



IP PARIS

Transfer Learning en Finance

Modélisations Statistiques et Applications (MSA)

Télécom SudParis

Écrit par : Siran Demir et Rebecca Laïk

Supervisé par : Prof. Hélène Halconruy

Date : 6 mars 2025

Table des matières

1	Introduction	2
2	Introduction Mathématiques au Transfer Learning :	5
2.1	Les Transformations dans le TL	5
2.2	Risque de Transfert et Mesures d'Efficacité	8
2.2.1	Risque de transfert dans des modèles à distribution gaussienne	9
3	Problème d'optimisation de portefeuille	17
3.1	En Pratique	18
3.2	Étape 1 : Optimisation directe sur les données cibles	18
3.3	Étape 2 : Optimisation sur les données sources	18
3.4	Étape 3 : Optimisation régularisée avec transfer learning	18
3.5	Le Risque de Transfert en pratique	19
3.5.1	Première composante : Qualité du Portefeuille Source (R_1)	19
3.5.2	Deuxième composante : Similarité des Portefeuilles (R_2)	19
3.5.3	Corrélation entre Risque et Sharpe Ratio	19
4	Méthodologie et Expériences	20
4.1	Transfert Intercontinental	20
4.1.1	Définition et Objectif	20
4.1.2	Méthodologie	21
4.1.3	Résultats et Analyse	23
4.1.4	Difficultés Rencontrées	24
4.2	Transfert Intersectoriel	24
4.2.1	Définition et Objectif	24
4.2.2	Méthodologie	25
4.2.3	Résultats et Analyse	25
4.3	Transfert Inter-Fréquences	25
4.3.1	Définition et Objectif	25
4.3.2	Méthodologie	26
4.3.3	Résultats et Analyse	26
5	Conclusion	27

Résumé

Dans cette étude, nous explorons le concept de Transfer Learning (TL) appliqué à l'optimisation de portefeuilles financiers, en nous appuyant sur les travaux de Cao et al. (2023), issus de deux études : "Risk of Transfer Learning and its Applications in Finance"[1] et "Transfer Learning for Portfolio Optimization"[2]. Ces articles introduisent le concept de risque de transfert, une métrique combinant la qualité du portefeuille source et la similarité des distributions de rendements entre les domaines source et cible. Notre objectif est de suivre leur démarche mathématique et de mettre en pratique leurs expérimentations.

Nous commençons par une analyse mathématique rigoureuse du cadre théorique du Transfer Learning (TL), en étudiant d'abord la transition d'un problème d'apprentissage automatique classique à un problème de TL. Nous développons ensuite la notion de risque de transfert de manière générale, avant de l'approfondir spécifiquement dans le cadre gaussien. Ensuite, nous appliquons ce cadre à des expériences pratiques, en étudiant trois types de transferts : intercontinental, intersectoriel et inter-fréquences.

Notre approche s'appuie sur le cadre théorique proposé par Cao et al. (2023) [1], qui formalise le risque de transfert comme une fonction de (i) la qualité du portefeuille source, mesurée par son ratio de Sharpe, et (ii) la similarité statistique entre les distributions de rendements des marchés source et cible, évaluée à l'aide de la distance de Wasserstein. Ce risque est supposé être un prédicteur de la performance des portefeuilles transférés, les marchés plus dissemblables induisant un transfert moins efficace.

Afin de tester cette hypothèse, nous avons mis en place une approche empirique où nous optimisons les portefeuilles sur les marchés cibles sous trois configurations : (1) sans transfert (optimisation directe sur les données cibles), (2) optimisation sur les données sources du marché américain, et (3) optimisation régularisée intégrant une contrainte de proximité avec le portefeuille source. Nous analysons ensuite la corrélation entre le risque de transfert et la performance des portefeuilles transposés.

Pour le transfert intercontinental, nous avons ainsi implémenté la transférabilité des portefeuilles optimisés sur le marché américain (S&P 500) vers des marchés cibles tels que le Royaume-Uni, le Brésil, l'Allemagne et Singapour. Les résultats montrent une corrélation négative entre le risque de transfert et la performance des portefeuilles transférés, ce qui entre en contradiction avec l'étude pratique menée dans l'article [2], suggérant ainsi que d'autres facteurs, tels que la sélection des actions ou la régularisation, pourraient influencer les résultats.

Cette étude souligne l'importance de la sélection des sources de données, de la similarité entre les domaines, et de la gestion du risque de transfert pour maximiser l'efficacité du TL en finance. Bien que les résultats actuels montrent des limites, ils ouvrent la voie à des améliorations futures, notamment par l'ajustement des paramètres de régularisation et l'exploration de méthodes d'adaptation dynamique. Le TL, bien que prometteur, nécessite une approche prudente et réfléchie pour être pleinement exploité dans l'optimisation de portefeuilles.

1 Introduction

Le Transfer Learning (TL) ou apprentissage par transfert consiste à exploiter des connaissances acquises dans un domaine "source" pour améliorer les performances d'un modèle dans un domaine "cible" qui lui est similaire, mais pas identique. Cette technique permet de réduire le besoin de grandes quantités de données spécifiques au domaine "cible", tout en construisant des modèles performants. Cela constitue un avantage significatif dans le domaine de la finance, où l'accès à des données fiables est souvent limité. Cette difficulté d'accès à la donnée peut freiner la réalisation de certaines tâches, comme la prédiction des prix d'actions ou

l'analyse de séries chronologiques financières. En réutilisant les schémas appris dans des contextes analogues, le TL permet d'éviter de recréer un modèle entièrement pour chaque nouvelle tâche et favorise une meilleure généralisation du modèle à de nouvelles situations.

Cependant, pour que le TL ait un impact positif, les domaines "source" et "cible" doivent être suffisamment similaires. Même dans ce cas, des similitudes peuvent parfois introduire des erreurs dans l'apprentissage, ce qui souligne l'importance de garantir une correspondance adéquate entre ces deux modèles avant d'engager le transfert.

Dans notre étude, nous approfondirons ces aspects en nous basant sur deux articles clés. Le premier, *Risk of Transfer Learning and its Applications in Finance* de Haoyang Cao, Haotian Gu, Xin Guo et Mathieu Rosenbaum [1] explore des applications concrètes du TL en finance, comme la prédiction des retours boursiers et l'optimisation de portefeuilles, en analysant les performances du transfert entre différents marchés, secteurs ou fréquences de trading. Il met en lumière la corrélation entre un risque de transfert faible et une meilleure performance du TL. L'approche fournit également des outils pour identifier les sources de données inadaptées avant d'engager le transfert, évitant ainsi le transfert négatif.

Le second, *Transfer learning for portfolio optimization* [2], rédigé par les mêmes auteurs, applique le TL au problème d'optimisation de portefeuille dans des environnements où les données sont limitées, comme les marchés émergents. Il propose une méthode où les portefeuilles optimaux sont construits sur un marché source (avec des données abondantes) et ajustés pour un marché cible via un transfert. L'étude explore différents types de transferts, comme les transferts intercontinentaux (transfert du marché européen au marché américain), intersectoriels (d'un secteur industriel à un autre) ou entre fréquences de trading (passage de stratégies à basse fréquence vers des stratégies intrajournalières).

Notre objectif est double : d'une part, nous analyserons la notion de risque de transfert décrite dans le premier article et son impact sur les performances du TL. D'autre part, nous étudierons l'algorithme d'optimisation de portefeuille du second article pour l'adapter et le tester sur des données financières. Cette approche nous permettra de mieux comprendre les interactions entre le risque de transfert et les résultats obtenus avec le TL dans un cadre d'optimisation de portefeuille.

État de l'art

Le transfer learning a plusieurs utilisations clés. La première est son application dans les stratégies d'investissement personnalisées. Les approches comme celle de Zhang et al. (2018) [3] utilisent des données issues du marché boursier pour construire des recommandations adaptées à des marchés moins conventionnels, facilitant des décisions d'investissement même avec des données rares.

Une autre application est le trading haute fréquence, comme dans l'article de Leal et al. (2022) [4]. Des modèles sont pré-entraînés sur des données simulées avant de les ajuster sur des trajectoires réelles, augmentant la robustesse des stratégies malgré la rareté des données authentiques.

Il est également possible de prédire des actions. Wu et al. (2022) [5] démontrent que le transfert d'informations entre secteurs industriels améliore la précision des modèles de prévision boursière.

Enfin, le transfer learning est utile dans l'analyse de textes financiers. En effet, des modèles comme *FinBERT* (Liu et al. 2020) [6] exploitent les avancées du NLP pour extraire des informations des rapports et articles financiers, aidant à la prise de décision ou à l'identification des risques.

Le transfert d'apprentissage transforme la manière dont les modèles sont développés et déployés en finance, permettant de surmonter plusieurs limitations traditionnelles.

La revue systématique de Weiss et al. (2016) [7] met en lumière les défis du transfert d'apprentissage, notamment le transfert négatif et les biais, tout en proposant des solutions via des ajustements spécifiques des modèles. Cette approche permet de réduire considérablement le besoin en données annotées, particulièrement utile dans des contextes financiers où l'annotation manuelle est coûteuse ou difficile.

Wang et al. (2019) [8] ont souligné que l'alignement inadéquat entre tâches peut effectivement dégrader les performances et amener à un transfert négatif, que l'on cherche absolument à éviter dans le domaine financier. Ce défi souligne l'importance d'une sélection minutieuse des modèles sources et d'une adaptation appropriée au domaine cible, particulièrement cruciale dans un secteur où les erreurs peuvent avoir des conséquences économiques importantes.

Un défi majeur en finance est la gestion des séries temporelles. Comme l'ont montré Jiang et al. (2020) [9], l'application du transfert d'apprentissage aux données financières temporelles nécessite des adaptations spécifiques pour capturer efficacement les dépendances temporelles et les changements de régime.

Dans le secteur financier, l'interprétabilité des modèles est cruciale pour la conformité réglementaire. Arrieta et al. (2020) [10] ont souligné l'importance de développer des techniques pour rendre les modèles de transfert d'apprentissage plus transparents et interprétables, un défi particulièrement important dans un domaine aussi réglementé que la finance.

Malgré ces défis, le transfert d'apprentissage en finance ouvre de nombreuses perspectives prometteuses, allant du développement de modèles spécialisés à l'intégration de données alternatives, en passant par l'inclusion de la causalité et l'amélioration de l'interprétabilité, chacune de ces avenues contribuant à repousser les limites de l'analyse financière et de la prise de décision basée sur l'IA.

Des modèles comme *FinBERT*, spécifiquement conçus pour le domaine financier, promettent une meilleure compréhension du langage financier et une amélioration des performances dans des tâches spécifiques, comme démontré par Araci (2019) [11].

L'incorporation de données provenant des médias sociaux et d'autres sources non traditionnelles offre également de nouvelles perspectives pour enrichir les modèles financiers, comme l'ont montré Xu et Cohen (2018) [12].

