



Estimación de la función de riesgo de rezago académico universitario: un enfoque basado en el análisis de datos funcionales

Luz Enith Restrepo Chavarría

Trabajo de grado para optar al título de Magíster en Economía

Directora: Edel Laura Sánchez Higueta, Magíster en Economía

Codirector: Santiago Alejandro Gallón Gómez, Doctor en matemáticas aplicadas

Universidad de Antioquia
Facultad de Ciencias Económicas
Maestría en Economía
Medellín
2023

Cita	(Restrepo, 2023)
Referencia	Restrepo, L. E. (2023). <i>Estimación de la función de riesgo de rezago académico universitario: un enfoque basado en el análisis de datos funcionales</i> Tesis de Maestría. Universidad de Antioquia, Medellín.
Estilo APA 7 (2020)	



Maestría en Economía, Cohorte 2017.
 Grupo de Investigación de Microeconomía aplicada.



Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director Jair Albeiro Osorio Agudelo.

Jefe departamento: Claudia Cristina Medina Palacios.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Índice

Resumen	5
Abstract	6
Introducción	7
1 Marco de referencia y revisión de la literatura.	10
2 Datos y métodos	14
3 Resultados.	18
4 Discusión y conclusiones	25
Referencias	28

Lista de tablas

- | | | |
|---|---|----|
| 1 | Estimación no paramétrica de Kaplan–Meier de la función de riesgo de rezago acumulado para los programas Economía y Medicina. | 20 |
|---|---|----|

Lista de figuras

- | | | |
|---|---|----|
| 1 | Ejemplo de la brecha de créditos acumulados en la versión 7 del programa Economía. En la parte superior de la línea está el número de semestres del programa, debajo de ella son los créditos por semestre y en negrilla está el valor de los créditos acumulados en la trayectoria académica | 18 |
| 2 | Tasa de créditos rezagados acumulados de la cohorte 2009:2. La línea negra representa la curva de rezago entre estudiantes y las líneas verticales indican el semestre teórico de finalización. | 19 |
| 3 | Ejemplo de tipos de curvas de riesgo de rezago acumulado. La línea vertical indica el semestre teórico de finalización en el programa Economía. | 21 |
| 4 | Curvas representativas (línea negra) de la función de riesgo de rezago de las cohortes 2005:1–2020:1. | 22 |
| 5 | Curvas de rezago predichas para tres cohortes de Economía y Medicina. La línea continua son los semestres observados en cada cohorte, y la línea punteada corresponde a la predicción. | 23 |
| 6 | Curvas de rezago predichas para cohortes de Biología, Derecho e Ingeniería eléctrica. La línea continua son los semestres observados en cada cohorte, y la línea punteada corresponde a la predicción. | 24 |

Resumen

El rezago académico es un factor asociado a las altas y crecientes tasas de graduación tardía en educación superior, las cuales revelan que los esfuerzos de política pública están concentrados en aumentar la cobertura y la permanencia sin crear estrategias que garanticen la finalización oportuna de los estudiantes universitarios. Aunque el tiempo adicional requerido para graduarse es motivo de preocupación entre académicos y hacedores de política, dado el aumento de los costos individuales, económicos y sociales, existe poca literatura y acciones para medir e identificar el comportamiento de los estudiantes a lo largo de su trayectoria académica. Por lo tanto, en esta investigación se propone aplicar la metodología de análisis de datos funcionales para caracterizar las funciones de riesgo de rezago académico acumulado a nivel agregado (cohortes y programas académicos) del ciclo académico de los estudiantes de la Universidad de Antioquia (Colombia). Los resultados demuestran la utilidad del método para identificar patrones de riesgo de rezago acumulado y pronosticar el comportamiento de rezago de futuros semestres de cohortes parcialmente observadas, aportando información para diseñar políticas educativas que disminuyan el tiempo de graduación por parte de las autoridades académicas.

Palabras clave: Economía de la educación; educación superior; rezago académico; funciones de riesgo de rezago acumulado; datos longitudinales dispersos; análisis de datos funcionales, análisis de componentes principales funcionales.

Abstract

Delay time-to-degree is associated with high and increasing delayed graduation rates in higher education, which reveal public policy efforts are focused in increasing coverage and permanence without creating strategies to guarantee university students'timely completion. Although time required to graduate is a concern among scholars and policy-makers, due to individual, economical and social costs rise, there is little literature and actions to measure and identify students'behavior throughout their academic trajectory. Therefore, this research aims at applying functional data analysis methodology in order to characterize the accumulative educational lag hazard rates at an added level (cohorts and academic programmes). The results show the methodology's usefulness to identify cumulative lag hazard patterns and prognosticate upcoming partially-observed cohorts'lag behavior, by providing information to design educational policies to reduce the time needed for graduation.

Keywords: Education economics; higher education; delay time-to-degree; cumulative lag hazard rates; sparse longitudinal data; functional data analysis; functional principal component analysis.

Introducción

Las matrículas en la educación superior en Colombia crecieron 83,14% entre 2006 y 2016 como resultado de estrategias que buscan aumentar el acceso a la educación terciaria (Ministerio de Educación Nacional, 2017). Sin embargo, existe un retraso en el logro de la graduación oportuna. Por ejemplo, de los 1.812.500 estudiantes que iniciaron su pregrado en 2012 solo 23,35% se graduaron a tiempo (Sistema Nacional de Información de la Educación Superior, 2021).

En la literatura se han identificado dos causas que explican la diferencia entre el número de estudiantes matriculados y graduados: la deserción y el rezago académico. La deserción (abandono voluntario del pregrado por parte de un estudiante) ha sido ampliamente estudiado siguiendo el enfoque de los trabajos seminales de Tinto (1975, 1982, 1993). Por su parte, el retraso en la graduación empezó a analizarse en la década de los noventa bajo lineamientos similares a los de la deserción (Robinson, 1990; Volkwein and Lorang, 1996; Löfgren and Ohlsson, 1999), no obstante, ha sido menos estudiado que la primera causa.

En tal sentido, se ha relegado la importancia de medir, entender y diferenciar las conductas de los estudiantes a pesar de la preocupación generalizada entre actores académicos y gubernamentales por las bajas tasas de finalización oportuna y el aumento en los costos individuales, económicos y sociales asociados al tiempo adicional requerido por los estudiantes para completar su programa académico y obtener el título (DesJardins et al., 2002; Belfield et al., 2016; Huntington-Klein and Gill, 2021). Esta situación agudiza la situación financiera principalmente de las universidades públicas encargadas de asumir total o parcialmente los costos financieros del represamiento (Brunello and Winter-Ebmer, 2003; Brandsma-Dieters, 2013).

Aunque la literatura tradicional de rezago se ha centrado en comprender las razones por las que los estudiantes no se gradúan a tiempo, son análisis que miden las bajas tasas de finalización una vez transcurren los cuatro o cinco años de duración de los programas (Aina et al., 2018). Por lo tanto, un sistema de alerta temprana de rezago académico es importante para una universidad porque les permite a los formuladores de política educativa monitorear conductas propensas al retraso en la graduación. Por ejemplo, caracterizar el número de créditos matriculados en el primer semestre es importante, ya que la carga inicial determina la trayectoria y progreso de la acumulación oportuna de créditos (Attewell et al., 2012; Attewell and Monaghan, 2016; Belfield et al., 2016; Chan, 2022). En Estados Unidos con información del proceso de monitoreo en las universidades se creó el programa *15 to*

Finish, a partir del cual aumentó el logro académico de los estudiantes y la probabilidad de graduación en cuatro años (Scott-Clayton, 2011; Chan, 2022).

Es así como este trabajo propone, primero, medir las funciones de riesgo de rezago acumulado en cinco programas de pregrado de la Universidad de Antioquia (sede Medellín) y, segundo, aplicar el análisis de datos funcionales como metodología para caracterizar los patrones históricos de comportamiento de dicho rezago en cada programa; un ejercicio de pronóstico de rezago se efectúa al final. Lo anterior se logra, haciendo uso de datos longitudinales dispersos e irregulares de seguimiento a la trayectoria académica de los estudiantes que iniciaron sus estudios entre 2005 y 2019.

