HAL open science

Analyses quantitatives de décisions de justice en matière de Prestation Compensatoire (PC) dans une perspective de justice prédictive

Bruno Jeandidier, Jean-Claude Ray, Julie Mansuy

To cite this version:

Bruno Jeandidier, Jean-Claude Ray, Julie Mansuy. Analyses quantitatives de décisions de justice en matière de Prestation Compensatoire (PC) dans une perspective de justice prédictive. 2020. hal-02973071

HAL Id: hal-02973071

https://hal.science/hal-02973071v1

Preprint submitted on 20 Oct 2020

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BETA
Bureau d'économie théorique et appliquée

Documents de travail

« Analyses quantitatives de décisions de justice en matière de Prestation Compensatoire dans une perspective de justice prédictive »

Auteurs

Bruno Jeandidier, Jean-Claude Ray, Julie Mansuy

Document de Travail n° 2020 – 05

Février 2020

Bureau d'Économie Théorique et Appliquée BETA

www.beta-umr7522.fr

@beta_economics

Contact :
jaoulgrammare@beta-cnrs.unistra.fr

Analyses quantitatives de décisions de justice en matière de Prestation Compensatoire dans une perspective de justice prédictive

Bruno Jeandidier, Jean-Claude Ray, Julie Mansuy
(BETA, UMR CNRS 7522, Université de Lorraine, INRAE, Université de Strasbourg, AgroParisTech)

Résumé :

A partir d'un corpus de décisions de justice de divorce de première instance codifiées dans une base de données, représentatif au niveau national français, nous menons une analyse économétrique qui a pour finalité de prédire les montants de prestation compensatoire fixés par les juges aux affaires familiales des Tribunaux de Grande Instance. Cet exercice nous donne l'occasion, premièrement, de souligner les difficultés méthodologiques, inhérentes à ce type de source juridique, liées à l'exercice prédictif, notamment : la nécessité de construire l'information non directement accessible dans les décisions à partir d'autres informations nombreuses et complexes ; l'absence de certaines informations dans les décisions qui pourtant, au regard du droit, devraient a priori constituer de bons prédicteurs ; l'existence dans les décisions d'informations entachées de données manquantes qui font courir le risque de biais ; la nécessité de décomposer l'analyse au regard de la procédure (i.e. avec ou non accord des parties). Deuxièmement, dans la mesure où les propositions des parties (montants de l'offre et de la demande de prestation compensatoire) expliquent l'essentiel du montant de PC fixé par le juge en cas de désaccord des parties et absorbent une large part du rôle effectif des déterminants listés dans le Code civil, nous développons une approche originale consistant à retenir les résidus des deux équations d'offre et de demande, et non pas les montants de ces dernières, dans l'équation du montant de PC fixé par le juge, ce qui permet de restituer le vrai rôle de ces déterminants. Ainsi, nous montrons que certains facteurs ont un rôle déterminant, notamment l'écart intra-couple de niveau de vie, la durée de mariage, le régime matrimonial ou encore la forme de la prestation compensatoire (rente ou capital). Troisièmement, à partir des coefficients estimés des deux régressions les plus élaborées (sur le cas des affaires sans accord des parties), nous mesurons l'ampleur des erreurs de prédiction en comparant les montants de prestation compensatoire prédits aux valeurs observées. Notre conclusion est que l'ampleur de ces erreurs (en euros comme en proportion) est telle qu'il serait très risqué de recourir à ces estimations pour effectuer de la justice prédictive à destination des couples en instance de divorce. En effet, dans un cas sur quatre, l'erreur de prédiction excède la moitié du montant observé de la PC.

Mots-clés : divorce, prestation compensatoire, justice prédictive, erreur de prédiction.

Codes JEL : K36, J12.

Travaux réalisés dans le cadre de l'atelier de recherche « E-juris » financé par la Maison des Sciences de l'Homme de l'Université de Lyon et avec le support financier du Pôle Scientifique SJPEG de l'Université de Lorraine (projet « Economie et Justice : de nouveaux défis méthodologiques »).

La perspective qu'ouvre la Loi pour une République numérique (dite Loi Lemaire, 2016) dans le champ du droit, à savoir la mise en accès libre de l'intégralité des décisions de justice françaises numérisées dans un avenir proche, donnerait une certaine crédibilité aux initiatives privées des quelques *Legal-Tech* françaises qui se sont lancées récemment dans l'activité de « justice prédictive ». Parallèlement, l'accès à cette masse considérable d'informations pourrait être de nature à modifier les méthodes de recherche en sciences sociales s'appliquant à étudier quantitativement les pratiques judiciaires par le biais d'analyses statistiques du contenu de ces décisions. On peut cependant s'interroger sur les réelles possibilités qu'offrira le recours à ces bases de données gigantesques. Pour cheminer dans cette interrogation, nous proposons dans cet article de rendre compte des analyses quantitatives que nous avons menées dans une perspective de « justice prédictive », à titre d'exemple, à partir d'un corpus de décisions, représentatif au niveau national, limité en nombre, codé et saisi manuellement. Celui-ci porte sur la Prestation Compensatoire (PC) fixée lors de procédures de divorce[1]. Ainsi, sans tenir compte des capacités spécifiques que les techniques de fouille de données et d'intelligence artificielle peuvent fournir, nous menons nos analyses traditionnellement et ce, en particulier, pour souligner les limites et difficultés que nous avons rencontrées et avec comme perspective de poser la question de savoir si ces techniques associées aux *big data* seraient susceptibles de dépasser de telles limites (ou si, au contraire, elles risquent d'être confrontées aux mêmes obstacles).

1. A quelles questions de prédiction répondre ?

Si l'on parcourt, autant que faire se peut gratuitement, les sites des quatre ou cinq principales Le*gal Tech* de « justice prédictive » française, on peut schématiquement présenter les questions de prédiction proposées selon trois types d'offre :

- estimer une **probabilité** (ou un pourcentage) **de succès** dans une procédure judiciaire ;
- estimer une **valeur** issue de la décision (par exemple un montant d'indemnité ou une durée d'emprisonnement) ;
- identifier les **arguments juridiques** (ou contextuels) qui font pencher la balance vers le succès.

Notons que souvent l'analyse est décomposée par niveau de juridiction, par lieu géographique de la juridiction et par année. Dans quelle mesure pouvons-nous nous prêter aux mêmes exercices à partir de la base de données à notre disposition ?

Tout d'abord, le niveau de juridiction (TGI) et l'année (2013) sont imposés par les données à notre disposition. Ensuite, la question de l'identification des arguments pertinents est, nous semble-t-il, pour l'essentiel hors de portée car la base de données n'a pas été conçue dans cette perspective. En revanche, la question de l'estimation d'un taux de succès peut être envisagée. En effet, la base de données, constituée de 5 453 décisions de divorce, contient 3 203 affaires pour lesquelles la question de l'octroi d'une PC s'est posée et dans seulement 2 678 d'entre elles *in fine* une PC a été fixée (ou homologuée en cas d'accord) par le juge. Il est donc possible de calculer un taux de succès et de répondre, en comparant les deux types d'affaires, à une question prédictive de type : « *compte tenu de ma situation et de celle de mon conjoint, si je demande une PC (si mon conjoint demande une PC), est-ce que j'ai une bonne chance de l'obtenir (est-ce que je risque de devoir la payer) ?* ». De plus, comme la base de donnée contient à la fois des affaires où la question de la PC se pose (3 203) et d'autres où elle ne se pose pas (2 250)[2], il est également envisageable, par comparaison des deux

[1] Il s'agit de la base de données COMPRES collectée à l'occasion d'une recherche menée antérieurement dans le cadre d'un projet ANR. La collecte est le fruit d'une collaboration entre le Bureau d'Economie Théorique et Appliquée (BETA, UMR CNRS 7522, Université de Lorraine, INRAE) et la Sous-direction de la Statistique et des Etudes du Ministère de la Justice. Le lecteur peut trouver une courte présentation de la méthodologie de construction de cette base de données dans BOURREAU-DUBOIS Cécile, JEANDIDIER Bruno, MANSUY Julie (2018), « Les enjeux redistributifs de la prestation compensatoire : une analyse statistique de 5 000 décisions de divorce », *In* Bourreau-Dubois C., Sayn I. (Dir.), *Le traitement juridique des conséquences économiques du divorce,* Ed. Bruylant, pp. 127-150.

[2] La base de données ayant été constituée selon une méthode d'échantillonnage à probabilités inégales, il ne faut pas conclure de ces effectifs qu'il y a plus de divorces avec demande de PC ; en corrigeant par la pondération qui tient compte de cette inégalité de tirage, la proportion de divorces avec demande de PC est estimée à 23 %.

sous-ensembles, de traiter une autre question prédictive de type : « *Je veux divorcer, est-ce que j'ai le « profil » pour demander une PC (compte tenu de la situation de notre couple, est-ce qu'il y a un risque que mon conjoint demande une PC) ?* ».

Enfin, notre base de données permet également de traiter une question en termes de prédiction d'une valeur quantitative, à savoir ici bien sûr le montant de PC fixé par le juge[3]. L'estimation de ce montant permet en effet de répondre à la question que peut se poser une personne en instance de divorce de type : « *Lorsque le juge va décider d'octroyer la PC, compte tenu de la situation de notre couple, combien puis-je espérer recevoir (combien puis-je craindre de devoir payer) ?* ». Nous verrons d'ailleurs *infra* que la question pourrait également se décliner en intégrant les montants des propositions des parties dans les caractéristiques de la situation du couple et ainsi permettre de répondre à une question de prédiction plus élaborée de type : « *Si je demande tant, et que mon conjoint propose tant, compte tenu de notre situation de couple, combien puis-je espérer recevoir (craindre devoir payer) ?* ».

Il faut souligner que la manière selon laquelle nous venons de formuler, à titre d'exemple, ces questions de prédiction repose sur une volonté d'estimation multicritères (*cf. supra* les incises « *compte tenu de la situation de notre couple* »). D'une certaine manière, cette volonté s'écarte de ce que proposent apparemment les *Legal-Tech* de justice prédictive qui, souvent, ne proposent que le calcul d'un unique montant moyen (ou d'un unique pourcentage de succès), éventuellement décliné par année et par juridiction. Or, nous pensons que ces indicateurs globaux sont peu informatifs tant, dans le cas de la PC, les décisions sont hétérogènes. Par exemple, si, dans notre base de données, la moyenne des PC fixées par le juge est égale à 59 370€, en fait les montants varient sensiblement (20 % des PC fixées par le juge sont inférieures à 10 000€ et 20 % sont supérieures à 70 000€).

