

# SEM pour l'estimation des coûts de fabrication : une étude de cas sur les moules à injection

Minh-Phuoc. DOAN<sup>1</sup>, Iragael. JOLY<sup>2</sup>

<sup>1</sup> Université de Bordeaux, CNRS, IMS, UMR 5218, 33405 Talence, France

<sup>2</sup> Université Grenoble-Alpes, CNRS, INRAE, Grenoble INP, GAEL, Grenoble, France

minh-phuoc.doan@u-bordeaux.fr ; iragael.joly@grenoble-inp.fr

## Résumé

*L'estimation du coût est un défi pour le chiffrage des produits. Le coût final d'un produit est constitué de différents éléments associés à la conception, à la fabrication, à la distribution et aux processus de gestion de la fin de vie. Les méthodes d'estimation des coûts basées sur des données, qui prédisent le coût d'un produit sur la base des coûts d'autres produits similaires, connaissent un potentiel croissant en raison de leur niveau de précision de prédiction. Toutefois, pour certaines de ces méthodes, telles que le réseau de neurone, la précision dépend largement de la quantité et de la qualité des données disponibles. De plus, dans l'application de ces méthodes, la relation entre le coût et la caractéristique reste inexplicée. Cet article propose un modèle explicatif d'estimation des coûts, basé sur l'approche des modèles à équations structurelles (SEM). Des applications numériques, réalisées sur une étude de cas d'une entreprise de moule à injection, montrent le potentiel de cette méthode en termes de précision de prédiction et d'explicabilité.*

## Mots-clés

*Modèles à équations structurelles ; Devis de produit ; Estimation des coûts ; Méthode fondée sur les données*

## 1 Introduction

L'estimation des coûts de production, en particulier à un stade précoce du processus de développement du produit, offre à l'entreprise les avantages suivants [1] : i) être compétitive en proposant un prix de produit attractif à ses clients, ii) identifier les principaux facteurs de coût pour améliorer les processus et réduire les coûts à long terme, et iii) optimiser la conception du produit en intégrant la méthode d'estimation des coûts dans le système de conception. Dans un second temps, lors de la phase de production, une meilleure connaissance des coûts de production est une donnée essentielle du suivi de l'activité, de la stratégie de tarification et plus largement de l'activité de commercialisation du produit.

Les méthodes d'estimation des coûts, qui sont basées sur l'intelligence artificielle (IA), connaissent un potentiel croissant [1].

La Figure 1 présente un cadre conceptuel d'intelligence

## Abstract

*Estimating the cost is a challenge for product quotation. The product final cost is made up of different cost components associated with the design, the fabrication, the distribution and end-of-life management processes. The data-driven cost estimation methods, which predicts a product cost based on the costs of other similar products see an increasing potential in recent time because of its prediction accuracy level. However, for some of these methods, such as neural network, the accuracy depends wisely on the quantity and quality of available data and the cost-feature relationship is inexplicable. This article proposes an explicable cost estimation framework, which is based on the structural equation modelling (SEM) approach. Numerical experiments, performed on a case study of an injection mould company, show the potential of this method in in terms of the prediction accuracy and the explainability.*

## Keywords

*Structural equation modelling; Product quotation; Cost estimation; Data-driven method*

artificielle pour l'estimation du coût des produits en fabrication [2]. Ce cadre se compose de deux modules : le développement de produits, qui se concentre sur les phases de conception du produit et des processus le soutenant ; et l'analyse des coûts et des caractéristiques du produit, qui cherche à quantifier par l'analyse de données historiques de l'entreprise, les relations entre le coût du produit et ses caractéristiques et celles des processus. Ce dernier module peut être divisé en trois blocs : la transformation et la sélection des caractéristiques, le regroupement ou la classification des familles de produits, et la cartographie des coûts et des caractéristiques.

Au tout début du développement d'un produit, seules les spécifications du produit sont généralement disponibles. Elles fournissent certaines caractéristiques géométriques et fonctionnelles pour l'estimation des coûts [3]. Dans un second temps, ces spécifications permettent au concepteur de construire les modèles CAO du produit. Ces modèles précisent les caractéristiques géométriques et améliorent l'estimation des coûts [4], [5].

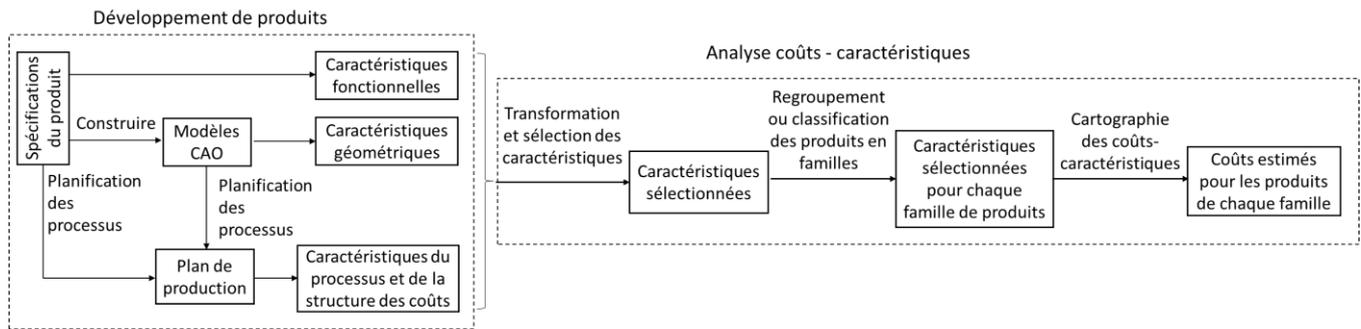


Figure 1. Représentation d'un système d'intelligence artificielle pour l'estimation du coût des produits dans l'industrie manufacturière

Les spécifications du produit et les modèles CAO sont ensuite utilisés pour créer le plan de production, en fournissant les caractéristiques du processus (par exemple, les technologies d'usinage) et les caractéristiques de la structure des coûts (par exemple, le coût des matériaux, le coût de la main-d'œuvre), ce qui est souvent appelé l'estimation des coûts à un stade ultérieur [5], [6].