El análisis de datos funcionales (*FDA* por sus siglas en inglés) se ha consolidado como un método alternativo a las herramientas tradicionales para analizar, modelar y predecir datos de trayectoria continua a partir de los trabajos seminales de Ramsay and Dalzell (1991); Ramsay and Silverman (2005); Ferraty and Vieu (2006); Ferraty and Romain (2010). La idea de este enfoque no paramétrico es tomar el conjunto de puntos discretos para agregarlos en una curva continua dentro de un espacio dimensional infinito, a partir de la cual se espera reconocer la estructura subyacente en la trayectoria observada, pues con métodos tradicionales puede ser imperceptible por el comportamiento irregular de los datos o la distorsión del tiempo (Ullah and Finch, 2013).

Adicionalmente, con las funciones se puede calcular la derivada para identificar el cambio de la función con una unidad de tiempo adicional (Ramsay and Silverman, 2005). Por ejemplo, al analizar el riesgo de rezago es posible reconocer la velocidad o la aceleración del riesgo en una cohorte ante la transición de un semestre a otro. Por otro lado, el método funcional permite usar la información de las funciones completamente observadas para predecir el comportamiento de las curvas futuras de funciones incompletas, que en términos de política contribuye a la toma de decisiones anticipadas para prevenir o potenciar el comportamiento predicho.

De acuerdo con los resultados de esta investigación se resaltan dos hallazgos. Primero, se proporciona evidencia sobre las diferencias entre los programas académicos. Esto sugiere que el riesgo de rezago, aunque tiende a aumentar a medida que los estudiantes avanzan en su ciclo académico responde a particularidades propias de cada programa. Segundo, el riesgo de rezago incrementa exponencialmente a partir del segundo semestre independiente de la cohorte académica, lo que indica la presencia de semestres cuello de botella que no han sido solucionados por los cambios de versión en los programas. Estos resultados señalan la

necesidad de profundizar en los factores de riesgo de rezago para formular las estrategias hacia la disminución en el tiempo de graduación, dado que los hallazgos señalan que hay aspectos institucionales que posiblemente inciden en la mala gestión académica de los estudiantes, sin embargo, también hay que delimitar las diferencias entre los estudiantes al interior de una cohorte y los factores individuales que aportan al retraso en la graduación.

En consecuencia, se espera que este trabajo contribuya a la escasa literatura sobre las conductas de retraso en la graduación, y a la toma de decisiones de política en las Unidades Académicas. Se parte de la premisa de incluir métodos alternativos y útiles, que enriquezcan las formas de uso de los datos para monitorear y hacer seguimiento a la gestión académica de forma pertinente (según sus conductas) y oportuna (desde el primer semestre hasta finalizar su permanencia). De esta manera, reducir y tratar el rezago académico con políticas educativas que integren el logro de la permanencia y la graduación oportuna, disminuyendo, por lo tanto, la financiación de un semestre adicional a los definidos por el programa académico, además de agilizar la vinculación laboral o la continuación de los estudios posgraduales (movilización social) de los estudiantes.

Este documento está organizado de la siguiente forma. La Sección 1 presenta la revisión de la literatura sobre el rezago académico. La Sección 2 define la metodología aplicada para estimar las funciones de riesgo de rezago acumulado como datos funcionales. Por último, se presentan los resultados y conclusiones generales en las Secciones 3 y 4, respectivamente.

1 Marco de referencia y revisión de la literatura

Los determinantes del rezago se pueden agrupar en cuatro categorías: i) sociodemográficos y económicos (e.g. género, grupo etario, ingresos familiares, situación laboral), ii) factores pre-institucionales (e.g. desempeño académico, habilidades y conocimientos obtenidos en la educación media, tipo de colegio, puntaje de pruebas estandarizadas), iii) factores institucionales (e.g. plan de estudios, infraestructura, ambiente institucional, vinculación docente-estudiante), y iv) factores individuales (e.g. hábitos de estudio, rendimiento académico, gestión de matrícula). Aunque el objetivo de la investigación no es explicar el rezago a través de sus determinantes es importante tener en cuenta el efecto, especialmente con los factores individuales e institucionales, sobre la formulación de estrategias o el patrón de comportamiento de los estudiantes.

Por ejemplo, las conductas de rezago pueden ser resultado de factores como la organización del plan de estudios (Swail, 2003; Jansen, 2004), la escasez de cursos (Kurlaender et al., 2014), una baja oferta disponible de tutorías (Aina et al., 2018; Pike and Robbins, 2020) y el incremento en la carga de estudio según la distribución de la evaluación en las materias (Jansen, 2004; Huntington-Klein and Gill, 2021). Sin embargo, hay evidencia que sugiere que aun cuando la coordinación del programa es oportuna y exitosa los estudiantes responden de diferente manera a los estímulos del clima organizacional universitario (Stage, 1988). En consecuencia, las decisiones de los estudiantes para administrar su ciclo académico y el rendimiento asociado al desempeño durante un semestre específico también inciden en comportamientos que aumentan la probabilidad de graduación tardía (Bettinger and Long, 2009; Bound et al., 2010).

La literatura que analiza las políticas para aumentar la tasa de graduación oportuna se centra en dos estrategias implementadas en Estados Unidos. La primera iniciativa aborda el impulso académico con la premisa que entre más créditos matriculen los estudiantes por semestre mayor será la tasa de graduación y la finalización oportuna (Adelman, 2004, 2006; Huntington-Klein and Gill, 2021). No obstante, Jansen (2004) muestra que entre mayor cantidad de curso paralelos matriculados mayor es la probabilidad de rezagarse por la menor disponibilidad de tiempo por curso, situación que puede inducir a la cancelación o la reprobación de las materias. Por otro lado, Huntington-Klein and Gill (2021) y Attewell and Monaghan (2016) encuentran que la preparación y los hábitos de estudio son los factores que determinan el éxito o el fracaso de una mayor carga de créditos por semestre.

Con lo expuesto previamente se creó la segunda estrategia relacionada en la literatura.

El programa *15 to Finish* implementado con el propósito de otorgar becas a los estudiantes de bajos ingresos con la condición de matricular y aprobar mínimo 15 créditos por semestre (Scott-Clayton, 2011; Chan, 2022). Chan (2022) encontró que una cohorte de primer año beneficiaria del programa en el estado de Indiana aumentó más rápido el número de créditos acumulados en comparación con una cohorte en condiciones similares excepto por la beca, con lo cual se demuestra que el programa logra que los estudiantes incrementen su esfuerzo y dedicación para acumular créditos de forma oportuna y graduarse a tiempo, lo que sugiere, a su vez, el éxito de los sistemas de monitoreo y seguimiento para diseñar políticas educativas que desincentiven el rezago académico.

A pesar de los resultados positivos de *15 to Finish* no es posible obviar la existencia de características de los estudiantes que limitan el programa como, por ejemplo, los estudiantes con dificultades económicas y los menos preparados académicamente, quienes son los que toman una menor carga de créditos (Attewell and Monaghan, 2016). En tal sentido, es posible que el programa logre cambiar el comportamiento de los menos preparados, pero el primer tipo de estudiante tiende a modificar su disponibilidad de tiempo para combinar el trabajo con sus estudios (Letkiewicz et al., 2014). En consecuencia, el esfuerzo del programa para incluirlos es insuficiente dadas las necesidades que inciden en la búsqueda de trabajo mientras estudian, ya que es una situación que incrementa el tiempo para obtener el título porque a mayor número de horas dedicadas al trabajo mayor es el efecto negativo sobre el número de créditos matriculados (Thomas, 2002; Jansen, 2004; Chan, 2022).

Con relación a los métodos de enfoque cuantitativo para analizar el rezago académico, se identifica que estos han sido tradicionalmente de tipo inferencial, y que los modelos más comunmente usados por la literatura, son los econométricos lineales, los de elección discreta y de conteo, y aquellos que usan el método de variables instrumentales para identificar la relación y el efecto entre los factores que retrasan la graduación a tiempo (Tagharrobi et al., 2013; Kurlaender et al., 2014; Letkiewicz et al., 2014; Theune, 2015; Attewell and Monaghan, 2016; Belfield et al., 2016). Otros métodos presentes, pero menos implementados en economía de la educación, son las diferencias en diferencias, que establece el efecto de las políticas educativas sobre la tasa de graduación oportuna (Umbricht et al., 2015; Birdsall, 2018; Chan, 2022) y la aplicación de una función de riesgo logit de tiempo discreto desarrollada por Yue and Fu (2017) para determinar los factores que afectan la graduación. En términos de uso de datos, los trabajos anteriores coinciden en que la unidad de observación es el estudiante, en su mayoría agrupados por cohortes, y usan la información académica cuya fuente de datos son los sistemas de información de las universidades para analizar causas, consecuencias o estrategias alrededor del rezago.

