

# L'ORIENTATION VERS LES COÛTS

---

Laurent Gille, Télécom ParisTech, 2012

## SOMMAIRE

1	Origine de l'orientation vers les coûts .....	2
1.1	Les ressources essentielles .....	3
1.2	Le dilemme de la régulation .....	5
1.3	Les options tarifaires .....	7
2	La notion de coût .....	8
2.1	Les coûts d'une firme monoproduit .....	8
2.2	Les coûts d'une firme multiproduits .....	9
2.3	Le coût incrémental .....	11
2.3.1	Le partage des coûts: approche stratégique .....	12
2.3.2	Les coûts pertinents: approche pragmatique .....	13
2.3.3	Les CMILT: approche technique .....	16
2.4	Le coût du capital .....	19
3	La mise en œuvre de l'orientation vers les coûts .....	22
3.1	Approche statique de l'orientation vers les coûts .....	23
3.2	Approche dynamique de l'orientation vers les coûts .....	25
4	Conclusions .....	26
5	Bibliographie .....	28

La plupart des réglementations inscrivent dans certaines de leurs dispositions une obligation tarifaire d'orientation vers les coûts. C'est notamment le cas en ce qui concerne la tarification des services offerts sur les marchés de gros (par exemple les tarifs d'interconnexion). Ce chapitre a trois objectifs: d'une part, comprendre d'où provient cette injonction réglementaire, et saisir la relation qui lie les coûts aux prix, d'autre part, appréhender la façon dont elle est mise en œuvre pratiquement et enfin, voir les incidences de ces pratiques.

## 1 ORIGINE DE L'ORIENTATION VERS LES COÛTS

Le cadre référentiel de la régulation économique des activités d'infrastructure est formé par la théorie néoclassique des marchés concurrentiels. Un marché est dit concurrentiel lorsqu'aucun acteur de ce marché ne peut influencer les prix des produits qui y sont écoulés. Sur un tel marché, prix et quantités se forment dans l'interaction des fournisseurs et des acheteurs et s'établissent à un niveau qui équilibre l'offre et la demande, sans qu'aucun agent de ce marché ne puisse influencer leur formation. On dit alors que les agents sont "preneurs" de prix: ils doivent fournir le marché en prenant le prix auquel se conforme l'ensemble des autres agents.

Nous n'examinerons pas ici les conditions sous lesquelles cet équilibre peut être raisonnablement atteint<sup>1</sup>, ni les mécanismes à travers lesquels il peut survenir<sup>2</sup>. On admettra ici l'idée qu'un nombre d'agents importants, tant du côté des fournisseurs que des acheteurs, garantit l'obtention d'un tel marché concurrentiel. Plus le nombre d'agents diminue, plus surgit la crainte qu'un agent ou un petit nombre d'entre eux, agissant de façon collusive, puissent manipuler les prix. Dans le cas limite, qui est celui d'un monopole de production, il est clair que l'agent en situation monopolistique peut établir les prix au niveau qui maximise son profit, et non au niveau du prix du marché concurrentiel. Dans un tel cas, le prix est supérieur au prix du marché concurrentiel (en fonction de l'élasticité de la demande), et la quantité offerte est inférieure à la quantité qui le serait sur un marché concurrentiel: on est alors d'autant plus éloigné de l'équilibre de marché concurrentiel que l'élasticité de la demande est faible: dans ces cas où les agents sont "faiseurs" de prix et non plus "preneurs" de prix, la régulation a pour objectif de tenter de remédier à cette situation en les obligeant à s'approcher d'un prix concurrentiel: la régulation est alors perçue comme un remède appliqué à un marché malade.

Lorsqu'on s'éloigne d'un marché concurrentiel, les agents économiques se trouvent donc en situation de dominance et ont la possibilité d'écouler leur production à un prix éloigné du prix de marché concurrentiel. Ils s'approprient dès lors une rente au détriment des consommateurs. La régulation vise à minorer cette rente tout en recherchant également une évolution dynamique favorable des paramètres de l'offre (prix, qualité, innovation...). Nous analyserons dans ce qui suit le dilemme de la régulation face à ce double objectif, puis nous considérerons le cas particulier des infrastructures qui supportent des services.

Prix et coûts peuvent être fixés de façon autonome. En effet, l'établissement des prix de détail par une firme peut prendre en compte les fréquentes disparités de solvabilité des consommateurs et favoriser les consommateurs les moins solvables (et les plus élastiques au prix) en leur proposant des prix inférieurs à ceux proposés aux consommateurs plus solvables. Cette discrimination des prix, largement pratiquée à travers la segmentation des produits et des clientèles, permet d'accroître le volume de production et le bien être des consommateurs. Par ailleurs, la satisfaction (l'utilité) que retire un consommateur de l'usage d'un produit peut être affectée par les consommations opérées par les autres

<sup>1</sup> Cf. Salanié B. (1998) Microéconomie, les défaillances du marché, Economica.

<sup>2</sup> Symbolisés par la main invisible d'Adam Smith, ces mécanismes, sur lesquels buttent les économistes, sont loin d'être triviaux.

consommateurs: dans une communication téléphonique, l'appel est payé par l'émetteur (principe de l'appelant payeur<sup>3</sup>), mais son correspondant est fréquemment soit gagnant (s'il tire profit de la communication) soit perdant (s'il subit une communication non désirée). Les utilités sont interdépendantes. Elles le sont également par exemple sur un réseau de transport du fait des phénomènes de congestion: le conducteur est gagnant s'il est le seul à fréquenter le réseau, perdant si beaucoup l'utilisent. Quand des externalités de consommation existent, il est bénéfique que les prix s'écartent des coûts du montant des incidences économiques de ces externalités<sup>4</sup>. Le rapport des prix aux coûts est donc complexe, et il n'est aucunement optimal, dans de nombreuses situations, de contraindre une orientation des prix vers les coûts, en tout cas sur le marché de détail: sur ces marchés, il est fréquemment souhaitable que les prix ne reflètent pas les coûts.

La question se pose différemment sur le marché de gros, c'est-à-dire sur le marché sur lequel les firmes d'un secteur s'échangent des prestations. On peut supposer que sur ces marchés les firmes sont totalement inélastiques aux prix: si elles doivent y acheter des prestations, c'est qu'elles ne peuvent pas faire autrement, et donc, que les situations de dominance sont particulièrement dommageables.

Dans cette première partie, nous supposerons que la notion de coût est clairement définie et que les coûts sont connus. Nous verrons dans la deuxième partie que le concept même de coût doit être précisé.

## 1.1 LES RESSOURCES ESSENTIELLES

La doctrine des ressources essentielles a un siècle: établie en 1912 aux Etats-Unis sur la seule base juridique du Sherman Act (1890), elle stipule que lorsqu'un acteur détient une ressource dont l'accès est essentiel pour développer une concurrence sur un marché aval, alors l'accès à cette ressource doit être rendu possible (sauf impossibilités techniques), et doit se faire de façon non discriminatoire et équitable, ce qui s'est traduit par un accès ouvert et à un prix orienté vers les coûts.

Cette doctrine, élaborée par la jurisprudence (ce qui traduit la difficulté à la formaliser sur des critères objectifs), initiée sur les réseaux ferroviaires, puis confortée sur de nombreux autres domaines, a concerné le champ des télécommunications lorsque les réseaux se sont ouverts à la concurrence et a franchi l'atlantique à la fin des années 70 pour s'introduire dans les textes européens. La ressource essentielle<sup>5</sup> se caractérise avant tout par la difficulté ou l'impossibilité qu'il y a pour un acteur à la dupliquer, que ce soit sur un plan technique ou économique, et par la nécessité d'y accéder pour pouvoir assurer des prestations sur un marché aval de cette ressource. Une compagnie maritime ou aérienne doit pouvoir accéder à des installations portuaires ou aéroportuaires, un fournisseur de services de télécommunications doit pouvoir accéder aux réseaux de ses concurrents pour acheminer les appels de ses propres abonnés.

Cette doctrine a été théorisée sur le plan économique dans les réflexions sur la forclusion (accaparement, préemption)<sup>6</sup>, et notamment les situations de forclusion verticale par intégration verticale ou accords d'exclusivité. Il peut y avoir également forclusion horizontale quand un agent

<sup>3</sup> CPP en anglais: Caller Pays Principle ou Calling Party Pays Principle.

<sup>4</sup> Il manque en effet un prix, qui serait celui (positif ou négatif) payé par l'utilisateur bénéficiaire ou l'utilisateur concurrent de la consommation. En d'autres termes, la question est de savoir qui doit "couvrir" le coût d'un produit: celui qui le paye, uniquement, ou ceux qui en bénéficient (en bien ou en mal)?

<sup>5</sup> Parfois qualifiée de goulot d'étranglement (bottleneck).

<sup>6</sup> On parle en anglais de foreclosure, terme dont la traduction littérale est la saisie. L'idée de foreclosure réside dans la capacité d'un acteur à préempter, à accaparer ou se saisir d'un marché au détriment de concurrents potentiels en mettant en œuvre différents stratagèmes.

économique contrôle un bien indispensable à l'usage d'un autre bien (dans le cadre par exemple de biens systèmes où deux biens sont étroitement liés, cas par exemple des applications associées à un système d'exploitation) (pour une analyse complète de ces situations, voir Rey et Tirole (2007)). Une situation de forclusion se produit donc quand une firme domine un marché et utilise ce pouvoir de marché pour réduire la production d'un autre marché de façon à y exercer également une dominance rémunératrice. Cette influence collatérale sur un autre marché peut prendre différentes formes, dont des restrictions à l'entrée (forclusion verticale) ou des tentatives d'éviction (forclusion horizontale).

Dans les situations de forclusion verticale, la firme qui détient la ressource essentielle peut s'intégrer verticalement sur le marché aval et exercer son pouvoir de monopole sur les marchés amont et aval. L'école de Chicago a soutenu que l'exercice du monopole sur la ressource essentielle suffisait à extraire la rente sur cette ressource et n'impliquait aucune stratégie d'intégration verticale. Cette vision ne tient pas compte du fait que le monopoleur ne peut totalement exercer son pouvoir de marché s'il ne pratique pas de stratégies d'exclusion en aval. En effet, la rente qu'il peut tirer de sa ressource essentielle suppose une certaine inélasticité du marché aval, qui est lié au fait que l'accès à cette ressource est contraint. Si tout le monde peut y accéder<sup>7</sup>, la demande aval devient très élastique et l'extraction de la rente difficile, ce qui est bien illustré par le cas des brevets (un détenteur de brevet n'exploite véritablement son brevet que s'il a la capacité d'en limiter les licenciés et de les choisir). Le détenteur d'une ressource essentielle a donc une véritable incitation à l'intégration verticale, à la discrimination et l'exclusion d'accès. On montre également que les situations ne sont pas les mêmes si la ressource est en amont du marché (la partie concurrentielle du marché se situe entre la ressource et le marché final) ou si la ressource est en aval du marché (en relation directe avec le marché final tandis que le marché concurrentiel se situe en amont)<sup>8</sup>.

Face à une situation d'exclusion du marché par intégration verticale, plusieurs remèdes sont disponibles. Le premier, radical, consiste à exiger le démantèlement de l'agent intégré avec un accès non discriminatoire, mesure très lourde d'application, et qui n'exclut aucunement le comportement monopoliste du détenteur de la ressource<sup>9</sup>. Une autre option est la gestion coopérative de la ressource essentielle, là encore avec un accès non discriminatoire et une exemption antitrust<sup>10</sup>. Enfin, il est possible de "renverser" le marché en mettant une ressource aval en amont ou réciproquement (en redessinant les relations commerciales possibles<sup>11</sup>).

Mais la plupart des remèdes tournent autour de la non discrimination de l'accès (obligation d'ouverture de la ressource de façon non discriminatoire) couplée à une orientation vers les coûts de la tarification de cet accès. Cette orientation vers les coûts peut se faire selon différentes règles, qui n'ont pas les mêmes incidences.

L'obligation d'interconnexion est typiquement une obligation d'accès à une ressource essentielle. Tout opérateur de réseau est en effet en situation de monopole (dans la plupart des cas) pour acheminer des appels aux usagers qu'il raccorde: il doit donner accès à son réseau aux autres opérateurs ou prestataires de façon à ce qu'ils puissent collecter ou acheminer des appels auprès de ses abonnés, que ce soit de façon temporaire (terminaison d'appel) ou de façon permanente (mise à disposition de tout ou partie de la boucle locale).

---

<sup>7</sup> De plus, le détenteur de la ressource a une forte incitation à fournir de façon à élargir ses revenus, mais ce faisant réduit le profit qu'il peut tirer de cette ressource (dilemme de l'engagement à fournir dénommé en anglais *commitment problem*).

