

Machine Learning and Economics

Alvaro J. Riascos Villegas

July 2019

Contenido

- 1 Main Themes
- 2 Causalidad y Predicción
- 3 Dependencia Parcial: Importancia y Causalidad
- 4 Prediction and Causality are usually Intertwined
- 5 Aplicación: Ajuste de Riesgo
 - Problemas
 - Resultados
- 6 References

Main Themes

- New data (administrative and corporate data): images, text, etc.
- Prediction.
- Causality.
- Demand estimation.
- Unbalanced panel data (matrix completion).
- Development.
- New markets (strategic algorithmic interaction).
- Testing theories.
- New problems: bias, discrimination, fairness, etc.

Contenido

- 1 Main Themes
- 2 Causalidad y Predicción**
- 3 Dependencia Parcial: Importancia y Causalidad
- 4 Prediction and Causality are usually Intertwined
- 5 Aplicación: Ajuste de Riesgo
 - Problemas
 - Resultados
- 6 References

Ejemplos

- Modelo de predicción de la temperatura de una ciudad dada la altura sobre el nivel del mar.
- Modelo de predicción del crimen: consumo de helados y crimen.
- Modelo de ocupación hotelera: correlación positiva entre precio y ocupación.

Ejemplos

- Modelo de predicción de la temperatura de una ciudad dada la altura sobre el nivel del mar.
- Modelo de predicción del crimen: consumo de helados y crimen.
- Modelo de ocupación hotelera: correlación positiva entre precio y ocupación.

Ejemplos

- Modelo de predicción de la temperatura de una ciudad dada la altura sobre el nivel del mar.
- Modelo de predicción del crimen: consumo de helados y crimen.
- Modelo de ocupación hotelera: correlación positiva entre precio y ocupación.

- Supongamos que estimamos una ecuación de la forma:

$$G_i = \alpha + \beta H_i + e_i$$

donde G_i es la nota final del estudiante i en una materia, H_i son las horas dedicadas a estudiar y e_i es el error.

- Intuitivamente, la causalidad va de H a G y no al contrario y la magnitud del efecto lo mide β siempre y cuando $E[e | H] = 0$ (i.e., H es exógena).
- Si estimamos la ecuación inversa es posible identificar la importancia de la variable G para predecir a H pero no es una relación de causalidad.

- Supongamos que estimamos una ecuación de la forma:

$$G_i = \alpha + \beta H_i + e_i$$

donde G_i es la nota final del estudiante i en una materia, H_i son las horas dedicadas a estudiar y e_i es el error.

- Intuitivamente, la causalidad va de H a G y no al contrario y la magnitud del efecto lo mide β siempre y cuando $E[e | H] = 0$ (i.e., H es exógena).
- Si estimamos la ecuación inversa es posible identificar la importancia de la variable G para predecir a H pero no es una relación de causalidad.

- Supongamos que estimamos una ecuación de la forma:

$$G_i = \alpha + \beta H_i + e_i$$

donde G_i es la nota final del estudiante i en una materia, H_i son las horas dedicadas a estudiar y e_i es el error.

- Intuitivamente, la causalidad va de H a G y no al contrario y la magnitud del efecto lo mide β siempre y cuando $E[e | H] = 0$ (i.e., H es exógena).
- Si estimamos la ecuación inversa es posible identificar la importancia de la variable G para predecir a H pero no es una relación de causalidad.

Especificación Lineal: Efecto tratamiento binario

- Para estimar el efecto que una intervención $W_i \in \{0, 1\}$ (programa de refuerzo en casa), tiene en una variable de resultado G_i usualmente especificamos una ecuación de la forma:

$$G_i = \alpha + \beta H_i + \gamma W_i + e_i$$

- Intuitivamente la magnitud del efecto lo mide γ siempre y cuando $E[e | W] = 0$ (i.e., W es exógena). Esto es una consecuencia de la hipótesis de independencia condicional (i.e., condicional a H , G y W son independientes).
- En aprendizaje de máquinas rara vez se especifica una función lineal.