Pour être plus informatif, il conviendrait donc de calculer des moyennes conditionnelles croisant différentes caractéristiques (par exemple selon la durée de mariage, le niveau de disparité de niveau de vie, etc.). Or, l'exercice n'est pas aisé. En effet, d'une part, il nécessite de détenir un nombre considérable de décisions, d'autre part, se pose la question du choix des caractéristiques à croiser. Imaginons que l'on retienne quatre caractéristiques binaires et quatre caractéristiques continues (revenus et propositions de deux conjoints) réduites à cinq classes de valeurs, cela amène à devoir calculer 10 000 moyennes conditionnelles. Pour que ces moyennes fassent sens, il convient que dans chaque cellule il y ait un nombre d'affaires minimum, disons 20 en moyenne, soit au total 200 000 décisions. Sachant qu'il y a environ 120 000 prononcés de divorces par an et que des PC sont décidées dans environ 20 % des cas, il conviendrait d'accumuler l'exhaustivité des décisions de PC pendant environ huit ans pour permettre de tels calculs de moyennes conditionnelles. Quant au choix des caractéristiques, il ne peut pas être que guidé par des hypothèses *a priori*, il requiert une analyse statistique permettant d'identifier les caractéristiques qui, prises en compte simultanément, sont effectivement discriminantes. Ce raisonnement milite donc plutôt pour un recours à une approche économétrique que pour une approche par moyennes conditionnelles.

2. Estimer, pour le prédire, le montant de Prestation compensatoire fixé par le juge

Pour mener cette analyse d'estimation, nous nous limitons aux seules affaires pour lesquelles un montant de PC a été fixé par le juge, soit 2 678 affaires. Plus exactement, nous avons limité notre analyse aux seules affaires où une PC a été fixée au bénéfice de l'épouse. Nous excluons donc 118 affaires où la PC en cause est au bénéfice de l'époux. En effet, nous avons pensé que ces rares affaires où c'est l'époux qui est créancier de la PC doivent être assez spécifiques et donc de nature à perturber l'estimation du montant de PC fixé par le juge[4]. Sur ce sous-ensemble d'affaires, envisageons successivement les informations qu'il conviendrait de mobiliser pour mener à bien une estimation de ce montant de PC fixé par le juge, à savoir la variable dépendante, puis les variables

[3] Il serait également possible d'estimer le montant des Contributions à l'Entretien et l'Education de l'Enfant (CEEE), de la pension alimentaire fixée lors de l'ordonnance de non-conciliation ou le montant des indemnités de dommages et intérêts.

[4] Cette hypothèse n'a pas été testée, mais si elle devait être invalidée, le fait d'ajouter 118 observations supplémentaires ne serait probablement pas de nature à modifier sensiblement les estimations et les conclusions que nous effectuons sur le sous-ensemble des épouses bénéficiaires d'une PC.

explicatives ou tout du moins les facteurs d'hétérogénéité pouvant constituer autant de prédicteurs pertinents.

2.1. La variable dépendante : le montant total de prestation compensatoire fixé par le juge

Mesurer le montant de PC fixé par le juge constitue une première difficulté dans la mesure où cette information n'est pas univoque dans les décisions ; c'est une information qu'il convient donc de construire en combinant plusieurs informations, ce qui, selon une approche automatisée peut constituer une réelle difficulté. En effet, si dans une proportion importante (90 %) la PC est fixée sous forme d'un capital unique (parfois selon une logique de soulte) assez facilement détectable dans les décisions, en revanche dans certains cas, la PC est fixée sous forme de rente mensuelle. Sauf à envisager des estimations séparées selon ces deux formes de PC[5], pour homogénéiser la mesure il convient donc, en présence de rente, dans le cas le plus simple, de combiner le montant de la rente mensuelle avec la durée du versement fixée par le juge. Mais dans certains cas, il s'agit d'une rente viagère, la durée n'est donc pas fixée et il convient alors de l'estimer (dans notre cas, nous avons eu recours à des statistiques d'espérance de vie en fonction de l'âge du bénéficiaire). Notons que dans quelques cas rares (2 %), la décision combine les deux formes de PC, ce qui complexifie la construction de la variable dépendante.

Comme on peut penser que la forme de la PC choisie par le juge (ou sollicité par les parties) ait un impact sur son montant, puisque rente et versement en capital ne répondent pas exactement à la même logique, il conviendra de tenir compte de cette information qualitative en termes de prédicteur du montant. Comme on le voit, mesurer le montant de PC à partir de décisions de justice n'est pas chose aisée et pose la question de savoir si une méthode automatisée appliquée à un corpus de très grande taille serait en capacité de réaliser cette opération de manière fine et fidèle[6].

Enfin, notons que la nécessité de combiner plusieurs informations (montant du capital, montant de de la rente, durée de versement de la rente, âge du bénéficiaire…) amène à nous confronter à l'existence de données manquantes partielles et à d'éventuels biais de sélection (*cf. annexe n°2*), les décisions n'étant pas toutes rédigées avec le même niveau de précision. Dans le cas présent, ces données manquantes relatives aux informations nécessaires à la construction de la variable dépendante nous amènent à devoir exclure 74 affaires.

2.2. Quelles variables indépendantes retenir pour estimer le montant de PC ?

En matière de prédiction, toute information permettant d'améliorer la qualité de la prédiction est bonne à retenir. On pourrait, sur la base de ce principe, mobiliser la totalité des informations (plusieurs centaines de variables) contenues dans la base de données, mais nous avons préféré procéder différemment en sélectionnant les informations qui nous semblaient analytiquement *a priori* les plus pertinentes avant d'en tester la pertinence statistique. Cette sélection repose principalement sur l'énoncé des deux principaux articles du Code civil relatifs à la PC. Reprenons-les successivement.

- Article 270 : « *L'un des époux peut être tenu de verser à l'autre une prestation destinée à compenser, autant qu'il est possible, la disparité que la rupture du mariage crée dans les conditions de vie respectives. (…) Toutefois, le juge peut refuser d'accorder une telle prestation si l'équité le commande, soit en considération des critères prévus à l'article 271, soit lorsque le divorce est prononcé aux torts exclusifs de l'époux qui demande le bénéfice de cette prestation, au regard des circonstances particulières de la rupture* ».

Outre la question des torts exclusifs de l'époux, qui porte plus sur l'éligibilité à la PC que sur le montant de la PC, l'information principale à retenir de l'article 270 du Code civil, et à rechercher dans les décisions, est celle relative à la « *disparité dans les conditions de vie respectives* ». Il convient donc de sélectionner de l'information permettant de mesurer une disparité d'une part, et des conditions

[5] Mais les PC sous forme de rente sont peu nombreuses et effectuer des estimations sur des échantillons de taille limitée est généralement peu pertinent en termes de robustesse des estimateurs.

[6] Un raisonnement identique à celui portant sur le montant de PC fixé par le juge pourrait être tenu à propos des montants de propositions de PC des parties.

de vie d'autre part. Comme l'article 271 (*cf. infra*) évoque par ailleurs des « besoins » de l'époux et des « ressources » de l'autre, il est assez explicite que la notion de conditions de vie peut être assimilée au concept économique de niveau de vie (revenu par unités de consommation, ces dernières exprimant des besoins). Pour autant, il ne serait pas inconsidéré de retenir les simples ressources des époux pour mesurer les conditions de vie. Concernant la disparité, on peut classiquement retenir soit la différence, soit le rapport intra-couple. De cette analyse, nous avons donc retenu plusieurs indicateurs : la différence ou le rapport de revenu total entre les deux époux, la différence ou le rapport de revenu d'activité entre les deux époux, la différence ou le rapport de niveau de vie[7] entre les deux époux[8]. C'est l'analyse statistique qui, ensuite, a permis de départager le meilleur indicateur en termes de pouvoir prédictif du montant de la PC (*cf. infra*). Notons que ces indicateurs sont construits au moment du divorce, la référence à « l'avenir prévisible » évoqué à l'article 271 ne pouvant pas être pris en compte finement à partir des informations contenues dans la base de données[9].

La construction de ces indicateurs nécessite de mobiliser des dizaines d'informations contenues dans les décisions : pour chacun des conjoints, il faut repérer tous les types de revenus mentionnés dans le descriptif de la situation (la grille de saisie COMPRES distingue dix types possibles de revenus pour chacun des conjoints et certains types agrègent des revenus qui peuvent prendre des dénominations différentes dans les décisions), vérifier l'unité de temps (mensuel ou annuel) des montants déclarés et les additionner de manière adéquate avant de procéder à la différence ou le rapport intra-couple ; et comme mentionné en note de bas de page, la construction des indicateurs de niveau de vie nécessite de mobiliser le nombre et l'âge des enfants, leur lieu d'hébergement et leur type de droit de visite et d'hébergement, etc. Il s'agit donc d'un ensemble d'opérations de détection et de combinaison d'informations que des traitements automatisés pourraient peut-être réaliser sur des échantillons de plus grosse taille, mais le pourraient-ils aisément ?

- Article 271 : « *La prestation compensatoire est fixée selon les besoins de l'époux à qui elle est versée et les ressources de l'autre en tenant compte de la situation au moment du divorce et de l'évolution de celle-ci dans un avenir prévisible. A cet effet, le juge prend en considération notamment*… (*cf.* tableau ci-dessous) ».

Outre les notions de ressources et de besoins des époux, dont nous avons déjà fait mention à propos de l'article 270 du Code civil mais qui ici ne sont pas exprimées en termes de disparité, l'article 271 propose une liste (non limitative, *cf.* « *notamment* ») de facteurs pouvant concourir à la détermination du montant de PC, ce sont donc a priori de bons prédicteurs de ce montant. Nous indiquons dans le tableau ci-dessous dans quelle mesure ces indicateurs peuvent être mobilisés à partir de notre base de données et donc à partir des décisions de justice.

« A cet effet, le juge prend en considération notamment…	
la durée du mariage,	Information plutôt bien renseignée dans les décisions ; à construire à partir de la date de mariage et de la date de la procédure de divorce.
l'âge et l'état de santé des époux,	L'âge des époux est également plutôt bien renseigné (date de naissance) ; elle est fortement corrélée avec la durée de

[7] L'indicateur de niveau de vie est le rapport entre le revenu total de l'individu et un nombre d'unités de consommation. Nous avons retenu la situation post-divorce, en considérant donc que les deux époux sont séparés, d'où l'attribution d'une unité de consommation à chacun d'entre eux. A cette unité de consommation sont éventuellement ajoutées des unités de consommation au titre des enfants (0,5 unité au-delà de 14 ans ; 0,3 unité en-dessous de 15 ans) en tenant compte du lieu d'hébergement principal décidé par le juge chez l'un ou l'autre des deux époux (en cas d'hébergement alterné ou lorsque qu'aucune indication sur l'hébergement d'un enfant majeur n'est mentionnée, les unités de consommation des enfants sont partagées à part égale entre les deux époux). Nous n'avons cependant pas tenu compte de l'éventuelle remise en couple, car si cette dernière est parfois précisée dans les décisions, en revanche il n'est pas mentionné si le nouveau conjoint a ou non des enfants ; l'information étant imparfaite, nous avons préféré nous abstenir.