Trois niveaux de décomposition du plan de production sont souvent considérés dans la littérature : 1) dans la décomposition simplifiée, le coût total est considéré comme la somme de groupes de coûts. Pour un produit assemblé, le plus souvent, le coût total du produit est considéré comme la somme des coûts de tous les composants et du coût d'assemblage (matériaux, main-d'œuvre, etc.) [7] ; 2) dans la décomposition traditionnelle, un assemblage est composé d'un ou de plusieurs éléments et chaque élément peut avoir des caractéristiques différentes, fabriquées par une ou plusieurs opérations d'usinage. Le coût estimé est donc la somme des coûts de toutes ces opérations de fabrication [7] ; 3) La décomposition détaillée (ou décomposition par activité) représente l'utilisation des ressources de l'organisation par les activités réalisées et associe ensuite le coût de ces activités aux résultats, tels que les produits et les services réalisés [8]. Les inducteurs de coûts sont identifiés pour chaque activité afin de calculer les coûts.

Dans la phase d'analyse coûts-caractéristiques, parmi les caractéristiques identifiées précédemment, certaines peuvent être transformées pour une meilleure comparaison des caractéristiques [9]. Ensuite, des techniques de sélection des caractéristiques, telles que la sélection par les experts [10] ou la sélection guidée par les données [11], sont appliquées pour identifier et sélectionner les caractéristiques les plus importantes ou ayant le plus d'impact pour l'étape suivante de l'estimation des coûts. Une méthode de regroupement ou de classification est parfois utilisée pour diviser les produits en familles présentant des caractéristiques similaires. Enfin, une méthode de mise en correspondance des coûts et des caractéristiques est ensuite utilisée pour estimer le coût du produit pour chaque famille. Le regroupement/la classification n'est pas toujours obligatoire pour l'estimation des coûts, mais il permet d'élaborer une méthode plus précise de mise en correspondance des coûts et des caractéristiques pour chaque famille de produits lorsqu'il existe différentes familles de

produits.

Cette analyse des relations coûts - caractéristiques peut être menée par l'appui des connaissances des experts du domaine ou par l'analyse des données historiques de produits similaires.

Les méthodes paramétriques et les méthodes analogiques sont deux familles courantes de méthodes basées sur les données. La méthode paramétrique est basée sur des formules mathématiques référencées, établies sur un grand nombre d'expériences statistiquement assez importantes pour pouvoir représenter la population de produits. Ces formules peuvent être construites à partir de l'exploration des données et/ou trouvées dans un manuel d'estimation des coûts, un manuel de référence ou une référence bibliographique [8], [10], [12]. La méthode analogique est basée sur l'analogie des données historiques de l'entreprise concernant des produits similaires. D'un point de vue statistique, il s'agit d'un échantillon d'une grande population de produits. Les méthodes courantes sont la régression linéaire [3], les forêts aléatoires [13], les k-plus proches voisins [12], [14] et les réseaux de neurones [9].

L'analyse des caractéristiques déterminantes du coût d'un produit et la prédiction de ce coût fondée sur les données repose sur une vision de la composition du coût de production. Plus particulièrement, les méthodes paramétriques formalisent les relations entre les caractéristiques des produits et des processus et les coûts de production. Ce formalisme est souvent hérité de la méthode d'estimation elle-même, comme par exemple la forme linéaire supposée entre caractéristiques et coût dans une régression linéaire, si ce n'est masqué ou trop complexe comme par exemple dans les réseaux de neurones multicouches.

Inspiré des trois niveaux de décomposition du plan de production présentés plus haut, cet article propose une discussion et une application d'une décomposition du coût d'un produit dans une approche fondée sur les données. Une attention particulière est donnée aux conséquences de la décomposition choisie et aux interrelations possibles entre les composants.

Le cas d'application est le chiffrage de moules pour injection plastique. Les données mises à disposition par une entreprise concernent les moules produits. Ils ont donc été chiffrés et les coûts ont ensuite été évalués lors de la production.

La méthode appliquée ici est la méthode SEM (ou modèle à équations structurelles), une méthode analogique axée sur les données, qui utilise un système de régressions linéaires multivariées pour modéliser les relations entre les composants du coût et les caractéristiques du produit [15]. Elle permet de tenir compte d'une décomposition a priori du coût du produit indiquée par les experts. De plus, la méthode SEM apporte une explication de la relation entre les composants du coût et les caractéristiques et peut être utilisée pour prédire simultanément les éléments de coût constituant le coût total, et pas seulement ce dernier.

La méthode SEM est couramment utilisée en biologie [16], en économétrie [17] et pour l'estimation des coûts des projets [18], mais à notre connaissance, cette méthode n'est pas encore utilisée comme méthode d'estimation des coûts des produits dans l'industrie.

**La principale contribution de cet article** est de proposer et d'évaluer la performance de plusieurs modèles d'estimation des coûts basés sur des données. Les modèles proposés correspondent à la régression linéaire et à plusieurs combinaisons de la méthode SEM avec plusieurs méthodes courantes de sélection des caractéristiques, y compris la sélection par un expert et par apprentissage automatique.

L'article est organisé comme suit. La section 2 présente la revue de la littérature sur les méthodes d'estimation des coûts basées sur les données. La section 3 présente la méthode SEM et quelques méthodes courantes de sélection des caractéristiques. La section 4 présente les données et l'application des méthodes proposées dans une étude de cas d'une entreprise de moules à injection. La section 5 présente quelques conclusions et perspectives pour la poursuite de la recherche.

## 2 Revue de littérature

La sélection des méthodes d'analyse des coûts et des caractéristiques dépend de différents facteurs, tels que le stade de développement du produit, la disponibilité des données et de la référence paramétrique, la précision de la prédiction et le besoin d'explicitation de la relation entre les coûts et les caractéristiques. Pour l'estimation des coûts à un stade précoce, il est rare qu'il existe une référence en matière d'estimation des coûts performante en termes de précision de prédiction. En l'absence de données historiques, la méthode intuitive est généralement le seul choix possible. Si des données historiques sont disponibles, la méthode analogique est la plus utilisée en raison de sa précision.

Dans son manuel, [19] propose une méthode paramétrique pour prédire le coût du moule d'injection plastique et de la pièce fabriquée par le moule. Le coût total du moule est la somme de trois éléments de coût : le coût de la base du moule, le coût de toutes les cavités et le coût de leur personnalisation. Chacun de ces trois éléments de coût est évalué à l'aide d'une formule mathématique fondée sur l'expertise. [8] utilise la décomposition des coûts par activité pour estimer le coût du produit. Le coût total est divisé en centres de coûts et chaque centre de coûts est représenté par des inducteurs de coûts. Une

formulation mathématique est établie entre chaque centre de coûts et les inducteurs de coûts associés.