Especificación Lineal: Efecto tratamiento binario

- Para estimar el efecto que una intervención $W_i \in \{0, 1\}$ (programa de refuerzo en casa), tiene en una variable de resultado G_i usualmente especificamos una ecuación de la forma:

$$G_i = \alpha + \beta H_i + \gamma W_i + e_i$$

- Intuitivamente la magnitud del efecto lo mide γ siempre y cuando $E[e | W] = 0$ (i.e., W es exógena). Esto es una consecuencia de la hipótesis de independencia condicional (i.e., condicional a H , G y W son independientes).
- En aprendizaje de máquinas rara vez se especifica una función lineal.

Especificación Lineal: Efecto tratamiento binario

- Para estimar el efecto que una intervención $W_i \in \{0, 1\}$ (programa de refuerzo en casa), tiene en una variable de resultado G_i usualmente especificamos una ecuación de la forma:

$$G_i = \alpha + \beta H_i + \gamma W_i + e_i$$

- Intuitivamente la magnitud del efecto lo mide γ siempre y cuando $E[e | W] = 0$ (i.e., W es exógena). Esto es una consecuencia de la hipótesis de independencia condicional (i.e., condicional a H , G y W son independientes).
- En aprendizaje de máquinas rara vez se especifica una función lineal.

- Existen fundamentalmente dos opciones:
 - 1 Hacer uso de las técnicas de aprendizaje de máquinas para mejorar las técnicas tradicionales (i.e., variables instrumentales, diferencias en diferencias con datos panel, etc.).
 - 2 Usar aproximaciones nuevas basadas principalmente en modelos no lineales (árboles causales, dependencias parciales, etc).

- Existen fundamentalmente dos opciones:
 - 1 Hacer uso de las técnicas de aprendizaje de máquinas para mejorar las técnicas tradicionales (i.e., variables instrumentales, diferencias en diferencias con datos panel, etc.).
 - 2 Usar aproximaciones nuevas basadas principalmente en modelos no lineales (árboles causales, dependencias parciales, etc).

Contenido

- 1 Main Themes
- 2 Causalidad y Predicción
- 3 Dependencia Parcial: Importancia y Causalidad**
- 4 Prediction and Causality are usually Intertwined
- 5 Aplicación: Ajuste de Riesgo
 - Problemas
 - Resultados
- 6 References

Dependencia Parcial

- Los modelos de aprendizaje de máquinas usualmente establecen relaciones altamente no lineales entre la variable dependiente y las variables independientes:

$$Y = f(X)$$

- Si escribimos $X = (X_S, X_C)$ donde X_S es un conjunto de variables que deseamos estudiar su efecto sobre Y y X_C es el complemento.
- La importancia de X_S en la predicción de Y en el punto x_S se puede estimar con la dependencia parcial, $f(x_S)$:

$$f(x_S) = E_{X_C}[f(x_S, X_C)]$$

donde la integral se hace sobre la distribución marginal de X_C

Dependencia Parcial

- Los modelos de aprendizaje de máquinas usualmente establecen relaciones altamente no lineales entre la variable dependiente y las variables independientes:

$$Y = f(X)$$

- Si escribimos $X = (X_S, X_C)$ donde X_S es un conjunto de variables que deseamos estudiar su efecto sobre Y y X_C es el complemento.
- La importancia de X_S en la predicción de Y en el punto x_S se puede estimar con la dependencia parcial, $f(x_S)$:

$$f(x_S) = E_{X_C}[f(x_S, X_C)]$$

donde la integral se hace sobre la distribución marginal de X_C

Dependencia Parcial

- Los modelos de aprendizaje de máquinas usualmente establecen relaciones altamente no lineales entre la variable dependiente y las variables independientes:

$$Y = f(X)$$

- Si escribimos $X = (X_S, X_C)$ donde X_S es un conjunto de variables que deseamos estudiar su efecto sobre Y y X_C es el complemento.
- La importancia de X_S en la predicción de Y en el punto x_S se puede estimar con la dependencia parcial, $f(x_S)$:

$$f(x_S) = E_{X_C}[f(x_S, X_C)]$$

donde la integral se hace sobre la distribución marginal de X_C

de la distribución conjunta de (Y, X)

- Esta medida de importancia de una variable en modelos no lineales (cajas negras) es estándar en la literatura (y muchas otras dependiendo de la función de aprendizaje: importancia en árboles, bosques, boosting, redes neuronales, etc.).
- Lo interesante es que esta medida bajo ciertas hipótesis (véase: Zhao, Q. y T. Hastie (2019). Causal Interpretation of Black-Box Models) puede interpretarse como el efecto causal de x_s en y .