[8] Et nous avons testé alternativement soit une simple relation linéaire soit une relation quadratique entre ces indicateurs de disparité et le montant de PC fixé par le juge.

[9] Au titre des revenus et des conditions de vie, nous avons également retenu comme information le fait que le créancier de la PC bénéficie, seul ou simultanément au débiteur, à taux plein ou à taux partiel, de l'Aide juridictionnelle ; cet indicateur permet de tenir compte spécifiquement de situations de besoins (très faibles ressources) assez prononcés.

	mariage. L'état de santé est une information très parcellaire dans les décisions au niveau du descriptif de l'affaire ; il est cependant possible d'identifier l'existence d'une dimension médicale en se fiant aux motivations du juge[10].
leur qualification et leur situation professionnelles,	Il est assez rare de trouver dans les décisions de l'information sur la qualification des époux et dans une moindre mesure sur leur situation professionnelle ; on peut cependant déduire des revenus d'activité le fait que les époux sont ou non en activité, mais les revenus souffrent d'un nombre important de données manquantes. Là encore, à défaut d'information systématique dans l'exposé de l'affaire, on peut se fier aux motivations du juge en ce qui concerne les qualifications[11].
les conséquences des choix professionnels faits par l'un des époux pendant la vie commune pour l'éducation des enfants et du temps qu'il faudra encore y consacrer ou pour favoriser la carrière de son conjoint au détriment de la sienne,	Dans les décisions, on ne retrouve pas cette information de manière explicite et précise ; elle est évoquée de manière plutôt implicite. Pour pallier cette difficulté, nous avons simplement retenu des indicateurs relatifs aux enfants : le nombre d'enfants total, le nombre d'enfants à charge au moment du divorce, la présence d'enfants en bas âge… Et l'on a, à nouveau, retenu les motivations du juge lorsqu'il évoque explicitement la trajectoire professionnelle de la femme[12].
le patrimoine estimé ou prévisible des époux, tant en capital qu'en revenu, après la liquidation du régime matrimonial,	Pour tenir compte du patrimoine, nous avons retenu une approche similaire à celle relative aux conditions de vie et avons calculé le montant de patrimoine de chacun des époux ainsi que la différence entre ces deux patrimoines[13]. Mais *in fine* cette information souffre de nombreuses données manquantes. Nous avons également retenu le régime matrimonial comme indicateur en considérant qu'un régime moins favorable à l'épouse (séparation de biens et participation aux acquêts) pouvait avoir un effet positif (compensatoire) sur le montant de la PC.
leurs droits existants et prévisibles,	Il est rare de trouver des indications quantitatives de ressources prévisibles dans les décisions, mais l'on peut identifier des informations plus qualitatives de type « prévision de variation des ressources à la hausse, ou à la baisse » pour chacun des conjoints.
leur situation respective en matière de pensions de retraite en ayant estimé, autant qu'il est possible, la diminution des droits à retraite qui aura pu être causée, pour l'époux créancier de la prestation compensatoire, par les circonstances visées au sixième alinéa ».	Comme pour les choix professionnels (*cf. supra*), dans les décisions, on ne retrouve pas cette information de manière explicite et précise. En conséquence nous avons retenu comme indicateur le fait que le juge souligne cette dimension dans ses motivations, soit en niveau (faibles droits à la retraite) soit en inégalité (droits moindres pour tel époux comparativement à l'autre époux).

[10] Le juge souligne dans ses motivations que la santé de l'époux/l'épouse est dégradée.

[11] Le juge souligne dans ses motivations que l'époux/l'épouse a (n'a pas) de formation et/ou de qualification et/ou d'expérience professionnelle.

[12] Le juge souligne dans ses motivations que l'épouse a consacré du temps à l'éducation de ses enfants, au détriment (ou non) de sa carrière, que l'épouse a fait un choix professionnel permettant de favoriser la carrière de son conjoint, que l'épouse a collaboré à l'activité de son conjoint, que l'épouse a fait un choix professionnel personnel, que l'épouse a fait un choix professionnel en commun avec son conjoint.

[13] Pour chaque époux, le patrimoine est égal à son patrimoine propre auquel nous ajoutons, d'une part, la part du patrimoine commun réparti et attribué à tel époux et, d'autre part, la moitié du patrimoine commun non réparti.

Les brefs commentaires rassemblés de ce tableau montrent donc que l'identification, dans les décisions, des facteurs *a priori* prédictifs n'est pas chose aisée dans un certain nombre de cas.

Outre les informations évoquées dans ces deux articles du Code civil, trois autres ensembles de facteurs peuvent être mobilisés dans une perspective prédictive. D'une part, les propositions des parties, d'autre part des facteurs contextuels et, enfin, les informations tirées des mesures provisoires.

Les propositions de montant de PC faites par les parties (offre et demande) sont déterminantes pour estimer le montant de PC dans la mesure où le juge doit décider *infra petita*. Dès lors qu'elles sont mentionnées dans les décisions, ce qui n'est pas toujours le cas, ces propositions, qui peuvent être exprimées – tout comme le montant de PC fixé par le juge – soit en termes de rente soit en termes de capital, doivent donc être construites à partir des informations de montant et éventuellement de durée dans le cas d'une rente. Leur prise en compte comme variables indépendantes de l'estimation du montant de PC fixé par le juge pose néanmoins un problème dans la mesure où les déterminants de ces montants sont très probablement, au moins pour partie, les mêmes déterminants que ceux qui seront valorisés par le juge. Ce problème nécessite un spécification économétrique particulière que nous expliciterons *infra*.

Au titre des facteurs contextuels, il convient de mentionner à minima la localisation du TGI, car on peut penser que, pour diverses raisons, les décisions peuvent varier significativement d'un juge ou d'une juridiction à l'autre. On pourrait certes retenir l'identification de chacun des juges de la base de données pour introduire un effet fixe individuel dans les estimations. Mais en termes de prédiction adressées à telle ou telle partie cela serait relativement inopérant, d'une part, parce que le justiciable ne connait pas à l'avance l'identité du juge qui traitera l'affaire, d'autre part, parce que l'identification dans la base de données est anonyme (codification par numéros). En revanche, les parties savent dans quel TGI leur affaire sera traitée ; la localisation du TGI peut donc être considérée comme un *proxy* de l'identité du juge, notamment lorsque l'on sait que localement parfois les juges se concertent pour tenter d'adopter une politique judiciaire homogène.

Même si l'on ne sait pas à l'avance quel juge traitera sa propre affaire, on peut tenter de tenir compte, comme facteur contextuel, du genre du juge, certains travaux[14] ayant montré des différences de comportement des juges selon leur sexe. Cela permettrait de tester si ce facteur est ou non déterminant et de proposer, dans le premier cas, deux estimations selon le genre du juge.

Enfin, nous pensons qu'il pourrait être utile d'ajouter comme facteur contextuel, un facteur de « climat de conflictualité » au sein du couple. Le degré de conflictualité est connu des parties, mais il ne transparaît pas explicitement dans les décisions. Nous avons donc tenté de construire des indicateurs indirects de conflictualité à partir des informations contenues dans les décisions. Il s'agit le plus souvent d'informations postérieures à la demande de PC (durant la procédure), alors que la prédiction en toute logique n'est utile qu'avant cette demande. Ce ne sont donc pas les faits eux-mêmes qui nous importent, mais le fait qu'indirectement ils peuvent indiquer qu'un conflit existait avant l'expression de la demande de PC. A ce titre, nous avons retenu et testé les indicateurs suivants :

- l'homme demande seul le divorce, ce qui peut être source d'un ressentiment de la part de la femme qui subit la décision de l'homme ;
- il s'agit d'un divorce pour faute, ce motif peut laisser penser que la situation est plus conflictuelle que lorsque la qualification du divorce est autre ;
- le divorce pour faute est aux torts de l'homme ; là encore on peut penser que cette qualification peut être corrélée à une situation conflictuelle ;
- le juge a ordonné une médiation ; s'il n'y avait pas de conflit, il est peu probable que le juge ait recouru à cette procédure ;
- le jugement est réputé contradictoire ; l'absence d'une des parties au procès pourrait être un signe de conflit entre les époux ;

[14] JEANDIDIER Bruno, BOURREAU-DUBOIS Cécile, RAY Jean-Claude, DORIAT-DUBAN Myriam (2016), "Does gender matter in the civil law judiciary? Evidence from French child support court decisions", *Working Paper du BETA*, n° 2016-55, 31 p.

- les parents ne se sont pas mis d'accord sur les modalités du divorce à propos des enfants ; ces désaccords peuvent être à l'origine du conflit ;
- seul l'époux s'est remis en couple ; là encore, l'hypothèse du ressentiment de la part de l'épouse peut venir de cette situation ;
- la qualification du divorce est modifiée en cours de procédure, signe que la situation était conflictuelle à tel point qu'un désaccord sur cet aspect central du divorce existait ;
- l'épouse obtient des dommages et intérêts au titre des articles 266 ou 1382 du code civil, ce qui traduit une absence de coopération de l'époux ;
- les deux parties ont chacune leur propre avocat.

L'analyse statistique permettra de déterminer lequel (voire lesquels) de ces indicateurs est le meilleur candidat pour tenir compte de l'information contextuelle que nous suggérons de retenir éventuellement pour relater le « climat conflictuel » du divorce dans l'estimation du montant de la PC.

Enfin, au titre des mesures provisoire pouvant constituer de bons prédicteurs du montant de PC, nous avons retenu le fait que la femme avait bénéficié d'une pension alimentaire fixée lors de l'ordonnance de non-conciliation. En effet, il est de notoriété que certains juges fixent un montant de PC égal à un multiple du montant de pension alimentaire. En revanche, nous n'avons pas retenu l'attribution temporaire de la résidence principale (mesure provisoire fréquente), car rien ne nous permet de penser que cette attribution ait une incidence, à niveau de patrimoine pris en compte, sur le montant de PC.

