

Comment passer des anciennetés aux durées ? Illustration à partir de l'Enquête Famille et Logement de 2011.

Domaine : économétrie des modèles de durées

Costemalle Vianney *.

* *Unité des études démographiques et sociales, Insee.*

Introduction En sciences économiques et sociales, mesurer les durées que passent les individus dans des situations particulières, comme la durée de chômage ou la durée de vie en couple, permet de mieux comprendre les parcours de vie ainsi que les dynamiques de transition d'un état à un autre. La plupart des modèles de durées sont basées sur des échantillonnages dans le flux et peu concernent des échantillonnages dans le stock qui pourtant ne sont pas rares puisqu'ils sont obtenus par enquêtes transversales. Nous nous intéressons ici à un cas particulier d'échantillonnage dans le stock, c'est-à-dire que seules les personnes qui sont dans la situation d'intérêt au moment de l'enquête sont observées, pour lequel toutes les durées sont censurées sur la droite. Ces durées, correspondant au temps entre le début de la situation et l'instant de l'enquête sont alors appelées anciennetés. Deux effets antagonistes structurent ces anciennetés, avec d'une part un effet de censure dont le biais sur les durées est négatif et d'autre part un effet de sélection, résultant du fait que les durées plus longues ont une plus grande probabilité d'être observées, dont le biais est positif. La problématique est ici de caractériser les durées réelles en fonction des seules anciennetés et de proposer une méthode permettant de rectifier ces deux biais.

Méthode et résultats Nous reprenons en la modifiant, une méthode développée par S.Nickell qui consiste à estimer par maximum de vraisemblance le risque instantané de sortir de la situation d'intérêt. Contrairement à l'approche de S.Nickell qui propose une estimation paramétrique du risque instantané, nous estimons un risque instantané constant par morceaux, ce qui offre plus de souplesse au modèle et permet alors de mieux estimer certaines données. De plus, des covariables sont ajoutées afin de mettre en évidence des différences de survie selon les caractéristiques individuelles retenues. Nous inspirant du modèle de Cox, les risques instantanés sont proportionnels entre eux permettant ainsi d'estimer un paramètre de régression du logarithme du risque instantané sur les covariables. Tous les paramètres du modèle sont alors estimés en maximisant la vraisemblance des observations. Il est néanmoins nécessaire d'estimer au préalable les flux d'entrées dans la situation d'intérêt ce qui nécessite en général une source alternative d'information. Une fois estimé le risque instantané, il est alors possible d'en déduire la fonction de survie, donnant la proportion d'individus toujours dans la situation au bout d'un temps donné. Dans le cas particulier d'un flux constant, on trouve alors que la densité de l'ancienneté est proportionnelle à la fonction de survie de la durée réelle, ce qui permet de caractériser directement cette durée. Si de plus la durée suit une loi exponentielle, on montre que l'ancienneté suit exactement la même loi. Cette méthode est ensuite appliquée aux données issues de l'Enquête Famille et Logement de 2011 (EFL) afin d'estimer la durée passée en famille monoparentale montrant alors que c'est la cause d'entrée en situation de monoparentalité qui a le plus fort impact sur la durée de cette situation, les personnes ayant eu des enfants hors couple restant plus longtemps que celles qui se sont séparées de leur conjoint ou dont le conjoint est décédé. La source annexe d'estimation des flux est l'Étude des Relations Familiales et Intergénérationnelles de 2005 (ERFI). De plus, en comparant le risque instantané estimé par une fonction continue par morceaux à partir de l'EFL et le risque instantané estimé par l'estimateur de Nelson-Aalen à partir d'ERFI, nous retrouvons des résultats semblables avec un risque instantané en forme de U.

Discussion Deux hypothèses sont nécessaires à l'estimation du modèle : la proportionnalité des risques instantanés et l'indépendance entre la variable de durée et la variable de flux. Nous n'avons pas développé de tests permettant de vérifier ces hypothèses, mais des simulations ont permis de voir qu'une corrélation négative entre la date de début de situation et la durée de cette situation entraîne une sous-estimation de la fonction de survie. De plus, les simulations permettent de vérifier la robustesse du modèle ainsi que la vitesse de convergence sur des cas très divers.