- Esta medida de importancia de una variable en modelos no lineales (cajas negras) es estándar en la literatura (y muchas otras dependiendo de la función de aprendizaje: importancia en árboles, bosques, boosting, redes neuronales, etc.).
- Lo interesante es que esta medida bajo ciertas hipótesis (véase: Zhao, Q. y T. Hastie (2019). Causal Interpretation of Black-Box Models) puede interpretarse como el efecto causal de x_s en y .

- En la práctica estimamos usamos este estimador muestral:

$$\hat{f}(x_S) = \frac{1}{n} \sum_{i=1}^n f(x_S, x_{iC})]$$

- En una representación estructural del modelo probabilístico como grafos acíclicos dirigidos (Pearl) Las hipótesis son intuitivamente:
 - 1 Ningún nodo de X_C es un descendiente de X_S (i.e., X_S no causa X_C).
 - 2 X_C bloquea cualquier camino *back-door* entre X_S y Y .

- En la práctica estimamos usamos este estimador muestral:

$$\hat{f}(x_S) = \frac{1}{n} \sum_{i=1}^n f(x_S, x_{I_C})]$$

- En una representación estructural del modelo probabilístico como grafos acíclicos dirigidos (Pearl) Las hipótesis son intuitivamente:
 - 1 Ningún nodo de X_C es un descendiente de X_S (i.e., X_S no causa X_C).
 - 2 X_C bloquea cualquier camino *back-door* entre X_S y Y .

Contenido

- 1 Main Themes
- 2 Causalidad y Predicción
- 3 Dependencia Parcial: Importancia y Causalidad
- 4 Prediction and Causality are usually Intertwined**
- 5 Aplicación: Ajuste de Riesgo
 - Problemas
 - Resultados
- 6 References

Prediction and causality are usually intertwined

- Kleinberg, Ludwig, Mullainathan, Obermeyer. 2015. Prediction Policy Problems. AER.
- Many interesting economic problems do not require the identification of causality relations (i.e., risk adjustment of public health expenditures).
- This paper provides a conceptual framework to think about causality and prediction.

Prediction and causality are usually intertwined

- Kleinberg, Ludwig, Mullainathan, Obermeyer. 2015. Prediction Policy Problems. AER.
- Many interesting economic problems do not require the identification of causality relations (i.e., risk adjustment of public health expenditures).
- This paper provides a conceptual framework to think about causality and prediction.

Prediction and causality are usually intertwined

- Kleinberg, Ludwig, Mullainathan, Obermeyer. 2015. Prediction Policy Problems. AER.
- Many interesting economic problems do not require the identification of causality relations (i.e., risk adjustment of public health expenditures).
- This paper provides a conceptual framework to think about causality and prediction.

Introduction

- Let y be the variable of interest. It depends on x_0 , which is exogenous (e.g., a policy variable) and x . To make explicit this dependence we write $y(x_0, x)$
- We want to maximize the known objective function $\Pi(x_0, y)$.
- The decision depends on:

$$\frac{d\Pi(x_0, y(x_0, x))}{dx_0} = \frac{\partial \Pi}{\partial x_0}(x_0, y) + \frac{\partial \Pi}{\partial y}(x_0, y) \frac{\partial y}{\partial x_0}(x_0, x)$$

- 1 First term: The effect on Π of x_0 depends on y (a prediction problem: given (x_0, x) predict y).
 - 2 Second term: The second term depends on how x_0 affects y (a causality problem) and y (a prediction problem).
- Both effects depend on y .
 - Therefore a decision maker has to solve both problems.

Introduction

- Let y be the variable of interest. It depends on x_0 , which is exogenous (e.g., a policy variable) and x . To make explicit this dependence we write $y(x_0, x)$
- We want to maximize the known objective function $\Pi(x_0, y)$.
- The decision depends on:

$$\frac{d\Pi(x_0, y(x_0, x))}{dx_0} = \frac{\partial \Pi}{\partial x_0}(x_0, y) + \frac{\partial \Pi}{\partial y}(x_0, y) \frac{\partial y}{\partial x_0}(x_0, x)$$

- 1 First term: The effect on Π of x_0 depends on y (a prediction problem: given (x_0, x) predict y).
 - 2 Second term: The second term depends on how x_0 affects y (a causality problem) and y (a prediction problem).
- Both effects depend on y .
 - Therefore a decision maker has to solve both problems.