2.3. A la recherche de la meilleure spécification économétrique pour une perspective prédictive

Avant d'en venir à la discussion sur l'identification des meilleurs prédicteurs du montant de PC fixé par le juge, revenons à la question de la prise en compte des propositions des parties dans ces estimations. Nous sommes partis d'un échantillon de 2 678 affaires où un montant de PC a été fixé par un juge à l'issue de la procédure, auxquelles nous avons retiré les 118 affaires où le créancier est l'époux, puis les 74 affaires souffrant de données manquantes pour la mesure de ce montant de PC. Si l'on veut introduire l'offre et la demande de PC comme facteur explicatif du montant décidé par le juge, il convient alors de retirer 63 autres affaires pour lesquelles soit l'offre, soit la demande, soit les deux ne sont pas mentionnés dans les décisions ; c'est donc sur un échantillon de 2 423 affaires que porte notre analyse.

Si, sur cet échantillon, l'on s'en tient à régresser le montant de la PC fixé par le juge sur les deux seules informations que sont les propositions des parties, on constate alors que l'offre et la demande expliquent presque parfaitement le montant de PC fixé. Nous pourrions donc en conclure qu'il suffit de connaître les propositions pour déterminer le montant retenu par le juge. En fait, cette quasi-parfaite estimation cache une réalité plus complexe. En effet, cet échantillon est constitué de deux cas de figure bien distincts. Le premier cas est constitué des 1 675 affaires où les parties se sont mises d'accord, dans ce cas l'offre est égale à la demande et ce montant est homologué dans plus de 99% des cas par le juge ; c'est cette homologation systématique qui explique la quasi-parfaite estimation du montant de la PC fixé par le juge. Le second cas est constitué des 748 affaires où les parties ne se sont pas mises d'accord sur le montant de la PC. C'est donc bien sur ce seul petit sous-échantillon que la question de l'estimation de la décision du juge est pertinente. Pour l'autre sous-échantillon, la question de prédiction est moins celle d'estimer le montant qui sera fixé par le juge que celle de tenter de prédire les propositions ayant la meilleure chance de déboucher sur un accord (qui sera homologué avec quasi-certitude) dans telle ou telle situation[15]. Il convient donc de mener séparément les estimations sur ces deux sous-échantillons.

En nous restreignant à ces 748 affaires, nous sommes alors confrontés au fait que les mêmes facteurs explicatifs sont susceptibles d'expliquer ou de prédire la variable dépendante (montant de PC fixé) et les deux variables indépendantes que sont les propositions des parties (les parties lorsqu'elle formulent leurs propositions tout comme le juge lorsqu'il prend sa décision finale

[15] Nous présentons néanmoins en annexe 3 à titre d'illustration une estimation du montant de PC portant sur ce sous-échantillon d'affaires où il y a accord des parties quant au montant de PC.

peuvent, par exemple, tenir compte du degré de disparition de conditions de vie dans le couple). Ne pas tenir compte de cela amènerait à minimiser la mesure de l'impact effectif de tel ou tel facteur parce qu'il serait pour partie « caché » par l'intermédiation de l'impact des propositions. En conséquence, nous proposons de recourir à une estimation en trois étapes : l'estimation de la demande, l'estimation de l'offre et, enfin, l'estimation du montant de PC fixé avec comme particularité d'introduire dans cette troisième estimation les résidus des deux premières équations. Ce faisant, dans cette troisième estimation, nous pouvons mesurer l'impact total des différents facteurs explicatifs puisque nous n'introduisons pas les propositions des parties, mais seulement les résidus de leurs estimations c'est-à-dire les « parts » des propositions qui ne sont pas expliquées par ces facteurs explicatifs. Ainsi, nous tenons bien compte des propositions, soit via les résidus, soit via l'impact direct des facteurs explicatifs (ce dernier impact combinant l'effet sur le juge et celui sur les parties)[16]. Nous montrons en annexe 1 l'intérêt de recourir à cette méthode en trois équations en comparant le résultat ainsi obtenu pour l'équation de montant de PC fixé par le juge avec ceux qui seraient obtenus avec une unique régression portant sur le montant de PC fixé par le juge dans laquelle sont introduites l'offre et la demande et non pas leurs résidus.

Venons-en maintenant à l'identification des meilleurs prédicteurs pour chacune des trois équations à estimer. Lorsque plusieurs indicateurs étaient en concurrence pour une même caractéristique d'affaire, nous avons sélectionné le meilleur candidat sur la base d'une analyse de corrélations linéaires simples, l'indicateur sélectionné est alors retenu comme variable indépendante de la régression et n'est gardé *in fine* dans la spécification que si l'effet estimé apparait suffisamment significatif (au seuil de 10%). Nous résumons ci-dessous cette démarche en nous attachant principalement à l'équation de montants de PC fixé et en renvoyant éventuellement en notes de bas de page lorsque la conclusion est différente pour les deux autres équations (offre et demande).

Ainsi, concernant la question de la disparité des conditions de vie, l'indicateur de différence de niveau de vie a été préféré aux autres indicateurs possibles et l'analyse a montré qu'une spécification quadratique n'était pas opportune. L'exclusion d'un indicateur en termes de rapport est justifiée principalement par le fait que l'indicateur n'est pas opérant lorsque le dénominateur est égal à zéro, en effet il amènerait à exclure plus d'un tiers des observations. Et pour tenir compte simultanément de la disparité et du niveau, l'indicateur de niveau retenu est le niveau de vie de l'époux, sans spécification quadratique et sans croisement entre niveau et disparité, ces spécifications plus complexes n'apportant que très peu de pouvoir explicatif.

Concernant l'indicateur d'enfants (qui tente de capter une information relative à l'incidence de la charge familiale sur la carrière de l'épouse), nous avons écarté plusieurs indicateurs (présence d'au moins un enfant en bas âge ; fratrie nombreuse ; nombre d'enfants à charge ; nombre d'enfants non à charge) pour ne garder qu'un indicateur, le plus pertinent statistiquement : le nombre d'enfants du couple, spécifié en *dummies* pour tenir compte d'éventuelles non linéarités.

Du point de vue de l'indicateur indirect de conflit qui s'appuie sur l'idée d'un changement de qualification du divorce en cours de procédure, nous avons testé la pertinence de plusieurs types de changements pour ne retenir *in fine* qu'une indicatrice identifiant le fait que, en cours de procédure, la qualification de faute est ajoutée ou supprimée. Ont donc été écartées, pour cause de moindre pouvoir explicatif, les indicatrices de changement, soit de manière générale entre tout type de divorce, soit du point de vue de l'identification des torts (à l'un, à l'autre, aux deux), soit plus spécifiquement dans le fait que la faute est attribuée à l'époux en cours de procédure[17].

En ce qui concerne les différentes manières d'appréhender la question du bénéfice de l'Aide juridictionnelle (AJ), nous avons écarté les indicatrices identifiant les situations de bénéfice à taux partiel, non discriminantes, et nous n'avons pas retenu la distinction entre le fait que l'épouse bénéficie seule de l'AJ ou simultanément avec son conjoint (la distinction n'étant pas statistiquement pertinente). Au final, c'est donc l'indicatrice qui oppose le fait que l'épouse bénéficie de l'AJ à taux plein aux autres situations (taux partiel ou absence d'AJ) qui a été privilégiée (sans tenir compte d'un éventuel bénéfice au profit de l'époux).

[16] Nous ne tenons pas compte du fait que l'offre peut influencer la demande et inversement.

[17] Pour l'équation d'offre une indicatrice peut être retenue en plus : le fait qu'au cours de la procédure la faute est attribuée à l'époux.

En ce qui concerne la prise en compte de l'âge des deux époux, nous avions à choisir entre une spécification retenant les deux âges ou une spécification ne retenant l'âge que d'un seul époux auquel serait ajouté une variable de différence d'âge intra-couple ; l'analyse des corrélations nous a amené à retenir l'âge de l'époux et la différence d'âge.

Quant à la prise en compte de mesures provisoires, à savoir la pension alimentaire, notre analyse nous conduit à privilégier la prise en compte du montant de pension alimentaire plutôt que le simple fait de percevoir ou non une telle pension alimentaire (indicatrice binaire), la corrélation simple entre le montant de PC et le montant de pension alimentaire étant de loin plus élevée que celle calculée entre l'indicatrice binaire et le montant de PC.

Les autres variables indépendantes ne font pas l'objet de choix de spécification (mesure unique) ; soulignons simplement, d'une part, que pour la variable de durée de mariage il s'avère qu'une spécification quadratique n'est pas nécessaire car cela n'apporte que très peu à la qualité de l'ajustement, d'autre part, que l'indicatrice de mesure indirecte de conflit identifiant le conjoint à l'origine de la requête ainsi que la différence intra-couple de patrimoine n'ont pas été retenues pour cause de données manquantes en nombre trop élevé. De même certaines indicatrices de motivation du juge n'ont pas été mobilisées car elles ne concernent aucune affaire dans l'échantillon retenu pour l'analyse (la mauvaise qualification professionnelle de l'époux ; le fait que les droits à la retraite de l'époux sont faibles ; le fait que les droits à la retraite de l'époux sont inférieurs à ceux de l'épouse).

Le tableau 1 ci-dessous présente les résultats obtenus selon cet ensemble de choix de spécification. Concernant l'équation de montant de PC fixé par le juge, comme attendu, l'effet associé à la disparité des niveaux de vie (12,05) montre un signe positif et très significatif ; on pourrait s'étonner de voir un signe négatif associé à la variable de niveau de vie de l'époux (-6,81). En fait, son interprétation ne peut pas se faire directement car le niveau de vie de l'époux (NDV_h)se trouve également dans la variable de différence de niveaux de vie, il faut donc effectuer une somme de coefficients et en tester la signification selon la logique algébrique suivante :

$$PC = a + b*NDV_h + c*(NDV_h-NDV_f) + d*X + \varepsilon$$
$$PC = a + (b+c)*NDV_h - c*NDV_f + d*X + \varepsilon$$

Ainsi, comme attendu, l'impact du niveau de vie de l'époux (12,05 – 6,81 = 5,24) est bien estimé positif et a été testé comme étant significatif : plus l'époux a un haut niveau de vie, plus la PC versée à l'épouse est élevée. De même, l'effet du niveau de vie de la femme (-12,05) est bien, comme attendu, négatif et significatif : plus l'épouse a un faible niveau de vie, plus la PC qu'elle reçoit est élevée, toutes choses égales par ailleurs. Quant aux impacts des autres facteurs explicatifs significatifs, ils sont dans la plupart des cas de signe attendu, il en va ainsi de la durée de mariage, du régime matrimonial moins favorable à la femme, du fait que la PC soit sous forme de rente, de la santé dégradée de l'épouse évoqué dans les motivations du juge, du montant de pension alimentaire versée à l'épouse lors de la fixation des mesures provisoires, de l'évocation dans les motivations du juge du fait que l'épouse a contribué à l'activité professionnelle de l'époux et de certains indicateurs de conflictualité (changement de qualification du point de vue de la faute ; médiation ; dommages et intérêts en faveur de l'épouse) qui montrent des coefficients estimés significatifs et positifs[18]. Plus curieusement, le signe négatif associé au bénéfice de l'Aide Juridictionnelle interpelle[19].