Es importante resaltar que la literatura no ha abordado el rezago estimando funciones de riesgo de rezago acumulado. El estudio de mayor proximidad al objeto de análisis es planteado por [Yue and Fu \(2017\)](#), quienes modelan la graduación a través de una función de riesgo logit de tiempo discreto para determinar cuáles son los factores que afectan la obtención del título. No obstante, a diferencia de estos autores, esta investigación usa la función de riesgo calculada con el estimador de Kaplan–Meier acumulando el comportamiento a lo largo del tiempo (semestres académicos).

El indicador de Kaplan–Meier ([Kaplan and Meier, 1958](#)) es un estimador no paramétrico que determina la probabilidad condicional de ocurrencia de un evento (rezago) en cada punto temporal t del seguimiento a las unidades de observación (estudiantes) hasta el instante $t - 1$, por medio de la función de riesgo h

$$h(t) = P(T = t | T \geq t)$$

La ventaja del estimador es que tiene en cuenta la presencia de observaciones incompletas (i.e. censuras, e.g. estudiantes que salen de la Universidad por deserción, insuficiencia académica o graduación), para las cuales no es posible determinar con precisión el riesgo del evento al finalizar el periodo de seguimiento. De modo que, la función de riesgo de rezago se calcula como

$$\hat{h}(t) = \prod_{j/t_j \leq t} \left(1 - \frac{d_j}{n_j} \right) \in [0, 1]$$

donde d_j es el número de unidades de observación que presentan el evento al finalizar el punto temporal de seguimiento y n_j el número de unidades de observación no censuradas en ese mismo periodo. Esta función representa la probabilidad de que una unidad de observación se rezague hasta el tiempo t_j (e.g. número total de semestres en un programa).

Este estimador, de acuerdo con la literatura, se ha usado principalmente en las áreas de bioestadística y epidemiología para identificar la supervivencia asociada a la exposición en tratamientos de enfermedades crónicas ([May, 2009](#); [Confavreux and Vukusic, 2014](#); [Stalpers and Kaplan, 2018](#); [Zee and Xie, 2018](#)), mercado laboral para estudiar los tiempos de duración de desempleo ([Hoffman, 1991](#); [Dolton et al., 2005](#); [García-Cabo and Madera, 2019](#)), y educación, especialmente en el análisis de la deserción estudiantil ([Booth and Satchell, 1995](#); [Castaño et al., 2006](#); [Vallejos and Steel, 2017](#); [Csalodi and Abonyi, 2021](#)).

Con las funciones de riesgo calculadas con el estimador de Kaplan–Meier se propone en este documento explorar el análisis de datos funcionales, el cual se ha usado principalmente en

literatura de biomedicina (Ullah and Finch, 2013) y demografía (Erbas et al., 2007; Hyndman and Shahid Ullah, 2007; Chiou and Müller, 2009; Wang et al., 2016; Shang, 2018). Este método tiene un enfoque no paramétrico y permite identificar el patrón de la trayectoria de los datos a lo largo del tiempo como muestra de funciones aleatorias en el espacio funcional $L^2(S)$. De manera que, el conjunto de funciones X son medibles en el intervalo S , tal que $X(s)^2 < \infty$. Entre las fortalezas del análisis se destaca su flexibilidad en el sentido de que los intervalos de tiempo en las observaciones no tienen que estar igualmente espaciados (datos dispersos) (Ullah and Finch, 2013), y permite analizar la dinámica de la muestra de funciones sin hacer supuestos sobre su distribución.

2 Datos y métodos

Este trabajo hizo uso de información académica (número de créditos matriculados y aprobados por semestre) de 13.881 estudiantes de los programas de Biología, Derecho, Economía, Ingeniería eléctrica y Medicina de la Universidad de Antioquia (sede Medellín) de las cohortes 2005:1–2019:2. Se obtuvo la información de los sistemas institucionales de la Universidad de Antioquia consolidados en la plataforma DataUdeA.

A partir de los datos anteriores se calculó un indicador de riesgo de rezago acumulado para cada cohorte de cada programa académico, y luego, usando el análisis de datos funcionales, se construyeron las curvas típicas históricas de rezago de cada programa. A partir de dichas curvas, se realiza un análisis descriptivo del comportamiento entre cohortes al interior de cada programa y entre todos los programas. Y también se lleva a cabo un ejercicio de predicción.

Para calcular el indicador de riesgo de rezago acumulado es necesario establecer y estimar una medida de rezago. Con base en la tasa de créditos rezagados acumulados (TCRZ) propuesta por [Gallón et al. \(2022\)](#) se mide la propensión de los estudiantes (agregados en cohortes) a rezagarse durante su permanencia en la universidad.

Con la TCRZ se puede identificar si un estudiante se encuentra rezagado en un programa específico de acuerdo al plan de estudios, en el cual se definen los cursos obligatorios y electivos por semestre del programa. La TCRZ compara el número de créditos aprobados acumulados esperados (i.e. definidos por programa) $\sum_{j=1}^t C_k^{(p)}(j)$ con el número de créditos aprobados acumulados observados $\sum_{j=1}^t CA_{i,k}^{(p)}(j)$ al finalizar el semestre t , con $t = 1, \dots, T$ (= número de semestres, para el estudiante i que pertenece a la cohorte k del programa p). De esta forma, la TCRZ se define como:

$$\text{TCRZ}_{i,k}^{(p)}(t) = \frac{\sum_{j=1}^t C_k^{(p)}(j) - \sum_{j=1}^t CA_{i,k}^{(p)}(j)}{\sum_{j=1}^t C_k^{(p)}(j)}.$$

La TCRZ es útil porque permite identificar el estado y la evolución de los estudiantes hacia la graduación oportuna desde el primer semestre académico, lo que facilita tomar decisiones oportunas para prevenir o tratar a tiempo el rezago de los estudiantes.

Una vez el indicador TCRZ se estima para cada estudiante o conjunto de ellos, la función de riesgo de rezago para cada cohorte se construye a partir del estimador no paramétrico de Kaplan-Meier, el cual permite medir la probabilidad de experimentar el rezago

en algún semestre del programa académico. La función de riesgo de rezago está definida por $RZ_k^{(p)}(t)$, y con A_t como el número de estudiantes activos de una cohorte al inicio del semestre t , los cuales son la diferencia entre el número de estudiantes matriculados y la cantidad de observaciones censuradas (estudiantes graduados, desertores o con rendimiento académico insuficiente), y AR_t como los estudiantes rezagados de la cohorte en el periodo t , calculada con base en la TCRZ:

$$RZ_k^{(p)}(t) = \prod_{k/t_j \leq t} \left(1 - \frac{AR_t}{A_t} \right) \in [0, 1]$$

$$AR_t = \sum_{i=1}^{n_k^{(p)}(t)} 1 \left(TCRZ_{i,k}^{(p)}(t) > 0 \right),$$

donde $n_k^{(p)}(t)$ es el número de estudiantes del programa p de la cohorte k al finalizar el semestre t , y $1(x_i)$ una variable indicadora de rezago,

$$1 \left(TCRZ_{i,k}^{(p)}(t) > 0 \right) = \begin{cases} 0, & \text{si } TCRZ_{i,k}^{(p)}(t) \leq 0 \\ 1, & \text{si } TCRZ_{i,k}^{(p)}(t) > 0. \end{cases}$$

Debe aclararse que la sumatoria en AR_t , otro de los indicadores propuestos en [Gallón et al. \(2022\)](#), no es sobre el total de estudiantes del programa y cohorte, $n_k^{(p)}$, ya que el rezago se da en estudiantes activos que avanzan semestralmente, excluyendo los que van saliendo por deserción, bajo rendimiento, o graduación.

Para construir las curvas típicas de rezago por programa se consideró que las funciones de riesgo de rezago son una muestra de datos funcionales, ya que las observaciones son vistas como trayectorias de un proceso estocástico continuo y suave $RZ = \{RZ^{(p)}(t), \dots, RZ^{(p)}(t) \in R : t \in T\}$, donde \mathcal{T} es el intervalo continuo que discretiza los datos en el momento que son observados, $(t_1, \dots, t_{n_i}) \in \mathcal{T}$, y es un espacio semiabierto asociado al continuo de los semestres académicos ([Ferraty and Vieu, 2006](#)).

