectiveness of their application is compared using standard tools. The results show a significant match between prediction and reality.

Keywords. Forecast, Prediction, Time series, ARIMA, ETS, Practical application

1 Introduction

Une des plus grosses pressions qui s'appliquent sur les collectivités territoriales en France est la diminution annuelle de la dotation de l'état [8]. Celles-ci doivent apprendre à gérer de mieux en mieux leurs dépenses, dont le pôle principal est la masse salariale. Ainsi, prédire ces montants avec justesse plusieurs années à l'avance est un enjeu capital.

Dans la fonction publique, de nombreux contrats différents peuvent être utilisés pour employer un individu à une tâche donnée, chacun ayant ses particularités. La difficulté est d'isoler les informations pertinentes, sans toutefois être obligé de faire un calcul par individu.

Nous comparons ici plusieurs méthodes de prédiction automatique, c'est-à-dire pour lesquelles aucune intervention humaine n'est nécessaire : les modèles ARIMA, et les modèles ETS. La modélisation par séries temporelles est la plus adaptée à ce problème, comme le montre Montgomery dans [6]. Les travaux de Rob Hyndman [5] reprenant des calculs précédents de Hannan & Rissanen [4] implémentent les fonctionnalités de bases qui sont utilisées ici.

Malgré des débats dans la littérature sur le sujet, l'erreur quadratique moyenne (ou *RMSE*) est dans notre situation la mesure la plus appropriée, comme le montrent [1, 3].

Afin de gérer la masse de calculs à effectuer, l'implémentation concrète est faite à l'aide d'outils de parallélisation reposant sur la technologie *Snow*, bien étudiée dans [2].

2 Particularités de la masse salariale

Le niveau de granularité des calculs doit répondre à deux exigences : d'une part, il est souhaitable de faire le moins de calculs possible, ce qui pousse à prédire des groupes d'individus plutôt que des individus. D'autre part, pour que les clients puissent établir leurs budgets, il leur faut une granularité la plus fine possible.

Le régime (privé ou publique), la période temporelle de référence pour la rémunération (mensuelle, journalière, horaire), les avantages fiscaux et les différents régimes indemnitaires sont autant de distinctions entre les contrats que peuvent passer les collectivités avec leurs employés. On compte au total plus d'une vingtaine de profils de rémunération différents, qu'il ne faut pas amalgamer. Au sein de chaque profils sont regroupés des corps de métiers différents, qu'il est également souhaitable de distinguer.

Le niveau de détail qui semble optimal est celui du *grade* des individus. Ceci permet de lisser les irrégularités qui apparaissent aux niveaux inférieurs (l'échelon, et l'individu), tout en gardant une cohérence en termes de métier, de niveau de responsabilité et de rémunération.

Deux phénomènes impactant la masse salariale intéressent particulièrement les collectivités : le Glissement Vieillissement Technicité (ou *GVT*), et l'effet Noria, qui témoignent de l'avancement et du remplacement des employés. Ainsi, ce n'est pas le montant total des dépenses qu'il faut étudier, car il ne permet pas à lui seul d'identifier ces phénomènes. Nous prédisons donc d'une part la quantité d'Équivalent Temps Plein (ou *ETP*) de chaque grade, et d'autre part le coût moyen par ETP (ou *CM*), afin de pouvoir expliquer le plus clairement possible les évolutions futures des dépenses.

Les variations de ces deux séries temporelles, et particulièrement celles du CM, sont fortement impactées par la conjoncture économique et par les législations successives, comme typiquement la réforme PPCR (voir [7]). Ceci invalide l'hypothèse de stationnarité des séries temporelles, qui est souvent admise par défaut. De même, l'hétéroscédasticité des données est parfois assez marquée, ce qui rend moins légitime l'utilisation de certains modèles.

3 Méthodologie

Soit $\{x_t \in \mathbb{Q}\}$ une série temporelle rationnelle, définie pour t de 1 à T.
Lorsque l'on souhaite prédire les valeurs d'une série temporelle, le modèle de référence est le modèle stationnaire, aussi appelé modèle naïf, qu'on note M_{naif}, et qui donne $\forall h \in \mathbb{N}^*, \hat{x_{T+h}} = x_T$.

Dans notre contexte, le réflexe naturel serait d'ajuster un processus ARMA aux données. Néanmoins, la non-stationnarité des séries nous indique que leur utilisation ne fournira pas de résultats satisfaisants.
On peut par contre étudier les différences d'ordre 1 ou plus : $\Delta^d x_t = (1-L)^d x_t$, ce qui se traduit par l'utilisation de la famille des modèles ARIMA.

Le second type de modèle étudié est la famille des modèles à lissage exponentiel. Elle s'est agrandie ces dernières années jusqu'à contenir 30 modèles différents, comme exposé dans [9]. On les note $ETS(\cdot,\cdot,\cdot)$, pour *Error*, *Trend*, *Season*, et chacun des trois paramètres indique la forme (Nulle, Additive ou Multiplicative) de chaque composante.