Enfin, la recherche sur des méthodes permettant aux modèles de s'adapter en temps réel aux changements de conditions de marché pourrait améliorer significativement la résilience des systèmes financiers basés sur l'IA, comme proposé par Ye et al. (2020) [13].

Ces avancées et perspectives soulignent le potentiel transformateur du transfert d'apprentissage dans le domaine financier, promettant des analyses plus précises, des décisions plus éclairées et une meilleure gestion des risques.

2 Introduction Mathématiques au Transfer Learning :

Comme mentionné précédemment, le TL comporte deux domaines :

- Un domaine D_S correspondant au domaine source.
- Un domaine D_T correspondant au domaine cible.

À chaque domaine est associé un espace caractéristique \mathcal{X} et une loi marginale $P(\mathcal{X})$. Ainsi, on notera $D. = (\mathcal{X}, P(\mathcal{X}))$.

Similairement, le TL comporte deux tâches :

- Une tâche source S .
- Une tâche cible T .

À chaque tâche est associé un couple (Y, f) , avec Y un espace de labels et f une fonction de décision de $\mathcal{X} \rightarrow Y$. On aura ainsi deux couples (D_S, T_S) et (D_T, T_T) à considérer. L'objectif du TL étant de tirer parti des connaissances acquises à partir de la tâche source S pour améliorer la performance de la tâche cible T , ou autrement dit d'apprendre f pour faire des prédictions en utilisant (D_S, T_S) .

En général, on transfère s'il y a une certaine similarité entre la source et la cible. Plusieurs cas existent :

- Si $D_T = D_S$ et $f_T \approx f_S$, le modèle est dit *inductif*. Les méthodes utilisées sont généralement des méthodes de *fine tuning* où l'apprentissage est réalisé sur le modèle source, et on ajuste les paramètres du modèle pré-entraîné en fonction des données spécifiques de la tâche cible pour améliorer sa performance sur cette dernière.
- Si $D_T \approx D_S$ et $f_T = f_S$, le modèle est dit *transductif*. On utilise des méthodes d'adaptation de domaine où l'on cherche à aligner les distributions des données de la tâche source et de la tâche cible, afin que le modèle pré-entraîné puisse généraliser directement aux données de la tâche cible sans nécessiter un réapprentissage complet.
- Si $D_T \approx D_S$ et $f_T \approx f_S$, le modèle est dit *non supervisé*. Il faut alors identifier et extraire des caractéristiques pertinentes des données sans utiliser de labels. Cela peut impliquer l'utilisation de techniques telles que la *clustering*, la réduction de dimension, ou l'apprentissage de représentations, afin de découvrir des structures cachées ou des relations sous-jacentes dans les données.

Ici, nous travaillons dans un cadre *inductif* où $D_T = D_S$ et $f_T \approx f_S$.

2.1 Les Transformations dans le TL

Définissons un espace de probabilité (Ω, \mathcal{F}, P) .

En notant $L_T : Y_T \times Y_T \rightarrow \mathbb{R}$ une fonction à valeurs réelles, alors, dans ce cadre, la tâche cible T , sans appliquer de méthode de Transfer Learning soit en transfert direct, est représentée comme un problème d'optimisation :

$$\min_{f \in A_T} \mathcal{L}_T(f_T) = \min_{f_T \in A_T} \mathbb{E}[L_T(Y_T, f_T(X_T))] \quad (1)$$

où $L_T(f_T)$ est une fonction de perte mesurant le modèle $f_T : \mathcal{X}_T \rightarrow Y_T$ pour la tâche cible T , et A_T désignant l'ensemble des modèles cibles tels que $A_T \subset \{f_T \mid f_T : \mathcal{X}_T \rightarrow Y_T\}$.

De manière similaire, la tâche source S , sans TL, serait définie comme un problème d'optimisation tel que :

$$\min_{f \in A_S} \mathcal{L}_S(f_S) = \min_{f_S \in A_S} \mathbb{E}[L_S(Y_S, f_S(X_S))] \quad (2)$$

où $L_S(f_S)$ est une fonction de perte mesurant le modèle $f_S : \mathcal{X}_S \rightarrow \mathcal{Y}_S$ et A_S désigne l'ensemble des modèles sources tels que $A_S \subset \{f_S \mid f_S : \mathcal{X}_S \rightarrow \mathcal{Y}_S\}$.

Dans le cadre d'un modèle de TL, appelons f_S^* le modèle optimal obtenu à partir de la tâche source. Notre objectif ici est d'utiliser f_S^* pour atteindre l'objectif d'optimisation (1), c'est-à-dire améliorer la performance de la tâche cible T . Pour faire cela, nous procédons en trois étapes.

Étape 1 : X_T n'est pas nécessairement inclus dans l'espace X_S , donc nous allons d'abord effectuer une adaptation appropriée de l'entrée X_T .

Pour cela, utilisons une fonction T^X , telle que $T^X \in \{f_{\text{entrée}} \mid f_{\text{entrée}} : \mathcal{X}_T \rightarrow \mathcal{X}_S\}$ qui est une application permettant de transformer les entrées par rapport au couple de tâches (S, T) . Celle-ci prend tout point de données de l'espace d'entrée cible X_T et le transforme en un point de l'espace d'entrée source X_S .

Ainsi, cette première étape peut être représentée comme ceci :

$$X_T \ni X_T \xrightarrow{\text{Étape 1. Entrées transformées par } T^X} T^X(X_T) \in X_S.$$

N'importe quel élément de X_T est transformé en un élément de X_S grâce à la fonction de transformation de transport des entrées.

Étape 2 : Après avoir appliqué une transformation des entrées T^X aux entrées cibles X_T , le modèle pré-entraîné f_S^* utilisera les données précédemment transformées, c'est-à-dire $T^X(X_T) \in X_S$, comme nouvelles entrées.

$$\mathcal{X}_S \ni T^X(X_T) \xrightarrow{\text{Étape 2. Appliquer } f_S^*} (f_S^* \circ T^X)(X_T) \in Y_S$$

avec $f_S^* \circ T^X$ qui désigne la sortie correspondante du modèle pré-entraîné f_S^* .

Étape 3 : Après avoir utilisé le modèle pré-entraîné source, le modèle résultant n'est probablement pas adapté à la tâche cible. C'est pourquoi, nous allons utiliser une application T^Y de transformation, cette fois-ci permettant de générer un élément de l'espace de sortie cible Y_T :

$$T^Y \in \{f_{\text{sortie}} \mid f_{\text{sortie}} : X_T \times Y_S \rightarrow Y_T\}$$

T^Y transforme les sorties par rapport au couple de tâches (S, T) si, pour un modèle source optimal $f_S^* : \mathcal{X}_S \rightarrow Y_S$, la fonction composée $T^Y(\cdot, f_S^*(\cdot)) \in A_T$.

La dernière étape est donc la suivante :

$$X_T \times Y_S \ni (X_T, (f_S^* \circ T^X)(X_T)) \xrightarrow{\text{Etape 3. Sorties transformées par } T^Y} T^Y(X_T, (f_S^* \circ T^X)(X_T)) \in Y_T.$$

En résumé, le TL vise à trouver une paire appropriée de transformations T^X et T^Y , avec :

T^X qui transforme des éléments de l'entrée cible X_T en des éléments de l'espace des entrées de la source X_S afin d'utiliser le modèle source optimal f_S^* .

T^Y qui convertit un modèle à valeurs dans Y_S en un modèle à valeurs dans Y_T .

$$\mathcal{X}_S \ni X_S \xrightarrow{\text{Modèle pré-entraîné } f_S^* \text{ par (2)}} f_S^*(X_S) \in \mathcal{Y}_S$$

$$T^X \uparrow \qquad \qquad \qquad \downarrow T^Y$$

$$\mathcal{X}_T \ni X_T \xrightarrow[\substack{\text{Apprentissage Direct (1)} \\ f_T^* \in \arg \min_{f \in A_T} \mathcal{L}_T(f_T)}}{f_T^*} f_T^*(X_T) \in Y_T$$

Pourquoi ces transformations ? Notre objectif initial était de minimiser $\mathcal{L}_T(Y_T, f_T(X_T))$, qui est la fonction de coût sans opération de transfert. Puisqu'il était difficile d'estimer directement la fonction optimale f_T^* , nous avons transféré, c'est-à-dire que nous sommes passés par l'utilisation des données sources.

En effet, ici nous remplaçons f_T^* par une fonction f_S^* que nous allons entraîner à partir des données sources. Cela revient à faire de l'apprentissage direct mais sur les sources, et donc à écrire :

$$f_T^*(X_T) = T^Y(X_T, f_S^* \circ T^X(X_T)).$$

Nous allons donc apprendre f_S^* , T^X et T^Y à la place de f_T^* , ce qui représente un coût plus faible.

Ainsi, pour résumer :

- La **première étape** permet d'écrire tout élément de X_S comme une transformation d'un élément de X_T , tel que $X_S = T^X(X_T)$.

- La **seconde étape** permet d'appliquer la fonction optimale f_S^* aux données sources, tel que $f_S^*(X_S) = f_S^*(T^X(X_T))$.

- La **dernière étape** permet de transformer cette dernière équation en un élément de Y_T en appliquant T^Y , c'est-à-dire en faisant $T^Y(X_T, f_S^*(X_S))$.

Dans la cadre d'un transfert direct, l'estimateur \hat{f}_T serait associé au risque $\mathcal{L}_T(Y_T, f_T^*(X_T))$. Ici, on remplace $f_T^*(X_T)$ par une transformation $T^Y(X_T, (f_S^* \circ T^X)(X_T))$.

Ainsi, $\mathcal{L}_T(Y_T, T^Y(X_T, (f_S^* \circ T^X)(X_T)))$ est désormais associé à l'estimateur \hat{f}_{ST} , qu'on appellera *estimateur intermédiaire de transfer learning*.

$$\mathcal{L}_T(Y_T, f_T^*(X_T)) \xrightarrow{\text{Apprentissage direct}} \hat{f}_T(X_T)$$

$$\mathcal{L}_T(Y_T, T^Y(X_T, (f_S^* \circ T^X)(X_T))) \xrightarrow{\text{Transfer Learning}} \hat{f}_{ST}(X_T)$$

En d'autres termes, le TL correspond au **problème d'optimisation** suivant :

$$\min_{T^X \in \mathcal{T}^X, T^Y \in \mathcal{T}^Y} \mathcal{L}_T(T^Y(\cdot, (f_S^* \circ T^X)(\cdot))) = \mathbb{E}[L_T(Y_T, T^Y(X_T, (f_S^* \circ T^X)(X_T)))] \quad (3)$$

La résolution de ce problème d'optimisation est souvent désignée comme du *fine-tuning*. Cette procédure consiste à choisir des transformations initiales T_0^X et T_0^Y pour dériver un modèle intermédiaire f_{ST} tel que $f_{ST} \in A_T$.