En el FDA clásico se asume que las curvas estimadas son observadas en su totalidad con una misma longitud. Sin embargo, existen datos longitudinales donde los puntos temporales son dispersos o espaciados de forma desigual ([James et al., 2000](#); [Yao et al., 2005](#); [Peng and Paul, 2009](#)). Este es el caso de la información académica analizada, que contiene datos de seguimiento de la trayectoria universitaria semestre a semestre, en el que las unidades de observación (estudiantes) no tienen la misma longitud porque, por ejemplo, hay estudiantes

que se retiran del programa voluntariamente (deserción), por rendimiento académico insuficiente (ver el Acuerdo 1 de 1981, por el cual se explide el reglamento estudiantil y de normas académicas de pregrado de la [Universidad de Antioquia \(2015\)](#)), o al cambiar de estado en el programa una vez culminan con el plan de estudios (i.e. egreso), entre otros.

El enfoque de este trabajo está basado en el análisis de componentes principales funcionales (FPCA, por sus siglas en inglés) para calcular la función representativa de la estructura de rezago intracohorte de un programa académico basada en curvas aleatorias disponibles.

Las curvas del programa académico se estiman mediante los valores discretos de la función de riesgo de rezago acumulado de la cohorte $k = \{2005 : 1, \dots, 2019 : 2\}$ en el semestre $t = \{1, \dots, T\}$. Sobre estos puntos se construye la función $RZ^{(p)}(t)$ suavizada para minimizar las desviaciones a corto plazo y, de esta forma, resaltar los patrones de comportamiento de cada curva.

La técnica usada para suavizar es B-Spline cúbico sobre la estructura de modelos de efectos mixtos, dado su carácter flexible, simple y de adaptación en funciones no periódicas ([Ullah and Finch, 2013](#)). Esta base es un conjunto de funciones polinómicas de M trozos de grado m en el intervalo T , donde los tramos o subsecciones de la curva de la cohorte se unen suavemente a través de nodos equiespaciados ([Hastie et al., 2009](#)).

$$RZ^{(p)}(t) = Y_i(t) = \mu(t) + \mathbf{f}(t)^\top \boldsymbol{\alpha}_i + \varepsilon_i(t), \quad i = 1, \dots, N,$$

donde $\mu(\cdot)$ es la curva representativa, $\mathbf{f}(\cdot)$ el vector de funciones de componentes principales, $\boldsymbol{\alpha}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ el vector de ponderaciones, y $\varepsilon_i(\cdot) \sim \mathcal{N}(0, \sigma^2)$

El método pretende aproximar cada curva para obtener la función representativa ($\mu(t)$) estimando los parámetros de suavización por máxima verosimilitud restringida con los algoritmos Newton-Raphson y Expectation Maximization ([James et al., 2000](#); [Peng and Paul, 2009](#)).

Finalmente, a partir del FDA se realizó un ejercicio de predicción usando el algoritmo de análisis principal por estimación condicional (PACE, por sus siglas en inglés) propuesto por [Yao et al. \(2005\)](#) para datos longitudinales dispersos o espaciados irregularmente, que puede ser aplicado mediante la librería `fdapace` ([Zhou et al., 2022](#)) en el lenguaje R ([R Core Team, 2021](#)).

El método se basa en la descomposición de las primeras K funciones de componentes principales de las curvas de rezago por medio del modelo

$$\begin{aligned} y_{ij} &= f_i(t_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n, j = 1, \dots, n_i, t_{ij} \in [0, T] \\ &= \mu(t_j) + \sum_{k=1}^K \beta_{ik} \phi_k(t_{ij}) + \varepsilon_{ij}, \end{aligned}$$

donde $\mu(t_j)$ es la función o perfil medio, $\{\phi_k(t_{ij})\}$ el conjunto de componentes principales funcionales (o bases ortonormales), $\{\beta_{ik}\}$ el respectivo conjunto de coeficientes (scores), y $\varepsilon_{ij} \stackrel{\text{iid}}{\sim} (0, \sigma^2)$ el término de error.

Con base en la información y_{ij} en los puntos de tiempo t_{ij} , $j = 1, \dots, n_i$, de una nueva cohorte o para la cual se tienen pocas mediciones, y las estimaciones de la función media, las primeras K componentes principales funcionales y scores $\{\hat{\mu}(t), \hat{\phi}_k(t), \hat{\beta}_{ik}, k = 1, \dots, K\}$, la predicción de la trayectoria $f_i(t)$ para la i -ésima cohorte está dada por

$$\hat{f}_i(t) = \hat{\mu}(t) + \sum_{k=1}^K \hat{\beta}_{ik} \hat{\phi}_k(t)$$

Con el algoritmo, se dividió aleatoriamente el 80% de las curvas de rezago en un conjunto de entrenamiento y el 20% restante en un conjunto de prueba. La trayectoria de las curvas predichas se construyó utilizando la estimación de la función de covarianza para obtener las componentes principales funcionales y luego calcular los scores bajos a través de un enfoque de estimación condicional con el supuesto de normalidad de β_{ik} y ε_{ij} .

3 Resultados

La Figura 1 muestra un ejemplo de la gestión académica de un estudiante matriculado en la versión 7 de Economía. El plan de estudios del programa tiene 150 créditos distribuidos en 16 créditos por semestre a excepción del último con seis créditos. Para cualquier pregrado, el estudiante tiene dos opciones: administrar sus créditos académicos según lo definido por el plan de estudios o de manera autónoma matricular mayor o menor número a lo estipulado. La TCRZ compara la situación esperada con la observada en cada estudiante. De acuerdo con el ejemplo, en el séptimo semestre de Economía el estudiante debería tener 112 créditos aprobados acumulados, sin embargo, solo tiene 94 créditos porque aprobó un menor número de créditos en los primeros semestres y fue acumulando el rezago semestre a semestre.

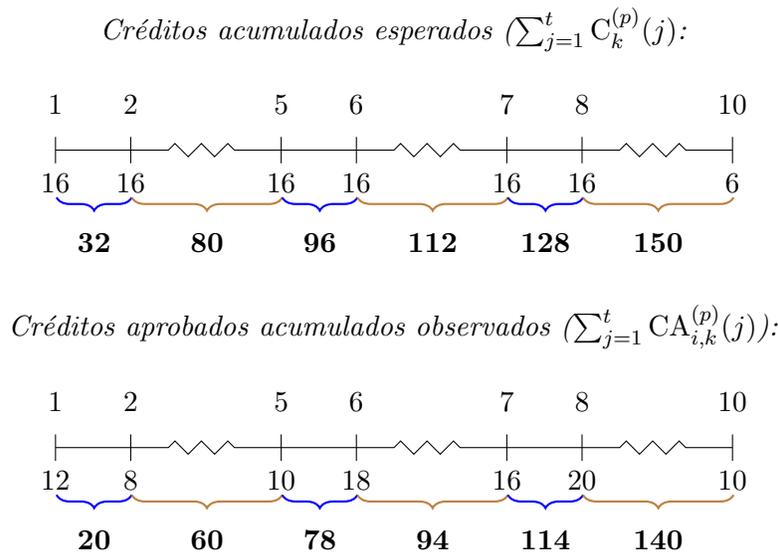


Figura 1
Ejemplo de la brecha de créditos acumulados en la versión 7 del programa Economía. En la parte superior de la línea está el número de semestres del programa, debajo de ella son los créditos por semestre y en negrilla está el valor de los créditos acumulados en la trayectoria académica

En términos agregados, la Figura 2 muestra la tasa de créditos rezagados acumulados, $TCRZ_{i,k}^{(p)}(t)$, de los estudiantes de la cohorte en los programas con su respectiva curva representativa de rezago (línea negra). Economía muestra una tendencia creciente de rezago y relativamente estable entre el tercer y noveno semestre (alrededor de 30%), luego la curva decrece con una tasa similar a la inicial (aproximadamente 18%) fluctuando con incrementos

leves hasta decaer en el semestre 16, donde alcanza una tasa negativa. En cambio, Medicina presenta tasas inferiores a 1% en la totalidad de la trayectoria observada pese a que el último estudiante de la cohorte tardó 19 semestres en cumplir el plan de estudios del programa.