<sup>8</sup> Cas par exemple des réseaux où le réseau d'accès porte la relation commerciale avec les abonnés (câblodistribution...)

<sup>9</sup> Comportement monopoliste que l'on retrouve fréquemment chez les gestionnaires d'infrastructures portuaires (ports, aéroports, gares...), ce qui en rend la privatisation aisée.

<sup>10</sup> Ce qui n'exclut pas encore ici le comportement monopoliste du détenteur (exemple de l'industrie spatiale des télécommunications pendant des décennies [Intelsat, Inmarsat, Eutelsat, Arabsat, etc.]).

<sup>11</sup> Par exemple, la possibilité offerte à un abonné à un réseau de choisir son opérateur longue distance (carrier selection) ou son fournisseur de services internet forme des illustrations de cette ouverture relationnelle.

Cette ouverture doit donc s'opérer de façon non discriminatoire et équitable, pour limiter le pouvoir de marché du détenteur de l'accès. La non discrimination signifie l'ouverture de la ressource à qui la demande, et sans discrimination entre demandeurs, l'équité réclame l'orientation vers les coûts de façon à ce que le détenteur de la ressource ne puisse disposer d'une rente qui lui permettrait de jouer le prédateur sur le marché aval concurrentiel en cas d'intégration.

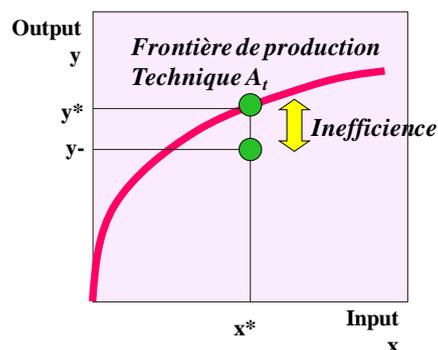
Nous supposons ici les coûts connus. La doctrine des ressources essentielles réclame donc un prix d'accès<sup>12</sup> à la ressource tel que la firme détenant cette ressource ne puisse extraire de son exploitation une rente qui formerait subvention croisée avec d'autres activités, et notamment les activités avec lesquelles elle se trouverait en concurrence avec les acteurs accédant à cette ressource.

## 1.2 LE DILEMME DE LA REGULATION

L'orientation vers les coûts trouve donc son origine dans la nécessité, notamment sur certains marchés de gros, d'éviter qu'un agent en situation dominante, exerce son pouvoir de marché en "faisant" les prix, c'est-à-dire en déconnectant le prix de ses prestations du prix d'un marché concurrentiel. Il n'est pas suffisant de se situer, dans de nombreux contextes, dans un cadre statique pour traiter cette question. Il est nécessaire également d'apprécier la dynamique des marchés, notamment sur les marchés qui bénéficient de progrès technologiques importants, c'est-à-dire où les coûts diminuent d'année en année du fait de ces progrès et où les performances évoluent également très rapidement. L'intérêt de la régulation est certes que les positions dominantes soient les moins fortes possibles, mais également que les coûts (et donc les prix) s'abaissent au rythme du progrès technique, avec amélioration de la qualité et de la performance des produits offerts.

Un rapide détour par la fonction de production s'impose, pour introduire la dynamique des coûts. Une fonction de production relie la quantité maximale de biens (outputs) produits  $y$  en utilisant une quantité donnée d'inputs  $x$ , à partir d'une technologie déterminée  $A_t$ :  $y=f(x)$ . Cette fonction  $f$  représente la frontière de production: une firme qui se situe sur cette frontière sera dite efficiente. Si pour  $x^*$  donné,  $f(x^*)=y^*$ , toute production inférieure à  $y^*$  caractérisera une situation d'inefficience.

Figure 1 - La frontière de production



Un monopole est enclin à tarifier sa production à un prix de monopole, au-dessus du prix d'un marché concurrentiel, mais il est également enclin à être inefficace, c'est-à-dire à ne pas se situer sur la frontière de production.

Si  $p$  est le prix de l'output et  $c$  le coût de l'input, le profit  $\Pi$  de la firme s'exprime comme:

<sup>12</sup> On parlera en anglais à ce propos d'access charge.

$$\Pi = py - c$$

a. Si la firme est en situation concurrentielle (p indépendant de y), elle maximise son profit en prenant un prix p égal au coût marginal c'. Si la firme est en situation de monopole (faiseuse de prix), alors p est une fonction de y tout comme c. La maximisation de  $\Pi$  donne, en dérivant cette expression par rapport

à y et si  $e = -\frac{\frac{dy}{y}}{\frac{dp}{p}}$ : caractérise l'élasticité de la demande au prix:

$\frac{dp}{dy} + p - c' = 0$ , soit,  $p = c' \cdot \frac{e}{e-1}$  ou, si  $L=1/e$ ,  $L = \frac{p-c'}{p}$ , L étant appelé l'indice de Lerner, inverse de l'élasticité. Un prix concurrentiel signifie un indice de Lerner égal à 0. Le prix concurrentiel est donc augmenté d'un facteur  $e/(e-1)$  en situation de dominance ( $e > 1$ ).

b. Un monopole est fréquemment enclin à être inefficace. Non soumis à concurrence, il laisse fréquemment dériver ses coûts, soit en payant plus cher ses inputs, soit en gaspillant ses inputs ou en mettant en œuvre des techniques de production non optimales. Cette inefficace des monopoles, appelée X-inefficace, fut étudiée par Harvey Leibenstein en 1966<sup>13</sup>. Le phénomène est bien connu chez les exploitants de télécommunications en situation de monopole: salaires élevés, effectifs pléthoriques, surinvestissement dû à des normes d'ingénierie disproportionnée aux exigences de qualité, exploitation-maintenance fréquemment déficiente et dispendieuse, laisser-aller du management sur les frais de structure, etc. Non soumis à concurrence, le monopole (ou la firme en situation de dominance) laisse dériver ses coûts et se trouve donc en situation d'inefficace. Une grande partie de la rente de monopole peut être "consommée" dans cette X-inefficace, laissant apparaître des profits modestes, la rente de monopole étant captée par les fournisseurs et les salariés.

Une des raisons de l'ouverture des marchés à la concurrence, y compris pour des monopoles naturels, réside dans cette X-inefficace: en soumettant les firmes à la concurrence, on les incite à supprimer ces sources d'inefficace pour lutter contre une concurrence qui serait, elle, efficace.

Le régulateur a donc deux soucis: réduire l'inefficace (abaisser les coûts) et extraire la rente de monopole. Mais, sur un marché concret, le régulateur ne connaît pas bien les coûts efficaces de production. Ceux-ci sont même parfois ignorés des firmes. Il y a donc méconnaissance des coûts efficaces, et fréquemment asymétrie de cette connaissance, la firme étant en principe, plus à même de les apprécier correctement. Quand un régulateur constate donc des recettes R, des coûts C et donc un profit  $\Pi=R-C$ , il ne sait pas évaluer le positionnement exact de la firme en termes d'efficace. Il a alors à sa disposition deux stratégies:

1. limiter la rente, c'est-à-dire fixer les tarifs de telle sorte que  $\Pi$  soit contenu dans certaines limites.
2. Peser sur les coûts, c'est-à-dire fixer les tarifs de telle sorte que la firme soit contrainte d'abaisser ses coûts.

La première stratégie est dénommée cost-plus: le monopole est autorisé à fixer les prix à un niveau lui permettant de rembourser ses coûts et de générer une marge raisonnable, mais ses coûts peuvent être très éloignés des coûts efficaces. La deuxième stratégie est dénommée price-cap: le monopole doit faire évoluer ses prix à la baisse selon un taux défini par le régulateur.

La théorie des incitations a bien analysé les avantages et les défauts de ces deux mécanismes. On ne peut avoir en effet simultanément :

- la suppression de la rente de monopole
- l'incitation maximale à réduire les coûts

<sup>13</sup> Leibenstein, Harvey (1966), "Allocative Efficiency vs. X-Efficiency", *American Economic Review* 56 (3): 392-415

Ce dilemme est équivalent à celui que connaissent toutes les entreprises dans la gestion des frais de leurs salariés (dilemme des notes de frais): si elles remboursent ces frais sur factures, les salariés seront tentés de choisir des prestations élevées, mais ne feront pas de bénéfice sur ces frais; si elles choisissent un remboursement forfaitaire, les salariés chercheront éventuellement des prestations inférieures à ce forfait, de façon à tirer un bénéfice de ce forfait. L'établissement du forfait est donc crucial, c'est-à-dire la connaissance des coûts réels pour une prestation "raisonnable".

### 1.3 LES OPTIONS TARIFAIRES

Pendant longtemps, les régulateurs ont appliqué une règle dite cost-plus pour la régulation des monopoles infrastructurels avant la refonte de la régulation des années 90. Sous ce mécanisme, on autorise le monopole à recouvrer ses coûts (sans contrôle d'efficacité) avec une rémunération "raisonnable" sur les capitaux investis: le monopole se rembourse sur factures en quelque sorte, avec dès lors une forte incitation à la dérive des coûts et notamment des coûts d'investissement (effet dit Averch-Johnson<sup>14</sup>). D'où un revirement important des règles tarifaires lorsque furent constatées les incitations perverses de cette régulation dite RoR ou RRR (Rate of Return Regulation).

La structure tarifaire optimale (dite first-best) pour un prix d'accès est le coût marginal de la ressource. Celui-ci est, sur de nombreuses infrastructures suffisamment dimensionnées, extrêmement faible. Une telle tarification ne permet pas l'équilibre de l'exploitant de la ressource qui doit être subventionné par l'Etat pour couvrir ses coûts fixes, solution irréaliste. Cette solution n'est guère orientée vers les coûts, au sens usuel où on peut l'entendre.

La deuxième solution, (dite second-best), consiste à faire payer l'ensemble des clients (sur les marchés de gros et les marchés de détail) en fonction de leur propension à payer (leur élasticité aux prix): les clients les moins élastiques aux prix paient plus cher que ceux qui sont très élastiques aux prix, ce qui permet de recouvrer les coûts fixes sur les contributeurs ne pouvant se passer des services offerts (en général le monde professionnel et le marché de gros). Cette solution, dite "Ramsey pricing", nécessite une information très fine sur les coûts et la demande, difficile à obtenir pour le régulateur. De plus, cette solution est discriminatoire, puisque dépendant à la fois des coûts de la ressource mais aussi de l'élasticité de la demande des réseaux, différente d'un réseau à un autre. Enfin, elle fait supporter l'essentiel des coûts fixes sur les utilisateurs des autres réseaux.

La troisième solution sépare la question du marché de gros du marché de détail, tenant les prix de détail fixés. Elle consiste à facturer l'accès à la ressource sur la base du coût d'opportunité, égal au prix de détail moins le coût de l'activité aval. Cette solution, dite ECPR ("Efficient Component Pricing Rule") ou "Baumog-Willig Rule", rend neutre l'allocation des ressources entre gros et détail pour le détenteur de la ressource, mais n'est optimale que si les tarifs de détail sont établis selon la règle de Ramsey.

Ces règles tarifaires sont statiques, en ce sens qu'elles ne tiennent pas compte d'éventuels progrès techniques qui contraignent à l'évolution des infrastructures. Des prix bas induisent une sur-utilisation de la ressource sans incitation à investir dans des réseaux concurrents, mais n'incitent pas le détenteur de la ressource à améliorer, voire maintenir la qualité de la ressource. Des prix hauts encouragent l'investissement alternatif, mais induisent une sous utilisation de la ressource. Il est donc nécessaire de préciser ce qu'on entend par coût lorsqu'on envisage une régulation basée sur l'orientation vers les coûts des prix d'une ressource essentielle.

<sup>14</sup> Averch, Harvey; Johnson, Leland L. (1962). "Behavior of the Firm Under Regulatory Constraint". *American Economic Review* 52 (5): 1052–1069

Ce qu'il en coûte de produire quelque chose est une appréciation extrêmement délicate. La notion de coût est en effet très imprécise en soi, et stipuler une orientation des prix vers les coûts ne dit pratiquement rien qui soit opérationnel.

Il nous faut donc préciser de quels coûts il s'agit quand on parle d'orientation vers les coûts. Deux grandes approches des coûts peuvent être mobilisées: une approche comptable et une approche économique. Nous nous situons dans un premier temps dans le contexte d'une firme monoproduit, produisant un seul service dont nous cherchons à apprécier les coûts.