Les méthodes analogiques ont aussi été utilisées dans la littérature, mais à notre connaissance rarement pour le chiffrage de moules à injection. [20] estime le coût des moules industriels en combinant la méthode de sélection de l'analyse factorielle avec trois méthodes d'analyse de la relation coût - caractéristiques : les forêts aléatoires, les k plus proches voisins, et les réseaux de neurones. Les modèles sont assez performants avec une incertitude satisfaisante dans la prédiction comprise entre 10% et 20% selon le type de moule. [21] estime les coûts des produits dans l'industrie automobile en combinant la méthode de sélection par étapes avec six méthodes analogiques différentes de mise en correspondance des coûts et des caractéristiques : arbre de décision, ensembliste, réseau de neurones artificiels, régression SVM, régression linéaire et k plus proches voisins. Ils ont constaté que tous les modèles donnent une précision élevée de l'estimation des coûts avec le meilleur  $R^2$  de 0,96.

Dans de nombreuses situations, l'explicabilité est, avec la précision de la prédiction, un autre facteur important pour le choix d'un modèle d'estimation des coûts. Pour améliorer l'explicabilité du modèle d'estimation des coûts, en particulier des modèles analogiques, les méthodes suivantes sont couramment utilisées dans la littérature : visualiser la relation entre le coût total et les caractéristiques associées [1], décomposer le coût total en groupes de coûts et établir la relation entre chacun de ces groupes et les caractéristiques du produit/processus [22], [23], ou décomposer le produit assemblé en composants, puis ces composants en caractéristiques/fonctions et estimer le coût de ces dernières [7], [24].

La sélection des caractéristiques relève de méthodes classées en deux catégories : la sélection par des experts et les méthodes d'apprentissage automatique (ou guidées par les données). Les méthodes de sélection par des experts utilisent les connaissances des experts pour sélectionner les caractéristiques [25]. [25] estiment le coût des pièces de moulage par injection de plastique. En interrogeant les experts du domaine, ils sont en mesure de redessiner le chemin que les experts ont suivi pour estimer le coût. En suivant ce chemin, ils peuvent identifier des facteurs de coût importants tels que le poids net du produit fini, la taille de la pièce moulée, etc.

La sélection par apprentissage automatique utilise l'analyse statistique des données pour sélectionner les variables. Une méthode courante est l'analyse de corrélation de Pearson, [3], l'analyse factorielle [26], l'analyse stepwise [11].

L'analyse de corrélation de Pearson est une méthode qui permet d'identifier les caractéristiques les plus étroitement corrélées avec la variable de sortie. [3] combine la sélection de la corrélation de Pearson et le réseau de neurones pour estimer le coût de fabrication des éléments de tuyauterie. Les corrélations de Pearson entre le coût et six caractéristiques, dont le poids, le type de soudage, le diamètre, la difficulté, les cavités et la classe, sont calculées. Seules quatre de ces caractéristiques dont le coefficient est supérieur à un seuil sont

sélectionnées.

L'analyse factorielle est une méthode qui permet de réduire un grand nombre de caractéristiques en un nombre plus restreint de facteurs importants [26]. [26] combine l'analyse factorielle avec l'optimisation par essais de particules (PSO) et le réseau de neurones artificiels pour donner une estimation du coût du moulage par injection plastique.

La méthode de régression stepwise est utilisée pour sélectionner et ajouter progressivement des caractéristiques au modèle. Chaque fois qu'une caractéristique est ajoutée au modèle, un test statistique (généralement le F-test associé à une mesure de qualité d'ajustement) est utilisé pour valider le modèle. Si une caractéristique initiale n'est plus significative après l'introduction de nouvelles caractéristiques, elle est supprimée du modèle. Ce processus est répété jusqu'à ce qu'aucune caractéristique significative ne puisse être introduite dans le modèle de régression [11]. Pour la méthode stepwise, différentes mesures de qualité d'ajustement peuvent être utilisées, telles que la p-valeur, l'AIC et le BIC [15].

Une des limitations de l'application des modèles de régression linéaire à l'estimation des coûts réside dans la prise en compte de la décomposition du coût total. Les méthodes SEM apparaissent avantageuses dans cette perspective. [27] utilisent les SEM pour estimer les coûts d'une mission d'exploitation minière d'un astéroïde. Les données ont été collectées par le biais d'entretiens avec des ingénieurs spatiaux et des scientifiques internationaux de différents niveaux et domaines professionnels. [28] utilise les SEM, la régression multiple et les réseaux de neurones pour estimer les coûts d'un projet de construction de tunnels routiers. Les résultats montrent que le modèle SEM donne une meilleure précision de prédiction que celle de la régression multiple et aussi précise que celle du réseau de neurones. En outre, ce modèle permet à l'utilisateur de représenter visuellement les voies par lesquelles plusieurs variables affectent le coût d'un projet.

**L'originalité de notre approche** est de proposer une estimation fondée sur des données du coût de moules à injection plastique. Pour cela plusieurs modèles d'estimation des coûts par apprentissage automatique sont évalués : la régression linéaire multiple et méthode SEM. Les caractéristiques introduites dans les modèles sont sélectionnées par application de plusieurs méthodes de sélection concurrentes : la sélection par un expert et deux modèles de sélection par apprentissage automatique (sélections stepwise sur l'AIC et le BIC).

### 3 Méthodes

L'estimation du coût pour le chiffrage des produits soulève plusieurs défis. L'estimation du coût final d'un produit se compose d'un certain nombre de sous-coûts qui sont interdépendants. Les coûts de conception et de R&D varient en fonction de la complexité du produit, ce qui a une incidence sur les techniques et les coûts de fabrication. Par conséquent,

la prédiction du coût total doit tenir compte des interdépendances entre ses composants.