Dans les deux cas, pour chaque modèle possible au sein d'une famille, les paramètres sont estimés par maximum de vraisemblance. Ensuite, la sélection de modèle dans une famille s'opère de manière standard à l'aide du Critère d'Information d'Akaike (AIC).

4 Mesure d'erreur

Dans le cas d'une série temporelle x_t, on peut mesurer l'erreur générée par un modèle M sur l'ensemble des valeurs prédites $\hat{x}_t$ jusqu'à une date h en calculant :

$$RMSE_M(x,h) = \sqrt{\frac{\sum_{t=T+1}^{T+h} (\hat{x}_t - x_t)^2}{h}}$$

Cette définition permet de comparer la qualité de deux modèles pour un même jeu de donnée. Néanmoins, la valeur réalisée de la RMSE dépend de l'ordre de grandeur des valeurs de x : elle n'est pas normalisée. Cette mesure seule n'est donc pas satisfaisante.

Nous cherchons ici à évaluer l'amélioration d'un modèle par rapport à un autre sur un large ensemble E de jeux de données distincts. On utilisera la notion d' "amélioration" d'un modèle par rapport à un autre, définie de la manière suivante :

$$A_h(M_1, M_2) = \frac{1}{Card(E)} \sum_{x \in E} \frac{RMSE_{M_1}(x,h)}{RMSE_{M_2}(x,h)} - 1$$

C'est-à-dire le ratio moyen de l'erreur commise par le modèle M_1 par rapport au modèle M_2 sur chacune des séries temporelles x d'un ensemble E, pour un horizon h fixé. Si cette mesure est positive, alors le modèle M_2 commet moins d'erreurs que le M_1, sinon c'est l'inverse.
Cette mesure a pour objectif de permettre une comparaison simple des différents modèles de prédiction ; on l'utilise en particulier en comparaison avec le modèle naïf.

5 Résultats

Les données en provenance de cinq établissements en particulier sont utilisées pour mesurer l'efficacité de chaque modèle : Metz, Versailles, Drancy, Pertuis et la communauté de communes de l'Île d'Oléron. La taille de chacune de ces collectivités est représentative du type de structure qui existe dans la base de données entière.

Notre meilleure stratégie consiste à utiliser un modèle ETS à chaque fois que l'historique le permet, et sinon ajuster un modèle ARIMA. Les mesures ont été effectuées sur la variable la plus importante, c'est-à-dire le Coût Moyen.

Pour les collectivités les plus larges et pour un horizon de prédiction de 24 mois, on a $A_{24}(Arima.seul, Modele.naif)$ qui vaut en moyenne -31%. C'est-à-dire que l'erreur de prédiction est 31% plus faible lorsqu'on utilise systématiquement un modèle ARIMA pour chaque série temporelle, par rapport à une prédiction naïve. Lorsqu'on considère une prédiction à 6 ans, soit 72 mois, le résultat s'améliore encore, et la diminution d'erreur atteint $A_{72} = 47\%$.
De même, on a $A(Ets.sinon.Arima, Arima.seul)$ qui vaut en moyenne -15% pour un horizon de 24 mois, ce qui est également une nette diminution de l'erreur. On a donc bien intérêt à ajuster un modèle de la famille ETS lorsque c'est possible.
Ces améliorations restent valables pour les collectivités de taille moindre, avec cependant d'importantes variations locales.

La validité des intervalles de confiance a été vérifiée expérimentalement. Lors de prédiction à 24 mois, en moyenne, il y a toujours plus de 90% des prédictions qui tombent effectivement dans l'intervalle de confiance à 90% ; il y a donc une légère sur-évaluation de la taille des intervalles.

6 Conclusion

Après avoir isolé les informations pertinentes et ciblé les méthodes à tester, nous arrivons à un compromis satisfaisant. En effet, le temps de calcul est suffisamment faible pour être supporté par les programmes de la société Adelyce, et les prédictions ont un niveau de précision et de détails approprié pour être utilisées par les clients.

Malgré tout, la famille des modèles ARIMA repose sur l'hypothèse d'homoscédasticité, ce qui est une approximation de la réalité que nous préfèrerions éviter.
Un second outil est en préparation, plus complexe, traitant les séries temporelles comme des données fonctionnelles, et reposant sur une classification non-supervisée en amont de la prédiction.

References

[1] J Scott Armstrong and Fred Collopy. Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, 8(1):69–80, 1992.

[2] Rich Calaway, Steve Weston, and Dan Tenenbaum. *Foreach Parallel Adaptor*. Revolution Analytics, 2015.

[3] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7(3):1247–1250, 2014.

[4] Edward J Hannan and Jorma Rissanen. Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, pages 81–94, 1982.

[5] Rob J Hyndman and Yeasmin Khandakar. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22, 2008.

[6] Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

[7] Assemblée nationale. Décret n 2016-907. 1er juillet 2016.

[8] Assemblée nationale. Loi de finances 2016. 2015.

[9] James W Taylor. Exponential smoothing with a damped multiplicative trend. *International journal of Forecasting*, 19(4):715–725, 2003.