Introduction

- Let y be the variable of interest. It depends on x_0 , which is exogenous (e.g., a policy variable) and x . To make explicit this dependence we write $y(x_0, x)$
- We want to maximize the known objective function $\Pi(x_0, y)$.
- The decision depends on:

$$\frac{d\Pi(x_0, y(x_0, x))}{dx_0} = \frac{\partial \Pi}{\partial x_0}(x_0, y) + \frac{\partial \Pi}{\partial y}(x_0, y) \frac{\partial y}{\partial x_0}(x_0, x)$$

- 1 First term: The effect on Π of x_0 depends on y (a prediction problem: given (x_0, x) predict y).
 - 2 Second term: The second term depends on how x_0 affects y (a causality problem) and y (a prediction problem).
- Both effects depend on y .
 - Therefore a decision maker has to solve both problems.

Introduction

- Let y be the variable of interest. It depends on x_0 , which is exogenous (e.g., a policy variable) and x . To make explicit this dependence we write $y(x_0, x)$
- We want to maximize the known objective function $\Pi(x_0, y)$.
- The decision depends on:

$$\frac{d\Pi(x_0, y(x_0, x))}{dx_0} = \frac{\partial \Pi}{\partial x_0}(x_0, y) + \frac{\partial \Pi}{\partial y}(x_0, y) \frac{\partial y}{\partial x_0}(x_0, x)$$

- 1 First term: The effect on Π of x_0 depends on y (a prediction problem: given (x_0, x) predict y).
 - 2 Second term: The second term depends on how x_0 affects y (a causality problem) and y (a prediction problem).
- Both effects depend on y .
 - Therefore a decision maker has to solve both problems.

Introduction

- Let y be the variable of interest. It depends on x_0 , which is exogenous (e.g., a policy variable) and x . To make explicit this dependence we write $y(x_0, x)$
- We want to maximize the known objective function $\Pi(x_0, y)$.
- The decision depends on:

$$\frac{d\Pi(x_0, y(x_0, x))}{dx_0} = \frac{\partial \Pi}{\partial x_0}(x_0, y) + \frac{\partial \Pi}{\partial y}(x_0, y) \frac{\partial y}{\partial x_0}(x_0, x)$$

- 1 First term: The effect on Π of x_0 depends on y (a prediction problem: given (x_0, x) predict y).
 - 2 Second term: The second term depends on how x_0 affects y (a causality problem) and y (a prediction problem).
- Both effects depend on y .
 - Therefore a decision maker has to solve both problems.

- Suppose on a morning you face the following problem: Take an umbrella just in case it rains or, make a rain dance looking forward to avoid raining.
- The first problem is probably a pure prediction problem since:

$$\frac{\partial y}{\partial x_0} = 0$$

- The second one, if $\frac{\partial \Pi}{\partial x_0}(x_0, y) = 0$, is a causality problem and prediction problem.

- Suppose on a morning you face the following problem: Take an umbrella just in case it rains or, make a rain dance looking forward to avoid raining.
- The first problem is probably a pure prediction problem since:

$$\frac{\partial y}{\partial x_0} = 0$$

- The second one, if $\frac{\partial \Pi}{\partial x_0}(x_0, y) = 0$, is a causality problem and prediction problem.

- Suppose on a morning you face the following problem: Take an umbrella just in case it rains or, make a rain dance looking forward to avoid raining.
- The first problem is probably a pure prediction problem since:

$$\frac{\partial y}{\partial x_0} = 0$$

- The second one, if $\frac{\partial \Pi}{\partial x_0}(x_0, y) = 0$, is a causality problem and prediction problem.