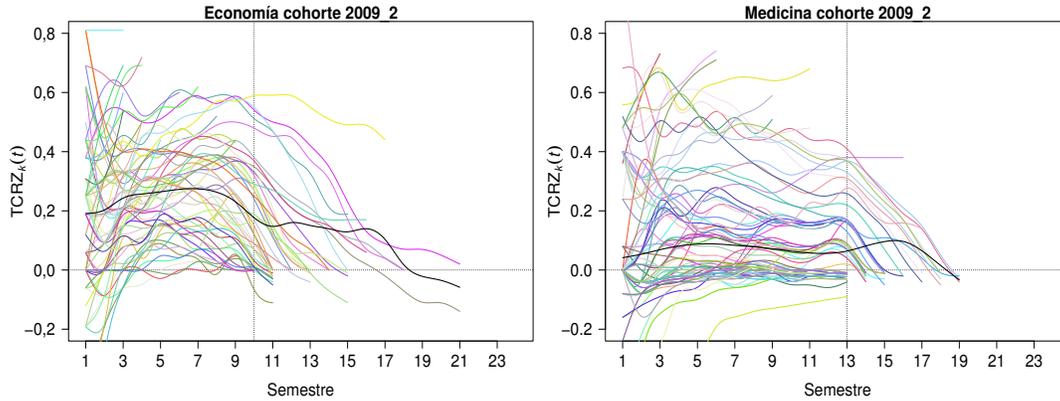


Figura 2

Tasa de créditos rezagados acumulados de la cohorte 2009:2. La línea negra representa la curva de rezago entre estudiantes y las líneas verticales indican el semestre teórico de finalización.

Es importante resaltar que las curvas con tasas de rezago negativas durante los primeros semestres o toda la carrera, en su mayoría en Medicina, se deben principalmente a tres situaciones. Primero, estudiantes con una mayor oferta para matricular créditos de los definidos en el plan de estudios dado su buen rendimiento académico. Segundo, estudiantes que validan créditos al presentar exámenes de clasificación y validación. Por último, estudiantes de transferencia o cambio de programa que validan los créditos aprobados una vez son estudiantes matriculados en el programa académico de referencia.

El patrón capturado por la curva representativa confirma que los estudiantes de Economía respecto a los de Medicina tienen mayor riesgo de rezago de los créditos acumulados observados versus los esperados, también demuestra que hay estudiantes que difieren del comportamiento promedio de la cohorte aun cuando los dos programas tienen alto riesgo de rezago en semestres superiores. Estos resultados indican la existencia de diferencias en la estructura de rezago entre programas, por lo que es posible que las variaciones de la situación educativa analizada tengan un patrón subyacente según el programa académico y las tipologías de estudiantes intracohorte, sobre todo con las curvas atípicas por encima de la función propia de rezago que exhiben desigualdades en la gestión académica durante la trayectoria del programa, lo que induce a bajas tasas de graduación oportuna.

La Tabla 1 presenta un ejemplo de la estimación de las funciones de riesgo de rezago acumulado, $RZ^{(p)}(t)$, de los programas $p^1 = \text{Economía y Medicina}$ para $T = 15$ semestres de la cohorte $k = 2009:2$. En ambos pregrados para la media de los estudiantes aumenta el riesgo de rezago de la cohorte conforme avanzan en la permanencia del programa. Sin embargo, Economía alcanza el nivel de riesgo máximo (100 %) en el cuarto semestre, que, a diferencia de Medicina, tiene un aumento progresivo limitado entre semestres dada la alta probabilidad de rezago que obtuvieron los estudiantes en el primer periodo académico.

Tabla 1

Estimación no paramétrica de Kaplan–Meier de la función de riesgo de rezago acumulado para los programas Economía y Medicina.

Programa	Variable	Semestre									
		1	2	3	4	5	6	7	8	9	10
Economía	A_t	96	74	68	62	55	53	52	51	48	46
	AR_t	73	54	60	58	52	51	51	49	43	36
	$D^{(p)}(t)$	13	3	6	5	1	0	1	2	2	0
	$M^{(p)}(t)$	9	3	0	2	1	1	0	1	0	0
	$G^{(p)}(t)$	0	0	0	0	0	0	0	0	0	8
	$RZ^{(p)}(t)$	0.76	0.94	0.99	1	1	1	1	1	1	1
Medicina	A_t	131	129	126	124	124	123	121	121	121	119
	AR_t	29	46	51	61	69	72	65	71	69	49
	$D^{(p)}(t)$	2	3	0	0	1	0	0	0	2	0
	$M^{(p)}(t)$	0	0	2	0	0	2	0	0	0	0
	$G^{(p)}(t)$	0	0	0	0	0	0	0	0	0	74
	$RZ^{(p)}(t)$	0.22	0.50	0.70	0.85	0.93	0.97	0.99	0.99	1	1

1. A_t = estudiantes activos,
2. AR_t = estudiantes rezagados,
3. $D^{(p)}(t)$ = estudiantes desertores,
4. $M^{(p)}(t)$ = estudiantes con insuficiencia académica,
5. $G^{(p)}(t)$ = número de graduados , y
6. $RZ^{(p)}(t)$ = tasa de riesgo de rezago

Un ejemplo de tres posibles tipos de funciones de riesgo acumulado en estudiantes de Economía se ilustra en la Figura 3. El estudiante con riesgo de rezago acumulado alto finaliza su primer semestre con una probabilidad de rezago superior a 80 %, en este caso el rezago persiste durante los semestres observados. En el riesgo moderado está el estudiante que en los primeros semestres tiene conductas de riesgo de rezago, pero logra disminuir la brecha entre el número de créditos esperados versus los observados, lo cual le permite graduarse oportunamente. Por su parte, el estudiante con riesgo de rezago acumulado bajo son principalmente casos donde la TCRZ es ≤ 0 .

¹En el documento se presentan solo los resultados de Economía y Medicina como ejemplo, sin embargo, los datos están disponibles para los tres programas académicos restantes, así como para todas las cohortes del periodo analizado

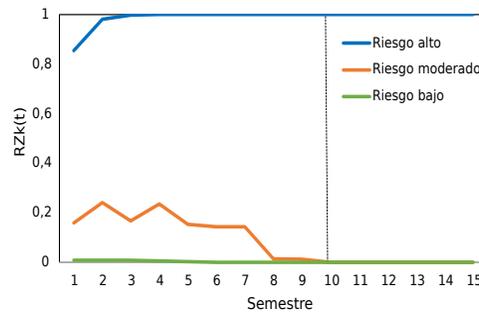


Figura 3

Ejemplo de tipos de curvas de riesgo de rezago acumulado. La línea vertical indica el semestre teórico de finalización en el programa Economía.

Las curvas típicas y representativas del riesgo de rezago acumulado de los cinco programas se presentan en la Figura 4. En todos los programas se observan diferencias en las curvas representativas. No obstante, tienen en común la tendencia creciente de la función de riesgo de rezago acumulado que aumenta a medida que los estudiantes avanzan en su ciclo académico. Medicina, seguido por Derecho, tiene probabilidades de riesgo de rezago bajas (aproximadamente 32 % y 43 %, respectivamente). Una característica que resalta es que Medicina a pesar de tener 13 semestres y mayor cantidad de créditos académicos es una carrera con menor riesgo de rezago, contrario a los otros programas, excepto Derecho, con diez semestres (Ingeniería eléctrica, Biología y Economía), quienes muestran una probabilidad de riesgo superior a 60 % con aumento a 100 % a partir del tercer semestre.

Con las gráficas desagregadas por cohorte se reconocen diferencias en las funciones de riesgo de rezago acumulado, donde la evolución semestral se aleja del patrón uniforme estructural del programa, lo que las convierte en curvas atípicas que, de hacerles seguimiento, alertan sobre factores que aumentan o disminuyen, según sea el caso, el rezago dentro de los programas. Ejemplos de esto se identifican en Biología con la cohorte 2015:1 que, pese a iniciar con una probabilidad de riesgo inferior a 70 %, salta abruptamente en el segundo semestre a 100 %, y en las cohortes 2013:2, 2014:1, 2009:1 y 2007:1 de Medicina que durante la trayectoria observada nunca alcanzaron el 100 % de probabilidad de riesgo de rezago acumulado.

Finalmente, el ejercicio de predicción se llevó a cabo para la TCRZ, útil por ejemplo, para predecir trayectorias de cohortes académicas en las que se tienen pocas mediciones (i.e., cohortes nuevas). La Figura 5 muestra tres curvas de rezago para Economía (los tres primeros gráficos) y Medicina, como ejemplo, en el conjunto de prueba, una cohorte con la tasa de rezago completa para identificar la capacidad de predicción y dos con las trayectorias previstas de los semestres futuros.