### 2.1 LES COUTS D'UNE FIRME MONOPRODUIT

Dans l'approche comptable, les coûts sont appréciés par la somme des dépenses encourues au cours du temps pour la production. Ces coûts sont sommés quelque soit la période de décaissement. On parle de coûts historiques. Les biens et services immobilisés sont amortis annuellement selon une règle de dépréciation fonction principalement des dispositions fiscales. Les règles comptables édictent de nombreuses dispositions en ce qui concerne la comptabilisation des dépenses, par exemple sur la production immobilisée, sur les consommations internes à l'exploitation, la valeur des productions en cours, etc. Certaines dépenses seront imputées à l'exploitation, et appartiendront donc aux coûts, d'autres interviendront à un autre niveau (par exemple les investissements ou les dividendes versés aux actionnaires) et ne relèveront pas des coûts d'exploitation. Ne sont comptabilisées que les dépenses effectives, non pas celles que peut par exemple engendrer l'exploitation en cause sur d'autres agents économiques (du fait d'externalités de production). Enfin, les coûts que l'entreprise encourt pour disposer de la ressource financière nécessaire à son exploitation diffèrent selon la nature de cette ressource, emprunt, fond propre ou toute forme mixte de ressource. Les règles comptables structurent donc fortement la nature des coûts de revient d'un produit (Riveline, 2005).

La comptabilité introduit une grande différence entre, d'une part les coûts variables, ceux qui sont fonction du volume de production, et d'autre part les coûts fixes, indépendants du volume de production (dans le contexte d'une unité de production installée). Les coûts variables comprendront principalement les matières, biens et services achetés en vue de la production, ainsi que le travail nécessaire à cette production, ils sont retracés au sein du compte d'exploitation. Les coûts fixes concernent principalement l'appareil de production, par exemple le réseau dans le contexte d'un opérateur de télécommunications. Ce vont être essentiellement des coûts amortissables, qui sont retracés dans le bilan (les immobilisations) de la firme.

L'approche économique, elle, considère la quantité des facteurs de production  $q_i$  nécessaires à une production, et leurs prix unitaires  $p_i$ . Leur somme,  $\sum p_i q_i$ , forme un coût total exprimé comme une fonction de coût du volume de production  $y$  du produit en cause:  $C(y)$ . Les questions soulevées précédemment demeurent, puisqu'il faut définir les produits  $i$  intervenant comme facteurs de production. Cette formulation introduit deux difficultés supplémentaires: la première concerne l'identification d'une unité volumétrique de la production  $y$ . Autant cette identification est généralement triviale dans le cas de biens (la "quantité" de biens produits), autant cette identification peut poser problème en matière de services: par exemple, dans le cas du service téléphonique, doit-on considérer le nombre d'appels ou la durée des appels? Enfin, cette formulation postule l'existence d'une fonction  $C(y)$  que, dans la plupart des cas, on va supposer dérivable, donc relativement continue, et supposer qu'il existe une fonction inverse  $C^{-1}$  telle que  $y=C^{-1}(q_i, p_i)=F(q_i)$ .

L'existence d'une fonction de coût permet d'enrichir notre approche des coûts de quatre notions:

- Celle d'efficience déjà évoquée: on dira d'une exploitation qui produit  $y$  à l'aide d'inputs  $\{q_i\}$ , qu'elle est efficiente si, dans le cadre des meilleures technologies de production, il est strictement impossible de produire  $y^o > y$  avec le même vecteur d'inputs. La fonction de production  $F=C^{-1}$  est alors désignée comme frontière de production
- Celle de coût moyen, égal à  $C(y)/y$
- Celle de coût marginal (coût de la dernière unité produite) égale à  $C'(y)$
- Celle de rendements d'échelle; si nous posons que  $F(\lambda q_i) = \lambda^\alpha \cdot F(q_i)$ , la production sera dite à rendements croissants si  $\alpha > 1$ , rendements constants si  $\alpha = 1$  et rendements décroissants si  $\alpha < 1$ .

Enfin, l'approche économique des coûts ne va pas sommer les coûts de différentes périodes sans différenciation. Les coûts ne seront additifs que s'ils sont actualisés, de façon à prendre en compte le fait qu'un euro de l'année  $n+1$  n'est pas tout à fait équivalent à un euro de l'année  $n$ , mais devra être dévalué du coût de l'argent sur une année<sup>15</sup>. Si  $k$  est ce coût, alors une somme  $S$  dépensée ou gagnée en  $n+1$  sera ramené à la valeur  $S/(1+k)$  à l'année  $n$ .

Dans cette perspective, l'amortissement économique va se distinguer assez sensiblement d'un amortissement comptable. La somme  $C$  annuelle qui permet de récupérer un investissement  $I$  fait en année 1 amortissable sur une durée de vie économique  $T$  (et non plus une durée de vie comptable) sera telle que:

$$I = \sum_{t=1}^T \frac{C}{(1+k)^t} = C \cdot \sum_{t=1}^T \frac{1}{(1+k)^t} = C \cdot \phi(k, T), \text{ soit } C = \frac{I}{\phi(k, T)}$$

Cette fonction  $\phi$  de  $k$  et de  $T$ , intervient ainsi comme un diviseur de  $I$  permettant d'obtenir un amortissement économique annuel. Ce diviseur intègre non seulement la durée de vie  $T$ , mais également le coût du capital  $k$  auquel la ressource financière est acquise.

Sachant que  $\phi(k, T)$  se développe analytiquement simplement:

$$\phi(k, T) = \sum_{t=1}^T \frac{1}{(1+k)^t} = \frac{(1+k)^T - 1}{k \cdot (1+k)^T}$$

Nous avons là un équivalent économique de la durée de vie comptable intégrant le coût du capital. A noter que lorsque  $T$  croît vers l'infini,  $\phi$  tend vers  $1/k$ <sup>16</sup>.

On le voit déjà ici dans le cadre d'une firme monoproduit, il existe une grande diversité de coûts, que ce soit dans une approche comptable ou une approche économique.

## 2.2 LES COÛTS D'UNE FIRME MULTIPRODUITS

Ces considérations ont été introduites dans le cadre d'une firme monoproduit, ce qui est rarement le cas. Il faut alors développer ces définitions dans le cadre d'une firme multiproduits. Une firme multiproduits se caractérise par l'existence d'un portefeuille de produits  $n$ , écoulés sur un marché. L'activité est caractérisée par un vecteur de productions  $y_n$ .

Comptablement, on distingue alors les dépenses directement attribuables à un produit  $k$  de celles qui ne le sont pas. Le plus fréquemment, les coûts directement attribuables seront des coûts variables, mais la

<sup>15</sup> En effet, si je place une somme  $S$  l'année  $n$  à un taux d'intérêt  $j$ , elle rapportera  $S \cdot (1+j)$  l'année  $n+1$ .

<sup>16</sup> Ainsi, pour un  $k$  de 10%,  $\phi$  tend vers 10, pour un  $k$  de 20%,  $\phi$  tend vers 5.

réciproque est loin d'être vraie. Les coûts fixes seront rarement des coûts directement attribuables, bien que certains le soient, quand un équipement est dédié à la production d'un seul produit.

Une firme multiproduits pose donc la question de la répartition des coûts non directement attribuables aux différents produits de la firme. Cette question est celle de la séparation comptable: comment répartir entre produits de façon "équitable" les coûts qui sont communs à différents produits?

La comptabilité répond à cette attente par un dispositif de comptabilité analytique. Plusieurs méthodes sont possibles, mais de façon générale, on divisera l'activité globale en activités élémentaires fournissant des produits intermédiaires (biens ou services), dotés d'une unité d'œuvre, qui seront consommés par des activités aval, et dont il est possible d'identifier les coûts. On déclinera ce processus jusqu'à atteindre les activités "finales", celles appartenant au portefeuille d'activités  $k$ , celles vendues sur le marché. Le coût des activités intermédiaires est alors "déversé" sur les activités aval qu'elles approvisionnent au prorata de leurs consommations respectives du produit de ces activités. Ces clés de répartition intermédiaires permettent ainsi de répartir l'ensemble des coûts sur les produits finaux. Toute la question est la détermination de ces clés de répartition, qui stipulent l'existence d'une unité d'œuvre de ces activités intermédiaires, répartissables entre les activités aval.

Les économistes privilégient une formalisation différente. Ils partent du principe que si différents produits appartiennent à un même portefeuille, c'est-à-dire sont produits de façon jointe, c'est qu'il existe vraisemblablement un intérêt économique à ce qu'il en soit ainsi<sup>17</sup>. On ne parle plus alors d'économies d'échelle, mais d'économies d'envergure, ce qui traduit le fait qu'il est plus économique de produire deux produits ensemble que de les produire séparément. Formellement, on parlera de coûts subadditifs, ce qui s'exprime de la façon suivante, dans le cas de deux produits 1 et 2:

$$C(y^1, y^2) < C(y^1, 0) + C(0, y^2)$$

Si donc on prend pour base le coût de production isolée des deux produits, les produire ensemble apporte une économie qu'il va falloir répartir entre les deux produits en question. Il n'existe pas de solution unique à cette répartition, mais un ensemble de solutions avec d'un côté l'attribution entière de cette économie au produit 1 et de l'autre l'attribution entière de cette économie au produit 2, toutes les solutions intermédiaires étant également possibles.

Cette formalisation donne naissance à la notion de coût incrémental. Si je considère le coût de production d'un portefeuille de produits  $n$ ,  $C^{n+1}$ , le coût incrémental du produit  $n+1$  qui serait ajouté à ce portefeuille, est le coût additionnel encouru pour la production de ce produit  $n+1$ , compte tenu du coût  $C^n$  déjà encouru: le coût incrémental se définit donc comme la différence  $C^{n+1} - C^n$ .

Ce n'est donc pas un coût marginal. Ce coût incrémental sera d'ailleurs fréquemment considéré comme un coût moyen, c'est-à-dire qu'on estimera le coût d'une unité du produit  $n+1$  comme la division de ce coût incrémental par le volume  $y^{n+1}$  du produit  $n+1$ .

Si le produit  $n+1$  est produit à partir d'un dispositif de production existant (tel qu'un réseau) et qu'il ne vient pas immédiatement redimensionner ce réseau, alors, le coût incrémental sera essentiellement composé de coûts variables. En d'autres termes, le coût incrémental stricto-sensu ne contribue pas à la couverture des coûts fixes du dispositif de production, coûts fixes supportés par les produits  $\{n\}$  du portefeuille existant. Ceci permet de saisir deux répercussions:

---

<sup>17</sup> Dans le cas où ce ne serait pas le cas, on parlera de firme conglomérale, dont la logique d'assemblage d'activités productives ne se justifie pas par des économies de production, mais plutôt par des économies financières ou d'autres raisons, par exemple la volonté d'amortir des cycles.

- La notion de coût incrémental est subordonnée à l'ordre d'entrée des produits dans le portefeuille de produits; le dernier produit entré supportera vraisemblablement des coûts incrémentaux plus faibles; le premier supportera une grande part des coûts fixes du dispositif de production<sup>18</sup>.
- Par contre, si le dernier produit entré dans le portefeuille entraîne un redimensionnement du dispositif de production, alors son coût incrémental peut être très élevé.

Pour échapper à cet arbitrage, on va considérer que le dispositif de production est redimensionné à chaque entrée d'un produit dans le portefeuille, de façon à éviter ainsi la question de l'ordre d'entrée et de redimensionnement: on parlera alors de coût incrémental de long terme (Long Run Incremental Cost en anglais - LRIC<sup>19</sup>). On rend alors variable en quelque sorte le dispositif de production, de façon à en répartir "équitablement" les coûts sur l'ensemble des produits du portefeuille.

Pour ramener ce coût à une valeur moyenne, il faut retenir une unité de production et considérer que la quantité de biens produits ne s'exprime que sur cette unité. On parlera à cet égard d'inducteur de coût (cost driver). Cet inducteur est particulièrement difficile à choisir, car tout bien possède différents attributs qui permettent de le caractériser. Le coût moyen incrémental de long terme (CMILT), qui résulte de ce choix, ne portera donc que sur un seul des attributs du produit. Ainsi, pour un service téléphonique, on peut considérer le nombre d'appels ou la durée des communications, et y ajouter des attributs liés à la qualité, codage utilisé par exemple, délai d'acheminement de l'appel, couverture du réseau d'accès, etc. Depuis 20 ans, la tarification de l'interconnexion de la terminaison d'appel a peu à peu réduit ses inducteurs de coût, principalement par souci de simplification, à la seule durée des communications, exprimée en minutes.

Le choix de cet inducteur de coût n'est pas totalement neutre dans la mesure où il va conditionner la nature de la causalité qui relie les services dont on cherche la répartition des coûts, à la nature de ces coûts: ce sont les coûts sensibles au trafic qui vont peu à peu s'imposer comme gérant cette causalité inductive des coûts incrémentaux.