La régression linéaire est un outil classique et simple d'apprentissage automatique utilisé dans les méthodes d'estimation des coûts basées sur les données. Cette régression suppose que les termes aléatoires du modèle sont aléatoires et indépendants. Néanmoins, cette hypothèse d'indépendance peut conduire à des estimations et des prédictions peu fiables [29] et ne tient pas compte des interdépendances entre composant du coût total. La spécification habituelle de la régression linéaire multiple est une simple équation,  $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \epsilon$ , où  $y$  est le coût du produit à prédire en fonction des caractéristiques  $x_1, x_2, \dots$  et les paramètres à estimer,  $\alpha, \beta_1, \beta_2, \dots$ . Les erreurs de prédiction sont modélisées par  $\epsilon \sim N(0, \sigma^2)$  qui sont supposées être indépendantes entre les produits.

Les techniques de régression multivariée, comme les modèles d'équations structurelles (SEM), permettent d'estimer le coût de fabrication et chaque composant ( $j=1\dots J$ ) à l'aide d'une équation spécifique incluant une éventuelle covariance entre les termes aléatoires entre les équations ( $cov(\epsilon_j, \epsilon_i) \neq 0$ ) (voir Figure 2).

$$\begin{aligned} y_1 &= \alpha_1 + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \epsilon_1 \\ y_2 &= \alpha_2 + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \epsilon_2 \\ &\dots \\ y_j &= \alpha_j + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \epsilon_j \end{aligned}$$

Un deuxième défi consiste à identifier et à sélectionner les caractéristiques pertinentes à inclure dans le modèle de régression. Les connaissances des experts donnent un aperçu de l'ensemble des variables à prendre en compte. Les méthodes itératives automatiques stepwise sont aussi appliquées ici pour sélectionner les caractéristiques en fonction de leur impact évalué sur un indicateur de performance (AIC ou BIC).

Dans cet article, l'analyse des données a été réalisée à l'aide du package Lavaan [30] de la version 4.3.1 de R (R Core Team, 2023). Nous comparons la régression linéaire multiple à la régression SEM avec différentes spécifications : en incluant ou non des interdépendances entre les équations de coût et avec différentes méthodes de sélection des caractéristiques. Cette comparaison est basée sur des mesures de performance classiques pour la prédiction quantitative. Les dépendances ou interdépendances des erreurs entre les équations de coût sont évaluées par un test de significativité de la matrice de variance et de covariance estimée des résidus.

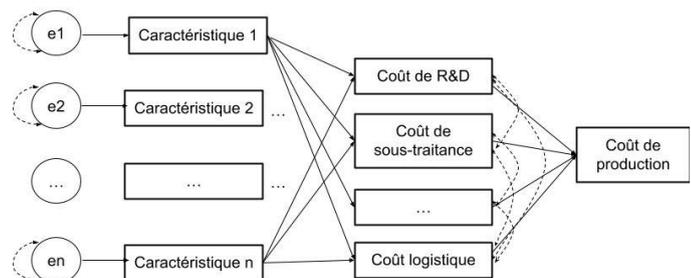


Figure 2. Structure générale du SEM du coût estimé

Dans le cadre des SEM, les indicateurs de performance habituels sont basés sur :

- Les  $R^2$  exprimés en termes de résidus quadratiques du modèle.
- L'indice comparatif d'adéquation (CFI) et l'indice d'adéquation de Tucker-Lewis (TLI), qui sont des indicateurs évaluant l'adéquation entre les matrices de covariance attendues et observées des données ajustées en fonction de la taille de l'échantillon. Le CFI est compris entre 0 et 1. Une valeur de CFI plus élevée indique une meilleure adéquation du modèle. Dans la pratique, un CFI proche de 0,95 ou supérieur indique une adéquation acceptable. Le TLI peut être supérieur à 1. Un TLI supérieur à 0,90 est acceptable [15].
- L'AIC et le BIC sont basés sur le logarithme de la vraisemblance du modèle et ne sont ni bornés, ni normés. Entre plusieurs modèles estimés sur un même jeu de données, les valeurs les plus faibles caractérisent les meilleurs modèles.

## 4 Expérimentation et résultats

### 4.1 Données

Nous présentons une étude de cas d'une entreprise de moules à injection, qui conçoit et fabrique des moules à injection pour des clients du monde entier. Actuellement, le devis des moules est réalisé manuellement par des experts en moules qui suivent une procédure intégrée dans un fichier Excel. L'ensemble des données de l'entreprise utilisées pour les expériences consiste en 131 devis de moules acceptés par les clients entre 2016 et 2022. Un traitement préliminaire des données avec l'aide des experts en devis, comprenant le nettoyage, la transformation et la sélection des données, permet d'obtenir un ensemble de données approprié de 97 devis de moules pour les expériences. Dans ce nouvel ensemble de données, pour chaque moule, 32 caractéristiques d'entrée (variables explicatives) et 24 caractéristiques de sortie (variables expliquées) sont disponibles.

Les caractéristiques d'entrée peuvent être regroupées en plusieurs catégories : caractéristiques de la catégorie de moule (par exemple, type, catégorie et date de sortie), caractéristiques du moule (types d'injection, d'éjection et de régulation, nombre de cavités), caractéristiques du mécanisme du moule (nombre et types de cylindres mécaniques, nombre et types de cylindres hydrauliques), caractéristiques du produit (par exemple, type de produit, qualité de la surface, matériau) et autres caractéristiques (par exemple, méthode d'essai, garantie et livraison).

Les caractéristiques de sortie comprennent les coûts estimés par l'expert et les coûts réels mis en œuvre pour chacune des onze activités induites par les coûts (par exemple, l'achat, la conception, l'usinage brut et la finition). Dans cette étude, nous nous concentrons uniquement sur les coûts réels des moules et, avec l'aide des experts de l'entreprise, nous combinons les onze activités de coûts en six groupes (voir Tableau 1).

Tableau 1. Explication des activités induites par les coûts

Activité induite par les coûts	Explication
Achats	Coûts liés à l'achat de matériel et d'équipements
Etude	Coûts liés au processus de conception du produit
Sous-Traitance	Coûts liés aux opérations externalisées
Méthode	Coûts liés à l'industrialisation du produit (processus de fabrication et outils nécessaires)
Ebauche	Coûts associés aux opérations d'usinage d'ébauche et de finition
Ajustage	Coûts liés aux tests et ajustements

### Définitions des variables explicatives (caractéristiques) :

*Nbre\_mvt\_meca\_simple* (Variable numérique) : nombre de mouvements "simples" effectués par le moule.