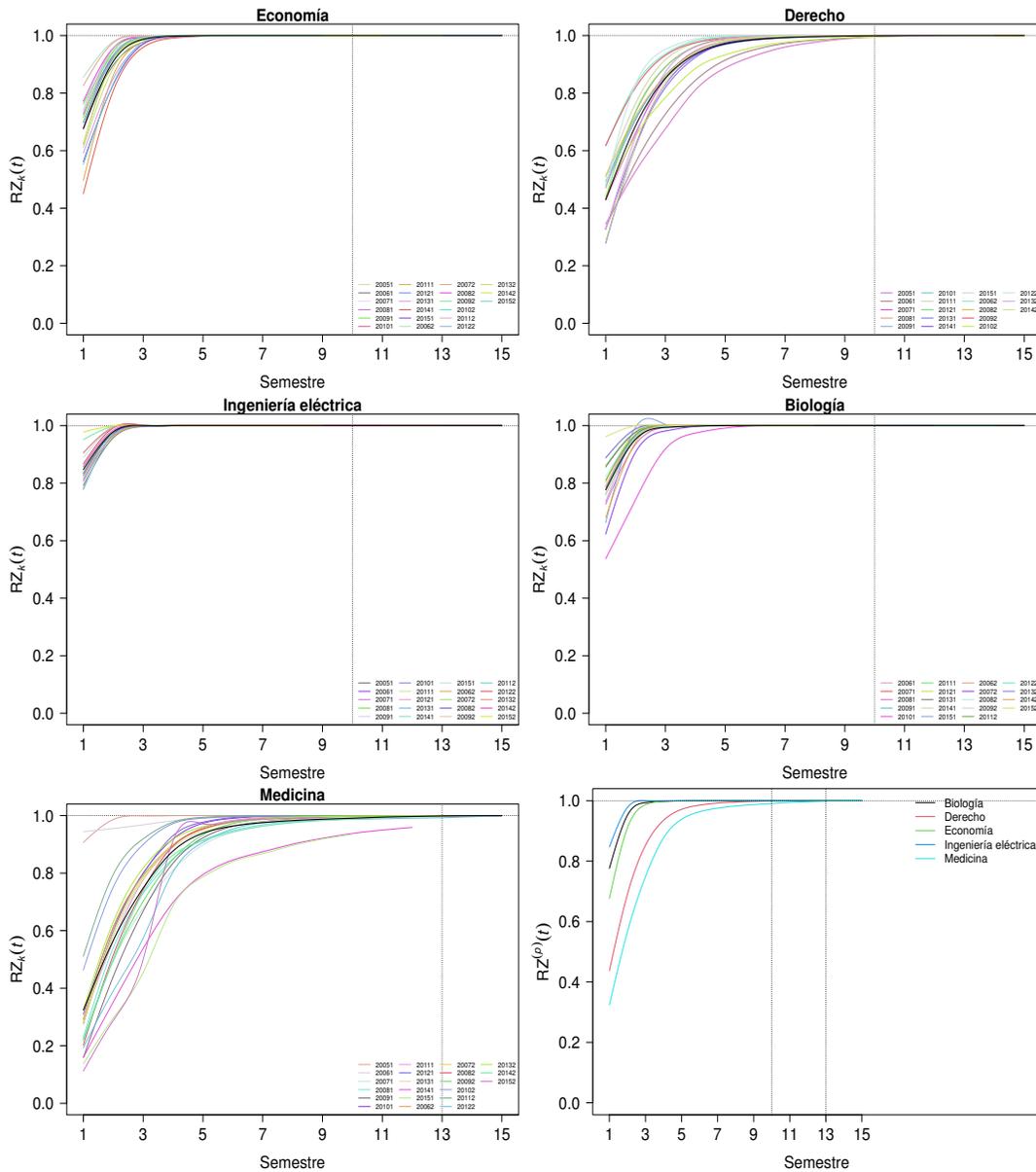


Figura 4

Curvas representativas (línea negra) de la función de riesgo de rezago de las cohortes 2005:1–2020:1.

La predicción para una cohorte de los otros programas se presenta en la figura 6. Las trayectorias de las cohortes siguen marcando las mismas diferencias entre programas. Medicina, seguido por Derecho, tiene tasas de rezago inferiores a 10 %, excepto en la cohorte 2016:2 de Derecho con un aumento inferior a 20 % entre el segundo y quinto semestre. Ingeniería eléctrica tiene mayores tasas de rezago acumuladas con la persistencia de una tasa positiva

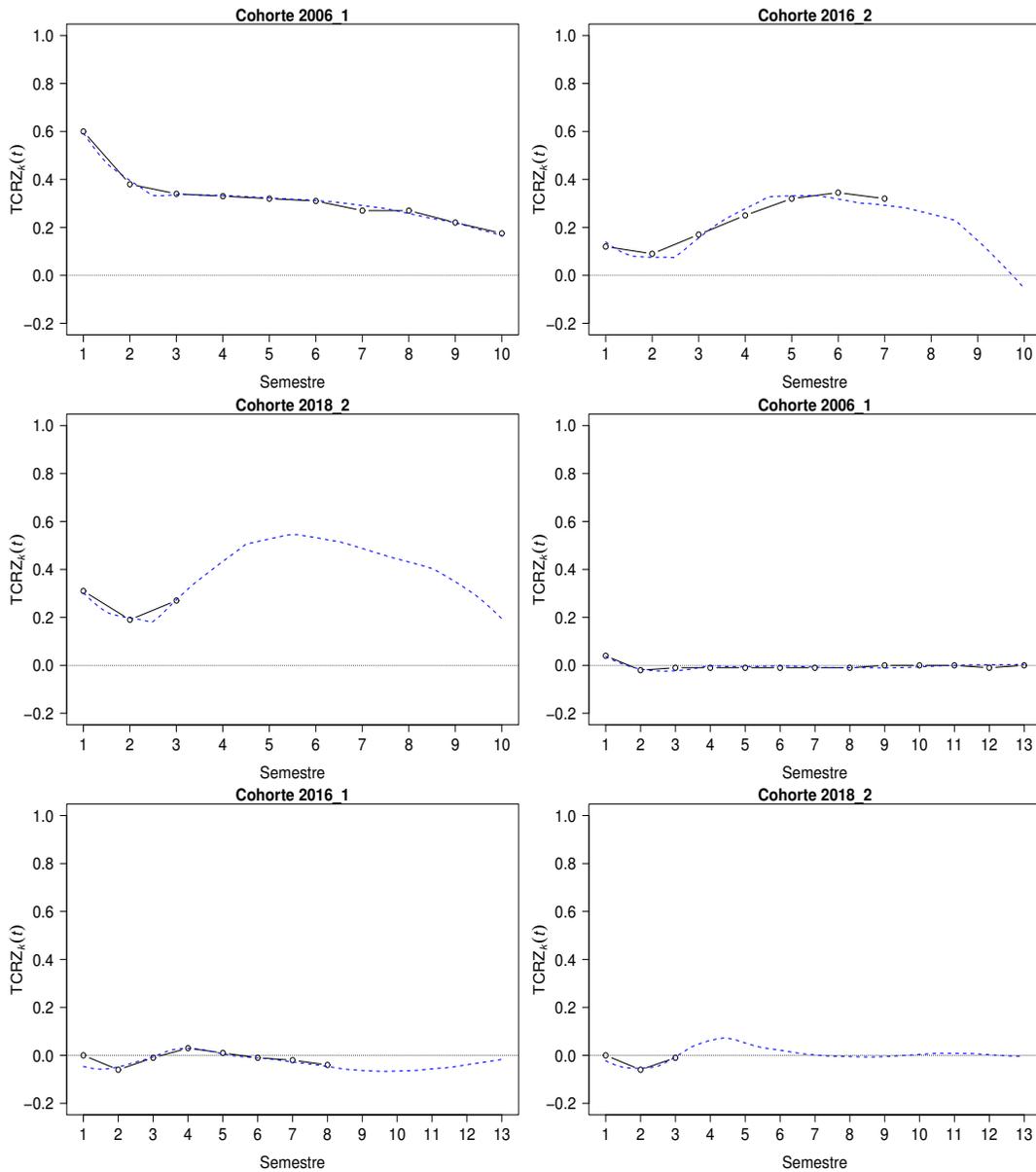


Figura 5

Curvas de rezago predichas para tres cohortes de Economía y Medicina. La línea continua son los semestres observados en cada cohorte, y la línea punteada corresponde a la predicción.

superior a 30% en el décimo semestre, seguido por Biología y Economía. En las trayectorias se nota que el rezago aumenta durante los primeros cinco semestres con tendencia decreciente a partir de este, que puede significar semestres cuellos de botella en los que los estudiantes tienden a encontrar dificultad para avanzar con el plan de estudios o la decisión de aumentar

su esfuerzo para ponerse al día con los créditos acumulados teóricos del programa.