Example: Hip replacement

- The utility for an individual of a hip or knee replacement depends the weather there is in fact a replacement (x_0) and how long he will live after replacement (y).
- We assume the intervention does not affect the chances of dying (a pure prediction problem). But welfare is improving if people leave long enough to get the benefits of replacement.
- 1,3 % of Medicare beneficiaries in 2010 had a replacement.
- Out of these 1,4 % die the first month. 4,2 % die between month 1 and 12.
- Is replacement futile?
- Based on these averages apparently not. But the relevant question is if among those that are predictably riskier, replacement was futile?

Example: Hip replacement

- The utility for an individual of a hip or knee replacement depends the weather there is in fact a replacement (x_0) and how long he will live after replacement (y).
- We assume the intervention does not affect the chances of dying (a pure prediction problem). But welfare is improving if people leave long enough to get the benefits of replacement.
- 1,3 % of Medicare beneficiaries in 2010 had a replacement.
- Out of these 1,4 % die the first month. 4,2 % die between month 1 and 12.
- Is replacement futile?
- Based on these averages apparently not. But the relevant question is if among those that are predictably riskier, replacement was futile?

Example: Hip replacement

- The utility for an individual of a hip or knee replacement depends the weather there is in fact a replacement (x_0) and how long he will live after replacement (y).
- We assume the intervention does not affect the chances of dying (a pure prediction problem). But welfare is improving if people leave long enough to get the benefits of replacement.
- 1,3 % of Medicare beneficiaries in 2010 had a replacement.
- Out of these 1,4 % die the first month. 4,2 % die between month 1 and 12.
- Is replacement futile?
- Based on these averages apparently not. But the relevant question is if among those that are predictably riskier, replacement was futile?

Example: Hip replacement

- The utility for an individual of a hip or knee replacement depends the weather there is in fact a replacement (x_0) and how long he will live after replacement (y).
- We assume the intervention does not affect the chances of dying (a pure prediction problem). But welfare is improving if people leave long enough to get the benefits of replacement.
- 1,3 % of Medicare beneficiaries in 2010 had a replacement.
- Out of these 1,4 % die the first month. 4,2 % die between month 1 and 12.
- Is replacement futile?
- Based on these averages apparently not. But the relevant question is if among those that are predictably riskier, replacement was futile?

Example: Hip replacement

- The utility for an individual of a hip or knee replacement depends the weather there is in fact a replacement (x_0) and how long he will live after replacement (y).
- We assume the intervention does not affect the chances of dying (a pure prediction problem). But welfare is improving if people leave long enough to get the benefits of replacement.
- 1,3 % of Medicare beneficiaries in 2010 had a replacement.
- Out of these 1,4 % die the first month. 4,2 % die between month 1 and 12.
- Is replacement futile?
- Based on these averages apparently not. But the relevant question is if among those that are predictably riskier, replacement was futile?

Example: Hip replacement

- The utility for an individual of a hip or knee replacement depends the weather there is in fact a replacement (x_0) and how long he will live after replacement (y).
- We assume the intervention does not affect the chances of dying (a pure prediction problem). But welfare is improving if people leave long enough to get the benefits of replacement.
- 1,3 % of Medicare beneficiaries in 2010 had a replacement.
- Out of these 1,4 % die the first month. 4,2 % die between month 1 and 12.
- Is replacement futile?
- Based on these averages apparently not. But the relevant question is if among those that are predictably riskier, replacement was futile?

Example: Hip replacement

TABLE 1—RISKIEST JOINT REPLACEMENTS

Predicted mortality percentile	Observed mortality rate	Futile procedures averted	Futile spending (\$ mill.)
1	0.435 (0.028)	1,984	30
2	0.422 (0.028)	3,844	58
5	0.358 (0.027)	8,061	121
10	0.242 (0.024)	10,512	158
20	0.152 (0.020)	12,317	185
30	0.136 (0.019)	16,151	242

Notes: We predict 1–12 month mortality using an L_1 regularized logistic regression trained on 65,395 Medicare beneficiaries undergoing joint replacement in 2010, using 3,305 claims-based variables and 51 state indicators. λ was tuned using ten-fold cross-validation in the training set. In columns 1 and 2 we sort a hold-out set of 32,695 by predicted risk into percentiles (column 1) and calculate actual 1–12 month mortality (column 2). Columns 3 and 4 show results of a simulation exercise: we identify a population of eligibles (using published Medicare guidelines: those who had multiple visits to physicians for osteoarthritis and multiple claims for physical therapy or therapeutic joint injections) who did not receive replacement and assign them a predicted risk. We then substitute the high risk surgeries in each row with patients from this eligible distribution for replacement, starting at *median* predicted risk. Column 3 counts the futile procedures averted (i.e., replaced with non-futile procedures) and column 4 quantifies the dollars saved in millions by this substitution.