## 2.3 LE COUT INCREMENTAL

Les régulateurs retiennent en général le coût moyen incrémental de long terme (CMILT) comme le bon signal de la tarification des ressources essentielles. La justification en est aisée: si je suis l'entrepreneur d'un réseau, je ne vais ajouter à mon portefeuille de produits un nouveau produit que si la recette incrémentale qu'il génère couvre a minima le coût incrémental que je subis, avec la possibilité de situer ce coût incrémental entre un coût incrémental strict (de court terme) et un coût incrémental de long terme, incluant éventuellement une participation à la couverture des coûts communs. La décision de l'entrepreneur de retenir telle ou telle conception du coût incrémental s'opérera vraisemblablement selon les effets externes de cette introduction (le nouveau produit entraîne-t-il des consommations accrues des produits existants ou au contraire risque-t-il de cannibaliser certains produits du portefeuille actuel?).

Le régulateur doit donc apprécier, comme doit le faire l'entrepreneur quand il prend une décision d'extension de son portefeuille de produits, quel coût incrémental il s'agit de couvrir par la tarification de la ressource essentielle.

<sup>18</sup> Pour échapper totalement à ce conditionnement par l'ordre des services considérés, il faut recourir à la valeur dite de Shapley qui est la valeur moyenne résultant de tous les arrangements possibles d'ordre des services; cette valeur possède de nombreuses propriétés intéressantes explicitées par exemple dans Boyer et al., 2006, qui lui confère une sorte d'équité dans la répartition des coûts.

<sup>19</sup> On parle en français de CMILT (Coût Moyen Incrémental de Long Terme) en prenant en considération le fait qu'il s'agit d'un coût moyen.

Interviennent à ce stade deux nouvelles considérations:

1. tout service de communication met en relation deux correspondants, mais en général un seul des deux paye: la couverture du coût doit-elle être partagée entre les deux correspondants d'une façon ou d'une autre?
2. quelle est l'assiette des coûts qui doivent être considérés comme incrémentaux?

### 2.3.1 LE PARTAGE DES COÛTS: APPROCHE STRATEGIQUE

Si la connexion sur un réseau génère d'importantes externalités (la capacité de tous les abonnés à joindre de nouveaux abonnés), l'appel génère lui aussi des externalités dans la mesure où les personnes appelées bénéficient en général des communications qu'elles reçoivent. Or, bien que la tarification des communications téléphoniques soit assise sur le principe de l'appelant payeur<sup>20</sup>, on constate fréquemment un bénéfice de l'appelé, du fait des informations, des nouvelles, de la relation... apportées dans la communication. Les communications non désirées, tels les appels publicitaires, sont généralement peu nombreuses et peuvent être aisément refusées.

Cet intérêt se manifeste notamment dans la régulation sociale des appels. Appellent celles ou ceux qui sont généralement les plus solvables: selon les sociétés et leurs usages, les parents appellent les enfants, les hommes appellent les femmes, les dirigeants appellent leurs employés, etc. Quand un usager ressent un besoin de communication mais est peu solvable, il "sonne" (beep) son correspondant pour qu'il le rappelle. Les mécanismes de call back ont formé à certaines époques d'autres manifestations de cet intérêt réciproque à la communication, bien que la charge de l'appel ne soit supportée que par un correspondant.

La justification de ce principe de l'appelant payeur réside dans la prévisibilité budgétaire des usagers. Si le récepteur d'un appel devait en supporter le coût, il ne pourrait en aucune façon maîtriser son budget de communications, et serait tenté de récuser tous les appels. La maîtrise budgétaire des usagers impose donc que la facturation accompagne la décision d'appel<sup>21</sup>.

La compréhension commune de l'interconnexion des réseaux veut alors que le prix payé par l'appelant couvre le coût de la terminaison d'appel sur le réseau de destination, et éventuellement, sur les réseaux de transit. On considère alors que les deux réseaux interconnectés, le réseau d'origine et le réseau de destination, doivent se partager la recette pour au moins couvrir leurs coûts, la charge de l'appel étant supportée par l'appelant.

On montre qu'en termes de bien être, cette solution n'est pas optimale, et que l'appelé, dès lors qu'il retire un bénéfice de la communication, devrait en supporter une partie des coûts. Ce n'est donc pas tant le partage des coûts entre les réseaux qu'il faut considérer, que le partage des coûts entre les correspondants. Qu'est-il juste de faire payer à chacun des deux correspondants?

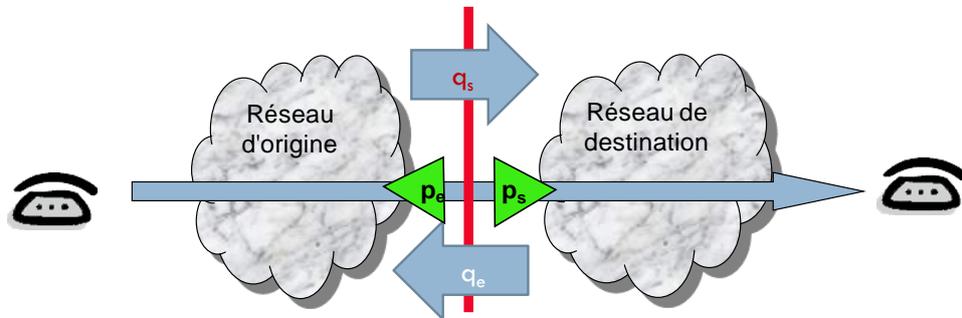
Cette considération s'exprime également dans l'alternative présentée par les deux grands régimes d'interconnexion:

- a. le premier régime est celui mis en place le plus fréquemment dans le monde de la téléphonie, connu généralement comme le régime de la terminaison d'appel (TA) ou le régime des charges d'accès. Dans ce régime, le réseau qui demande la terminaison d'un appel sur un autre réseau supporte la charge d'acheminement vers le correspondant final.

<sup>20</sup> Le CPP (Caller Pays Principle) déjà évoqué.

<sup>21</sup> Sauf quand le coût de l'appel est incertain, ce qui est le cas dès lors que le correspondant sort du contexte tarifaire usuel; ainsi, les appels internationaux sont partiellement pris en charge par les appelés en situation de roaming.

b. le second régime est celui mis en place historiquement dans le monde internet, connu sous le nom de peering ou de "bill and keep" (BAK), dans lequel chaque réseau garde les recettes qu'il perçoit sans opérer aucun reversement aux autres réseaux avec lesquels il est interconnecté. Dans ce régime, on peut considérer en quelque sorte que la TA est égale à zéro<sup>22</sup>, et que chaque réseau ne se rémunère que sur les appels sortants dont le prix doit couvrir en quelque sorte à la fois le demi-coût des appels sortants et le demi-coût des appels entrants<sup>23</sup>.



Si on résume la situation d'interconnexion aux flux sortants ( $q_s$ ) et entrants ( $q_e$ ) entre les deux réseaux, et aux prix de ces trafics ( $p_s$  et  $p_e$ ), alors, la différence économique induite par le régime de la TA par rapport au régime BAK (qui forme une différence de profit), et que l'on peut nommer la balance d'interconnexion, pour le réseau d'origine, est constituée de la recette issue du trafic entrant minorée du coût généré par le trafic sortant<sup>24</sup>:

$$\text{Balance d'interconnexion} = p_e \cdot q_e - p_s \cdot q_s = p_e \cdot (q_e - q_s) + q_s \cdot (p_e - p_s)$$

Cette relation montre que si les trafics entrants et sortants sont équilibrés et si les tarifs de TA sont similaires, alors économiquement, le régime de la TA et le régime BAK sont équivalents.

De façon à maximiser l'apport économique du régime de la TA, l'opérateur a donc intérêt à recevoir plus de trafic qu'il n'en émet vers les réseaux interconnectés et à sur-tarifier sa TA. L'un ne va pas généralement avec l'autre, et les équilibres sont donc plus complexes, que nous ne développerons pas ici. Ce développement montre l'impact que peut avoir une différenciation des terminaisons d'appel: si celles-ci sont égales à zéro, ou proches de zéro, la différence économique s'annule ou s'atténue grandement, quelque soit la différence de trafics.

Orienter  $p_e$  et  $p_s$  vers les coûts a donc une importance stratégique et économique tout à fait considérable. Cette expression montre que le niveau des prix a son incidence (facteur des différences de trafic) ainsi que l'asymétrie des prix.

Cette analyse illustre également le fait qu'entre un régime de la TA où le prix reflète des coûts d'assiette très large, et le régime BAK, où le prix reflète des coûts nuls, existe un curseur de prix qui va partager les coûts, non seulement entre deux réseaux, mais également entre deux correspondants.

Ce curseur va refléter in fine la façon dont va être déterminée l'assiette des coûts à prendre en compte, celle qui définit ce qu'est un coût incrémental.

### 2.3.2 LES COÛTS PERTINENTS: APPROCHE PRAGMATIQUE

<sup>22</sup> On ne peut pas considérer le régime du peering comme une forme asymptotique du régime de la TA dans lequel la TA sera égale à zéro, car les modalités juridiques des deux types d'interconnexion sont radicalement différentes.

<sup>23</sup> On se reportera à Tera et al. (2009) pour une analyse substantielle de ces deux régimes.

<sup>24</sup> Dans les deux régimes, les opérateurs reçoivent en sus les recettes issues de leurs abonnés (abonnements et communications).

Cette assiette est ce qu'on appelle généralement les coûts pertinents de l'interconnexion. Le coût incrémental d'un service est le coût engendré par l'ajout de ce service, sur le court terme ou sur le long terme, au portefeuille de services, ou réciproquement d'ailleurs, l'économie de coût réalisée en ôtant le service du portefeuille de services<sup>25</sup>. Le coût incrémental est en quelque sorte "causé" par l'ajout du service au portefeuille de services.

Il faut donc qu'il y ait un lien de causalité entre le service dont on cherche à mesurer le coût et le coût à considérer. Ce lien de causalité s'apprécie de deux façons:

1. sur le court terme, le coût ne serait pas supporté si le service n'est pas offert ou inversement, supprimer le service affecte un coût pertinent et aucunement un coût non pertinent
2. sur le long terme, le coût est modifié si le service est offert. Dans le cas d'un réseau, cela signifie que les coûts sont liés d'une façon ou d'une autre au trafic écoulé.

Ce lien de causalité, qui définit l'assiette des coûts pertinents (ce que l'on nomme aussi parfois l'incrément), est de façon pragmatique assez arbitraire, et permet au régulateur, comme à l'entrepreneur, dans sa prise de décision d'offre de nouveaux services, de gérer le curseur des coûts entre les deux régimes d'interconnexion décrits précédemment, c'est-à-dire de répartir ces coûts entre les deux correspondants et entre les deux réseaux.

Ces coûts pertinents sont définis par le régulateur dans des décisions spécifiques. Elles sont fréquemment modifiées dans le temps en fonction de la position stratégique des acteurs et des régulateurs. La Commission Européenne a ainsi fait constamment évoluer sa position en la matière depuis la fin des années 90.

Arrêtons-nous sur quelques-uns des enjeux associés à cette définition.

La question des coûts, dans le champ des télécommunications, a d'abord été soulevée par la terminaison d'appel sur les réseaux fixes: stratégiquement, cette TA sur réseaux fixes devait être faible si on voulait soutenir le développement des réseaux mobiles, de façon à restreindre l'écart des tarifs entre les communications des réseaux fixes et celles des réseaux mobiles. Or, dans les réseaux fixes, une part importante des coûts est représentée par le coût de la boucle locale. Celle-ci écoule l'ensemble des trafics et ne peut être en aucune façon redimensionnée en fonction du trafic écoulé, entrant ou sortant. La boucle locale, et tous les équipements "dédiés" en quelque sorte à l'utilisateur final de la ligne (y compris par exemple la carte d'abonné qui permet de le desservir), ne sont aucunement dimensionnés par le trafic écoulé sur la ligne. Leurs coûts ont dès lors été considérés comme non pertinents pour l'interconnexion: aucune relation de causalité n'existe entre la constitution de ces coûts et le trafic d'aucun des services offerts. Inversement, le soutien économique aux réseaux mobiles a longtemps justifié des terminaisons d'appel sur réseaux mobiles très élevées, bien au-dessus des coûts encourus.

Il existe de nombreux autres coûts qui ne sont aucunement pertinents pour l'interconnexion. Tous les coûts commerciaux supportés par un opérateur sont non pertinents pour l'interconnexion, car ils sont induits par les abonnés d'un réseau et d'aucune façon, par les services d'interconnexion. Supprimer les services d'interconnexion ou les ajouter au portefeuille de services ne modifie en rien ces coûts commerciaux, qui comprennent les coûts de recrutement des clients, les coûts de gestion des abonnements (incluant les différentes plateformes de services prépayés ou postpayés), les coûts des services d'assistance, etc. Aucune causalité n'existe entre les services d'interconnexion et le niveau de ces coûts.