2.2 Risque de Transfert et Mesures d'Efficacité

Dans le cadre du TL, il existe deux sources de risque de transfert pour le modèle intermédiaire f_{ST} .

Risque de transfert de sortie

Soit $\mathcal{E}^O : A_T \rightarrow \mathbb{R}$ une fonction à valeurs réelles définie sur l'ensemble des modèles cibles.

Pour tout $f_{ST} \in A_T$, $\mathcal{E}^O(f_{ST})$ est appelé un risque de transfert de sortie du modèle f_{ST} si les conditions suivantes sont satisfaites :

1. $\mathcal{E}^O(f_{ST}) \geq 0$, c'est-à-dire que le transfert d'apprentissage engendre toujours un effort non négatif.
2. $\mathcal{E}^O(f_{ST}) = 0$ si, et seulement si, $\mathbb{P}_T = \mathbb{P}_{ST}$, où $\mathbb{P}_T := \mathcal{L}(f_T(X_T))$ et $\mathbb{P}_{ST} := \mathcal{L}(f_{ST}(X_T))$. Autrement dit, le risque de transfert de sortie est nul lorsque le modèle intermédiaire f_{ST} reproduit exactement la distribution de la tâche cible optimale. Ainsi, plus ce risque de sortie est faible, plus le transfert est efficace avec le modèle f_{ST} .

Risque de transfert d'entrée

Soit $\mathcal{E}^I : \mathcal{T}^X \rightarrow \mathbb{R}$ une fonction à valeurs réelles définie sur l'ensemble \mathcal{T}^X . Étant donné une transformation initiale $T_0^X \in \mathcal{T}_0^X$, $\mathcal{E}^I(T_0^X)$ est le risque de transfert d'entrée si les conditions suivantes sont satisfaites :

1. $\mathcal{E}^I(T_0^X) \geq 0$, c'est-à-dire que le transfert d'apprentissage engendre toujours un effort non négatif.
2. $\mathcal{E}^I(T_0^X) = 0$ si, et seulement si, $T_0^X \# \mathcal{L}(X_T) = \mathcal{L}(X_S)$. Cela signifie que lorsqu'on applique T_0^X à une variable X_T ayant la distribution $\mathcal{L}(X_T)$, on obtient une nouvelle variable dont la distribution correspond exactement à $\mathcal{L}(X_S)$.

Risque de transfert

Pour une procédure de TL caractérisée par le 6-uplet $(S, T, T_0^X, T_0^Y, T^X, T^Y)$ dans (3), le risque de transfert de la tâche source S vers la tâche cible T est défini comme $C : \mathbb{R} \rightarrow \mathbb{R}$ avec $C(0, 0) = 0$, tel que :

$$C(S, T) = \inf_{f_{ST} \in \mathcal{I}} C(S, T \mid f_{ST})$$

Ici, pour un modèle intermédiaire de TL donné $f_{ST} = T_0^Y(\cdot, (f_S^* \circ T_0^X)(\cdot)) \in \mathcal{I}$, $C(S, T \mid f_{ST}) \geq 0$, le risque de transfert a les propriétés suivantes :

1. $C(S, T \mid f_{ST}) = C(\mathcal{E}^O(f_{ST}), \mathcal{E}^I(T_0^X))$ est non décroissant par rapport à $\mathcal{E}^O(f_{ST})$ pour une $\mathcal{E}^I(T_0^X)$ fixée, et non décroissant par rapport à $\mathcal{E}^I(T_0^X)$ pour une $\mathcal{E}^O(f_{ST})$ fixée.
2. $C(S, T \mid f_{ST})$ est K -lipschitzienne. Ainsi en définissant une autre procédure de transfert caractérisée par $(\bar{S}, \bar{T}, T_0^X, T_0^Y, T^X, T^Y)$ et son modèle intermédiaire $\bar{f}_{ST} = T_0^Y(\cdot, (f_S^* \circ T_0^X)(\cdot))$, il existe une constante $K > 0$ telle que :

$$|C(S, T \mid f_{ST}) - C(\bar{S}, \bar{T} \mid \bar{f}_{ST})| \leq K(|\mathcal{E}^O(f_{ST}) - \mathcal{E}^O(\bar{f}_{ST})| + |\mathcal{E}^I(T_0^X) - \mathcal{E}^I(T_0^Y)|).$$

Ces définitions des risques impliquent que l'on ait des applications initiales T_0^X et T_0^Y .

Un exemple simple de risque de transfert est donné par :

$$C^\lambda(S, T \mid f_{ST}) = \mathcal{E}^O(f_{ST}) + \lambda \mathcal{E}^I(T_0^X),$$

où $\lambda > 0$ est un paramètre préspecifié modulant le poids du risque d'entrée. .

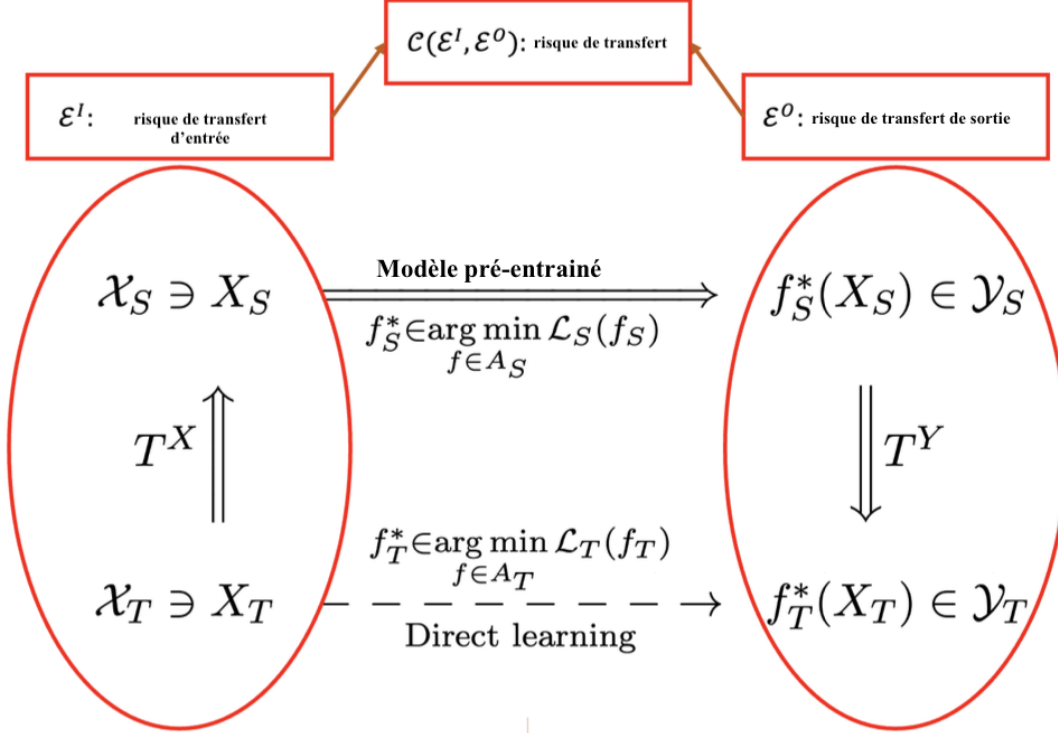


FIGURE 1 – Relation entre le risque de transfert et le cadre du transfer learning

2.2.1 Risque de transfert dans des modèles à distribution gaussienne

Le but ici est d'établir un lien entre le risque de transfert $C(S, T)$ et les performances de l'apprentissage par transfert dans des modèles gaussiens. Nous allons nous concentrer sur le risque de transfert basé sur la distance de Wasserstein, idéale pour les modèles d'apprentissage supervisé.

Considérons une tâche source S et une tâche cible T avec un espace d'entrée $X_S = X_T = \mathbb{R}^d$ et un espace de sortie $\mathcal{Y}_S = \mathcal{Y}_T = \mathbb{R}$. Nous supposons que les données source et cible suivent une distribution gaussienne de dimension $(d + 1)$, telle que $(X, Y) \sim \mathcal{N}(\mu, \Sigma)$, avec :

$$\mu = \begin{pmatrix} \mu_{\cdot, X} \\ \mu_{\cdot, Y} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{\cdot, X} & \Sigma_{\cdot, XY} \\ \Sigma_{\cdot, YX} & \Sigma_{\cdot, Y} \end{pmatrix},$$

où $\mu_{\cdot, Y}$ et $\Sigma_{\cdot, Y} \in \mathbb{R}$, $\mu_{\cdot, X}$ et $\Sigma_{\cdot, XY} \in \mathbb{R}^d$, $\Sigma_{\cdot, YX} = \Sigma_{\cdot, XY}^\top$, et $\Sigma_{\cdot, X} \in \mathbb{R}^{d \times d}$. Supposons que pour la tâche source et la tâche cible, nous avons $A_S = A_T = \{f : \mathbb{R}^d \rightarrow \mathbb{R}\}$ et que, pour tout $f \in A_S = A_T$, les fonctions

de perte sont données par :

$$\mathcal{L}_S(f) = \mathbb{E}\|Y_S - f(X_S)\|_2^2, \quad \mathcal{L}_T(f) = \mathbb{E}\|Y_T - f(X_T)\|_2^2.$$

Pour que $X_S = X_T = \mathbb{R}^d$ et que $\mathcal{Y}_S = \mathcal{Y}_T = \mathbb{R}$, nous considérons un cadre où les applications d'entrées T^X et de sorties T^Y sont des singletons ne contenant que l'identité sur \mathbb{R}^d (resp. \mathbb{R}). Par conséquent, l'ensemble des estimateurs intermédiaires \mathcal{I} est un singleton donné par $\mathcal{I} = \{f_S^*\}$. Ainsi, le schéma d'apprentissage du TL se simplifie en une étape :

$$\mathcal{X}_S \ni X_S \xrightarrow{\text{Modèle pré-entraîné } f_S^*} f_S^*(X_S) \in \mathcal{Y}_S$$

Nous allons définir le risque de transfert comme le risque de transport des sorties sur la distance de Wasserstein.

$$C_W(S, T) = C_W(S, T|f_S^*) = \mathcal{E}_0^W(f_S^*).$$

avec

$$\mathcal{E}_W^O(f_S^*) := \mathcal{W}_p(\mathbb{P}_{ST}, \mathbb{P}_T)^p := \inf_{\gamma \in \Pi(\mathbb{P}_{ST}, \mathbb{P}_T)} \int_{\mathbb{R}^{d_O, T} \times \mathbb{R}^{d_O, T}} \|x - y\|_{\mathcal{Y}_T}^p d\gamma(dx, dy),$$

où :

- $\mathcal{W}_p^p(\mathbb{P}_{ST}, \mathbb{P}_T)$ est la p -ème puissance de la distance de Wasserstein d'ordre p ,
- $\Pi(\mathbb{P}_{ST}, \mathbb{P}_T)$ désigne l'ensemble des couplages possibles (i.e., distributions jointes γ) entre les deux mesures \mathbb{P}_{ST} et \mathbb{P}_T ,
- $\|x - y\|_{\mathcal{Y}_T}$ norme définie dans l'espace cible \mathcal{Y}_T ,
- $\gamma(x, y)$ est une mesure jointe de \mathbb{P}_{ST} et \mathbb{P}_T .