[18] L'hypothèse relative aux indicateurs de conflit étant que le mécontentement de l'épouse pourrait se traduire par une plus forte revendication en matière de PC.

[19] On peut penser que l'hypothèse d'un effet de très bas revenus que nous faisons soit compensé par un effet d'homogamie de ressources dans le couple (lorsque l'épouse à des bas revenus, l'époux est dans une situation comparable), d'où un signe négatif (bien que nous contrôlions le niveau de vie de l'époux). Une autre interprétation pourrait être défendue selon laquelle les bénéficiaires de l'Aide Juridictionnelle seraient moins bien défendus par leurs avocats.

Tableau 1 : estimation par moindres carrés ordinaires du montant de PC fixé par le juge selon une spécification en trois équations, lorsque les parties ne sont pas d'accord

	Offre de PC		Demande de PC		Montant de PC fixé	
Constante	-16 356	***	-1 357	NS	-9 199	NS
PC sous forme uniquement de rente	/		/		16 603	*
Différence de niveau de vie intra-couple (€/mois)	4,33	NS	16,81	*	12,05	***
Niveau de vie de l'époux (€/mois)	-0,90	NS	-2,59	NS	-6,81	*
L'épouse bénéficie de l'AJ à taux plein		NS	-27 874	**	-9 570	**
Régime matrimonial moins favorable à l'épouse		NS	61 007	***	13 468	**
Durée de mariage en années	497	***	1 770	***	918	***
Désaccord des parents sur le montant de CEEE		NS	-17 858	*		NS
Changement dans la qualification en faute		NS	45 930	**	14 187	*
Médiation ordonnée par le juge		NS		NS	25 071	**
Hausse prévisible des revenus de l'époux		NS	54 140	#		NS
Montant pension alimentaire de l'épouse (€/mois)	39,6	***	119	***	56,8	***
Montant de dommages et intérêts pour l'épouse		NS	6,30	**	2,26	*
Motivation du juge : mauvaise santé de l'épouse	11 947	**	41 234	***	20 282	***
Motivation du juge : l'épouse collabore à l'activité de l'époux		NS		NS	16 232	*
Résidus de l'équation d'offre de PC	/		/		0,66	***
Résidus de l'équation de demande de PC	/		/		0,18	***
Qualité de l'ajustement : R^2	**31,7%**		**49,8%**		**83,6%**	
Qualité de l'ajustement : R^2 ajusté	**31,0%**		**48,7%**		**77,6%**	
Effectifs	**569**		**545**		**541**	
Caractéristiques d'affaire non retenues dans les trois équations pour cause de non-significativité :						
Age de l'épouse	X		X		X	
Différence d'âge intra-couple	X		X		X	
Nombre d'enfants du couples (dummies)	X		X		X	
L'époux est en emploi	X		X		X	
L'épouse est en emploi	X		X		X	
L'épouse est au chômage	X		X		X	
Hausse prévisible des revenus de l'épouse	X		X		X	
Baisse prévisible des revenus de l'épouse	X		X		X	
Baisse prévisible des revenus de l'époux	X		X		X	
Remise en couple de l'époux	X		X		X	
Remise en couple de l'épouse	X		X		X	
Désaccord des parents sur le lieu d'hébergement	X		X		X	
Désaccord des parents sur le DVH	X		X		X	
Les époux ont des avocats différents	X		X		X	
L'épouse demande des dommages et intérêts	X		X		X	
Montant de dommages et intérêts pour l'époux	X		X		X	
Le jugement est réputé contradictoire	X		X		X	
Divorce pour faute	X		X		X	
Divorce pour faute aux torts de l'époux	X		X		X	
Changement dans l'attribution des torts	X		X		X	
Médiation ordonnée par le juge	X		X		X	
Le juge est un homme	X		X		X	
Motivations du juge :						
- *Mauvaise santé de l'époux*	X		X		X	
- *Choix de carrière de l'épouse pour les enfants*	X		X		X	
- *Choix de carrière personnel de l'épouse*	X		X		X	
- *Choix de carrière de l'épouse décidé en commun*	X		X		X	
- *Choix de carrière de l'épouse pour favoriser celle du conjoint*	X		X		X	
- *Bonne qualification de l'épouse*	X		X		X	
- *Mauvaise qualification de l'épouse*	X		X		X	
- *Bonne qualification de l'époux*	X		X		X	
- *Droits à la retraite faibles de l'épouse*	X		X		X	

Source : base COMPRES, 2013. AJ : Aide Juridictionnelle. DVH : Droit de Visite de d'Hébergement. CEEE : Contribution à l'Entretien et l'Education de l'Enfant. Dans l'équation de montant de PC fixé par le juge est ajouté un jeu d'indicatrices identifiant les différents TGI. NS : non significatif au seuil de 10%. # : significatif au seuil de 10%. * : significatif au seuil de 5%. ** : significatif au seuil de 1%. *** : significatif au seuil de 0,1%. / : facteur non introduit dans l'équation.

Au total, 16,4% de la variance du montant de PC ne sont pas expliqués par la conjonction des propositions des parties et par les facteurs principalement liés aux dispositions du Code civil, mais il faut souligner que ce qui, dans les propositions des parties, n'est pas expliqué par les caractéristiques d'affaires[20] (c'est-à-dire les résidus des équations expliquant l'offre et la demande) est responsable d'une bonne part de la qualité de l'ajustement puisque, sur cet échantillon, en régressant le montant de PC sur uniquement ces deux variables de résidus, le R^2 est égal à 35%. D'un autre côté, si à l'inverse on ne retient que les autres facteurs explicatifs en excluant les résidus d'offre et de demande, le R^2 est égal à 66%, ce qui est à la fois beaucoup en soi mais aussi bien trop peu dans la perspective de la justice prédictive (*cf. infra* § 2.4) comme dans la perspective d'une utilisation ayant pour finalité de construire un barème « observé » de PC. Cela signifie que le pouvoir explicatif (et prédictif) de l'ensemble des facteurs que nous avons sélectionnés au regard principalement des deux articles du Code civil est prépondérant, mais qu'une partie non négligeable du comportement décisionnel en matière de PC est due à des facteurs que nous ne savons pas observer à partir des décisions de justice et qui transitent par l'expression de l'offre et de la demande ; ces facteurs inobservés constituent une importante source d'hétérogénéité des montants de PC observés. On notera également que la liste des prédicteurs exclus pour cause de non-significativité est longue, cette longueur exprime le fait que l'on a bien du mal à identifier les facteurs pertinents (et notifiés dans les décisions écrites) permettant de comprendre sur la base de quels déterminants objectifs les juges prennent leur décision.

Pour autant, pour tenter d'améliorer notre modèle prédictif, nous proposons d'effectuer deux estimations alternatives en recourant à deux autres outils économétriques. D'une part, nous testons une régression par moindres carrés ordinaires exponentielle (les variables dépendantes sont alors exprimées en logarithmes), d'autre part nous testons une régression par quantile qui recoure aux moindres valeurs absolues (et toujours avec des variables dépendantes exprimées en logarithmes). Dans les deux cas, ces méthodes procurent deux avantages : elles évitent d'estimer d'éventuelles valeurs négatives (qui dans notre cas n'auraient aucun sens) et elles réduisent l'impact éventuel des observations extrêmes (*outliers*). Mais en contrepartie elles impliquent une difficulté. En effet, dans un nombre de cas très important (568 affaires) le montant de l'offre est nul et ne peut donc pas être exprimé en logarithme, or il est totalement inconcevable d'exclure un si grand nombre d'observations. Cela nous oblige donc à retirer le résidu de l'offre dans notre spécification et donc de passer d'une spécification à trois équations à une spécification à deux équations. Nous verrons cependant (tableau 2) que cette exclusion d'un facteur explicatif ne réduit pas sensiblement la capacité explicative globale des modèles (perte de 5 points de pourcentage en termes de R^2 ajusté, ou de 9 points en termes de R^2) comparativement à la modélisation présentée au tableau 1 *supra*.

A cette exception près, les deux estimations alternatives sont effectuées selon la même logique de spécification que celle retenue précédemment au tableau 1. Le tableau 2 présente les résultats obtenus à l'aide de ces outils. Au titre des facteurs explicatifs significatifs au seuil de 10%, on peut souligner quelques différences selon l'outil économétrique retenu. Dans les modèles exponentiels, quelques motivations du juge relatives aux choix de carrière deviennent significatifs (surtout dans la régression par quantile), ainsi que l'âge des conjoints. A l'inverse, deux indicateurs de conflictualité perdent en significativité : le fait que l'épouse perçoive des dommages et intérêts et le fait que la qualification de la faute ait changé en cours de procédure. Ces différences peuvent tenir à l'outil économétrique ou au fait que l'échantillon d'affaires est légèrement différent (le fait d'exclure la variable d'offre de PC permet en effet de récupérer quelques observations exclues initialement pour cause de données manquantes pour cette variable d'offre).

[20]Non expliqué parce que nous ne disposons pas, dans les décisions de justice, des facteurs pertinents ou parce que ces derniers sont affectés par des données manquantes, ou encore parce que nous les intégrons mais de façon non pertinente (spécification inadaptée, erreurs de mesure...).