Chaque poste de coût doit dès lors faire l'objet d'une analyse pour savoir si son niveau est conditionné ou non par les services dont on cherche à évaluer les coûts. Cette analyse est assez simple pour les coûts directement attribuables, qu'ils soient fixes ou variables; elle l'est également pour les éléments

---

<sup>25</sup> On considère donc implicitement que les coûts sont évitables.

d'infrastructure qui ne sont pas figés une fois pour toute, mais dont le dimensionnement dépend du volume de trafic écoulé.

Evidemment, cette analyse de causalité ou d'incrémentalité trouve ses limites dans l'interprétation du caractère dimensionnant du trafic. Ainsi, la Commission Européenne a changé en 2009 son interprétation en matière de coûts pertinents pour les réseaux mobiles. Constatant que les terminaisons d'appel sur les réseaux mobiles restaient élevées et désavantageaient les petits opérateurs (et donc les nouveaux entrants), la Commission a réduit l'assiette des coûts pertinents pour l'interconnexion en considérant que le réseau était constitué de deux composantes: une composante de couverture territoriale et une composante de capacité. Du fait des obligations de couverture associées à leur cahier des charges, les opérateurs sont contraints de déployer leur réseau sur le territoire, indépendamment du trafic qui y passe: les stations de base, équipées au minimum (un TRX par secteur), ne sont dès lors plus considérées comme des coûts pertinents pour l'interconnexion, car non conditionnées par le volume de trafic, mais par les engagements pris en termes de déploiement du réseau. Seules, les capacités additionnelles, elles justifiées par le volume du trafic, restent dans l'assiette des coûts pertinents pour l'interconnexion. Ce changement d'interprétation de l'incrément permet ainsi de réduire drastiquement le niveau du coût de la terminaison d'appel, puisque le coût du réseau d'accès, majeur dans tout réseau, sort en très grande partie de l'assiette des coûts pertinents<sup>26</sup>.

D'autres types de coûts forment parfois litige. Il en est ainsi par exemple du coût des licences. Il est clair que le coût des licences n'est, d'aucune façon, conditionné par le volume du trafic. C'est un coût d'accès au marché perceptible dès lors qu'il y a *numerus clausus* des entrées sur le marché. Si l'entrée sur le marché est libre, il est très difficile d'exiger des prix de licence dépassant le niveau des coûts de traitement des dossiers administratifs. Ce que payent les opérateurs est le droit d'entrée sur un marché avec un faible nombre de concurrents. Ils gagnent ainsi le droit de recruter des clients. Il est donc extrêmement difficile de défendre l'idée que le coût des licences puisse faire partie d'un incrément en matière d'interconnexion.

Défendre la réduction de l'assiette des coûts pertinents pour l'interconnexion, c'est implicitement admettre dans le portefeuille de services, un service d'accès offert aux abonnés du réseau, dont le coût rassemble l'ensemble des coûts sortis des incréments des services de communication à proprement parler.

Comme dans le management de l'entreprise, la conception des coûts retenue par le régulateur est un mixte de considérations économiques et stratégiques. Les balances d'interconnexion des opérateurs (dont la somme est nulle pour le trafic domestique) forment un élément clé de ces approches.

Ces balances, sur un marché parfait (dont les consommateurs se répartiraient sans discrimination entre les opérateurs, avec la même probabilité d'appel et dont les opérateurs seraient efficaces), devraient être quasi-nulles.

$$\text{Balance d'interconnexion} = p_e \cdot q_e - p_s \cdot q_s = p_e \cdot (q_e - q_s) + q_s \cdot (p_e - p_s)$$

Quand elles ne le sont pas, il est tentant de corriger la discrimination introduite par des flux de trafic asymétriques par une asymétrie des tarifs d'interconnexion. En effet, si

$$\frac{q_e - q_s}{q_s} = - \frac{p_e - p_s}{p_e}$$

<sup>26</sup> Cf la recommandation 2009/396/EC de la Commission du 7 mai 2009 sur le traitement réglementaire des tarifs de terminaison d'appels fixe et mobile dans l'UE (disponible sur <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:124:0067:0074:fr:PDF>)

alors, la balance d'interconnexion est nulle. Si les terminaisons d'appel sont symétriques,  $p_e - p_s = 0$ , seul l'abaissement du niveau de  $p_e$  permet d'abaisser le niveau de la balance. D'où le jeu sur le niveau des TA.

Abaisser le niveau d'une terminaison d'appel ne précarise aucunement la situation économique d'un opérateur et ne doit influencer d'ailleurs, contrairement à une idée répandue, sur le prix de détail des communications<sup>27</sup>. En effet, abaisser ce prix revient à laisser à l'appelé soit la charge du coût de réception, qui ne sera pas perçu sur les appels reçus, mais sur les appels émis, soit la charge d'un coût de raccordement, perçu comme tel ou amorti sur les appels émis. En d'autres termes, si les usagers émettent autant de communications qu'ils en reçoivent, ils payent en appelant, l'usage de leur réseau à la fois pour leurs appels sortants et pour leurs appels entrants. Si l'abaissement de la TA ne résulte que d'une modification d'assiette des coûts pertinents, cela ne doit avoir aucune influence sur le niveau des prix de détail. Si l'abaissement de la TA s'opère à assiette constante, cela signifie que le niveau des coûts globalement baisse et que cette baisse devrait également toucher les prix de détail.

### 2.3.3 LES CMILT: APPROCHE TECHNIQUE

Le principe économique est donc d'utiliser comme critère d'orientation vers les coûts le coût moyen incrémental de long terme (CMILT). Ce CMILT s'apprécie selon la définition d'un incrément (assiette de coûts pertinents) qui reste à l'appréciation du régulateur à travers des décisions ad-hoc et dont nous avons vu le caractère éminemment stratégique. Une fois cet incrément défini, restent quelques aspects techniques à traiter.

Les coûts considérés sont des coûts économiques au sens où nous les avons définis. Ils intègrent donc, pour les coûts qui sont amortissables, un taux d'actualisation qui est en première analyse le coût d'accès à la ressource financière, soit le coût  $k$  du capital, sur lequel nous reviendrons ci-après. Une durée de vie technique (la durée de vie pendant laquelle il est possible de faire fonctionner cet équipement) associée à ce coût du capital, permet comme nous l'avons vu de déduire un coût économique annuel relatif à un investissement  $I$  qui se distingue d'un amortissement comptable en ce sens qu'il prend en considération la "juste" rémunération du capital. Mais, dans les systèmes techniques à progrès rapide, cette durée de vie technique est fréquemment écourtée par l'obsolescence technique des équipements, soit du fait de nouveaux matériels plus performants et moins coûteux, soit du fait de changements de génération d'équipements. Si on prend comme indicateur de cette obsolescence l'évolution des prix  $j$  des matériels, on montre (Babusiaux, 1990) que le taux d'actualisation à prendre en compte, si  $k$  représente le coût du capital, devient:

$$1 + i = (1 + k) \cdot (1 - j)$$

Si  $j$  est négatif (baisse des prix), il y a obsolescence technologique et le taux d'actualisation est accru de  $1-j$ . Au contraire, si  $j$  est positif (hausse des coûts, par exemple du génie civil), il y a intérêt à faire vivre plus longtemps l'équipement et le taux d'actualisation est abaissé de  $1-j$ <sup>28</sup>.

<sup>27</sup> Ce constat n'est valable que dans des situations de type two-way access. On distingue ici les situations dans lesquelles les agents économiques sont en situation de réciprocité (chacun dispose d'une ressource essentielle, two-way access, cas typique de l'interconnexion) de celles dans lesquelles un seul des agents détient une ressource essentielle (one-way access, cas du dégroupage par exemple).

<sup>28</sup> D'autres méthodes plus complexes peuvent être utilisées. Cf. par exemple la Tilted Annuity Method (TAM) utilisée en Belgique (Description du modèle des coûts top-down pour le calcul des tarifs d'interconnexion – IBPT, disponible à <http://www.ibpt.be/ShowDoc.aspx?objectID=2060&lang=fr>)

Le deuxième aspect technique à considérer a trait au prix retenu pour les équipements. Dans les comptes des exploitants, comme nous l'avons évoqué, les coûts sont des coûts historiques qui, compte tenu du progrès technique extrêmement rapide des technologies de l'information et de la communication, baissent de façon très sensible<sup>29</sup>. Le bon signal de coût, celui sur lequel sera prise la bonne décision d'investissement, n'est pas le coût historique, mais le coût actuel des équipements. On entend par coût actuel le coût des mêmes équipements (ou équivalents modernes de ces équipements si ceux-ci ne sont plus produits) offerts commercialement sur le marché par les équipementiers au moment ou s'opère l'évaluation des coûts.

Cette actualisation des coûts pose une autre question relative à l'efficacité d'un opérateur. Quand on est confronté à l'évaluation des coûts d'un opérateur, se pose fréquemment la question de savoir si l'opérateur ne pourrait pas être plus efficace s'il exploitait, non seulement des matériels plus modernes, question résolue ci-dessus, mais également des architectures plus modernes ou des organisations plus efficaces du travail. Cette question de l'efficacité d'un opérateur est particulièrement délicate, car toute amélioration de l'efficacité peut se faire au détriment de la qualité ou de la sécurité du réseau. Ainsi, on peut par exemple alléger très sensiblement les coûts en réduisant le nombre de points de commutation, qui résultent des limites technologiques historiques, mais au prix éventuellement de questions de sécurité, ou en réduisant les sources d'énergie secondaire, au prix d'une certaine détérioration de la disponibilité du réseau si l'alimentation primaire se trouve prise en défaut. Pour éviter des débats sans fin sur ces arbitrages coût/qualité-sécurité, on retient généralement la topologie ou l'architecture historique du réseau, celle qui s'est imposée pour des raisons techniques, financières, historiques, voire politiques à l'exploitant<sup>30</sup>. Mais, on n'exclut pas d'éliminer certaines inefficiences de l'opérateur, issus de décisions managériales manifestement malheureuses (par exemple des sureffectifs) dont les concurrents n'ont pas à pâtir.

Reste la question des coûts communs. L'approche des coûts incrémentaux, stricto-sensu, concerne les coûts "causés" par le nouveau service. Il s'agit en général de coûts de court terme (variables). Pour tenir compte des coûts fixes, on définit un incrément de long terme qui intègre les coûts fixes dimensionnés par le trafic (i.e. le volume de production). Mais, une entreprise est aussi constituée de fonctions dites générales (direction générale, service financier, gestion des ressources humaines, etc.) qui sont plus ou moins dimensionnées par le volume de production, mais dont il est difficile d'affecter précisément le coût aux différentes lignes de production, au moyen d'une clé de ventilation "technique" robuste. Ces coûts communs sont structurels, c'est-à-dire qu'ils ne sont affectables à aucun des incréments relatifs aux services présents dans le portefeuille de services ; en d'autres termes, la somme des coûts incrémentaux de tous les services offerts ne couvre pas la totalité des coûts de l'entreprise. Pour assurer cette couverture, il convient donc de répartir ces coûts communs (ou coûts joints si les coûts communs ne concernent qu'une sous partie de l'entreprise) au moyen d'un mark-up sur les coûts incrémentaux, de façon à ce que la totalité des coûts soit bien prise en compte<sup>31</sup>.

Techniquement, deux grandes approches sont mises en œuvre:

- Une approche descendante [top-down (TD)] qui consiste à partir d'une comptabilité analytique de façon à parvenir à l'appréciation des CMILT (dite en anglais TD-LRIC).
- Une approche ascendante [bottom-up (BU)] qui consiste à estimer les CMILT à partir d'une maquette simulant le réseau à analyser (dite en anglais BU-LRIC).

<sup>29</sup> Les coûts des équipements électroniques peuvent baisser de 5 à 15% par an selon la nature de l'équipement. A -12% par an, valeur courante, le coût d'un équipement de 100 en année 1 s'abaisse à 28 en année 10.

<sup>30</sup> On parle alors d'approche "scorched node" (par opposition à une approche "scorched earth" qui consiste à reconstruire entièrement le réseau sur une architecture optimale).

<sup>31</sup> On parle alors de TELRIC en anglais (Total Element LRIC).

L'approche top-down nécessite une comptabilité analytique orientée régulation, notamment pour la bonne prise en compte des incréments définis pour les services régulés. Une fois cette comptabilité disponible, un certain nombre de traitements doivent être effectués pour aboutir à l'estimation des CMILT "top-down":

- a. remplacer les prix des matériels anciens par les prix actuels de matériels équivalents modernes
- b. calculer leur annuité économique fonction du coût du capital, de la durée de vie des matériels et de leur obsolescence, en substitution des amortissements intervenant dans les comptes
- c. repérer les éventuelles inefficiences et en corriger les coûts
- d. éliminer les coûts non pertinents

La difficulté de cette approche réside dans la disponibilité d'une comptabilité analytique fournissant les bonnes assiettes de pertinence des coûts (de préférence auditée) suffisamment détaillée au niveau des investissements pour pouvoir procéder au redressement des coûts historiques.