Par ailleurs, définissons le regret. Celui-ci correspond à l'écart entre le TL et l'apprentissage direct :

$$\mathcal{R}(S, T) := \mathcal{L}_T(f_S^*) - \mathcal{L}_T(f_T^*).$$

Nous allons maintenant démontrer que le risque de transfert constitue une borne inférieure du regret.

Théorème : Dans un modèle gaussien de TL, le regret par rapport au modèle intermédiaire $\mathcal{R}(S, T)$ est borné par le risque de transfert basé sur la distance de Wasserstein, tel que :

$$C_W(S, T) \leq \mathcal{R}(S, T).$$

Ce théorème permet ainsi d'avoir une première indication sur l'efficacité du TL, en particulier pour éliminer les tâches sources inadaptées si le risque de transfert est trop élevé.

Démonstration : Supposons que les données sources et cibles ont les mêmes dimensions. Alors, l'objectif est de trouver le modèle optimal f_S^* qui minimise la fonction de perte $\mathcal{L}_S(f)$, définie comme l'erreur quadratique moyenne :

$$\mathcal{L}_S(f) = \mathbb{E}\|Y_S - f(X_S)\|_2^2.$$

Ainsi, le problème d'optimisation s'écrit :

$$f_S^* \in \arg \min_{f \in A_S} \mathcal{L}_S(f),$$

où A_S représente l'ensemble des modèles admissibles. Dans ce cadre, on suppose que les modèles $f \in A_S$ sont linéaires, ce qui correspond à une régression linéaire classique.

La solution optimale de régression linéaire est une fonction affine donnée par :

$$f_S^*(x) = w_S^\top x + b_S,$$

où w_S est le vecteur des poids, et b_S est un biais scalaire. Ainsi les prédictions sont bien des combinaisons linéaires pondérées des entrées x .

Pour obtenir les coefficients w_S et b_S , nous allons annuler les dérivées partielles de $\mathcal{L}_S(f)$ par rapport à w et b .

La fonction de perte est définie comme suit :

$$\mathcal{L}_S(f) = \mathbb{E} [\|Y_S - (w_S^\top X_S + b_S)\|_2^2],$$

où $f(X_S) = w_S^\top X_S + b_S$ est le modèle linéaire.

En développant, on a :

$$\|Y_S - (w_S^\top X_S + b_S)\|_2^2 = (Y_S - w_S^\top X_S - b_S)^2.$$

Ainsi, la fonction de perte devient :

$$\mathcal{L}_S(f) = \mathbb{E} [(Y_S - w_S^\top X_S - b_S)^2].$$

Calculons la dérivée partielle de $\mathcal{L}_S(f)$ par rapport à w_S :

$$\frac{\partial \mathcal{L}_S(f)}{\partial w_S} = \mathbb{E} [-2X_S (Y_S - w_S^\top X_S - b_S)].$$

En isolant les termes, cela devient :

$$\frac{\partial \mathcal{L}_S(f)}{\partial w_S} = -2\mathbb{E}[X_S Y_S] + 2\mathbb{E}[X_S X_S^\top] w_S + 2\mathbb{E}[X_S] b_S.$$

Pour annuler cette dérivée, on résout :

$$\mathbb{E}[X_S X_S^\top] w_S + \mathbb{E}[X_S] b_S = \mathbb{E}[X_S Y_S].$$

Calculons la dérivée partielle de $\mathcal{L}_S(f)$ par rapport à b_S :

$$\frac{\partial \mathcal{L}_S(f)}{\partial b_S} = \mathbb{E} [-2(Y_S - w_S^\top X_S - b_S)].$$

Cela devient :

$$\frac{\partial \mathcal{L}_S(f)}{\partial b_S} = -2\mathbb{E}[Y_S] + 2\mathbb{E}[w_S^\top X_S] + 2b_S.$$

Pour annuler cette dérivée, on résout :

$$b_S + w_S^\top \mathbb{E}[X_S] = \mathbb{E}[Y_S].$$

En introduisant les moments statistiques :

- $\mu_{S,X} = \mathbb{E}[X_S]$,
- $\mu_{S,Y} = \mathbb{E}[Y_S]$,
- $\Sigma_{S,X} = \mathbb{E}[(X_S - \mu_{S,X})(X_S - \mu_{S,X})^\top] = \mathbb{E}[X_S X_S^\top] - \mu_{S,X} \mu_{S,X}^\top$,
- $\Sigma_{S,XY} = \mathbb{E}[(X_S - \mu_{S,X})(Y_S - \mu_{S,Y})] = \mathbb{E}[X_S Y_S] - \mu_{S,X} \mu_{S,Y}$,

En substituant les moments statistiques, on obtient :

$$\begin{aligned}\mathbb{E}[X_S X_S^\top] &= \Sigma_{S,X} + \mu_{S,X} \mu_{S,X}^\top, \\ \mathbb{E}[X_S Y_S] &= \Sigma_{S,XY} + \mu_{S,X} \mu_{S,Y}.\end{aligned}$$

Ainsi :

$$\mathbb{E}[X_S X_S^\top] w_S + \mathbb{E}[X_S] b_S = \mathbb{E}[X_S Y_S].$$

En remplaçant avec les moments statistiques, on obtient :

$$(\Sigma_{S,X} + \mu_{S,X} \mu_{S,X}^\top) w_S + \mu_{S,X} b_S = \Sigma_{S,XY} + \mu_{S,X} \mu_{S,Y}.$$

Développons :

$$\Sigma_{S,X} w_S + (\mu_{S,X} \mu_{S,X}^\top w_S) + \mu_{S,X} b_S = \Sigma_{S,XY} + \mu_{S,X} \mu_{S,Y}.$$

En regroupant les termes contenant $\mu_{S,X}$, on a :

$$\Sigma_{S,X} w_S + \mu_{S,X} (\mu_{S,X}^\top w_S + b_S) = \Sigma_{S,XY} + \mu_{S,X} \mu_{S,Y}.$$

Puisque $\mu_{S,X}^\top w_S + b_S = \mu_{S,Y}$, cela réduit l'équation à :

$$\Sigma_{S,X} w_S = \Sigma_{S,XY}.$$

En isolant w_S , on obtient :

$$w_S = \Sigma_{S,X}^{-1} \Sigma_{S,XY}.$$

Reprenons l'équation :

$$b_S + \mathbb{E}[w_S^\top X_S] = \mathbb{E}[Y_S].$$

En isolant b_S , on obtient :

$$b_S = \mu_{S,Y} - w_S^\top \mu_{S,X}.$$

Les coefficients optimaux sont donc donnés par :

$$\begin{aligned}w_S &= \Sigma_{S,X}^{-1} \Sigma_{S,XY}, \\ b_S &= \mu_{S,Y} - w_S^\top \mu_{S,X}.\end{aligned}$$

De manière similaire, le modèle optimal pour la tâche cible est donné par :

$$f_T^*(x) = w_T^\top x + b_T, \quad \forall x \in \mathbb{R}^d,$$

où :

$$w_T = \Sigma_{T,X}^{-1} \Sigma_{T,XY}, \quad b_T = \mu_{T,Y} - \Sigma_{T,YX} \Sigma_{T,X}^{-1} \mu_{T,X}.$$

La distribution des sorties P_T est donnée par :

$$P_T = \mathbb{E}[Y_T|X_T] = \mathcal{N}(w_T^\top \mu_{T,X} + b_T, w_T^\top \Sigma_{T,X} w_T),$$

En effet, dans un modèle linéaire gaussien, la moyenne conditionnelle est donnée par :

$$\mathbb{E}[Y_T|X_T] = w_T^\top X_T + b_T,$$

où $w_T = \Sigma_{T,X}^{-1} \Sigma_{T,XY}$ et $b_T = \mu_{T,Y} - w_T^\top \mu_{T,X}$. Puisque $X_T \sim \mathcal{N}(\mu_{T,X}, \Sigma_{T,X})$, une combinaison linéaire de X_T suit également une loi normale. Par conséquent, $w_T^\top X_T + b_T$ suit une loi normale avec :

$$(w_T^\top X_T + b_T) \sim \mathcal{N}(w_T^\top \mu_{T,X} + b_T, w_T^\top \Sigma_{T,X} w_T).$$

En effet, pour toute transformation affine $Z = a^\top X + b$ d'une variable normale $X \sim \mathcal{N}(\mu, \Sigma)$, la variance de Z est donnée par :

$$\text{Var}(Z) = a^\top \Sigma a,$$

où a est un vecteur et Σ est la matrice de covariance de X . Dans notre cas :

$$Z = w_T^\top X_T + b_T, \quad a = w_T, \quad \text{et} \quad \Sigma = \Sigma_{T,X}.$$

Étant donné que $\mu_T = w_T^\top \mu_{T,X} + b_T$, on a donc :

$$P_T \sim \mathcal{N}(\mu_T, w_T^\top \Sigma_{T,X} w_T).$$

La distribution \mathbb{P}_{ST} correspond à la loi de $f_S^*(X_T)$, où $X_T \sim \mathcal{N}(\mu_{T,X}, \Sigma_{T,X})$:

$$\mathbb{P}_{ST} = f_S^* \# \mathcal{N}(\mu_{T,X}, \Sigma_{T,X})$$

En appliquant la transformation f_S^* à une variable aléatoire suivant une loi normale $\mathcal{N}(\mu_{T,X}, \Sigma_{T,X})$, et sachant qu'on a : $f_S^*(x) = w_S^\top x + b_S$, où $w_S = \Sigma_{S,X}^{-1} \Sigma_{S,XY}$ et $b_S = \mu_{S,Y} - w_S^\top \mu_{S,X}$, la distribution de $f_S^*(X_T)$ est donnée par :

$$f_S^*(X_T) \sim \mathcal{N}(w_S^\top \mu_{T,X} + b_S, w_S^\top \Sigma_{T,X} w_S).$$

On remarque que $\mathbb{P}_T \ll \mathbb{P}_{ST}$, ce qui signifie que la mesure \mathbb{P}_T est absolument continue par rapport à \mathbb{P}_{ST} . Ceci implique l'existence de la densité de Radon-Nikodym $h_{ST}(y)$ telle que :

$$\frac{d\mathbb{P}_T}{d\mathbb{P}_{ST}}(y) = h_{ST}(y).$$

Les densités respectives des lois normales \mathbb{P}_T et \mathbb{P}_{ST} sont données par :

$$p_{\mathbb{P}_{ST}}(y) = \frac{1}{\sqrt{2\pi w_S^\top \Sigma_{T,X} w_S}} \exp \left\{ -\frac{(y - (w_S^\top \mu_{T,X} + b_S))^2}{2w_S^\top \Sigma_{T,X} w_S} \right\},$$

$$p_{\mathbb{P}_T}(y) = \frac{1}{\sqrt{2\pi w_T^\top \Sigma_{T,X} w_T}} \exp \left\{ -\frac{(y - (w_T^\top \mu_{T,X} + b_T))^2}{2w_T^\top \Sigma_{T,X} w_T} \right\}.$$