*Nbre\_mvt\_meca\_double* (Variable numérique) : nombre de mouvements "doubles" (plus complexes).

*Nbre\_mvt\_rotatif* (Variable numérique) : nombre de mouvements "rotatifs".

*Nbre\_total\_point* (Variable numérique) : nombre total de points d'injection pour le moule.

*Type.de.produit\_HT* (valeur de référence) : l'entreprise réalise le produit pour la mise en production.

*Type.de.produit\_MIX* : L'entreprise réalise en partie la mise en place de la production.

*Type.de.produit\_PROTO* : L'entreprise réalise le prototype uniquement.

*Type.injection\_CAROTTE* (Variable binaire) : le système utilise une carotte qui permet de traverser le bloc empreinte fixe jusqu'au plan de joint. Sa conicité favorise le démoulage. Sa valeur de référence est *Type.injection\_Bloc\_chaud* qui donne l'information que le système utilise un bloc chaud qui permet de maintenir la matière à température.

*Ejection\_Autre* : la forme du vérin d'éjection est tubulaire ou en lame. Sa valeur de référence est *Ejection\_Cylindrique* décrivant que la forme du mécanisme d'éjection est cylindrique.

*Commande.ejection\_Verin* : Le mécanisme d'éjection se fait par vérin. Sa valeur de référence est *Commande.ejection\_Asservissement\_presse* qui décrit que le mécanisme est déclenché par Asservissement presse.

*Regulation\_Normal* : La régulation de la température pour refroidir le système se fait de manière "normale" donc à plus basse température que la valeur de référence *Regulation\_Chaud*.

*Acier.moulant\_H11* (Valeur de référence) : acier avec une ténacité légèrement supérieure au aciers classiques.

*Acier.moulant\_110KG* : acier plus souple.

*Acier.moulant\_Autre* : acier non-conventionnel (pas 100KG, H11 ou H13).

*Acier.moulant\_H13* : l'acier le plus commun pour les outillages.

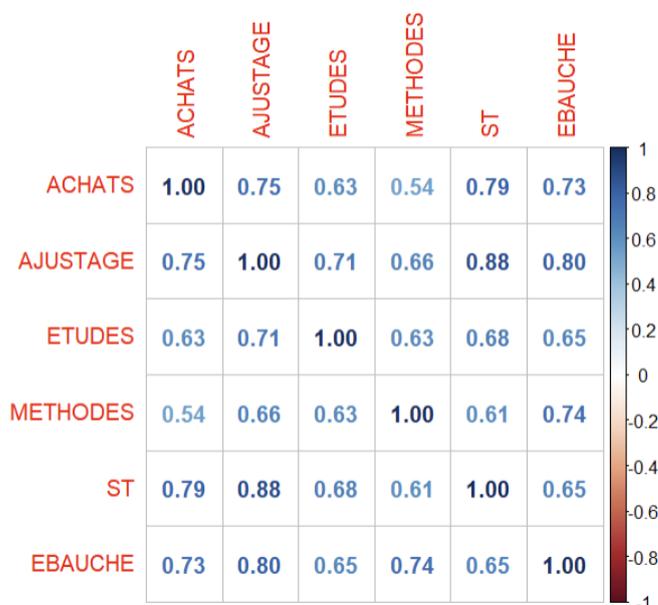


Figure 3. Matrice de corrélation entre les groupes de coûts

La Figure 3 montre que les 6 éléments de coût covarient principalement dans un schéma de corrélation positive, ce qui suggère que leurs prédictions peuvent ne pas être totalement indépendantes les unes des autres.

## 4.2 Résultats

La première étape consiste à sélectionner les caractéristiques pertinentes pour l'estimation et la prédiction des coûts. La sélection est basée sur l'expertise des ingénieurs de l'entreprise (modèle "M.Expert") et sur deux méthodes guidées par les données, une sélection stepwise basée sur les critères AIC et BIC (modèles "M.AIC" et "M.BIC").

Le modèle "Expert" comprend trois à cinq caractéristiques en fonction de l'élément de coût. Le nombre de mouvements mécaniques simples, de mouvements rotatifs et la taille sont des caractéristiques communes pour les 6 composantes de coûts.

Les modèles M.AIC et M.BIC comprennent jusqu'à sept caractéristiques, parmi lesquelles les caractéristiques sélectionnées par l'expert sont communes aux 6 composantes de coût.

La deuxième étape consiste à évaluer les deux spécifications avec et sans résidus interdépendants. Chaque modèle est estimé comme si les équations du système étaient indépendantes. Cela conduit à des modèles M.Expert.0, M.AIC.0, M.BIC.0 qui sont estimés avec la contrainte de résidus indépendants entre les équations.

Le Tableau 2 présente les indices d'adéquation globale des 6 modèles. En ce qui concerne la sélection des caractéristiques par l'expert ou par les données, nous observons que la méthode basée sur les données semble plus précise, avec un CFI et un TLI plus élevés, et un SRMS, un AIC et un BIC plus faibles. Les méthodes basées sur les données complètent les connaissances de l'expert en détectant des informations supplémentaires à partir des données.

L'indépendance des termes d'erreur entre les équations n'est pas confirmée par les résultats. Chaque modèle contraint à des covariances résiduelles nulles est associé à un ajustement plus faible par rapport à sa version concurrente comprenant une matrice de covariance non restreinte. Les tableaux 4, 5 et 6 (en annexe) présentent les matrices de covariances estimées dans les 3 modèles. Les covariances sont significativement différentes de zéro entre de nombreuses équations de coûts, ce qui confirme les dépendances entre ces estimations. Par exemple, les erreurs de prédictions des coûts d'achat, d'ajustage et d'études sont positivement reliées.

Tableau 2. Indicateurs de performance

	cfi	tli	srmr	aic	bic
M.Expert	0,93	0,68	0,09	1027	1158
M.Expert0	0,66	0,35	0,10	1208	1301
M.AIC	1,00	1,02	0,02	933	1103
M.AIC0	0,78	0,60	0,04	1097	1228
M.BIC	0,99	0,97	0,06	951	1072
M.BIC0	0,75	0,59	0,08	1120	1202

Le Tableau 3 compare les  $R^2$  de chaque équation de coût et suggère que les méthodes basées sur les données améliorent de manière significative l'ajustement et la capacité de prédiction par rapport à la spécification du modèle expert.