L'approche bottom-up suppose la disponibilité d'un modèle reconstruisant les coûts du réseau (dit cost-proxy model en anglais). Cette maquette du réseau peut être extrêmement détaillée, elle n'en est pas moins fréquemment réductrice de la complexité d'un réseau. Il s'agit donc d'une approche simplifiée du réseau qui présente l'avantage de pouvoir optimiser un certain nombre de caractéristiques architecturales du réseau en fonction de la nature de la demande. La difficulté de cette approche réside dans la capacité à alimenter en informations techniques et commerciales (par exemple localisation de la demande) une maquette la plus détaillée possible.

Les deux approches devraient aboutir à des résultats voisins et peuvent s'alimenter mutuellement. Notamment, les mark-up de coûts communs peuvent être issus d'une analyse comptable, même sommaire.

Dans la pratique, les deux approches sont de facto très similaires. Une bonne comptabilité analytique recourt le plus fréquemment à une modélisation des coûts qui est plus ou moins identique à la modélisation présente dans une maquette de réseau. C'est donc le contexte qui guide la meilleure approche. Les opérateurs qui ont engagé la construction d'une comptabilité analytique préfèrent le plus souvent fournir cette base aux régulateurs pour étayer leurs propositions de tarifs orientés vers les coûts, ce qui suppose de la part du régulateur la capacité d'auditer cette comptabilité. Dans certains cas, notamment dans les pays en développement, les comptes des opérateurs sont trop incertains pour permettre l'élaboration d'une telle comptabilité analytique<sup>32</sup>, les régulateurs trop imprécis sur leurs attentes en la matière, pour que l'approche top-down soit possible. Une approche bottom-up est alors privilégiée avec un modèle adapté à la situation économique et informationnelle de l'opérateur.

Les premiers modèles de coût ont été développés à la fin des années 90 pour les réseaux fixes, qui posaient les problèmes les plus cruciaux en termes d'interconnexion. Ces modèles ont été bâtis sur de vrais modèles d'ingénierie de réseau, nécessitant des informations très précises sur la localisation de la demande de façon à pouvoir optimiser un réseau dans le contexte de pays où tous les ménages et établissements étaient desservis par le réseau téléphonique (Gasmi et al., 2002). Ces modèles ont permis de simuler la fonction de coût des opérateurs et de déterminer ses caractéristiques en fonction du déploiement et du développement du réseau. Quand la téléphonie mobile est devenue significative, une nouvelle série de modèles plus légers ont été proposés, notamment pour les pays en développement par la Banque Mondiale et l'UIT, de façon à aider les opérateurs et régulateurs à mieux apprécier la réalité de leurs coûts (cf. par exemple Banque Mondiale, 2003).

---

<sup>32</sup> Notamment en ce qui concerne les opérateurs historiques dont les bilans d'ouverture, au moment de changements institutionnels (par exemple séparation d'avec la poste), ont été faits à la hâte et ne présentent pas des immobilisations précisément établies.

Qu'il s'agisse d'une approche comptable (descendante - top-down) ou d'une approche de reconstruction des coûts (ascendante - bottom-up), il revient aux autorités de régulation de préciser la façon dont le déversement ou la reconstruction des coûts s'opère, quels sont les coûts pertinents et les modèles d'ingénierie utilisés<sup>33</sup>.

De façon générale, ces approches, après avoir défini la délimitation des coûts pertinents et la façon de les ajuster pour tenir compte des considérations précédentes, affectent ces coûts aux différents services offerts par l'opérateur sur la base de l'utilisation des éléments de réseau et du recours aux services de l'opérateur par chaque catégorie de services. Cette affectation passe par le recours à des facteurs de routage qui spécifient pour chaque type de service l'usage qu'ils font du réseau. Implicitement, ces facteurs de routage normalisent l'occupation du réseau par les différentes catégories de service, basée de ce fait sur le seul trafic<sup>34</sup>.

## 2.4 LE COUT DU CAPITAL

Un élément important du calcul des coûts d'un service est formé par le coût du capital ( $k$  dans ce qui suit). Celui-ci représente grossièrement le taux de rémunération des capitaux investis dans l'exploitation. Il est établi par les régulateurs au niveau auquel l'opérateur paye sa ressource financière.

La ressource financière d'un opérateur est résumée par le passif du bilan de l'entreprise. Elle est constituée de deux composantes majeures: les capitaux propres (CP) et les dettes (D) contractées auprès de différentes catégories de prêteurs. Le coût de cette ressource financière est donc un coût pondéré du coût des capitaux propres (CP) et du coût des dettes (D):

$$k = k_{CP} \cdot \frac{V_{CP}}{V_{CP} + V_D} + k_D \cdot \frac{V_D}{V_{CP} + V_D}$$

Où  $V_{CP}$  et  $V_D$  représentent respectivement la valeur des capitaux propres et la valeur des dettes.

La valeur des capitaux propres retenue ( $V_{CP}$ ) n'est pas la valeur comptable présente au bilan, mais la valeur de l'entreprise en bourse (sa capitalisation boursière) si elle est cotée. Une estimation pourra en être faite si l'entreprise n'est pas cotée. Ceci ne vaut que sur un marché financier rationnel, c'est-à-dire sur lequel n'existe aucune bulle spéculative.

La valeur des dettes ( $V_D$ ) peut prêter à discussion. En effet, traditionnellement, sont considérées ici les dettes contractées auprès d'établissements financiers (emprunts) ou de marchés financiers (obligations). L'examen des bilans des opérateurs montrent qu'ils bénéficient fréquemment de ressources financières non "orthodoxes", qui, à notre sens, doivent prises en compte dans le montant de la valeur des dettes:

- a. Les crédits fournisseurs obtenus auprès des équipementiers (rémunérés ou non) du fait de leur mise en concurrence
- b. Les crédits obtenus auprès de fournisseurs d'exploitation, notamment les réseaux interconnectés, par désaccord sur les facturations présentées (généralement non rémunérés)

<sup>33</sup> Cf. par exemple la décision n° 2007-0128 du 5 avril 2007 de l'Arcep portant sur la spécification des obligations de comptabilisation et de restitution des coûts des réseaux mobiles. Cf. également le *position paper* du régulateur de Bahrain synthétisant les grandes options considérées à travers le monde (Development, implementation and use of bottom-up fixed and mobile network cost models in the Kingdom of Bahrain – 2011, disponible à <http://www.tra.org.bh/en/pdf/MCD1011144PositionPaperonBU-LRICcostmodels.pdf>).

<sup>34</sup> Il convient alors de spécifier des taux de remplissage des éléments d'infrastructure dédiés à certains usages (par exemple, les lignes louées).

- c. Les crédits implicites (mais apparaissant au bilan) associés aux différés de paiement des licences d'exploitation (non rémunérés)

Ces dettes (éventuellement corrigées de montants similaires pouvant se trouver en actif circulant) représentent dans certains cas une véritable ressource financière pour l'entreprise, à un coût très faible, voire nul, qui doit, nous semble-t-il, être comptabilisée dans la valeur de dettes.

Ce coût moyen pondéré du capital (CMPC en français ou WACC [Weighted Average Cost of Capital] en anglais) est corrigé des effets fiscaux. En effet, il y a un premier avantage à l'endettement, qui est la possibilité de déduire la rémunération de la dette (les intérêts payés) du bénéfice imposable, réduisant d'autant l'impôt sur les bénéfices. Si IS représente l'impôt sur les bénéfices, le coût pondéré du capital avant impôt ressort à :

$$k = \frac{k_{CP}}{1 - IS} \cdot \frac{V_{CP}}{V_{CP} + V_D} + k_D \frac{V_D}{V_{CP} + V_D}$$

Une fois estimées les valeurs des capitaux propres et des dettes, il reste à estimer le coût de ces deux composantes du coût de la ressource financière,  $k_{CP}$  et  $k_D$ .

Le coût des dettes est connu de l'entreprise, comme la somme pondérée du coût de chaque dette par son montant: il ne présente donc pas de difficultés particulières. Pour les firmes très importantes qui ne sauraient calculer ce coût moyen, on retient une classe de risque (associé à une notation) et on ajoute la prime de dette correspondante au taux sans risque<sup>35</sup>.

Le coût des fonds propres est plus délicat. C'est grossièrement la rémunération que l'entreprise doit verser à ses actionnaires pour qu'ils souscrivent à son capital. Le monde financier, pour évaluer un actif, recourt le plus fréquemment à la théorie du portefeuille, qui, même si elle est contestée et sans doute contestable, n'a pas pour le moment de substitut simple à partir desquels évaluer ce coût des capitaux propres.

Les marchés financiers classent les entreprises par classe de risque. La valeur de l'entreprise  $V$  est posée comme l'espérance de ses gains  $[E(X)]$  divisé par le coût du capital associé au risque  $r$ , ou de façon similaire, le coût du capital est bien le rapport de l'espérance de gains à la valeur de l'entreprise :

$$V = \frac{E(X)}{k_r} \text{ ou } k_r = \frac{E(X)}{V}$$

La valeur de l'entreprise étant la somme de la valeur de ses capitaux propres et de la valeur de ses dettes, pour un risque  $r$ , nous avons :

$$E(X) = k \cdot V = k \cdot (V_{CP} + V_D) = k_{CP} \cdot V_{CP} + k_D \cdot V_D$$

De là découle la formule du coût moyen pondéré du capital, mais aussi l'expression de l'effet de levier de l'endettement :

$$k_{CP} = k + (k - k_D) \cdot \frac{V_D}{V_{CP}}$$

Si  $k_D < k$ , alors  $k_{CP} > k$ . Les actionnaires attendent une rentabilité supérieure aux prêteurs, car le risque qu'ils prennent est supérieur à celui des prêteurs (en cas de difficultés financières de la firme, ils seront remboursés avant les actionnaires). Il est donc intéressant pour une firme de s'endetter (jusqu'à un certain point) de façon à abaisser le coût de son capital.

La théorie du portefeuille stipule qu'il est intéressant de constituer un portefeuille d'actifs en cherchant à neutraliser les risques présentés par les firmes. Nous n'allons pas développer ici cette théorie présente dans tous les manuels de finance, mais indiquer qu'à la suite des travaux de Sharpe, prix Nobel d'économie en 1990, elle débouche sur une expression simple du coût des capitaux propres d'une firme

<sup>35</sup> Pour les opérateurs français, la prime de dette se situe autour de 2%.

i dénommée MEDAF (Modèle d'évaluation des actifs financiers) ou CAPM (Capital Asset Pricing Model); ce coût s'exprime en fonction de la valeur du marché M sur lequel cette firme est cotée:

$$k_{CPi} = k_F + \beta_i \cdot (k_M - k_F) \text{ avec } \beta_i = \frac{\sigma_{iM}}{\sigma_M^2}$$

Où:  $k_F$  est le taux sans risque du marché financier,  $k_M$  est le taux du marché financier et  $\beta_i$  (dénommé le béta) est le rapport d'une covariance entre la valeur de i et la valeur du marché M rapportée à la variance de ce marché M.

Trois paramètres permettent ainsi de calculer le coût des capitaux propres:

- Le taux  $k_F$ , taux sans risque du marché, qui est le taux servi sur des actifs jugés ne présentant aucun risque, et qui est pris comme le taux servi sur les obligations d'Etat des pays notés AAA (par exemple l'Allemagne). Ce taux, qui oscillait entre 4 et 5% préalablement à la crise de 2008, a aujourd'hui considérablement chuté et se situe en dessous de 2%. Certains régulateurs prennent également en compte le niveau d'inflation.
- Le taux  $k_M$  du marché M est le taux servi en moyenne sur le marché M: c'est en général le taux de l'indice de référence de ce marché (DJII, Nasdaq, Nikkei, FTSE, DAX, CAC, etc.). C'est en fait la prime de marché ( $k_M - k_F$ ) (dite ERP – Equity Risk Premium en anglais) qui est utilisée. Pour les pays en développement, on tient compte également du risque pays.
- Le béta traduit la volatilité de l'action de la firme i par rapport au marché M: si l'action i amplifie les fluctuations du marché M, son béta sera supérieur à 1; si au contraire, l'action i est moins volatile que le marché M, son béta sera inférieur à 1.