Tableau 2 : estimation, selon trois méthodes économétriques, du montant de PC fixé par le juge selon une spécification en deux ou trois équations, lorsque les parties ne sont pas d'accord

	MCO, modèle à trois équations (rappel)	Modèle MCO exponentiel, à deux équations	Modèle par quantile (MVA) exponentiel, à deux équations
	Montant de PC fixé	**Logarithme de montant de PC fixé**	**Logarithme de montant de PC fixé**
Constante	-9 199 NS	8 82 ***	9,06 ***
PC sous forme uniquement de rente	16 603 *	0,28 **	0,30 **
Différence de niveau de vie intra-couple (€/mois)	12,05 ***	0,14 **	0,23 ***
Niveau de vie de l'époux (€/mois)	-6,81 *	-0,04	-0,11 **
L'épouse bénéficie de l'AJ à taux plein	-9 570 **	-0,29 ***	-0,34 ***
Régime matrimonial moins favorable à l'épouse	13 468 **	0,32 ***	0,29 ***
Durée de mariage en années	918 ***	0,03 ***	0,03 ***
Changement dans la qualification en faute	14 187 *	0,18 *	0,02 NS
Hausse prévisible des revenus de l'époux	25 071 **	-0,29 #	-0,37 *
Montant de pension alimentaire épouse (€/mois)	56,8 ***	0,60 ***	0,55 ***
Montant de dommages et intérêts pour l'épouse	2,26 *	0,04 *	0,01 NS
Age de l'époux	/ NS	0,01 #	0,002 NS
Ecart d'âge intra-couple	/ NS	-0,02 **	-0,008 *
Motivation du juge : mauvaise santé de l'épouse	20 282 ***	0,18 **	0,14 *
Motivation du juge : l'épouse collabore à l'activité du conjoint	16 232 *	/ NS	/ NS
Motivation du juge : l'épouse a fait le choix de favoriser la carrière de l'époux	/ NS	0,15 #	0,15 *
Motivation du juge : l'épouse a fait un choix de carrière décidé en commun avec l'époux	/ NS	0,19 #	0,24 **
Motivation du juge : bonne qualification professionnelle de l'époux	/ NS	0,20 #	0,33 **
Motivation du juge : bonne qualification professionnelle de l'épouse	/ NS	-0,20 #	-0,30 **
Résidus de l'équation d'offre de PC	0,66 ***	/	/
Résidus de l'équation de demande de PC	0,18 ***	0,64 ***	0,62 ***
Qualité de l'ajustement : R^2	**83,6%**	**74,4%**	**/**
Qualité de l'ajustement : R^2 ajusté	**77,6%**	**72,3%**	**/**
Effectifs	**541**	**582**	**582**

Source : base COMPRES, 2013. AJ : Aide Juridictionnelle. Est ajouté un jeu d'indicatrices identifiant les différents TGI. NS : non significatif au seuil de 10%. # : significatif au seuil de 10%. * : significatif au seuil de 5%. ** : significatif au seuil de 1%. *** : significatif au seuil de 0,1%. / : facteur non introduit dans l'équation. Dans les modèles exponentiels, les résidus de l'équation de demande sont exprimés en logarithmes et les variables explicatives relatives à des montants sont exprimées en milliers d'euros.

2.4. Applications illustratives de ces estimations dans une perspective prédictive

Peut-on faire confiance à nos estimations dans une perspective prédictive ? Une variabilité (mesurée par la variance, en l'occurrence) des montants de PC expliquée à hauteur d'au moins 70% est-elle suffisante pour proposer des valeurs prédites crédibles ? Pour répondre à ces questions légitimes, nous commençons par donner une idée de l'ampleur et du sens du biais associé à chacun de nos deux modèles : pour cela nous comparons (en termes de moyennes mais aussi de médianes) les montants de PC observés pour les 582 affaires sans accord pour lesquelles une estimation a pu être réalisée (colonne n°1 du tableau 3) avec les montants prédits (pour ces mêmes 582 affaires) selon le modèle MCO exponentiel, d'une part (colonne n°2), et selon le modèle exponentiel par quartile, d'autre part (colonne n°3).

Ce qui ressort de ce tableau 3 c'est que si, en termes de moyenne, le modèle MCO souffre d'un léger biais de sous-estimation (car la PC moyenne prédite se monte à 34 660 € tandis que la PC observée

moyenne vaut 37 400 €[21]), en revanche le modèle exponentiel par régression sur quantile n'enregistre qu'un biais négligeable (prévision moyenne : 37 369 €). En termes de médiane, nos deux modèles surestiment légèrement seulement (et presque autant : 20 440 € pour le MCO, et 20 638 € pour la régression par quantile) les PC observées (dont la médiane se situe à 20 000 €).

Tableau 3 : montants moyens et médians de PC observés et prédits pour les 582 affaires sans accord pour lesquelles une estimation a pu être réalisée

	Montants observés	Montants estimés selon le modèle MCO exponentiel	Montants estimés selon le modèle exponentiel par quantile
Moyenne	37 400	34 660	37 369
Médiane	20 000	20 440	20 638

Source : base COMPRES, 2013.

Alors que le tableau 3 a fourni des informations sur la tendance centrale de la distribution des montants observés pour les 582 affaires, puis a fait de même pour les distributions des montants prédits par les deux modèles, le tout sans comparaison, affaire par affaire, des différences « PC observée – PC prédite », et du rapport de cet écart au montant de PC observée, cette comparaison est précisément l'objet du tableau 4.

Pour mesurer la pertinence de nos modèles prédictifs, on calcule donc d'abord, pour chacune des 582 affaires pour lesquelles les valeurs des divers prédicteurs du montant de PC sont connues (i.e. absence de données manquantes), la valeur absolue[22] de la différence entre la valeur observée et la valeur prédite ; et on complète ensuite cette appréciation par une mesure de l'ampleur relative de cette erreur de prédiction (en rapportant la valeur absolue de l'erreur au montant de PC observé). Il en résulte (pour chacun des deux modèles exponentiels : MCO et par régression sur quantile) deux distributions d'erreurs : celle des montants d'erreur en euros et celle de l'ampleur relative (en %) des erreurs en question.

Pour caractériser et comparer les distributions de ces erreurs et taux d'erreur on peut utiliser, pour commencer, le résumé très partiel fourni par les valeurs de la moyenne et de la médiane, qui sont des indicateurs de la tendance centrale de ces distributions (tableau 4). Ce qui ressort de ce tableau 4 c'est que nos deux modèles exponentiels prédisent le montant de PC de façon très insatisfaisante.

En effet, si on utilise l'outil « moyenne », on constate que l'ampleur de l'erreur absolue de prédiction est importante puisque l'erreur absolue moyenne est, selon le modèle retenu, soit égale à 14 141 euros (dans le cas des MCO, avec un taux moyen d'erreur valant 45%), soit égale à 16 508 euros (dans le cas de la régression par quantile, avec un taux moyen d'erreur peu différent de celui des MCO : 46%), ce qui est considérable.

Si on se sert de l'outil « médiane », on observe que l'ampleur absolue de l'erreur est (deux à trois fois) moindre qu'avec l'outil « moyenne » ; quant au rapport de cette erreur absolue à la PC observée, si sa médiane est plus faible (d'un bon quart) que sa moyenne, et cela à la fois pour l'estimation par MCO et pour l'estimation par régression sur quantile, elle reste considérable : au moins 30%, ce qui signifie que le taux d'erreur absolue de prédiction pour la moitié des 582 affaires excède 30%.

[21] Ce montant moyen de PC observé (37 400 €), qui correspond aux seules 582 affaires dont aucun prédicteur n'a de valeur manquante, est très nettement inférieur (d'un sixième) au montant moyen de PC observé pour l'ensemble des 748 affaires sans accord (44 836 €). En revanche, les montants médians des deux échantillons sont très voisins : 20 580 € pour les 748 affaires *versus* 20 000 € pour les 582 affaires.

[22] Comme les erreurs positives compensent les erreurs négatives, il faut raisonner non pas sur les erreurs de prédictions signées mais sur leurs valeurs absolues.

Tableau 4 : moyennes et médianes des valeurs absolues des écarts « PC observée – PC prédite » et des rapports de ces valeurs au montant observé de PC

	Montants estimés selon le modèle MCO exponentiel		Montants estimés selon le modèle exponentiel par quantile	
	Moyenne	**Médiane**	**Moyenne**	**Médiane**
Valeur absolue de l'erreur de prédiction	14 141	5 931	16 508	5 577
Valeur absolue du taux d'erreur de prédiction	45%	32%	46%	30%

Source : base COMPRES, 2013.

Ces médianes d'erreur de prédiction, nettement plus faibles que les moyennes, laissent penser que ces dernières sont fortement déterminées par des écarts extrêmes. Une analyse de la distribution des écarts absolus s'avère donc utile (tableau 5).

Effectivement, on observe qu'il existe des écarts absolus très élevés qui, pour partie, expliquent l'ampleur des moyennes commentées précédemment, puisque le décile 9 se situe près de 30 000 euros. Cela signifie donc que nos estimations génèrent pour 10% des affaires des erreurs supérieures à environ 30 000 euros et qu'elles génèrent des erreurs supérieures à environ 13 000 euros dans un quart des affaires (quartile 3). Comparativement aux valeurs observées et donc en valeur relative, dans un cas sur quatre l'erreur est supérieure à 50% environ quelle que soit la méthode d'estimation, c'est-à-dire que dans un cas sur quatre l'erreur de prédiction excède la moitié du montant observé de la PC[23].

Tableau 5 : Distribution des valeurs absolues des écarts « PC observée – PC prédite » et des rapports de ces valeurs au montant observé de PC

	Montants estimés selon le modèle MCO exponentiel		Montants estimés selon le modèle exponentiel par quantile	
	Valeur absolue de l'erreur de prédiction	**Valeur absolue du taux d'erreur de prédiction**	**Valeur absolue de l'erreur de prédiction**	**Valeur absolue du taux d'erreur de prédiction**
Décile 9 (90%)	30 898	83%	29 561	91%
Quartile 3 (75%)	13 907	49%	13 557	52%
Médiane (50%)	5 931	32%	5 577	30%
Quartile 1 (25%)	2 544	15%	2 163	12%
Décile 1 (10%)	896	5%	306	2%

Source : base COMPRES, 2013.

Conclusion

Dans la perspective prochaine de l'accès électronique à l'exhaustivité des décisions de justice de première instance, nous nous sommes interrogés sur la potentialité que pourrait représenter cette perspective dans une finalité d'analyse relevant d'une logique de justice prédictive. Pour mener cette réflexion nous avons mobilisé un corpus de décisions de justice représentatif des TGI relatif à la prestation compensatoire lors de divorces et nous avons mené une analyse économétrique traditionnelle, c'est-à-dire une analyse sans mobiliser les outils de fouille automatisée de données et/ou d'intelligence artificielle. Ainsi avons-nous pu mettre en lumière un certain nombre de

[23] Un taux de 50% en valeur absolue peut signifier que la valeur prédite est deux fois plus petite que la valeur observée (ex. |(200 - 100) / 200| = 0,5, soit 50%) ou que la valeur prédite est une demie fois supérieure à la valeur observée (ex. |(200 - 300) / 200| = 0,5, soit 50%). Et lorsque l'on observe les queues de distribution des écarts en valeur absolue, on s'aperçoit que les erreurs peuvent être vraiment considérables. Avec le modèle MCO exponentiel, les 20 écarts en valeur absolue les plus élevés (3% de l'échantillon) s'étalent de 58 491 euros à 619 093 euros et les 20 écarts relatifs en valeur absolue s'étalent de 174% à 559%. Avec le modèle exponentiel par quantile, les chiffres sont respectivement de 60 850 euros à 1 593 138 euros et de 198% à 760%. Ces chiffres sont sans conteste assez alarmants, même si, dans les queues de distribution des écarts relatifs en valeur absolue, les affaires en question concernent presque exclusivement des montants de PC modestes (rarement supérieurs à 10 000 euros).

difficultés qui constituent autant de questionnements adressés aux spécialistes des méthodes d'analyse des *big data* qui se proposent d'investir ce nouveau domaine d'information.