La comparaison du  $R^2$  entre les spécifications de la matrice de covariance avec et sans contrainte montre une légère diminution des indices d'ajustement lorsque les covariances estimées sont non nulles. Cependant, la comparaison des  $R^2$  entre régression univariée et régression multivariée reste difficile. En effet, la tâche de prédiction d'un seul composant indépendant des autres est plus facile que la prédiction simultanée du système de composants. La performance devrait être évaluée sur l'ensemble du système.

Tableau 3. Comparaison des  $R^2$

	Achat	Etude	Sous-Traitance	Méthode	Ebauche	Ajustage
M.Expert	0,61	0,37	0,61	0,47	0,47	0,47
M.Expert.0	0,69	0,43	0,72	0,49	0,58	0,50
M.AIC	<b>0,78</b>	<b>0,69</b>	<b>0,44</b>	<b>0,59</b>	<b>0,74</b>	<b>0,66</b>
M.AIC.0	0,79	0,70	0,45	0,59	0,75	0,67
M.BIC	0,74	0,63	0,36	0,51	0,68	0,63
M.BIC.0	0,77	0,67	0,38	0,53	0,71	0,65

La précision générale de la prédiction peut être évaluée avec les autres indicateurs présentés dans le tableau 2. Les modèles où les résidus interdépendants sont plus performants.

La comparaison des  $R^2$  entre équations dans un système indique les qualités relatives de prédiction entre les composantes du coût. Pour le modèle final, M.AIC, la prédiction du coût d'achats est la plus précise ( $R^2=0,78$ ), avec le coût d'ébauche ( $R^2=0,74$ ), suivi du coût d'étude ( $R^2=0,69$ ), d'ajustage ( $R^2=0,66$ ), de sous-traitance ( $R^2=0,44$ ).

Nous pouvons observer que l'estimation du coût d'étude est améliorée par les méthodes de sélection des caractéristiques basées sur les données, alors que les connaissances des experts semblent être plus précises pour le coût de sous-traitance.

Dans le meilleur modèle : M.AIC, les caractéristiques ont un impact significatif sur les différents composants du coût (Tableau 7 - Annexe).

La longueur du moule, le nombre de mouvements mécaniques "simples" sont des caractéristiques significatives dans tous les composants, avec des impacts positifs. Les variables continues sont standardisées, ce qui permet de comparer la taille des impacts. La longueur du moule a un impact plus important sur le coût de sous-traitance (ST) et des achats et plus faible sur le coût d'étude. Le nombre de mouvements mécaniques "simples" a un impact plus important sur le coût d'ajustage et d'étude et plus faible sur celui de la sous-traitance.

Le fait que le mécanisme d'éjection se fait par vérin ou non (*commande.ejection.verin*) a un impact significatif sur le coût d'ébauche, ce qui est cohérent pour l'expert, car il s'agit d'une option plus coûteuse pour l'usinage, qui est réalisée, soit pour des contraintes techniques, soit à la demande du client.

Pour le coût de la méthode, l'acier de moulage est significatif. On peut supposer que le H13 est un matériau aux propriétés similaires à celles du matériau le plus couramment utilisé, ce qui peut entraîner moins de difficultés pour la phase FAO et donc une réduction du coût de cette activité.

Il est surprenant de constater que les variables du type de régulation du système de refroidissement n'ont pas d'impact significatif sur les coûts. On s'attendait à ce que les coûts de cette activité soient plus élevés pour le type normal que pour le type chaud.

Le modèle estime que le type d'acier de moulage classique (110 KG et H13) a également un impact négatif significatif sur le coût d'étude par rapport au type de référence H11. Cependant, il est difficile d'établir un lien réel entre cette information et l'estimation selon la vision de l'expert.

La majorité des autres relations significatives sont cohérentes avec la vision d'expert.

## 5 Conclusion et perspectives

Dans cet article, nous proposons et évaluons plusieurs modèles d'estimation des coûts par apprentissage automatique correspondant aux combinaisons de la méthode SEM et de plusieurs méthodes courantes de sélection des caractéristiques, y compris la sélection par l'expertise et deux modèles de sélection par apprentissage automatique (sélection stepwise AIC et stepwise BIC). Des expériences numériques, réalisées sur une étude de cas d'une entreprise de moules à injection, ont montré tout d'abord que les modèles basés sur la combinaison des méthodes SEM avec la sélection des caractéristiques stepwise AIC donnent la meilleure estimation selon les mesures de performance classiques, telles que CFI, TLI, SRMR, AIC et BIC. Deuxièmement, la précision de la prédiction est différente pour chacun des six composants considérés du coût total.

Ensuite, certaines caractéristiques sont statistiquement importantes pour tous les éléments de coût (longueur du moule, le nombre de mouvements mécaniques "simples"), tandis que d'autres ne le sont que pour un ou plusieurs éléments. Ensuite, les inducteurs de coûts pour chaque élément de coût sont souvent cohérents avec la vision des experts en devis. D'autres recherches pourraient être menées sur les aspects suivants : 1) tester le modèle dans le contexte des big data en tenant compte de l'ajustement des hyperparamètres, 2) intégrer une méthode de regroupement comme K-mean, LCM ou Kamila dans le cadre de l'apprentissage automatique, 3) comparer la méthode SEM avec d'autres méthodes classiques d'apprentissage automatique et d'apprentissage profond comme les forêts aléatoires et les réseaux de neurones, et 4) prendre en compte d'autres facteurs de coûts liés à des phénomènes économiques comme l'inflation, la rareté des matériaux et la fluctuation des prix sur le marché.

L'accompagnement des experts de l'entreprise dès le début du développement de ces méthodes d'estimation des coûts nous permet de bien sélectionner les modèles pertinents, de bien préparer les données pour entraîner, tester, et valider les modèles, ainsi que d'interpréter les résultats obtenus. De plus, cet accompagnement permet à l'entreprise une compréhension complète de la méthode proposée. Les travaux futurs envisagés pour améliorer la performance de prédiction des modèles actuels sont de développer un outil plus convivial avec une interface visuelle facile à utiliser pour les opérateurs/techniciens. Enfin, la documentation de la méthode proposée devra être faite pour la rendre accessible aux personnels qui n'ont pas de compétence sur l'IA.