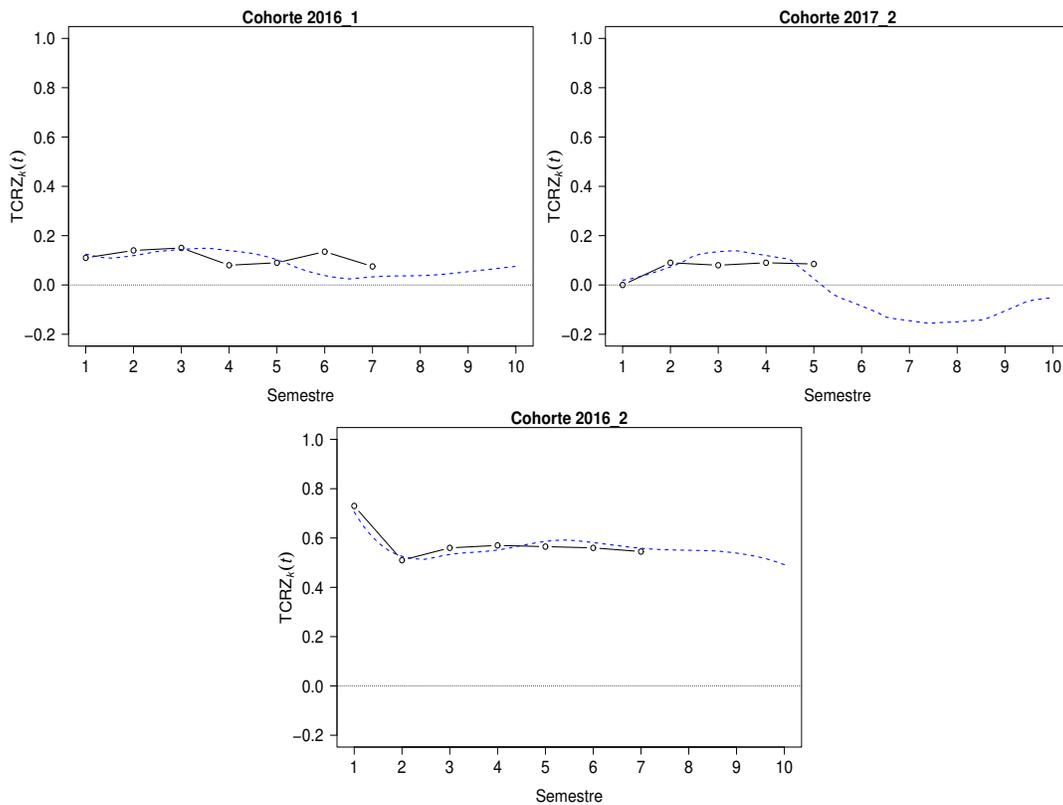


Figura 6

Curvas de rezago predichas para cohortes de Biología, Derecho e Ingeniería eléctrica. La línea continua son los semestres observados en cada cohorte, y la línea punteada corresponde a la predicción.

4 Discusión y conclusiones

Con el enfoque de análisis de componentes principales funcionales para estimar las funciones de riesgo de rezago acumulado como datos funcionales se encontró una metodología útil para analizar las tendencias de rezago acumulado al centrarse en la evolución del riesgo para calcular la curva representativa que refleja el comportamiento subyacente de cada programa académico. Además, es aplicable al tipo de datos de la información académica de las cohortes, que es una base longitudinal dispersa e irregular porque la observación de un estudiante es independiente de sus compañeros de cohorte, lo que genera diferencias en la longitud de las trayectorias. Estos resultados son un insumo para las Unidades Académicas en el diseño de políticas de graduación oportuna para crear medidas correctivas y preventivas del rezago académico.

Los hallazgos sugieren que el comportamiento estructural del riesgo de rezago acumulado es independiente del programa, es decir, a medida que los estudiantes aumentan su permanencia en el programa incrementa el riesgo de rezago. Sin embargo, al analizar tanto la curva representativa, que captura la evolución del riesgo de rezago por semestre, como las curvas de las cohortes se comprueba que el patrón subyacente difiere entre programas. Medicina y Derecho son los pregrados con mejor comportamiento frente al rezago académico, mientras que Ingeniería eléctrica, seguido por Biología y Economía, tiene peor comportamiento. Este resultado es compatible con la conclusión en la literatura sobre la asociación positiva entre los programas con mayores tasas de retorno y la tasa de graduación a tiempo (DesJardins et al., 2002; Adelman, 2006). En Colombia, Medicina es de las carreras con mayor demanda laboral y mejor remuneración, sus egresados ganan en promedio \$3.455.866 (Observatorio laboral para la Educación, 2020) y, de acuerdo con el resultado en esta investigación, las cohortes académicas del programa tienen mayor probabilidad de graduarse como mínimo en los 13 semestres reglamentarios.

Por otro lado, con la evidencia del comportamiento diferenciado entre programas surge la hipótesis que entre mayor sea el riesgo de rezago acumulado de un programa menor es la presencia de cohortes atípicas, por lo tanto, aunque el programa académico realice, por ejemplo, cambios de versión del plan de estudios no habrá incidencia en la disminución del rezago académico, dando lugar a la persistencia de la graduación tardía independiente de la cohorte y el año. También es perceptible que las principales cohortes que difieren de la curva representativa del programa, en el sentido de una evolución menor del riesgo de rezago son las del primer semestre del año, no obstante, para validar la observación se plantean

dos investigaciones futuras. La primera es analizar los determinantes del rezago académico incluyendo estrategias de identificación diferenciadoras para clasificar factores según tipos de estudiantes y especificar las razones que conducen al patrón de rezago intracohorte. La segunda es el análisis de clustering para datos funcionales dispersos, con la cual se establecería si existen cambios intergeneracionales y efectos en el riesgo de rezago acumulado según el calendario de la cohorte académica (primer o segundo semestre).

En esta investigación no fue posible incorporar en las funciones de riesgo la información de los estudiantes que reingresan al programa académico luego de cuatro semestres de ausencia por la inestabilidad que generan en el indicador. Para seguir adelante en la implementación de la metodología se excluyeron del estudio bajo la definición de deserción de Castaño et al. (2006), en la que un estudiante es desertor siempre y cuando no matricule créditos académicos por tres semestres consecutivos.

Aunque, se alerta sobre la necesidad de construir un indicador robusto en futuras investigaciones que capture la variabilidad de los reingresos sin sobreestimar el patrón de rezago académico. La literatura relaciona que esta flexibilidad del reglamento estudiantil como medida para evitar la deserción genera mayor probabilidad de rezago en los estudiantes (DesJardins et al., 2002; Stinebrickner and Stinebrickner, 2006), puesto que existe correspondencia con el incremento del tiempo para graduarse, asimismo, de haber reformas académicas deben ingresar bajo los nuevos parámetros establecidos en el programa y validar los créditos académicos aprobados en la anterior versión del pensum, lo que usualmente les implica cursar más materias de las que les faltaban antes de abandonar temporalmente el programa.

Finalmente, la información de las funciones de rezago se usó para predecir el comportamiento de las cohortes con semestres faltantes para anticipar la trayectoria de rezago. Las curvas predichas para las cohortes recientes también reflejan los semestres cuello de botella previamente identificados, los cuales se ubican entre el tercer y quinto semestre de los programas. No obstante, una posible limitación asociada a los resultados de la predicción son las variaciones en el patrón de rezago en los semestres de 2020 y 2021, dados los cambios en las condiciones académicas. Por lo tanto, a futuro también es conveniente analizar el efecto de la pandemia en el patrón de rezago de los programas.

Las conclusiones de este trabajo son relevantes para los administradores académicos de la educación terciaria y en particular, para aquellas universidades que se financian directamente con recursos del Estado o en las cuales la matrícula es subsidiada. Los estudiantes enfrentan mayores riesgos de prolongar su permanencia en el programa académico

al no asumir los costos marginales de un semestre adicional (Brunello and Winter-Ebmer, 2003), mientras que los estudiantes de universidades privadas son motivados a completar sus estudios en el tiempo definido por los costos de la matrícula, que son asumidos por ellos o sus familias (Garibaldi et al., 2012).