Le rapport des deux densités donne :

$$h_{ST}(y) = \sqrt{\frac{w_S^\top \Sigma_{T,X} w_S}{w_T^\top \Sigma_{T,X} w_T}} \exp \left\{ -\frac{(y - (w_T^\top \mu_{T,X} + b_T))^2}{2w_T^\top \Sigma_{T,X} w_T} + \frac{(y - (w_S^\top \mu_{T,X} + b_S))^2}{2w_S^\top \Sigma_{T,X} w_S} \right\}.$$

Calculons le risque de transfert $C_W(S, T)$ sur la distance de Wasserstein. Le risque de transport basé sur Wasserstein est défini comme :

$$\mathcal{E}_W^O(f_S^*) := \mathcal{W}_2^2(\mathbb{P}_{ST}, \mathbb{P}_T) = \inf_{\gamma \in \Pi(\mathbb{P}_{ST}, \mathbb{P}_T)} \int_{\mathcal{Y}_T \times \mathcal{Y}_T} \|x - y\|^2 d\gamma(x, y),$$

Ici, \mathbb{P}_{ST} et \mathbb{P}_T sont des distributions normales, on peut calculer explicitement cette distance [14]. Pour deux lois normales $\mathcal{N}(\mu_1, \Sigma_1)$ et $\mathcal{N}(\mu_2, \Sigma_2)$, la distance de Wasserstein d'ordre 2 est donnée par :

$$\mathcal{W}_2^2(\mathcal{N}(\mu_1, \Sigma_1), \mathcal{N}(\mu_2, \Sigma_2)) = \|\mu_1 - \mu_2\|_2^2 + \text{Tr}(\Sigma_1 + \Sigma_2 - 2(\Sigma_2^{1/2} \Sigma_1 \Sigma_2^{1/2})^{1/2}).$$

Les distributions considérées sont :

- $\mathbb{P}_{ST} \sim \mathcal{N}(w_S^\top \mu_{T,X} + b_S, w_S^\top \Sigma_{T,X} w_S)$,
- $\mathbb{P}_T \sim \mathcal{N}(\mu_T, w_T^\top \Sigma_{T,X} w_T)$.

Substituons dans la formule de Wasserstein. La différence des moyennes donne :

$$\|\mu_T - (w_S^\top \mu_{T,X} + b_S)\|_2^2.$$

Or $\mu_T = w_T^\top \mu_{T,X} + b_T$, donc on peut réécrire la norme :

$$\|w_T^\top \mu_{T,X} + b_T - (w_S^\top \mu_{T,X} + b_S)\|_2^2$$

De plus, on a

$$\begin{aligned} w_S &= \Sigma_{S,X}^{-1} \Sigma_{S,XY}, \\ b_S &= \mu_{S,Y} - w_S^\top \mu_{S,X}, \\ w_T &= \Sigma_{T,X}^{-1} \Sigma_{T,XY}, \\ b_T &= \mu_{T,Y} - \Sigma_{T,YX} \Sigma_{T,X}^{-1} \mu_{T,X} \end{aligned}$$

On peut alors calculer la norme :

$$\|\Sigma_{T,XY} \Sigma_{T,X}^{-1} \mu_{T,X} + \mu_{T,Y} - \Sigma_{T,YX} \Sigma_{T,X}^{-1} \mu_{T,X} - \Sigma_{S,YX} \Sigma_{S,X}^{-1} \mu_{S,X} - \mu_{S,Y} + \Sigma_{S,YX} \Sigma_{S,X}^{-1} \mu_{S,X}\|_2^2$$

Ce qui donne en simplifiant :

$$\|\mu_{T,Y} - \mu_{S,Y} - \Sigma_{S,YX} \Sigma_{S,X}^{-1} (\mu_{T,X} - \mu_{S,X})\|_2^2$$

Correspondant ainsi au premier terme de $C_W(S, T)$.

Pour le second terme, on a :

$$\text{Tr}(w_S^\top \Sigma_{T,X} w_S + w_T^\top \Sigma_{T,X} w_T - 2 \left(\sqrt{w_T^\top \Sigma_{T,X} w_T} \cdot w_S^\top \Sigma_{T,X} w_S \cdot \sqrt{w_S^\top \Sigma_{T,X} w_S} \right)^{1/2}).$$

Posons :

$$\Sigma_1 = w_S^\top \Sigma_{T,X} w_S, \quad \Sigma_2 = w_T^\top \Sigma_{T,X} w_T.$$

Ainsi ce second terme peut s'écrire :

$$\text{Tr}(\Sigma_1 + \Sigma_2 - 2(\Sigma_2^{1/2} \Sigma_1 \Sigma_2^{1/2})^{1/2}),$$

où :

- Σ_1 et Σ_2 sont les matrices de covariance des deux lois normales,
- $(\Sigma_2^{1/2} \Sigma_1 \Sigma_2^{1/2})^{1/2}$ est la racine carrée de la matrice obtenue par produit matriciel.

Or, on remarque que Σ_1 et Σ_2 sont des scalaires (ils commutent donc), ce qui nous permet de simplifier la trace :

$$\text{Tr}(\Sigma_1 + \Sigma_2 - 2(\Sigma_2^{1/2} \Sigma_1 \Sigma_2^{1/2})^{1/2}) = \Sigma_1 + \Sigma_2 - 2(\Sigma_2^{1/2} \Sigma_1 \Sigma_2^{1/2})^{1/2} = \Sigma_1 + \Sigma_2 - 2\sqrt{\Sigma_2 \Sigma_1}$$

Ce qui peut être réécrit sous la forme :

$$\left(\sqrt{\Sigma_1} - \sqrt{\Sigma_2} \right)^2.$$

Dans notre cas, nous avons :

$$\Sigma_1 = w_S^\top \Sigma_{T,X} w_S, \quad \Sigma_2 = w_T^\top \Sigma_{T,X} w_T.$$

Ainsi, le second terme de la distance de Wasserstein devient :

$$\left(\sqrt{w_S^\top \Sigma_{T,X} w_S} - \sqrt{w_T^\top \Sigma_{T,X} w_T} \right)^2.$$

En combinant les deux termes, on obtient :

$$C_W(S, T) = \left[\mu_{T,Y} - \mu_{S,Y} - \Sigma_{S,YX} \Sigma_{S,X}^{-1} (\mu_{T,X} - \mu_{S,X}) \right]^2 + \left(\sqrt{w_S^\top \Sigma_{T,X} w_S} - \sqrt{w_T^\top \Sigma_{T,X} w_T} \right)^2.$$

Et en réécrivant $w_S = \Sigma_{S,X}^{-1} \Sigma_{S,XY}$ et $w_T = \Sigma_{T,X}^{-1} \Sigma_{T,XY}$, on obtient $C_W(S, T)$.

Le regret est défini comme :

$$\mathcal{R}(S, T) := \mathcal{L}_T(f_S^*) - \mathcal{L}_T(f_T^*),$$

La perte quadratique est donnée par :

$$\mathcal{L}_T(f) = \mathbb{E}[(Y_T - f(X_T))^2].$$

Pour $f_S^*(x) = w_S^\top x + b_S$, on a :

$$\mathcal{L}_T(f_S^*) = \mathbb{E}[(Y_T - (w_S^\top X_T + b_S))^2].$$

Développons le carré :

$$(Y_T - (w_S^\top X_T + b_S))^2 = Y_T^2 - 2Y_T(w_S^\top X_T + b_S) + (w_S^\top X_T + b_S)^2.$$

En prenant l'espérance :

$$\mathcal{L}_T(f_S^*) = \mathbb{E}[Y_T^2] - 2\mathbb{E}[Y_T(w_S^\top X_T + b_S)] + \mathbb{E}[(w_S^\top X_T + b_S)^2].$$

Calcul du premier terme : $\mathbb{E}[Y_T^2]$

On sait que $Y_T \sim \mathcal{N}(\mu_{T,Y}, \sigma_{T,Y}^2)$. Par conséquent :

$$\mathbb{E}[Y_T^2] = \text{Var}(Y_T) + (\mathbb{E}[Y_T])^2 = \sigma_{T,Y}^2 + \mu_{T,Y}^2.$$

Calcul du second terme : $\mathbb{E}[Y_T(w_S^\top X_T + b_S)]$

$$\mathbb{E}[Y_T(w_S^\top X_T + b_S)] = \mathbb{E}[Y_T w_S^\top X_T] + \mathbb{E}[Y_T] b_S.$$

Or, $\mathbb{E}[Y_T w_S^\top X_T] = w_S^\top \mathbb{E}[Y_T X_T]$ et $\mathbb{E}[Y_T X_T] = \Sigma_{T,XY} + \mu_{T,Y} \mu_{T,X}$. Donc :

$$\mathbb{E}[Y_T w_S^\top X_T] = w_S^\top (\Sigma_{T,XY} + \mu_{T,Y} \mu_{T,X}).$$

Ainsi :

$$\mathbb{E}[Y_T(w_S^\top X_T + b_S)] = w_S^\top \Sigma_{T,XY} + w_S^\top \mu_{T,X} \mu_{T,Y} + \mu_{T,Y} b_S.$$

Calcul du troisième terme : $\mathbb{E}[(w_S^\top X_T + b_S)^2]$

Développons le carré :

$$(w_S^\top X_T + b_S)^2 = (w_S^\top X_T)^2 + 2w_S^\top X_T b_S + b_S^2.$$

En prenant l'espérance de chaque terme :

$\mathbb{E}[(w_S^\top X_T)^2] = w_S^\top \Sigma_{T,X} w_S + (w_S^\top \mu_{T,X})^2$, car pour $Z = w_S^\top X_T$, nous avons $\mathbb{E}[Z^2] = \text{Var}(Z) + (\mathbb{E}[Z])^2$, avec

$\text{Var}(Z) = w_S^\top \Sigma_{T,X} w_S$ et $\mathbb{E}[Z] = w_S^\top \mu_{T,X}$

$\mathbb{E}[2w_S^\top X_T b_S] = 2b_S w_S^\top \mu_{T,X}$,

$\mathbb{E}[b_S^2] = b_S^2$.