On peut résumer ces difficultés ou ces points de vigilance ainsi. Premièrement, l'information à mobiliser pour mener une analyse pertinente ne se trouve pas toujours directement dans les décisions de justice, il convient en fait de la construire en combinant différentes informations (parfois plusieurs dizaines). Deuxièmement, certaines informations qui en toute logique devraient être retenues comme facteurs de prédiction, notamment au regard des règles de droit inscrites dans le Code civil, peuvent être totalement absentes des décisions, parce qu'elles demeurent implicites ou parce que le juge n'a pas souhaité les évoquer par écrit. Troisièmement, la rédaction des décisions n'est pas homogène, certaines décisions sont assez détaillées, d'autre pas ; il en résulte que pour tout un ensemble de facteurs, l'analyse est confrontée à des données manquantes qui obligent à écarter une partie du corpus et ainsi à prendre le risque de subir des biais de sélection. Quatrièmement, parce que les affaires sont toutes assez spécifiques et donc assez hétérogènes et parce que le juge s'applique à juger sur le cas d'espèce, il devient essentiel de mobiliser *a priori* un très grand nombre d'informations pour cerner au mieux chacun des cas et donc de mettre ensuite en œuvre des méthodes pour faire le tri dans cette masse d'information pour se limiter aux seules informations les plus pertinentes dans une perspective de prédiction réplicable. Cinquièmement, une bonne connaissance du contentieux s'avère nécessaire si l'on ne veut pas, par méconnaissance, développer des analyses non pertinentes. En effet, dans notre cas de figure, ne pas distinguer les affaires avec accord des affaires sans accord entre les parties, nous aurait mené à une impasse analytique ; or à la simple lecture de quelques décisions sans expertise de ce type de contentieux, rien ne prédispose *a priori* à penser qu'il faille effectuer cette distinction.

Enfin, sixièmement, il convient de s'interroger sur le degré de précision des prédictions qui peuvent émerger de ces travaux d'analyse approfondis. En la matière, notre expérience s'avère plutôt inquiétante. Certes, la mobilisation de très nombreuses informations et la mise en œuvre d'outils économétriques diversifiés nous a permis d'aboutir à des estimations qui, en sciences humaines, sont de qualité globalement satisfaisante (notamment au regard de la qualité d'ajustement des régressions, mesuré par le coefficient de R^2 ajusté, qui est de l'ordre de 70% à 75%), mais lorsque l'on utilise ces estimations pour prédire les montants de prestation compensatoire des affaires du corpus mobilisé (du moins du sous-échantillon des affaires sans accord qui ont fait l'objet de l'analyse la plus fouillée) et que l'on compare ces prédictions aux montants effectifs (ceux décidés par les juges et notifiés dans les décisions), on s'aperçoit que pour une proportion non négligeable d'affaires, les erreurs de prédiction sont considérables. Peut-on par exemple faire confiance à un outil qui, dans un cas sur quatre, propose une prédiction qui, en fait, génère une erreur qui excède la moitié du montant observé de la PC ?

Mais, à l'inverse, nos analyses montrent que la perspective d'avoir accès prochainement plus facilement que par le passé à de vastes corpus de décisions de justice sur supports électroniques constitue une perspective très stimulante pour la recherche en sciences juridiques. En effet, les difficultés que nous avons rencontrées et l'insatisfaction finale qui résulte de l'exercice si on se place uniquement dans une logique de prédiction, sont déjà riches en enseignements, enseignements qui alors pourraient peut-être être renforcés par des travaux à partir d'un corpus plus large, voire exhaustif. Au titre de ces enseignements, sur le cas spécifique de la prestation compensatoire, d'une part, nous avons pu identifier quelques facteurs hautement significatifs comme, par exemple, la différence de niveau de vie intra-couple, la durée de mariage ou encore la forme de PC et, d'autre part, nous sommes arrivés, dans le cas des affaires sans accord, à expliquer une part non négligeable de la variance des montants de prestation compensatoire.

Au total, si l'on additionne nos trois principaux résultats, à savoir,

- premièrement, le fait que pour les affaires avec accord la qualité de l'estimation du montant de prestation compensatoire est médiocre signifiant que les facteurs objectifs que nous avons retenus expliquent peu le comportement des parties qui se mettent d'accord,
- deuxièmement, le fait que pour les affaires sans accord au contraire l'estimation du montant de prestation compensatoire à l'aide de ces mêmes facteurs est de bien meilleure qualité (ce

qui pourrait signifier que le juge se conforme davantage en cas de désaccord qu'en cas d'accord à l'argumentaire qui est celui du Code civil) mais qu'une part non négligeable de l'explication de la variance des prestations compensatoires est à attribuer à des facteurs que nous ne savons pas observer à partir des décisions et qui transitent par l'expression des propositions des parties,

- et troisièmement, le fait que même dans le cas où les estimations sont de bonne qualité (affaires sans accord) ces dernières génèrent des erreurs de prédiction importantes dans nombre de cas,
- nos travaux permettent de conclure qu'il n'existe pas vraiment de barème implicite de prestation compensatoire (ou du moins qu'il est très difficile d'en identifier un), au sens d'un barème qui serait partagé, même inconsciemment du fait de l'accumulation de pratiques par mimétisme, par les juges aux affaires familiales et les avocats qui conseillent les parties.

Il faut cependant reconnaître que l'exemple d'application que nous avons choisi, à savoir l'estimation du montant de prestation compensatoire fixé par les juges, est probablement un exemple où la modélisation et la prédiction ne sont pas des plus aisées ; d'autres contentieux relevant du droit civil sont certainement plus homogènes et donc meilleurs candidats à la prédiction. En effet, il est notoire que les décisions en matière de prestation compensatoire sont très hétérogènes et que les juges sont souvent hésitants pour prendre leurs décisions en la matière parce qu'ils sont dépourvus d'outils d'aide à la décision convaincants et officiels. Pour s'en convaincre on peut se référer aux travaux[24] qui ont recensé les multiples barèmes de prestation compensatoire non officiels qui sont utilisés dans les juridictions, barèmes qui, lorsqu'on les compare par simulation sur des cas fictifs, aboutissent à des montants de prestation compensatoire très différents.

Références

BOURREAU-DUBOIS Cécile, JEANDIDIER Bruno, MANSUY Julie (2018), « Les enjeux redistributifs de la prestation compensatoire : une analyse statistique de 5 000 décisions de divorce », *In* Bourreau-Dubois C., Sayn I. (Dir.), *Le traitement juridique des conséquences économiques du divorce,* Ed. Bruylant, pp. 127-150.

JEANDIDIER Bruno, BOURREAU-DUBOIS Cécile, RAY Jean-Claude, DORIAT-DUBAN Myriam (2016), "Does gender matter in the civil law judiciary? Evidence from French child support court decisions", *Working Paper du BETA*, n° 2016-55, 31 p.

SAYN Isabelle (2016), « Compenser les inégalités économiques des époux après divorce ? Des critères légaux aux outils d'aide à la décision », *Revue Canadienne Droit et Société*, vol. 31, n° spécial, p. 219-241.

[24] SAYN Isabelle (2016), « Compenser les inégalités économiques des époux après divorce ? Des critères légaux aux outils d'aide à la décision », *Revue Canadienne Droit et Société*, vol. 31, n° spécial, p. 219-241.

Annexe 1 : estimation directe *versus* estimation en trois étapes

On peut apprécier le choix fait d'estimer le montant de PC fixé par le juge selon une procédure en trois équations, en comparant le résultat obtenu et présenté dans la dernière colonne du tableau 1 *supra* avec le résultat obtenu lorsque l'on n'effectue qu'une seule équation dans laquelle sont introduites l'offre et la demande et non pas leurs résidus.

	MCO, modèle à trois équations		**MCO, modèle à une seule équation**	
Constante	-9 199	NS	5 830	NS
PC sous forme uniquement de rente	16 603	*	16 662	**
Différence de niveau de vie intra-couple (€/mois)	12,05	***	6,29	*
Niveau de vie de l'époux (€/mois)	-6,81	*	-5,82	*
L'épouse bénéficie de l'AJ à taux plein	-9 570	**	-4053	NS
Régime matrimonial moins favorable à l'épouse	13 468	**	821	NS
Durée de mariage en années	918	***	294	*
Changement dans la qualification en faute	14 187	*	7 287	NS
Médiation ordonnée par le juge	25 071	**	25 793	**
Hausse prévisible des revenus de l'époux	25 071	**	/	NS
Montant de pension alimentaire épouse (€/mois)	56,8	***	9,1	#
Montant de dommages et intérêts pour l'épouse	2,26	*	/	NS
Motivation du juge : mauvaise santé de l'épouse	20 282	***	/	NS
Motivation du juge : l'épouse collabore à l'activité du conjoint	16 232	*	15 404	*
Résidus de l'équation d'offre de PC	0,66	***	/	
Résidus de l'équation de demande de PC	0,18	***	/	
Offre de PC	/		0,64	***
Demande de PC	/		0,18	***
Qualité de l'ajustement : R^2	**83,6%**		**83,6%**	
Qualité de l'ajustement : R^2 ajusté	**77,6%**		**77,7%**	
Effectifs	**541**		**541**	

Source : base COMPRES, 2013. AJ : Aide Juridictionnelle. Est ajouté un jeu d'indicatrices identifiant les différents TGI. NS : non significatif au seuil de 10%. # : significatif au seuil de 10%. * : significatif au seuil de 5%. ** : significatif au seuil de 1%. *** : significatif au seuil de 0,1%. / : facteur non introduit dans l'équation.