Referencias

- Adelman, C. (2004). *Principal indicators of student academic histories in postsecondary education*. Washington, DC: US Department of Education.
- Adelman, C. (2006). *The Toolbox Revisited: Paths to Degree Completion from High School through College*. Washington, DC: US Department of Education.
- Aina, C., E. Baici, G. Casalone, and F. Pastore (2018). The economics of university dropouts and delayed graduation: a survey. Technical Report 189.
- Attewell, P., S. Heil, and L. Reisel (2012). *Educational Evaluation and Policy Analysis*. 34(1), 27–44.
- Attewell, P. and D. Monaghan (2016). *Research in Higher Education*. 57(6), 682–713.
- Belfield, C., D. Jenkins, and H. Lahr (2016). Momentum: The Academic and Economic Value of a 15-Credit First-Semester Course Load for College Students in Tennessee.
- Bettinger, E. and B. T. Long (2009). *Journal of Human Resources*. 44(3), 736–771.
- Birdsall, C. (2018). *Public Performance & Management Review*. 41(4), 669–695.
- Booth, A. L. and S. E. Satchell (1995, sep). *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. 158(2), 297–318.
- Bound, J., M. F. Lovenheim, and S. Turner (2010). *American Economic Journal: Applied Economics*. 2(3), 129–157.
- Brandsma-Dieters, A. (2013). Factors influencing student delay in higher education: A case study at Stenden University of Applied Sciences. Master's thesis, University of Twente.
- Brunello, G. and R. Winter-Ebmer (2003). *Economics Letters*. 80(2), 247–253.
- Castaño, E., S. Gallón, K. Gómez, and J. Vásquez (2006). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. *Lecturas de Economía* 65, 9–36.
- Chan, R. Y. (2022). *Do Credit Momentum Policies Through the 15 to Finish Improve Academic Progression and Completion of Low-Income, First-Generation Students? Evidence from a College Promise Program*. Research in Higher Education.

-
- Chiou, J.-M. and H.-G. Müller (2009). *Journal of the American Statistical Association*. 14(486), 572–585.
- Confavreux, C. and S. Vukusic (2014). *Chapter 15 - The clinical course of multiple sclerosis*. In D. S. Goodin (Ed.), *Multiple Sclerosis and Related Disorders, Volumen 122 of Handbook of Clinical Neurology*, (pp. 343–369). Elsevier.
- Csalodi, R. and J. Abonyi (2021). *Mathematics*. 9(5).
- DesJardins, S., D. Ahlburg, and B. McCall (2002). *A temporal investigation of factors related to timely degree completion*. *The Journal of Higher Education* 73, 555–581.
- Dolton, P., R. Audas, and E. Berde (2005). *Youth unemployment and labour market transitions in Hungary*. *Education Economics* 13, 1–25.
- Erbas, B., R. J. Hyndman, and D. M. Gertig (2007). *Forecasting age-specific breast cancer mortality using functional data models*. *Statistics in medicine* 26, 458–470.
- Ferraty, F. and Y. Romain (2010). *The Oxford Handbook of Functional Data Analysis*. Oxford: Oxford University Press.
- Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis*. In *Springer Series in Statistics*. New York: Springer.
- Gallón, S., L. E. Restrepo, and L. Sánchez (2022). *Medición y análisis del rezago académico universitario*. *Universidad de Antioquia*.
- García-Cabo, J. and R. Madera (2019). *The Self-Employment Option in Rigid Labor Markets: An Empirical Investigation*.
- Garibaldi, P., F. Giavazzi, A. Ichino, and E. Rettore (2012). *The Review of Economics and Statistics*. 94(3), 699–711.
- Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. In *Springer Series in Statistics* (2nd. editi ed.). New York: Springer.
- Hoffman, E. P. (1991). *Eastern Economic Journal*. 17(4), 393–401.
- Huntington-Klein, N. and A. Gill (2021). *Research in Higher Education*. 62(5), 623–650.
- Hyndman, R. J. and M. Shahid Ullah (2007). *Computational Statistics & Data Analysis*. 51(10), 4942–4956.

-
- James, G., T. Hastie, and C. Sugar (2000). Principal component models for sparse functional data. *Biometrika* 87, 587–602.
- Jansen, E. (2004). Higher Education. 47(4), 411–435.
- Kaplan, E. and P. Meier (1958). Journal of the American statistical association. 53(282), 457–481.
- Kurlaender, M., J. Jackson, J. S. Howell, and E. Grodsky (2014). College course scarcity and time to degree. *Economics of Education Review* 41, 24–39.
- Letkiewicz, J., H. Lim, S. Heckman, S. Bartholomae, J. J. Fox, and C. P. Montalto (2014). The Path to Graduation: Factors Predicting On-Time Graduation Rates. *Journal of College Student Retention: Research, Theory & Practice* 16, 351–371.
- Löfgren, C. and H. Ohlsson (1999). *Economics of Education Review*. 18(1), 79–88.
- May, W. L. (2009). Kaplan–Meier Survival Analysis, (pp. 1590–1593). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ministerio de Educación Nacional (2017). *Educación superior. Technical report*.
- Observatorio laboral para la Educación (2020). *Mirada regional de programas académicos de pregrado destacados por la mayor cantidad de graduados que cotizan al sector formal de la economía y mejores salarios. Technical report, Ministerio de Educación Nacional*.
- Peng, J. and D. Paul (2009). *Journal of Computational and Graphical Statistics*. 18(4), 995–1015.
- Pike, G. R. and K. R. Robbins (2020). *Research in Higher Education*. 61(4), 485–509.
- R Core Team (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Ramsay, J. O. and C. J. Dalzell (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* 53, 539–572.
- Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (2nd edn ed.). New York: Springer.
- Robinson, T. (1990). *The Journal of Negro Education*. 59(2), 207–218.
- Scott-Clayton, J. (2011). *Journal of Human Resources*. 46(3), 614–646.

-
- Shang, H. L. (2018). Visualizing rate of change: application to age-specific fertility. *Journal of the Royal Statistical Society Series A (Statistics in Society)* 182.
- Sistema Nacional de Información de la Educación Superior (2021). Estadísticas desagregadas de educación superior. Technical report.*
- Stage, F. (1988). *Student Typologies and the Study of College Outcomes*. Review of Higher Education, 11, 247–257.
- Stalpers, L. J. A. and E. L. Kaplan (2018). *BSHM Bulletin: Journal of the British Society for the History of Mathematics*. 33(2), 109–135.
- Stinebrickner, R. and T. R. Stinebrickner (2006). *Journal of Public Economics*. 90(8), 1435–1454.
- Swail, W. S. (2003). Retaining minority students in higher education: A framework for success.
- Tagharrobi, Z., N. Masoudi Alavi, E. Fakharian, F. Mirhoseini, S. A. Rasoulinejad, H. Akbari, and H. Ameli (2013, dec). *Nursing and midwifery studies*. 2(4), 89–96.
- Theune, K. (2015). *Higher Education*. 70(4), 725–752.
- Thomas, L. (2002). *Journal of Education Policy*. 17(4), 423–442.
- Tinto, V. (1975). *Review of Educational Research*. 45(1), 89–125.
- Tinto, V. (1982). Defining dropout: A matter of perspective. *New Directions for Institutional Research 1982*, 3–15.
- Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago. London : University of Chicago Press.
- Ullah, S. and C. F. Finch (2013). *BMC Medical Research Methodology*. 13(43).
- Umbricht, M. R., F. Fernandez, and J. C. Ortagus (2015, dec). *Educational Policy*. 31(5), 643–673.
- Universidad de Antioquia (2015). Acuerdo 1 de 1981. Por el cual se expide el reglamento estudiantil y de normas académicas de pregrado de la Universidad de Antioquia.*
- Vallejos, C. A. and M. F. J. Steel (2017, feb). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 180(2), 613–631.

Volkwein, J. F. and W. G. Lorang (1996). *Research in Higher Education*. 37(1), 43–68.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). *Annual Review of Statistics and Its Application*. 3(1), 257–295.

Yao, F., H.-G. Müller, and J.-L. Wan (2005). *Functional Data Analysis for Sparse Longitudinal Data*. *Journal of the American Statistical Association* 100, 577–590.

Yue, H. and X. Fu (2017). *Research in Higher Education*. 58(2), 184–213.

Zee, J. and S. X. Xie (2018). *Biostatistics & epidemiology*. 2(1), 23–33.

Zhou, Y., S. Bhattacharjee, C. Carroll, Y. Chen, X. Dai, J. Fan, A. Gajardo, P. Z. Hadjipantelis, K. Han, H. Ji, C. Zhu, H.-G. Müller, and J.-L. Wang (2022). *fdapace: Functional Data Analysis and Empirical Dynamics*. R package version 0.5.9.