Ainsi :

$$\mathbb{E}[(w_S^\top X_T + b_S)^2] = w_S^\top \Sigma_{T,X} w_S + (w_S^\top \mu_{T,X})^2 + 2b_S w_S^\top \mu_{T,X} + b_S^2.$$

En combinant tous les termes, on obtient :

$$\mathcal{L}_T(f_S^*) = \sigma_{T,Y}^2 + \mu_{T,Y}^2 - 2(w_S^\top \Sigma_{T,XY} + w_S^\top \mu_{T,X} \mu_{T,Y} + \mu_{T,Y} b_S) + w_S^\top \Sigma_{T,X} w_S + (w_S^\top \mu_{T,X})^2 + 2b_S w_S^\top \mu_{T,X} + b_S^2.$$

Pour $f_T^*(x) = w_T^\top x + b_T$, le même calcul donne :

$$\mathcal{L}_T(f_T^*) = \sigma_{T,Y}^2 + \mu_{T,Y}^2 - 2(w_T^\top \Sigma_{T,XY} + w_T^\top \mu_{T,X} \mu_{T,Y} + \mu_{T,Y} b_T) + w_T^\top \Sigma_{T,X} w_T + (w_T^\top \mu_{T,X})^2 + 2b_T w_T^\top \mu_{T,X} + b_T^2.$$

En soustrayant $\mathcal{L}_T(f_T^*)$ de $\mathcal{L}_T(f_S^*)$, on obtient pour les termes liés aux poids (w_T, w_S) :

$$\|\Sigma_{T,X}^{1/2}(w_T - w_S)\|_2^2 = w_T^\top \Sigma_{T,X} w_T + w_S^\top \Sigma_{T,X} w_S - 2w_T^\top \Sigma_{T,X} w_S.$$

puis pour les termes liés aux biais (b_T, b_S) :

$$\left[\mu_{T,Y} - \mu_{S,Y} - \Sigma_{S,YX} \Sigma_{S,X}^{-1} (\mu_{T,X} - \mu_{S,X}) \right]^2.$$

Ainsi, finalement :

$$\mathcal{R}(S, T) = \|\Sigma_{T,X}^{1/2}(w_T - w_S)\|_2^2 + \left[\mu_{T,Y} - \mu_{S,Y} - \Sigma_{S,YX} \Sigma_{S,X}^{-1} (\mu_{T,X} - \mu_{S,X}) \right]^2.$$

3 Problème d'optimisation de portefeuille

Considérons un marché de capitaux composé de d actifs.

- Les rendements annualisés sont représentés par le vecteur aléatoire $\mathbf{r} = (r_1, \dots, r_d)^\top \sim \mathbb{P}$.
- Un vecteur d'allocation de portefeuille est défini tel que $\phi = (\phi_1, \dots, \phi_d)^\top$. C'est un vecteur de dimension d appartenant à l'ensemble $\mathcal{X} := \{\phi \in \mathbb{R}^d : \sum_{i=1}^d \phi_i = 1\}$, avec ϕ_i le pourcentage du capital disponible investi dans l'actif i pour chaque $i = 1, \dots, d$. Cela signifie que les poids du portefeuille sont positifs et leur somme est égale à 1.

Ainsi, le rendement annualisé d'un portefeuille ϕ est donné par :

$$\phi^\top \mathbf{r}.$$

C'est une combinaison linéaire pondérée des rendements des actifs, pondérée par les proportions ϕ .

Notre objectif est de trouver le portefeuille optimal, c'est-à-dire celui ayant le **ratio de Sharpe le plus élevé**, ce qui revient à résoudre le problème d'optimisation suivant :

$$\hat{\phi} = \arg \max_{\phi \in \mathcal{X}} \frac{\mathbb{E}[\phi^\top \mathbf{r}]}{\text{Std}(\phi^\top \mathbf{r})} = \arg \max_{\phi \in \mathcal{X}} \frac{\mu_{\mathbb{P}}^\top \phi}{\sqrt{\phi^\top \Sigma_{\mathbb{P}} \phi}},$$

où :

- $\mu_{\mathbb{P}}$ et $\Sigma_{\mathbb{P}}$ sont respectivement l'espérance et la matrice de covariance des rendements \mathbf{r} ,
- $\text{Std}(\phi^\top \mathbf{r})$ est l'écart-type du rendement du portefeuille, mesurant la volatilité des rendements,
- $\mathbb{E}[\phi^\top \mathbf{r}]$ est l'espérance du rendement du portefeuille, soit la moyenne théorique des rendements.

En effet, le ratio de Sharpe est le rapport entre le rendement espéré $\mathbb{E}[\phi^\top \mathbf{r}]$ et le risque $\text{Std}(\phi^\top \mathbf{r})$. Ce ratio représente ainsi l'efficacité d'un portefeuille à générer des rendements pour chaque unité de risque prise.

Dans les marchés, il peut être difficile d'estimer $\mu_{\mathbb{P}}$ et $\Sigma_{\mathbb{P}}$ à cause du manque de données. Le TL nous permet alors d'utiliser des données issues de marchés similaires (source) pour estimer $\mu_{\mathbb{P}}$ et $\Sigma_{\mathbb{P}}$.

3.1 En Pratique

L'objectif est d'exploiter les informations des rendements d'actifs sources (marché américain, S&P 500) pour optimiser des portefeuilles sur des marchés cibles (Royaume-Uni, Brésil, Allemagne et Singapour). Ce processus se déroule en trois étapes principales :

1. **Optimisation directe sur les données cibles.**
2. **Optimisation sur les données sources.**
3. **Optimisation régularisée avec transfer learning.**

3.2 Étape 1 : Optimisation directe sur les données cibles

Dans cette première étape, nous optimisons directement le portefeuille pour les données cibles. Les poids du portefeuille, notés ϕ_{direct} , sont obtenus en maximisant le ratio de Sharpe défini comme suit :

$$\phi_{\text{direct}} = \underset{\phi \in X}{\operatorname{argmax}} \frac{\mu_T^\top \phi}{\sqrt{\phi^\top \Sigma_T \phi}} \quad (4)$$

où :

- μ_T est le vecteur des rendements moyens des actifs cibles.
- Σ_T est la matrice de covariance des rendements des actifs cibles.
- ϕ est le vecteur des poids du portefeuille.

La contrainte est que la somme des poids doit être égale à 1 et chaque poids positif :

$$\sum_{i=1}^d \phi_i = 1, \quad \phi_i \geq 0 \quad \forall i.$$

3.3 Étape 2 : Optimisation sur les données sources

Dans cette étape, nous réalisons une optimisation similaire, mais cette fois sur les données sources. Les poids optimaux ϕ_S sont obtenus en maximisant :

$$\phi_S = \underset{\phi \in X}{\operatorname{argmax}} \frac{\mu_S^\top \phi}{\sqrt{\phi^\top \Sigma_S \phi}} \quad (5)$$

où μ_S et Σ_S sont respectivement le vecteur des rendements moyens et la matrice de covariance des rendements des actifs sources. Les mêmes contraintes sur ϕ s'appliquent.

3.4 Étape 3 : Optimisation régularisée avec transfer learning

Dans cette dernière étape, nous combinons les informations des rendements sources et cibles pour optimiser un portefeuille sur les actifs cibles. Nous introduisons un terme de régularisation qui encourage les poids optimaux ϕ_T à être proches des poids sources ϕ_S :

$$\phi_T = \underset{\phi \in X}{\operatorname{argmax}} \frac{\mu_T^\top \phi}{\sqrt{\phi^\top \Sigma_T \phi}} - \lambda \|\phi_S - \phi\|_2^2 \quad (6)$$

où :

- λ est un hyperparamètre contrôlant la force de la régularisation.
- $\|\phi - \phi_S\|_2^2$ est la distance quadratique entre les poids optimaux cibles et sources.

Les contraintes sur ϕ restent identiques : somme des poids égale à 1 et positivité des poids.

Risque de transfert

Pour évaluer le risque associé à l'apprentissage par transfert dans le problème d'optimisation de portefeuille, nous définissons le risque de transfert, noté R_{trans} . Ce risque prend en compte deux aspects essentiels : la qualité et la pertinence du portefeuille source sélectionné.

3.5 Le Risque de Transfert en pratique

Le risque de transfert est défini comme la somme de deux composantes :

$$R_{\text{trans}} = R_1 + R_2. \quad (7)$$

3.5.1 Première composante : Qualité du Portefeuille Source (R_1)

La première composante R_1 est liée à la performance du portefeuille source et est donnée par :

$$R_1 = \left(\frac{\mu_S^\top \hat{\phi}_S}{\sqrt{\hat{\phi}_S^\top \Sigma_S \hat{\phi}_S}} \right)^{-1}. \quad (8)$$

Cette expression est inversement proportionnelle au ratio de Sharpe de la tâche source. Elle représente donc un indicateur de risque lié au choix d'une tâche source avec une faible performance de portefeuille.

3.5.2 Deuxième composante : Similarité des Portefeuilles (R_2)

La deuxième composante R_2 mesure la similarité entre les portefeuilles source et cible en comparant leurs distributions de rendements respectives.

Nous approchons les distributions de rendements des tâches source et cible par leurs moyennes et covariances respectives (μ_S, Σ_S) et (μ_T, Σ_T) . Ces distributions sont modélisées comme des distributions normales multivariées $\mathcal{N}(\mu_S, \Sigma_S)$ et $\mathcal{N}(\mu_T, \Sigma_T)$.

Ainsi, la mesure de dissimilarité entre ces distributions est donnée par la distance de Wasserstein-2 :

$$R_2 = W_2(\mathcal{N}(\mu_S, \Sigma_S), \mathcal{N}(\mu_T, \Sigma_T)). \quad (9)$$

Cette mesure quantifie à quel point les distributions des portefeuilles source et cible diffèrent en termes de moyennes et de covariances.

3.5.3 Corrélation entre Risque et Sharpe Ratio

En réitérant la tâche d'optimisation du portefeuille cible avec régularisation un grand nombre de fois, on peut ainsi obtenir une corrélation entre le risque de transfert et le Sharpe Ratio lié au transfert. En effet, à chaque itération, 10 assets sources sont sélectionnés aléatoirement et les scores de risque de transfert ainsi que de sharpe ratio lié au transfert sont stockés. Cela permet d'établir une corrélation entre ces deux grandeurs

à partir des nombreux résultats d'expérience. Les résultats obtenus permettent d'analyser l'impact du risque de transfert R_{trans} sur la performance des portefeuilles transférés. En particulier, nous étudions la corrélation entre R_{trans} et le ratio de Sharpe du portefeuille cible dans plusieurs marchés.

Une corrélation négative significative entre le risque de transfert et le ratio de Sharpe du portefeuille cible signifie que plus le risque de transfert est élevé, plus la performance du portefeuille transféré est faible.

Cela est cohérent avec la définition de R_1 , qui est inversement proportionnel au ratio de Sharpe du portefeuille source. En d'autres termes, si le portefeuille source a une faible performance (faible ratio de Sharpe), alors le risque de transfert est élevé et le portefeuille cible a également plus de chances d'afficher une performance médiocre après le transfert. De plus, la similarité des distributions de rendements (R_2) joue un rôle crucial : si le portefeuille source est trop différent du portefeuille cible en termes de caractéristiques statistiques, le transfert sera moins efficace.