Le coût des capitaux propres est ainsi la somme du taux sans risque plus  $\beta$  (voisin de 1) facteur des primes de risque subis par la firme, essentiellement de la prime de risque du marché M (égale à  $k_M - k_F$ ), mais aussi la prime de risque pays dans le contexte des pays en développement. Désormais, peut également s'y ajouter une prime de risque associée à des investissements dans de nouveaux réseaux prenant place sur des marchés dotés de fortes incertitudes.

$$k_{CPi} = k_F + \sum_r \beta_i \cdot \text{prime de risque}_r = k_F + \beta_i \cdot \sum_r \text{prime de risque}_r$$

$k_F$  et  $k_M$  sont donnés par le marché M. Le béta est fréquemment calculé sur les sites financiers<sup>36</sup> et souvent même publié par les firmes cotées. Ces paramètres peuvent être relativement volatiles. Les régulateurs retiennent généralement des valeurs de long terme, pour lisser les évolutions du coût du capital. Ainsi, en France, la prime de marché est prise égale à 5% par l'Arcep, considérée comme une estimation de long terme, relativement stable. Le taux sans risque utilisé en France, même estimé de long terme, baisse progressivement, étant désormais passé sous la barre des 4%. Enfin, le béta retenu pour les activités plutôt monopolistiques est généralement inférieur à 1 (de l'ordre de 0.8-0,9). Avec un taux d'imposition des bénéfices qui oscille entre 30 et 40% selon les pays, on obtient un coût du capital voisin de 10% ( $\pm 2\%$ ). Le coût des capitaux propres est donc supérieur, entre 12 et 15%, celui de la dette très inférieur, fréquemment entre 4 et 8%.

Pour les firmes non cotées, il est nécessaire d'estimer un béta. On collecte alors les béta de firmes comparables, cotées, qui permettront d'obtenir une estimation du béta sous l'hypothèse d'une analogie de situations (par moyenne de ces comparables). Pour que la comparaison soit pertinente, le béta étant supposé fonction de la structure  $V_{CP}/V_D$  de la ressource financière de la firme, il est nécessaire de corriger les béta des comparables de cette structure (dits béta unlevered), en faire la moyenne et recalculer le béta de la firme analysée en lui appliquant la correction liée à sa structure financière:

<sup>36</sup> Cette information est par exemple fréquemment fournie sur les sites tels que Google Finance ou Yahoo! Finance pour les entreprises ayant un historique de cotation substantiel.

$$\beta_{unlevered} = \beta \cdot \frac{1}{1 + (1 - IS) \cdot \frac{V_D}{V_{CP}}}$$

L'estimation du coût du capital s'opère pour la firme dans son ensemble, car l'usage du capital appelé par la firme reste à sa discrétion. Toutefois, la nature de l'activité régulée, si elle est relativement "protégée", peut conduire à retenir un bêta plus faible que celui qui résulte de l'ensemble des activités de l'exploitant. Pareil raisonnement peut être tenu pour la prime de dette.

Rappelons qu'il existe un lien direct entre le coût du capital et la valeur de la firme:  $V = \frac{E(X)}{k_r}$ , X étant les bénéfices futurs. Plus k est élevé (plus la firme est risqué), plus faible est la valorisation de la firme. Plus k est faible (moins la firme est risqué), plus élevé est sa valorisation. Il est ainsi possible, à partir des cessions d'entreprises d'en déduire le coût implicite du capital révélé par la valeur de cession: ces valeurs confirment peu ou prou généralement la validité du coût du capital calculé selon le MEDAF.

### 3 LA MISE EN ŒUVRE DE L'ORIENTATION VERS LES COÛTS

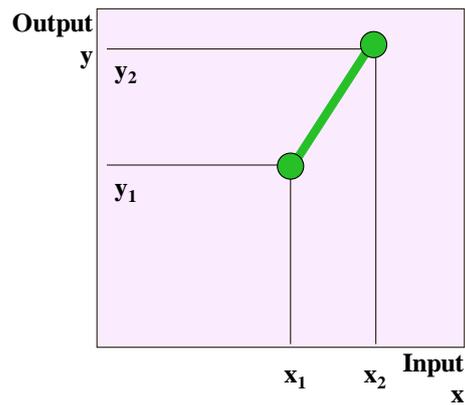
Nous avons identifié les soubassements théoriques de l'orientation vers les coûts et le dilemme auquel font face les régulateurs: accroître l'efficacité des opérateurs tout en limitant la rente de monopole. Pour ce faire, les régulateurs doivent réduire l'asymétrie d'information qui caractérise fréquemment leur situation vis-à-vis des opérateurs, en développant une véritable expertise sur les coûts de leurs régulés. Ils peuvent alors tenter de fixer un prix qui reflète les coûts, contraignant les opérateurs à être plus efficaces s'ils veulent conserver un peu de rente de monopole. Pour que l'incitation soit efficace, il faut sans doute abandonner une partie de cette rente aux opérateurs.

Cette estimation des coûts, faite donc sur la base des coûts moyens incrémentaux de long terme, telle que précisée précédemment, peut soit s'opérer annuellement, de façon à suivre pas à pas l'industrie, soit s'opérer à intervalle régulier (3 ou 5 ans) de façon à donner une vision prospective des prix de gros plus stable aux opérateurs. Dans ce dernier cas, on parlera stricto-sensu de "price cap", orientation générale des prix de gros à partir d'une référence fixée en année 0, qui se traduira par une évolution de ces prix sur la période considérée fixée par le régulateur. Cette solution présente deux difficultés: la première est de savoir si l'évolution concerne chaque tarif pris isolément, ou un panier de services qu'il faut alors définir (avec le risque que les masses de chaque service d'un tel panier évoluent rapidement, influant sur l'action commerciale des opérateurs pour tirer les recettes vers une structure favorable dans le contexte du price cap établi). La deuxième difficulté est d'évaluer correctement ce price cap, qui peut être positif ou négatif et qui borne les évolutions tarifaires des prix de gros.

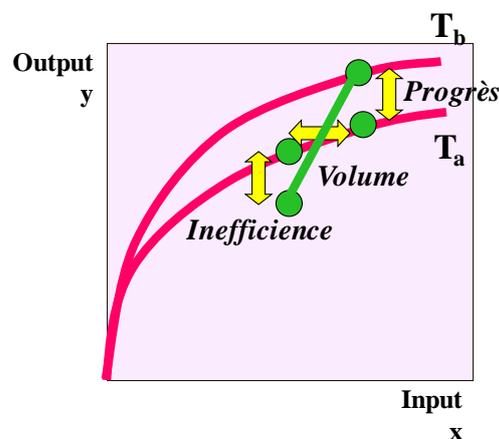
Il nous faut revenir ici à l'analyse dynamique des productivités. Soit deux périodes 1 et 2 sur lesquelles on observe une progression de la productivité d'une firme, mesurée par le rapport des outputs produits y aux inputs utilisés x:

$$r = \frac{\frac{dy}{y}}{\frac{dx}{x}} = \frac{\frac{y_2 - y_1}{y_1}}{\frac{x_2 - x_1}{x_1}} = \frac{\frac{y_2 - y_1}{x_2 - x_1}}{\frac{y_1}{x_1}}$$

Cette progression se constate ainsi sur un graphe de production:



mais, doit s'analyser en tenant compte des trois composantes de cette production repérées sur la représentation suivante:



Si l'on représente sur le graphe de production les deux fonctions de production efficiente aux deux périodes a et b, de technologie  $T_a$  et  $T_b$ , on peut distinguer trois effets:

1. Un gain d'efficacité qui rapproche la production de la frontière de production
2. Un gain lié au volume de production (effets d'échelle)
3. Un gain lié au déplacement de la frontière de production (progrès technique)

L'analyse de l'évolution des coûts doit donc faire la part de ces trois composantes: qu'est-ce qui est dû à l'échelle de production, à l'évolution des technologies et aux efforts de la firme? Dans la réflexion sur le partage de ces gains de productivité entre le consommateur (via la fixation des tarifs) et l'opérateur, doit entrer en ligne de compte la répartition de ces gains entre ces trois composantes.

### 3.1 APPROCHE STATIQUE DE L'ORIENTATION VERS LES COÛTS

La mise en œuvre concrète de l'orientation vers les coûts des tarifs de gros des marchés de télécommunications conduit donc à gérer un curseur entre les deux options présentées par le dilemme du régulateur: forcer l'efficacité ou extraire la rente. De façon générale, la tendance suivie par les régulateurs au fil, des ans a plutôt été de durcir la recherche d'efficacité en établissant des objectifs de coût contraignants.

Ce faisant, l'orientation vers les coûts présente dans les réglementations a conduit au calcul des coûts de façon individualisée pour chaque opérateur présent sur le marché. Cette évaluation individuelle des

coûts a pu conduire à une dispersion des coûts calculés selon les opérateurs, dispersion qui a soulevé de nouveaux problèmes dans la gestion des tarifs régulés. Assez logiquement en effet, compte tenu notamment des effets d'échelle présents sur les réseaux, même si ceux-ci sont minorés sur les réseaux mobiles par rapport à ce qu'ils pouvaient être sur les réseaux fixes, les coûts présentés par les petits réseaux, souvent les nouveaux entrants, s'avèrent plus élevés que ceux présentés par des réseaux établis depuis longtemps: aux conditions limites (le jour de l'ouverture d'un réseau), le coût de la première communication est quasi-infini! Faut-il alors retenir des coûts asymétriques entre réseaux ou privilégier des coûts symétriques ?

Nous avons évoqué l'importance de cette question sur les balances d'interconnexion. Elle est par ailleurs reliée à un autre débat repéré comme le "waterbed effect" (effet lit à eau): ce débat porte sur la question de savoir si une variation des tarifs régulés entraîne ou non une variation compensatrice des tarifs de détail d'un opérateur (y a-t-il vases communicants entre le niveau des tarifs de gros et de détail?)<sup>37</sup>.

Ces questions conduisent à des réponses évidemment liées à la nature concurrentielle du marché et au fonctionnement de ce marché. Si on considère que la recette provenant de chaque usager est la pondération des recettes issues des appels qu'il émet et qu'il reçoit, et que la concurrence se joue sur un tel prix moyen fictif, concurrentiel et donc "pris" par les opérateurs, alors, il existe un risque de compensation décrit par le waterbed effect. Si on considère que le marché concurrentiel est celui des appels émis (la concurrence se jouant sur le prix nominal de la communication), alors, ce tarif est "pris" par les opérateurs et ne peut être modifié par une baisse des recettes issue des tarifs régulés; dans le cas contraire, cela signifierait qu'il n'est pas "pris", mais "fait".

La constatation empirique du waterbed effect est difficile à vérifier, même s'il est probable qu'il s'est manifesté à certains moments du développement des réseaux. Si la littérature défend que cet effet survient en situation de concurrence, on peut défendre l'hypothèse inverse qu'il survient quand les opérateurs sont à même de "faire" les prix de détail puisque le waterbed effect suppose la capacité de relever un prix de détail face à une baisse des tarifs de gros.

L'analyse que nous avons proposée des balances d'interconnexion (dans un contexte dit two-ways, où il y a réciprocité des accès) suggère qu'un opérateur équilibre les tarifs régulés entre eux plutôt qu'avec les tarifs de détail: la recette des appels émis doit être minorée du coût des appels sortants, compensé par la recette des appels entrants. Une baisse des tarifs régulés correspond, soit à une réelle baisse des coûts, soit à une réduction des coûts pertinents pour l'interconnexion qui va de pair avec une baisse symétrique du coût des appels sortants, comme nous l'avons vu. Dans ces deux cas, le waterbed effect n'a pas lieu d'être.

Il ressort de ces considérations que la fixation de tarifs orientés vers les coûts doit tenir compte de la situation du marché et notamment de son climat concurrentiel, des disparités de situation des opérateurs présents (répartition de la dominance associée notamment au contrôle des externalités) et du niveau de développement du marché (saturation ou non).

Les effets d'échelle semblent s'atténuer assez rapidement dans l'univers technologique des réseaux d'aujourd'hui et deux opérateurs efficaces de taille différente peuvent présenter des coûts (assiette des coûts pertinents pour l'interconnexion) assez voisins. Ce qu'il est plus difficile d'atténuer est la captation des effets externes, qui se révèlent par exemple dans le taux des appels on-net sur le total des appels. Les opérateurs capables de capter ces effets externes présentent dès lors des balances de trafic très favorables difficiles à contrer. Une asymétrie tarifaire peut aider à nuancer cette captation de la valeur

---

<sup>37</sup> Cf. sur ce sujet Valletti et Genakos (2009)

des effets externes, mais elle ne peut durer, car elle revient à subventionner les abonnés du réseau discriminé par les abonnés des autres réseaux<sup>38</sup>.

En d'autres termes, la fixation d'un tarif symétrique, basé sur les coûts d'un opérateur efficient, conduit à donner une rente aux opérateurs en situation plus favorable (ce qui ne veut pas obligatoirement dit plus efficient) et une pénalité aux opérateurs en situation moins favorable (ce qui ne signifie pas automatiquement moins efficient).