Ces modèles ont été estimés par régression *backward* (à seuil de 10%), en partant d'un ensemble de facteurs explicatifs plus large que celui mentionné dans le tableau ci-dessus, où on a omis ceux qui ont été écartés du modèle final faute d'avoir exhibé un effet significatif au seuil de probabilité choisi. Au sein de ce vaste ensemble de facteurs on a forcé la procédure à conserver, qu'ils soient ou non significatifs, des variables jugées d'un intérêt particulier[25]. Par ailleurs on a imposé à la procédure de conserver tous les TGI, ainsi que les deux variables concernant les niveaux de vie des époux dès lors que l'une des deux variables avait un effet ; on l'a fait en raison des liens mécaniques qu'introduit la variable de différence (et en raison du test spécifique dès lors nécessaire à l'interprétation de l'effet du niveau de vie du mari).

Lorsqu'on passe du modèle à équation unique au modèle à trois équations on ne gagne certes rien en termes de R^2, dans les deux cas les modèles expliquent en effet plus des trois quarts (78% de R^2 ajusté et 84% de R^2) de la variabilité des montants de PC décidés par le juge, puisque ce sont les mêmes facteurs explicatifs qui sont mobilisés dans les deux régressions (explicitement dans le modèle à trois équations, pour partie explicitement et pour partie implicitement via l'offre et la demande dans le modèle à une équation). Mais, en revanche et surtout, avantage déterminant du modèle triple, les effets de nombreux facteurs, qui apparaissaient faibles ou inexistants avec le modèle unique (car la présence, dans ce modèle, de l'offre et de la demande phagocytent en partie le rôle de ces facteurs), deviennent significatifs dès lors que, en tant que facteurs explicatifs du

[25] Forme de la PC, niveaux de vie, Aide Juridictionnelle, régime matrimonial, durée de mariage, indicateur de conflictualité en termes de changement de qualification en faute, médiation ordonnée par le juge, offre et demande de PC, indicateur de conflictualité en termes de désaccord de CEEE, sexe du juge, ces deux derniers facteurs n'ayant *in fine* pas été retenu dans le tableau ci-dessus du fait de la non-significativité de leurs coefficients estimés.

montant de PC fixé par le juge, l'offre et la demande sont remplacés par les résidus non expliqués des équations d'offre et de demande.

Annexe n° 2 : données de revenus manquantes et biais d'estimation

La source principale de données manquantes sont les variables de revenus qui contribuent au calcul de la différence intra-couple de niveaux de vie. On peut donc légitimement s'interroger sur le risque de biais de l'échantillon du fait de ces exclusions.

Pour nous rendre compte si l'exclusion d'observations du fait de données manquantes dans les variables utilisées pour construire les indicateurs de revenus et de niveau de vie retenus comme variables indépendantes de l'estimation du montant de PC, nous avons créé une indicatrice identifiant les affaires pour lesquelles au moins une information de ce type est manquante et nous avons cherché à savoir, en régressant cette indicatrice sur tout un ensemble de caractéristiques d'affaires, si cette absence d'information était systématiquement dépendante de certaines caractéristiques d'affaire ou au contraire plutôt due au hasard. L'analyse porte sur l'ensemble des affaires pour lesquelles une PC sans accord des parties a été fixée et pour lesquelles les caractéristiques dont nous testons le lien avec l'absence d'information en matière financière ne sont pas elles-mêmes manquantes.

L'analyse montre que peu de caractéristiques d'affaire sont statistiquement liées à l'absence d'information en matière de niveaux de vie. Elle montre cependant une plus forte absence d'information significative, toutes choses égales par ailleurs, d'une part, lorsqu'il s'agit d'affaires impliquant un couple sans enfant et, d'autre part, à mesure que le montant de PC demandé s'élève. Si l'on exclut de la régression les caractéristiques d'affaire sans lien significatif avec l'indicatrice de données manquantes de niveaux de vie, ce qui permet de réintégrer dans l'analyse les quelques affaires exclues du fait de données manquantes en matière de caractéristiques d'affaire, seul le lien entre l'absence d'information de niveaux de vie et le niveau de PC demandé demeure significatif. Le coefficient d'ajustement est de 7%, ce qui est faible mais pas totalement négligeable, ce qui n'autorise donc pas à exclure toute existence de biais de sélection.

Variable dépendante : au moins une information pour le calcul des niveaux de vie est manquante

Constante	0,13 NS	0,10 ***
Montant de PC demandé (en €/mois)	5,72 ***	4,66 ***
Nombre d'enfants à charge	-0,04 *	-0,004 NS
Autres caractéristiques d'affaires (24 variables)	NS	/
Qualité de l'ajustement : R^2	**15,3%**	**7,2%**
Effectifs	**572**	**747**

Source : base COMPRES, 2013. NS : non significatif au seuil de 10%. * : significatif au seuil de 5%. *** : significatif au seuil de 0,1%. / : facteur non introduit dans l'équation.

Annexe n° 3 : estimation du montant de PC homologué par le juge, lorsque les parties sont d'accord sur le montant de PC

L'analyse principale présentée dans le corps du texte portait sur les affaires où les époux ne sont pas d'accord sur le montant de PC ; qu'en est-il lorsqu'il y a accord ? Dans ce cas de figure, estimer le montant de la PC fixé est équivalent à estimer le montant de PC offert ou le montant de PC demandé, puisqu'ils sont égaux. Le tableau ci-dessous rassemble les résultats issus de cette analyse en cas d'accord (colonne n°3), analyse qui a été menée, du point de vue de la sélection des variables indépendantes, comme précédemment. Les colonnes n°1 et n°2 ne sont qu'un rappel de ce que nous avons présenté et commenté *supra* à l'annexe 1 (et dans le corps du texte pour ce qui concerne la colonne n°1).

	Les parties ne sont pas d'accord		Avec accord
	MCO, modèle à trois équations	**MCO, modèle à une seule équation**	**MCO, modèle à une seule équation**
Constante	-9 199 NS	5 830 NS	-100 961 #
PC sous forme uniquement de rente	16 603 *	16 662 **	21 662 NS
Différence de niveau de vie intra-couple (€/mois)	12,05 ***	6,29 *	26,74 ***
Niveau de vie de l'époux (€/mois)	-6,81 *	-5,82 *	-6,00 NS
L'épouse bénéficie de l'AJ à taux plein	-9 570 **	-4053 NS	-9 529 NS
Régime matrimonial moins favorable à l'épouse	13 468 **	821 NS	27 666 **
Durée de mariage en années	918 ***	294 *	2 433 ***
Changement dans la qualification en faute	14 187 *	7 287 NS	62 317 NS
Médiation ordonnée par le juge	25 071 **	25 793 **	-14 218 NS
L'épouse est en emploi	/ NS	/ NS	-28 433 **
Hausse prévisible des revenus de l'époux	25 071 **	/ NS	/ NS
Montant de pension alimentaire épouse (€/mois)	56,8 ***	9,1 #	/
Montant de dommages et intérêts pour l'épouse	2,26 *	/ NS	/ NS
Motivation du juge : mauvaise santé de l'épouse	20 282 ***	/ NS	/ NS
Motivation du juge : l'épouse collabore à l'activité du conjoint	16 232 *	15 404 *	/ NS
Résidus de l'équation d'offre de PC	0,66 ***	/	/
Résidus de l'équation de demande de PC	0,18 ***	/	/
Offre de PC	/	0,64 ***	/
Demande de PC	/	0,18 ***	/
Qualité de l'ajustement : R^2	**83,6%**	**83,6%**	**37,7%**
Qualité de l'ajustement : R^2 ajusté	**77,6%**	**77,7%**	**28,4%**
Effectifs	**541**	**541**	**1176**

Source : base COMPRES, 2013. AJ : Aide Juridictionnelle. Est ajouté un jeu d'indicatrices identifiant les différents TGI. # : significatif au seuil de 10%. * : significatif au seuil de 5%. ** : significatif au seuil de 1%. *** : significatif au seuil de 0,1%. NS : non significatif au seuil de 10%. / : facteur non introduit dans la régression.

Le parallèle entre les deux modèles à équation unique, très similaires en termes de facteurs explicatifs (à la pension alimentaire de l'épouse, exclue pour cause de données manquantes trop importantes dans le modèle portant sur les accords, et aux variables d'offre et de demande près[26]), relatifs aux cas d'accord et aux cas de désaccord montre ceci :

- Le pouvoir explicatif du modèle relatif aux cas de désaccord (R^2 ajusté = 78%) est presque trois fois supérieur à celui concernant les cas d'accord (R^2 ajusté = 28%). Ceci peut s'expliquer par le fait qu'en excluant l'offre et la demande, on ne tient pas compte de l'effet de tout un ensemble de déterminants inobservés qui, dans les modèles sans accord sont pris en compte, soit directement dans les variables d'offre et de demande (colonne n°2), soit dans les résidus de ces deux variables (colonne n°1). Quoi qu'il en soit des explications envisageables, cette faiblesse de la qualité d'ajustement jette le doute sur la capacité de ce modèle à générer des prédictions susceptibles d'aider les avocats à conseiller leurs clients quant au montant de PC à proposer, d'un commun accord, au juge.
- Certains facteurs explicatifs du modèle relatif aux cas d'accord se retrouvent parmi ceux qui contribuent le plus à l'explication du montant de PC en cas de désaccord (selon le modèle à trois équations puisque le modèle à équation unique voit les facteurs offre et demande de PC s'approprier une partie du rôle des facteurs énumérés dans le Code civil notamment). Ainsi,

[26] Dans le modèle avec accord la variable « épouse en emploi » est la seule variable qui apparait significative alors qu'elle ne l'était pas dans le modèle sans accord.

dans les modèles relatifs à ces deux catégories d'affaires (affaires avec et affaires sans accord), l'effet le plus important est celui de la différence de niveau de vie entre époux : le coefficient standardisé vaut 0,63 pour les cas d'accord et 0,50 pour les cas de désaccord – des cas de désaccord pour lesquels les résidus des équations d'offre et de demande, qui ne figurent pas dans le modèle appliqué aux accords, jouent également des rôles majeurs (leurs coefficients standardisés valent en effet, respectivement, 0,38 et 0,26).

- Facteur très significatif pour les cas d'accord comme pour les cas de désaccord[27], la durée de mariage a, pour les premiers, un impact supérieur de plus de moitié à ce qu'il représente pour les seconds (918 *versus* 2 433).
- Quant aux autres facteurs, ils diffèrent en termes d'ampleur de leurs effets et surtout du fait que la plupart de ceux qui sont significatifs pour les cas de désaccord ne le sont plus pour les cas d'accord.
- Somme toute le modèle pour les cas d'accord ne présente pas beaucoup d'intérêt dans une perspective de justice prédictive.

[27] Et un effet, comparativement à ceux des autres facteurs, relativement élevé (bien que moindre que celui de la différence de niveau de vie ou de ceux des propositions des parties) : coefficient standardisé égal à 0,19 en cas d'accord et à 0,17 en cas de désaccord.