4 Méthodologie et Expériences

Dans cette section, nous détaillons la méthodologie utilisée et l'algorithme que nous avons mis en place pour appliquer le **Transfer Learning (TL)** à l'optimisation de portefeuilles financiers, en nous concentrant sur le transfert intercontinental. Nous expliquons également comment les données ont été préparées, comment le code a été implémenté, et les raisons derrière les choix méthodologiques. Chaque sous-section présente une démarche claire, en s'appuyant sur les équations déjà définies précédemment.

4.1 Transfert Intercontinental

4.1.1 Définition et Objectif

Le transfert intercontinental consiste à transférer un portefeuille optimisé sur un marché boursier d'un pays (source) vers un marché boursier d'un autre pays (cible). Par exemple, transférer un portefeuille optimisé sur le marché américain (S&P 500) vers des marchés internationaux comme le Royaume-Uni, le Brésil, l'Allemagne ou Singapour. L'objectif est de tirer parti des données abondantes et diversifiées disponibles sur un marché mature (comme le marché américain) pour améliorer les performances d'un portefeuille sur un marché émergent ou moins développé, où les données peuvent être limitées ou bruyantes. Pour mettre en oeuvre l'algorithme d'optimisation, il a fallu tout d'abord trouver les données utilisées dans l'étude. Pour cela, nous avons fait le choix d'utiliser une API de données financières et en particulier celle du site Yahoofinance. Cela nous permet de tirer les cours des marchés financiers sans télécharger les données localement. L'API se connecte directement au site YahooFinance et en extrait les données demandées.

La démarche suivie dans l'article repose sur plusieurs étapes clés, en particulier dans la préparation et l'utilisation des données.

- **Sélection des Marchés et des Actions** : Le marché source est défini comme le marché américain (S&P 500), tandis que les marchés cibles sont choisis parmi plusieurs pays, tels que le Royaume-Uni, le Brésil, l'Allemagne et Singapour. Pour chaque marché cible, les 10 actions avec les plus grandes capitalisations boursières sont sélectionnées comme actifs cibles [15] . Pour le marché source (États-Unis), 10 actions sont sélectionnées aléatoirement parmi les composantes du S&P 500.
- **Préparation des Données** :
 - Les données sont divisées en trois ensembles pour chaque marché cible :

- **Données d'entraînement source** : Retours journaliers des 10 actions sélectionnées sur le marché américain (source), sur une période donnée, par exemple de février 2000 à février 2020.
- **Données d'entraînement cible** : Retours journaliers des 10 actions du marché cible, sur une période plus récente, par exemple de février 2015 à février 2020.
- **Données de test cible** : Retours journaliers des 10 actions du marché cible, sur une période de test, par exemple de février 2020 à septembre 2021.

```
import pandas as pd
import numpy as np
import yfinance as yf
from scipy.optimize import minimize
from scipy.stats import pearsonr
import random

# Global parameters
risk_free_rate = 0.05 / 252 # Daily risk-free rate (5% annual)
source_start_date = "2000-02-01"
source_end_date = "2020-02-01"
target_start_date = "2015-02-01"
target_end_date = "2020-02-01"
test_start_date = "2020-02-01"
test_end_date = "2021-09-01"
lambda_reg = 0.2 # Regularization parameter

# Fixed target assets (e.g., top 10 from FTSE 100 for UK)
target_assets = ['AZN.L', 'SHEL.L', 'HSBA.L', 'ULVR.L', 'BHP.L', 'REL.L', 'BATS.L', 'BP.L', 'LSEG.L', 'RR.L']

# Function to get S&P 500 tickers
def get_sp500_tickers():
    url = "https://en.wikipedia.org/wiki/List_of_S%26P_500_companies"
    tables = pd.read_html(url)
    sp500_df = tables[0]
    return sp500_df["Symbol"].tolist()

# Function to randomly select n tickers
def select_random_tickers(n=10):
    tickers = get_sp500_tickers()
    if not tickers:
        return "Impossible to retrieve tickers."
    return random.sample(tickers, n)
```

FIGURE 2 – Importation et préparation des données

Les 'tickers' correspondent aux codes des entreprises dans un marché donné. Par exemple, le ticker de Shell est SHEL.L ('.L' montrant son appartenance au FTSE 100). Pour chaque pays, les target assets changent : ils représentent les 10 premières entreprises en termes de capitalisation boursière d'un pays. Nous avons choisi ces entreprises selon les classements actuels.

4.1.2 Méthodologie

La méthodologie pour le transfert intercontinental repose sur trois étapes clés :

1. **Optimisation Directe (Cible)** : Nous commençons par optimiser directement le portefeuille cible en maximisant le ratio de Sharpe, comme décrit dans l'équation (4). Cette étape sert de référence pour comparer les performances du transfert.

2. **Optimisation avec Transfert (Source)** : Ensuite, nous optimisons le portefeuille source en utilisant l'équation (5), qui maximise le ratio de Sharpe sur les données d'entraînement source.

```
# Function to optimize the source portfolio (phi_S)
def optimize_source_portfolio(mu_S, Sigma_S):
    def sharpe_ratio_source(phi, mu_S, Sigma_S):
        numerator = np.dot(mu_S, phi)
        denominator = np.sqrt(np.dot(phi.T, np.dot(Sigma_S, phi)))
        return -numerator / denominator # Negative because we minimize

    result_source = minimize(sharpe_ratio_source, np.ones(len(mu_S)) / len(mu_S), args=(mu_S, Sigma_S),
                             method='SLSQP', bounds=[(0, 1) for _ in range(len(mu_S))],
                             constraints=[{'type': 'eq', 'fun': lambda phi: np.sum(phi) - 1}])
    return result_source.x
```

FIGURE 3 – Optimisation du portefeuille source

3. **Optimisation avec Régularisation (Cible)** : Enfin, nous transférons les connaissances du portefeuille source vers le portefeuille cible en utilisant l'équation (6), qui inclut un terme de régularisation pour rapprocher les poids du portefeuille cible de ceux du portefeuille source.

```
# Function to optimize the target portfolio with regularization (phi_T)
def optimize_target_portfolio(mu_T, Sigma_T, phi_S, lambda_reg):
    def regularized_sharpe_ratio(phi, mu_T, Sigma_T, phi_S, lambda_reg):
        numerator = np.dot(mu_T, phi)
        denominator = np.sqrt(np.dot(phi.T, np.dot(Sigma_T, phi)))
        sharpe_ratio = numerator / denominator
        regularization = lambda_reg * np.linalg.norm(phi_S - phi) ** 2 # Regularization term
        return -(sharpe_ratio - regularization) # Negative because we minimize

    result_target = minimize(regularized_sharpe_ratio, np.ones(len(mu_T)) / len(mu_T),
                             args=(mu_T, Sigma_T, phi_S, lambda_reg), method='SLSQP',
                             bounds=[(0, 1) for _ in range(len(mu_T))],
                             constraints=[{'type': 'eq', 'fun': lambda phi: np.sum(phi) - 1}])
    return result_target.x, -result_target.fun
```

FIGURE 4 – Optimisation du portefeuille cible avec régularisation

Pour calculer le risque de transfert, on utilise la fonction ‘computetransferrisk’, qui reprend les équations (7), (8) et (9).

```

# Function to compute transfer risk
# Function to compute R1 (Transfer Risk)
def compute_R1(phi_T, mu_T, Sigma_T):
    numerator = np.dot(mu_T, phi_T)
    denominator = np.sqrt(np.dot(phi_T.T, np.dot(Sigma_T, phi_T)))
    sharpe_ratio_T = numerator / denominator
    R1 = 1.0 / sharpe_ratio_T
    return R1

def compute_transfer_risk(phi_T, source_data, target_test_data):
    mu_S = np.array(source_data.mean())
    Sigma_S = np.array(source_data.cov())
    mu_T_test = np.array(target_test_data.mean())
    Sigma_T_test = np.array(target_test_data.cov())

    # Compute R1 (inverse of source portfolio Sharpe ratio)
    R1 = compute_R1(phi_T, mu_S, Sigma_S)

    # Compute R2 (Wasserstein-2 distance between source and target distributions)
    R2 = np.linalg.norm(mu_S - mu_T_test) + np.linalg.norm(Sigma_S - Sigma_T_test, ord=2)
    return R1 + R2

```

FIGURE 5 – Implémentation du Risque de Transfert

Afin d’observer la corrélation entre le Sharpe Ratio et le risque de transfert, la tâche d’optimisation est répétée 1000 fois, avec pour chaque itération, un set de tickers sources différents (il est chargé aléatoirement, comme montré dans la fonction ‘selectrandomtickers’ de la figure 2).

4.1.3 Résultats et Analyse

Pays	Corrélation	p-value	Interprétation
Singapour	0.073	0.0426	Corrélation faiblement positive mais statistiquement significative (p inférieur à 0.05).
Allemagne	-0.042	0.247	Corrélation négligeable et non significative (p supérieur à 0.05).
Brésil	-0.0247	0.495	Aucune corrélation détectée (p-value très élevée).
Royaume-Uni	0.0129	0.718	Corrélation quasi-nulle et non significative.

TABLE 1 – Corrélations entre le Ratio de Sharpe (Transfert) et le Risque de Transfert

Tendance Générale : Aucune corrélation significative n’est observée dans nos résultats, contrairement aux fortes corrélations négatives rapportées dans l’article (-0.60 à -0.67). Cela suggère que le risque de transfert n’a pas eu d’impact significatif sur la performance du portefeuille transféré dans ces expériences. Cette faible corrélation pourrait être due à des facteurs tels que la sélection aléatoire des actions ou la valeur du paramètre de régularisation λ .

4.1.4 Difficultés Rencontrées

Lors de l'implémentation, nous avons rencontré plusieurs difficultés :

- **Manque d'informations sur les données utilisées** : Les études de Cao et Al. [1] [2] ne fournissent pas suffisamment d'informations sur les données utilisées, ce qui a rendu la réplique des expériences difficile. Par exemple, les périodes exactes des données et les méthodes de sélection des actions ne sont pas clairement spécifiées. Par exemple, les entreprises des marchés cibles sont définies comme étant le top 10 des entreprises en termes de capitalisation boursière mais elles ne sont jamais nommées précisément. De plus, le taux sans risque que nous avons utilisé est arbitraire et classique, mais ce taux peut avoir des conséquences sur nos résultats. Le paramètre λ dans le terme de régularisation est fixé à 0.2 dans l'article étudié mais il semble nécessaire de l'adapter selon nos résultats. Enfin, comme la méthode d'optimisation que nous avons utilisé pour notre algorithme est SLSQP (Sequential Least Squares Programming) mais nous aurions pu utiliser d'autres méthodes classiques plus adaptées.
- **Temps de calcul** : L'optimisation de portefeuilles, en particulier avec la régularisation, demande un temps de calcul important (une heure par pays), surtout lorsque l'on travaille avec des données intrajournalières ou des portefeuilles de grande taille. C'est pourquoi il est difficile de tester différents paramètres et méthodes.
- **Ajout de NaN dans les Données Cibles** : Nous avons essayé d'ajouter des valeurs manquantes (NaN) dans les données cibles pour simuler des conditions de marché plus réalistes. Cependant, cela n'a pas fonctionné car les algorithmes d'optimisation ne gèrent pas bien les valeurs manquantes, ce qui a entraîné des erreurs de convergence et des résultats incohérents.