## 6 Remerciements

Cette recherche est financée par le projet Accel 4.0 Grenoble INP - Génie Industriel, le Programme Investissements Avenir, et le Plan de Relance, Nous tenons également à remercier Romain Hoarau, Maxime Cacace-Soret et Eloi Peretti pour leur aide dans la transformation et le nettoyage des données.

## 7 Références

- [1] S. Yoo and N. Kang, "Explainable artificial intelligence for manufacturing cost estimation and machining feature visualization," *Expert Syst. Appl.*, vol. 183, no. April, p. 115430, 2021, doi: 10.1016/j.eswa.2021.115430.
- [2] M. Doan, N. Pelouzet, and I. Joly, "Artificial intelligence for product cost estimation in manufacturing: review and perspectives," in *1er congrès SAGIP*, 2023, pp. 3–7.
- [3] O. Duran, J. MacIel, and N. Rodriguez, "Comparisons between two types of neural networks for manufacturing cost estimation of piping elements," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 7788–7795, 2012, doi: 10.1016/j.eswa.2012.01.095.
- [4] D. T. Masel, W. A. Young, and R. P. Judd, "A rule-based approach to predict forging volume for cost estimation during product design," *Int. J. Adv. Manuf.*

- Technol.*, vol. 46, no. 1–4, pp. 31–41, 2010, doi: 10.1007/s00170-009-2108-6.
- [5] H. Zhang *et al.*, “A novel method based on a convolutional graph neural network for manufacturing cost estimation,” *J. Manuf. Syst.*, vol. 65, no. June, pp. 837–852, 2022, doi: 10.1016/j.jmsy.2022.10.007.
- [6] Z. Ding, Z. Jiang, Y. Liu, Y. Wang, and C. Li, “A Big Data based Cost Prediction Method for Remanufacturing End-of-Life Products,” *Procedia CIRP*, vol. 72, pp. 1362–1367, 2018, doi: 10.1016/j.procir.2018.03.129.
- [7] F. H’mida, P. Martin, and F. Vernadat, “Cost estimation in mechanical production: The Cost Entity approach applied to integrated product engineering,” *Int. J. Prod. Econ.*, vol. 103, no. 1, pp. 17–35, 2006, doi: 10.1016/j.ijpe.2005.02.016.
- [8] D. Ben-Arieh and L. Qian, “Activity-based cost management for design and development stage,” *Int. J. Prod. Econ.*, vol. 83, no. 2, pp. 169–183, 2003, doi: 10.1016/S0925-5273(02)00323-7.
- [9] F. Bodendorf, Q. Xie, P. Merkl, and J. Franke, “A multi-perspective approach to support collaborative cost management in supplier-buyer dyads,” *Int. J. Prod. Econ.*, vol. 245, no. November 2021, 2022, doi: 10.1016/j.ijpe.2021.108380.
- [10] S. Cavalieri, P. Maccarrone, and R. Pinto, “Parametric vs. neural network models for the estimation of production costs: A case study in the automotive industry,” *Int. J. Prod. Econ.*, vol. 91, no. 2, pp. 165–177, 2004, doi: 10.1016/j.ijpe.2003.08.005.
- [11] X. CHEN, J. HUANG, and M. YI, “Development cost prediction of general aviation aircraft using combined estimation technique,” *Chinese J. Aeronaut.*, vol. 34, no. 4, pp. 32–41, 2021, doi: 10.1016/j.cja.2020.07.024.
- [12] N. Sajadfar and Y. Ma, “A hybrid cost estimation framework based on feature-oriented data mining approach,” *Adv. Eng. Informatics*, vol. 29, no. 3, pp. 633–647, 2015, doi: 10.1016/j.aei.2015.06.001.
- [13] T. Buechler, M. Kolter, L. Hallweger, and M. F. Zaeh, “Predictive cost comparison of manufacturing technologies through analyzing generic features in part screening,” *CIRP J. Manuf. Sci. Technol.*, vol. 38, pp. 299–319, 2022, doi: 10.1016/j.cirpj.2022.04.012.
- [14] P. C. Chang, J. J. Lin, and W. Y. Dzan, “Forecasting of manufacturing cost in mobile phone products by case-based reasoning and artificial neural network models,” *J. Intell. Manuf.*, vol. 23, no. 3, pp. 517–531, 2012, doi: 10.1007/s10845-010-0390-7.
- [15] R. H. Hoyle, *Handbook fo Structural Equation Modelling*. The Guilford Press, 2023.
- [16] J. B. GRACE *et al.*, “Guidelines for a graph-theoretic implementation of structural equation modeling,” vol. 3, no. August 2012, pp. 1–44.
- [17] H. Li, Y. Liu, and K. Peng, “Characterizing the relationship between road infrastructure and local economy using structural equation modeling,” *Transp. Policy*, vol. 61, no. 1, pp. 17–25, 2018, doi: 10.1016/j.tranpol.2017.10.002.
- [18] F. M. Ottaviani and A. De Marco, “Multiple Linear Regression Model for Improved Project Cost Forecasting,” *Procedia Comput. Sci.*, vol. 196, no. 2021, pp. 808–815, 2021, doi: 10.1016/j.procs.2021.12.079.
- [19] D. O. Kazmer, *Injection Mold Design Engineering*. 2007. doi: 10.3139/9783446434196.fm.
- [20] S. Borzel and J. Frochte, “Case study on model-based application of machine learning using small CAD databases for cost estimation,” *IC3K 2019 - Proc. 11th Int. Jt. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag.*, vol. 1, no. Ic3k, pp. 258–265, 2019, doi: 10.5220/0007979802580265.
- [21] F. Bodendorf and J. Franke, “A machine learning approach to estimate product costs in the early product design phase: A use case from the automotive industry,” *Procedia CIRP*, vol. 100, pp. 643–648, 2021, doi: 10.1016/j.procir.2021.05.137.
- [22] M. Relich and R. Wójcik, “Decision Support Within Customized Products Using Computational Intelligence,” *IFAC-PapersOnLine*, vol. 55, no. 2, pp. 211–216, 2022, doi: 10.1016/j.ifacol.2022.04.195.
- [23] J. Y. Jung, “Manufacturing cost estimation for machined parts based on manufacturing features,” *J. Intell. Manuf.*, vol. 13, no. 4, pp. 227–238, 2002, doi: 10.1023/A:1016092808320.
- [24] E. M. Shehab and H. S. Abdalla, “Manufacturing cost modelling for concurrent product development,” *Robot. Comput. Integr. Manuf.*, vol. 17, no. 4, pp. 341–353, 2001, doi: 10.1016/S0736-5845(01)00009-6.
- [25] H. S. Wang, Y. N. Wang, and Y. C. Wang, “Cost estimation of plastic injection molding parts through integration of PSO and BP neural network,” *Expert Syst. Appl.*, vol. 40, no. 2, pp. 418–428, 2013, doi: 10.1016/j.eswa.2012.01.166.
- [26] Z. H. Che, “PSO-based back-propagation artificial neural network for product and mold cost estimation of plastic injection molding,” *Comput. Ind. Eng.*, vol. 58, no. 4, pp. 625–637, 2010, doi: 10.1016/j.cie.2010.01.004.
- [27] A. Probst, C. Nitzl, F. Kraus, and R. Förstner, “Cost estimation of an asteroid mining mission using partial least squares structural equation modelling (PLS-SEM),” *Acta Astronaut.*, vol. 167, no. November 2019, pp. 440–454, 2020, doi: 10.1016/j.actaastro.2019.07.032.
- [28] K. Petrousatou and S. Lambropoulos, “Road tunnels construction cost estimation: A structural equation model development and comparison,” *Oper. Res.*, vol. 10, no. 2, pp. 163–173, 2010, doi: 10.1007/s12351-009-0061-7.
- [29] W. H. Greene, *Econometric Analysis*, vol. 8th editio. Pearson Education, 2018. doi: 10.1007/978-3-030-56239-7\_5.
- [30] Y. Rosseel, “lavaan: An R Package for Structural Equation Modeling,” *J. Stat. Softw.*, vol. 48, no. 2, pp. 1–93, 2012.