La persistance de distorsions importantes de parts de marché ne peut donc justifier des asymétries importantes en matière de tarifs régulés. La solution provient plutôt de la baisse du niveau des tarifs, qui atténue les effets de ces distorsions en réduisant le niveau des balances d'interconnexion. Cette réduction est obtenue par la redéfinition des coûts pertinents pour l'interconnexion comme nous l'avons évoqué. Néanmoins, il faut avoir conscience des effets incitatifs produits par ces deux régimes alternatifs d'interconnexion que constituent d'un côté le régime des terminaisons d'appel et de l'autre le régime du BAK (peering):

1. abaisser les terminaisons d'appels, voire les annuler (BAK), incite les opérateurs à recruter préférentiellement des abonnés à forte contribution, en leur proposant éventuellement des tarifs binômes, voire des forfaits presque absolus; ce régime tend donc à délaisser les segments les moins solvables du marché, et délaisser également la couverture territoriale correspondant à ces segments; c'est un régime qui pousse à l'extension des trafics, moins à l'extension des raccordements et au déploiement des réseaux.
2. augmenter les terminaisons d'appel, voire étendre les coûts pertinents pour leur calcul jusqu'à la totalité des coûts imputables, incite les opérateurs à étendre leur marché aux populations les moins solvables, et aux territoires qui les desservent, mais incite moins au développement du trafic de ses abonnés; c'est un régime qui pousse à l'extension des raccordements, moins à l'extension des trafics.

Par ailleurs, l'abaissement des terminaisons d'appel, à travers la réduction de l'assiette des coûts pertinents, soulève la question des prix de raccordement aux réseaux: plus l'assiette des coûts pertinents est large, plus il est possible d'abaisser, voire d'annuler les prix de raccordement; plus cette assiette est réduite, plus se pose la question de la perception de prix de raccordement. Questions stratégiques et réglementaires se mêlent alors: un faible prix de raccordement permet à des populations faiblement solvables de se raccorder, comme nous venons de le souligner, mais peut gonfler les coûts commerciaux pour fidéliser des clients alors enclins à "zapper" entre réseaux; un faible prix de raccordement allège par ailleurs les politiques de service universel qui sont bâtis sur le niveau de ce prix. La répartition des coûts entre correspondants soulève donc également la problématique du recouvrement de ces coûts entre une composante fixe (raccordement et abonnement) et une composante variable (facturation des appels). La régulation de l'interconnexion ne peut totalement être dissociée de la régulation du service universel.

Le choix des tarifs régulés va donc bien au-delà de l'analyse des coûts. La fixation de ces tarifs est à cet égard une décision complexe de régulation concurrentielle des marchés, qui met en jeu deux curseurs corrélés, celui qui gère la force de la régulation (dilemme de la régulation) et celui qui gère l'alternative entre deux régimes d'interconnexion.

## 3.2 APPROCHE DYNAMIQUE DE L'ORIENTATION VERS LES COÛTS

<sup>38</sup> Le débat qui a entouré la discrimination positive en faveur de Bouygues dans la fin des années 2000 illustre les limites de cette discrimination (cf. l'analyse de l'Arcep - [http://www.arcep.fr/uploads/tx\\_gspublication/consult-ta-mobile-suite.pdf](http://www.arcep.fr/uploads/tx_gspublication/consult-ta-mobile-suite.pdf))

L'approche statique ne s'avère plus aujourd'hui suffisante pour réfléchir sur l'orientation vers les coûts. En effet, les réseaux se renouvellent, et l'investissement dans les nouveaux réseaux doit être pris en considération. Que ce soit dans les réseaux mobiles, avec l'apparition des générations successives (3G puis 4G et bientôt 5G), ou dans les réseaux fixes avec la substitution du cuivre par la fibre et l'émergence de réseaux très haut-débit, il s'avère que la tarification retenue pour les anciens réseaux influe fortement sur l'incitation à investir dans les nouveaux réseaux.

La situation est différente selon que l'on détient ou non des capacités dans l'ancienne technologie et selon que cette détention est symétrique ou non. Intuitivement, un accès peu onéreux aux anciens réseaux défavorise l'investissement de ceux qui bénéficient de cet accès, et favorise l'investissement de ceux qui subissent cette tarification faible. Inversement, un tarif d'accès élevé à l'ancien réseau défavorise l'investissement de ses détenteurs et favorise l'investissement de ceux qui y recourent. Ce raisonnement doit toutefois être nuancé selon la régulation de l'accès sur le nouveau réseau (ouverture et prix) et selon la difficulté éventuelle à couvrir certaines zones non rentables<sup>39</sup>.

Ces raisonnements ne tiennent que dans des situations dites one way, c'est-à-dire où la réciprocité des accès n'existe pas. Le niveau des tarifs est donc un niveau absolu et non un niveau relatif par rapport à des tarifs symétriques.

Les règles qui se mettent en place<sup>40</sup> conservent l'orientation vers les coûts des anciens réseaux, mais sans doute faudra-t-il mieux tenir compte des effets externes qui se jouent dans la migration. Là encore, le régulateur tient, dans ses décisions de tarification, des clés importantes liées à la vitesse de migration, au déploiement des réseaux et à la structure concurrentielle du secteur.

## 4 CONCLUSIONS

L'exigence d'orientation vers les coûts doit s'accommoder dans la pratique avec les objectifs assignés au régulateur, aménagement du territoire, évolution dynamique des réseaux, promotion de la concurrence, incitation à l'efficacité productive, etc.

La façon de concilier cette accommodation réside dans la conception des coûts retenus. L'évolution de la régulation sur les 20 dernières années montre en effet que les instances de régulation n'ont pas hésité à modifier leur conception des coûts pour permettre l'ajustement des tarifs orientés vers les coûts à des niveaux ou selon des structures permettant de concilier cette approche avec les autres objectifs de la régulation poursuivis.

Pendant des années, les réseaux fixes ont subventionné les réseaux mobiles (Canoy et al., 2006, p. 146) dans le sens où les tarifs d'interconnexion furent très largement supérieures aux "coûts", même fréquemment supérieurs aux coûts de détail des services. Dans le cadre d'une tarification à la Ramsey, cette "subvention" peut être vue comme une subvention croisée entre utilisateurs, les abonnés fixes étant moins élastiques que les autres usagers des réseaux mobiles. Elle peut aussi être vue comme le

<sup>39</sup> Il faut noter que ces zones non rentables peuvent ne pas être celles qui sont habituellement considérées comme telles, à savoir les zones de faible densité de raccordement. La demande semble en effet être relativement forte dans ces zones où le débit offert par les anciens réseaux est bien moindre.

<sup>40</sup> Cf. par exemple la recommandation de la Commission des Communautés Européennes du 20 septembre 2010 sur l'accès réglementé aux réseaux d'accès de nouvelle génération (NGA) disponible sur <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:251:0035:0048:FR:PDF>

signe de la dominance des opérateurs mobiles, les abonnés du fixe n'ayant pas d'autre solution que de payer (ce qui confirme leur inélasticité)<sup>41</sup>.

Les efforts de la Commission Européenne pour contraindre l'orientation vers les coûts des tarifs régulés trouve son terme dans l'objectif assigné aux régulateurs de parvenir à cette orientation au plus tard le 31 décembre 2012 pour les régulateurs efficaces (1<sup>er</sup> juillet 2014 pour les autres). Bien que cet objectif soit loin d'être atteint, la Commission tient une position extrêmement ferme sur ces sujets<sup>42</sup>.

Pour concilier cette "harmonisation" avec le soutien d'un marché concurrentiel, le niveau des coûts retenus dans le cadre des services régulés symétriques (two-way) a été peu à peu abaissé sans pour autant franchir le pas du peering, dont la Commission avait exigé au début des années 2000, l'abandon. Compte tenu de la maturité du marché, cet abaissement atténue les distorsions concurrentielles qui peuvent résulter de pouvoirs de marché associés aux effets club générés par les opérateurs.

Cette évolution s'est également accompagnée de la disparition progressive de la discrimination souvent établie à l'origine selon le statut de l'opérateur interconnecté et notamment de l'ampleur des investissements qu'il effectuait dans son infrastructure: un opérateur de réseau public disposait ainsi de tarifs plus favorables qu'un simple opérateur de services, par exemple un opérateur de réseau virtuel. L'accès est ainsi devenu indifférencié pour tous les opérateurs de services de télécommunications. Reste néanmoins une zone grise qui est celle des gros utilisateurs de réseau pouvant avoir intérêt dans certaines situations à demander le statut d'opérateur.

Enfin, il faut noter que toute cette régulation tarifaire résulte de l'existence d'une ressource essentielle non duplicable. Celle-ci résulte dans le champ des réseaux du constat qu'un usager ne se raccorde que sur un seul réseau, car ce raccordement est coûteux. Or, le déploiement des réseaux d'accès radio tend à rendre le coût de ce raccordement négligeable: on le constate dans les pays en développement dont les marchés se développent principalement sur des raccordements à prépaiement et dans lesquels il n'est pas rare de voir une proportion non négligeable du marché multiraccordée. Il n'en reste pas moins que l'usager reste joignable via un numéro qui représente son raccordement sur un réseau donné. Cet abonné peut néanmoins potentiellement être joint sur plusieurs numéros. Il suffirait alors de constituer une base de données formant annuaire des personnes avec les différents numéros sur lesquels ils peuvent être joints. Grâce à une telle base de données, un réseau cherchant à délivrer un appel pourrait mettre en concurrence les réseaux de raccordement et retenir celui offrant le meilleur prix, celui vraisemblablement couvrant son coût incrémental. Cette option toute théorique permet d'illustrer la notion de ressource essentielle et saisir à rebours la justification d'orientation vers les coûts, et l'importance de la caractérisation de ces coûts.

Comme tout dispositif réglementaire, l'orientation vers les coûts des tarifs régulés, ceux de l'accès à une ressource essentielle, reste subordonnée aux objectifs fréquemment contradictoires assignés aux régulateurs. L'orientation vers les coûts traduit donc un arbitrage stratégique entre ces objectifs et en ce sens, l'analyse des coûts se traduit in fine en signaux adressés aux marchés.

---

<sup>41</sup> Ils pourraient autrement, soit ne pas passer leurs appels, soit souscrire à un mobile, le prix de détail on-net sur les réseaux mobiles étant alors fréquemment inférieur au prix de terminaison du fixe sur les mobiles.

<sup>42</sup> Cf. la communication de la Commission sur l'examen de la situation réglementaire (Com(2010)271 du 1/06/2010) disponible sur [http://ec.europa.eu/information\\_society/policy/ecom/doc/implementation\\_enforcement/eu\\_consultation\\_procedures/communication\\_art7\\_en.pdf](http://ec.europa.eu/information_society/policy/ecom/doc/implementation_enforcement/eu_consultation_procedures/communication_art7_en.pdf)

- Babusiaux D. (1990) Décision d'investissement et calcul économique dans l'entreprise, *Economica*
- Banque Mondiale (2003) Modèle de détermination des tarifs d'interconnexion, Washington, PPIAF-WB.
- Boyer M., M. Moreaux et M. Truchon (2006) Partage des coûts et tarification des infrastructures, Centre Interuniversitaire de recherche en Analyse des Organisations (CIRANO), Montréal, disponible à <http://www.cirano.qc.ca/pdf/publication/2006MO-01.pdf>
- Canoy M., P. de Bijl & R. Kemp (2004) *Access to telecommunications network*, in *The Economics of Antitrust and regulation in Telecommunications*, Ed. by Pierre Buigues & Patrick Rey, Edward Elgar Publishing, pp. 135-168
- Gasmi F., D.M. Kennet, J.J. Laffont & W.W. Sharkey (2002) *Cost Proxy Models and Telecommunication Policy*, A new empirical approach to regulation, The MIT Press
- Genakos C. et T. Valletti (2009) Testing the “Waterbed” Effect in Mobile Telephony, disponible à <http://www.sel.cam.ac.uk/Genakos/Genakos%20Valletti-Testing%20Waterbed%20Effect.pdf>
- Laffont J.J. et J. Tirole (2000) *Competition in Telecommunications*, The MIT Press
- Rey P. et J. Tirole (2007) A primer on foreclosure, in *Handbook of Industrial Organization*, Vol. 3, edited by M. Armstrong and R. Porter, Elsevier, pp. 2145-2220.
- Riveline C. (2005) *Evaluation des coûts*, *Eléments d'une théorie de la gestion*, Les Presses Mines Paris
- Tera Consultants & Hogan Lovells (2010) *Study on the future interconnection charging methods*, étude pour la Commission Européenne, disponible à [http://ec.europa.eu/information\\_society/policy/ecomm/doc/library/ext\\_studies/2009\\_70\\_mr\\_final\\_study\\_report\\_F\\_101123.pdf](http://ec.europa.eu/information_society/policy/ecomm/doc/library/ext_studies/2009_70_mr_final_study_report_F_101123.pdf) (rapport rédigé par Michel Rogy).