L'article "Transfer Learning for Portfolio Optimization" [2] explore, en plus du transfert intercontinental, le transfert intersectoriel et inter-fréquences. La suite de notre rapport se base sur les résultats de l'article afin de comprendre comment l'algorithme d'optimisation peut être utilisé pour ces deux autres tâches.

4.2 Transfert Intersectoriel

4.2.1 Définition et Objectif

Le transfert intersectoriel consiste à transférer un portefeuille optimisé dans un secteur d'activité (source) vers un autre secteur (cible) au sein du même marché. Par exemple, transférer un portefeuille du secteur des technologies de l'information vers le secteur de la santé sur le marché américain. L'objectif est de comprendre comment les corrélations entre différents secteurs d'un marché peuvent influencer l'efficacité du transfert de connaissances. Ici, 10 secteurs différents sont considérés, tels que le secteur de la santé, les services de communication, l'énergie, l'industrie, les technologies de l'information, les services financiers ou encore l'immobilier. Pour chaque secteur, 10 actions sont sélectionnées aléatoirement parmi les composantes du S&P500 ou d'autres indices boursiers. Des paires de secteurs source-cible sont créées, par exemple, Technologie de l'information \rightarrow Santé ou Energie \rightarrow Services financiers, ce qui donne un total de $10 \times 10 = 100$ paires.

- **Préparation des Données** :
 - Pour chaque paire source-cible, les données sont divisées en trois ensembles :
 - **Données d'entraînement source** : Retours journaliers des 10 actions du secteur source sur une période donnée, par exemple de février 2000 à février 2020.
 - **Données d'entraînement cible** : Retours journaliers des 10 actions du secteur cible sur une période plus récente, par exemple de février 2015 à février 2020.

- **Données de test cible** : Retours journaliers des 10 actions du secteur cible sur une période de test, par exemple de février 2020 à septembre 2021.

4.2.2 Méthodologie

La méthodologie pour le transfert intersectoriel suit la même structure que celle du transfert intercontinental, mais avec un focus sur les différences sectorielles.

Comme pour le transfert intercontinental, les équations (4), (5) et (6) sont utilisées pour optimiser les portefeuilles source et cible, en tenant compte des caractéristiques spécifiques à chaque secteur :

- **Optimisation du Portefeuille Source** :
 - Pour chaque secteur source, le portefeuille est optimisé en maximisant le ratio de Sharpe sur les données d'entraînement source, en utilisant l'équation (5).
- **Optimisation du Portefeuille Cible avec Transfert** :
 - Pour chaque secteur cible, les connaissances du portefeuille source sont transférées en utilisant une régularisation L_2 pour rapprocher les poids du portefeuille cible de ceux du portefeuille source. Cette optimisation est réalisée en résolvant l'équation (6).
- **Évaluation des Performances** :
 - Pour chaque paire source-cible, la performance du portefeuille transféré est évaluée en calculant le ratio de Sharpe sur les données de test cible. Cette procédure est répétée 500 fois pour chaque paire source-cible, en sélectionnant aléatoirement les actions à chaque itération, afin d'obtenir des résultats robustes.
- **Calcul du Risque de Transfert** :
 - Pour chaque paire source-cible, le risque de transfert R_{trans} est calculé en utilisant les équations (7), (8) et (9).

4.2.3 Résultats et Analyse

Les expériences ont montré que certains secteurs, comme celui de la santé ou des technologies de l'information, présentent une forte corrélation négative entre le risque de transfert et le ratio de Sharpe. En revanche, pour des secteurs comme celui de l'immobilier ou les services aux collectivités, la corrélation est moins significative, probablement en raison de facteurs spécifiques à ces secteurs (comme les régulations ou les cycles économiques).

4.3 Transfert Inter-Fréquences

4.3.1 Définition et Objectif

Le transfert inter-fréquences consiste à transférer un portefeuille optimisé à une fréquence de trading (source) vers une autre fréquence (cible). Par exemple, transférer un portefeuille optimisé pour des transactions quotidiennes (1-day) vers des transactions intrajournalières (1-minute, 5-minute, etc.). L'objectif est d'explorer la possibilité de transférer des stratégies d'investissement entre différentes fréquences de trading.

- **Sélection des Fréquences** :
 - Plusieurs fréquences de trading sont considérées, allant de la basse fréquence (1 jour) à la haute fréquence (1 minute, 5 minutes, etc.).

- Des paires de fréquences source-cible sont créées, par exemple, 1 jour \rightarrow 130 minutes ou 65 minutes \rightarrow 10 minutes, ce qui donne un total de $7 \times 7 = 49$ paires.
- **Préparation des Données :**
 - Pour chaque paire source-cible, 10 actions sont sélectionnées parmi les 50 plus grandes entreprises américaines par capitalisation boursière.
 - Les données sont divisées en trois ensembles :
 - **Données d’entraînement source :** Retours des 10 actions à la fréquence source sur une période donnée, par exemple de février 2016 à septembre 2019.
 - **Données d’entraînement cible :** Retours des 10 actions à la fréquence cible sur la même période.
 - **Données de test cible :** Retours des 10 actions à la fréquence cible sur une période de test, par exemple de septembre 2019 à février 2020.

4.3.2 Méthodologie

La méthodologie pour le transfert inter-fréquences repose sur l’adaptation des portefeuilles à différentes fréquences de trading. Les équations (1), (2) et (3) sont utilisées pour optimiser les portefeuilles source et cible, en tenant compte des différences de volatilité et de dépendances temporelles entre les fréquences.

- **Optimisation du Portefeuille Source :**
 - Pour chaque fréquence source, le portefeuille est optimisé en maximisant le ratio de Sharpe sur les données d’entraînement source, en utilisant l’équation (2).
- **Optimisation du Portefeuille Cible avec Transfert :**
 - Pour chaque fréquence cible, les connaissances du portefeuille source sont transférées en utilisant une régularisation L_2 pour rapprocher les poids du portefeuille cible de ceux du portefeuille source. Cette optimisation est réalisée en résolvant l’équation (3).
- **Évaluation des Performances :**
 - Pour chaque paire source-cible, la performance du portefeuille transféré est évaluée en calculant le ratio de Sharpe sur les données de test cible.
 - Cette procédure est répétée 200 fois pour chaque paire source-cible, en sélectionnant aléatoirement les actions à chaque itération, afin d’obtenir des résultats robustes.
- **Calcul du Risque de Transfert :**
 - Pour chaque paire source-cible, le risque de transfert R_{trans} est calculé en utilisant les équations (4), (5) et (6).

4.3.3 Résultats et Analyse

Les résultats de l’étude montrent que le transfert d’un portefeuille basse fréquence (1-day) vers des fréquences plus élevées (intrajournalières) entraîne un risque de transfert élevé et une performance médiocre. En revanche, le transfert entre fréquences intrajournalières (par exemple, de 65-minute à 130-minute) donne des résultats plus robustes, en raison de la similarité des propriétés statistiques.

Ces deux expérimentations pratiques explorées par les auteurs permettent de mettre en avant l’utilité de l’algorithme d’optimisation lorsqu’il est question d’étudier l’impact du transfer learning sur des données financières.

5 Conclusion

Cette étude met en évidence les enjeux théoriques et pratiques du Transfer Learning appliqué à l'optimisation de portefeuilles financiers. À travers l'analyse rigoureuse du risque de transfert, nous avons démontré que la performance du TL dépend non seulement de la qualité du portefeuille source, mais aussi de la similarité statistique entre les domaines source et cible. Théoriquement, nous avons démontré que le risque de transfert constitue une borne inférieure du regret entre l'optimisation directe sur le marché cible et l'optimisation basée sur un transfert depuis le marché source. Cette relation suggère que plus le risque de transfert est élevé, plus la performance du portefeuille transféré devrait être dégradée.

Nos expériences empiriques, bien que conformes aux intuitions théoriques, révèlent une corrélation plus faible que prévue entre le risque de transfert et la performance des portefeuilles transposés. Cette divergence suggère que des paramètres non pris en compte, tels que les effets de sélection des actifs, la méthode d'optimisation et certains paramètres d'optimisation (terme de régularisation) jouent un rôle crucial. De nombreuses contraintes en termes de données (manque d'informations précises dans l'étude) ainsi que de temps de calcul (grosse tâche d'optimisation) causent des difficultés pour avoir des résultats probants. Néanmoins, l'exploration de cette étude sur le Transfer Learning nous a permis de mieux comprendre les mécanismes sous-jacents du transfert de connaissances entre marchés financiers, tout en soulignant les défis pratiques et les perspectives d'amélioration pour des applications futures plus robustes et efficaces.

Références

- [1] H. Cao, H. Gu, X. Guo, and M. Rosenbaum, *Risk of Transfer Learning and its Applications in Finance*, 2023.
- [2] H. Cao, H. Gu, X. Guo, and M. Rosenbaum, *Transfer Learning for Portfolio Optimization*, 2023.
- [3] Z. Zhang, et al., *Ensemble Learning for Stock Trend Prediction with Multi-source Data*, 2020.
- [4] Y. Leal, et al., *Transfer Learning for High-Frequency Trading Strategies*, 2022.
- [5] D. Wu, et al., *Cross-Sector Transfer Learning for Stock Market Prediction*, 2022.
- [6] Y. Liu, et al., *FinBERT : A Pre-trained Financial Language Representation Model for Financial Text Mining*, 2020.
- [7] D. Weiss, et al., *A Systematic Review of Transfer Learning in Machine Learning*, 2016.
- [8] J. Wang, et al., *Negative Transfer in Transfer Learning : A Survey*, 2019.
- [9] Y. Jiang, et al., *Time Series Transfer Learning for Financial Data*, 2020.
- [10] A. B. Arrieta, et al., *Explainable Artificial Intelligence (XAI) : Concepts, Taxonomies, Opportunities and Threats*, 2020.
- [11] D. Araci, *FinBERT : A Pretrained Language Model for Financial Communications*, 2019.
- [12] J. Xu and W. W. Cohen, *Modeling the Impact of Social Media on Financial Markets*, 2018.
- [13] J. Ye, et al., *Dynamic Adaptation in Financial Time Series Forecasting*, 2020.
- [14] D. Chafaï, *Wasserstein Distance Between Two Gaussians*, 2010.
- [15] TradingView, *Bourse du Brésil – Actions à grande capitalisation les plus actives*, 2025. Disponible à : <https://fr.tradingview.com/markets/stocks-brazil/market-movers-large-cap/>.