## 8 Annexe

Tableau 4. Matrice de Variance covariance estimées du modèle M.Expert

	<b>ACHATS</b>	<b>AJUSTAGE</b>	<b>ETUDE</b>	<b>METHODE</b>	<b>ST</b>	<b>EBAUCHE</b>
<b>ACHATS</b>	0,33***	0,21***	0,16***	0,000	0,16***	0,18***
<b>AJUSTAGE</b>		0,61***	0,23***	0,19***	0,23***	0,29***
<b>ETUDE</b>			0,33***	-0,006	0,07	0,30***
<b>METHODE</b>				0,51***	0,20***	0,07
<b>ST</b>					0,46***	0,28***
<b>EBAUCHE</b>						0,52***

\*\*\* : pvalue <0.001 \*\* pvalue <0.01 \* pvalue < 0.10

Tableau 5. Matrice de Variance covariance estimées du modèle M.AIC

	<b>ACHATS</b>	<b>AJUSTAGE</b>	<b>ETUDE</b>	<b>METHODE</b>	<b>ST</b>	<b>EBAUCHE</b>
<b>ACHATS</b>	0,22***	0,08***	0,13***	-0,03	0,12***	0,06***
<b>AJUSTAGE</b>		0,31***	0,19***	0,06	0,18***	0,16***
<b>ETUDE</b>			0,55***	0,15***	0,18***	0,14***
<b>METHODE</b>				0,41***	0,006	0,14***
<b>ST</b>					0,26***	0,02
<b>EBAUCHE</b>						0,33***

\*\*\* : pvalue <0.001 \*\* pvalue <0.01 \* pvalue < 0.10

Tableau 6. Matrice de Variance covariance estimées du modèle M.BIC

	<b>ACHATS</b>	<b>AJUSTAGE</b>	<b>ETUDE</b>	<b>METHODE</b>	<b>ST</b>	<b>EBAUCHE</b>
<b>ACHATS</b>	0,24***	0,08***	0,16***	-0,004	0,11***	0,07***
<b>AJUSTAGE</b>		0,34***	0,21***	0,08	0,21***	0,16***
<b>ETUDE</b>			0,62***	0,20***	0,20***	0,16***
<b>METHODE</b>				0,47***	0,01	0,18***
<b>ST</b>					0,29***	0,01***
<b>EBAUCHE</b>						0,35***

\*\*\* : pvalue <0.001 \*\* pvalue <0.01 \* pvalue < 0.10

Tableau 7. Résultat d'estimation des 6 équations de coûts du modèle M.AIC

Variable /Composant de coût	ACHATS	AJUSTAGE	ETUDE	METHODE	ST	EBAUCHE
Dim.outi.long	0.51***	0.46***	0.23*	0.46***	0.66***	0.37***
Nbre_mvt_meca_simple	0.21***	0.32***	0.32***	0.23**	0.15*	0.30***
Nbre_mvt_meca_double	-	0.16*	0.24**	0.23**	0.13*	0.27***
Nbre_mvt_rotatif	-	-	0.11	-	0.06	-
Nbre_total_point	0.13**			0.09	-	0.17***
Type.de.produit_HT	ref	ref	ref	ref	ref	ref
Type.de.produit_MIX	-	0.72**	-	-	0.76***	-
Type.de.produit_PROTO	-	-	-	-	-	-
Type.injection_BLOC_CHAUD	ref	ref	ref	ref	ref	ref
Type.injection_CAROTTE	-0.47***	-	-	-	-	-
Ejection_Cylindrique	ref	ref	ref	ref	ref	ref
Ejection_Autre	-	-0.07	-	-0.37**	-	-
Commande.ejection_Presse	ref	ref	ref	ref	ref	ref
Commande.ejection_Verin	0.66***	0.50***	-	-	0.17	0.55***
Regulation_Chaud	ref	ref	ref	ref	ref	ref
Regulation_Normal	-	-	-	-	-	-
Acier.moulant_H11	ref	ref	ref	ref	ref	ref
Acier.moulant_110KG	-	-	-0.28	-	-	-
Acier.moulant_Autre	-	-0.40*	-	-0.51*	-	-0.48*
Acier.moulant_H13	-0.33**	-	-0.33*	-0.41*	-	-0.26*
R2	0.78	0.69	0.44	0.59	0.74	0.67

\*\*\* : pvalue <0.001 \*\* pvalue <0.01 \* pvalue < 0.10