Contenido

- 1 Main Themes
- 2 Causalidad y Predicción
- 3 Dependencia Parcial: Importancia y Causalidad
- 4 Prediction and Causality are usually Intertwined
- 5 Aplicación: Ajuste de Riesgo**
 - Problemas
 - Resultados
- 6 References

Introducción

- Risk Adjustment Revisited using Machine Learning Techniques. Riascos, Romero, Serna.
- La ley 100 de 1993 transformó el sistema colombiano de salud en un mercado de aseguramiento competitivo.
- Elementos fundamentales en la organización del mercado: POS, IPS, EPS, UPC.
- Usando un modelo de regresión lineal, el gobierno distribuye más de COP 24 billones de pesos entre las EPSs.
- Esto se hace usando como principal insumo la base de suficiencia que registra todas las atenciones, servicios, hospitalizaciones, etc. de 20 millones de Colombianos en el POS (tiene más de 450 millones de registros).

Introducción

- Risk Adjustment Revisited using Machine Learning Techniques. Riascos, Romero, Serna.
- La ley 100 de 1993 transformó el sistema colombiano de salud en un mercado de aseguramiento competitivo.
- Elementos fundamentales en la organización del mercado: POS, IPS, EPS, UPC.
- Usando un modelo de regresión lineal, el gobierno distribuye más de COP 24 billones de pesos entre las EPSs.
- Esto se hace usando como principal insumo la base de suficiencia que registra todas las atenciones, servicios, hospitalizaciones, etc. de 20 millones de Colombianos en el POS (tiene más de 450 millones de registros).

Introducción

- Risk Adjustment Revisited using Machine Learning Techniques. Riascos, Romero, Serna.
- La ley 100 de 1993 transformó el sistema colombiano de salud en un mercado de aseguramiento competitivo.
- Elementos fundamentales en la organización del mercado: POS, IPS, EPS, UPC.
- Usando un modelo de regresión lineal, el gobierno distribuye más de COP 24 billones de pesos entre las EPSs.
- Esto se hace usando como principal insumo la base de suficiencia que registra todas las atenciones, servicios, hospitalizaciones, etc. de 20 millones de Colombianos en el POS (tiene más de 450 millones de registros).

Introducción

- Risk Adjustment Revisited using Machine Learning Techniques. Riascos, Romero, Serna.
- La ley 100 de 1993 transformó el sistema colombiano de salud en un mercado de aseguramiento competitivo.
- Elementos fundamentales en la organización del mercado: POS, IPS, EPS, UPC.
- Usando un modelo de regresión lineal, el gobierno distribuye más de COP 24 billones de pesos entre las EPSs.
- Esto se hace usando como principal insumo la base de suficiencia que registra todas las atenciones, servicios, hospitalizaciones, etc. de 20 millones de Colombianos en el POS (tiene más de 450 millones de registros).

Eficiencia

- Contención del gasto.
- Solución:
 - 1 Pagar con anterioridad a la prestación de los servicios (pago ex-ante).
 - 2 UPC debe reflejar el gasto esperado de salud de los afiliados.

Eficiencia

- Contención del gasto.
- Solución:
 - 1 Pagar con anterioridad a la prestación de los servicios (pago ex-ante).
 - 2 UPC debe reflejar el gasto esperado de salud de los afiliados.

Eficiencia

- Contención del gasto.
- Solución:
 - 1 Pagar con anterioridad a la prestación de los servicios (pago ex-ante).
 - 2 UPC debe reflejar el gasto esperado de salud de los afiliados.

Eficiencia

- Contención del gasto.
- Solución:
 - 1 Pagar con anterioridad a la prestación de los servicios (pago ex-ante).
 - 2 UPC debe reflejar el gasto esperado de salud de los afiliados.

- Descreme del mercado mediante estrategias sutiles: calidad del servicio, largas colas, tiempos prolongados para obtener citas, etc.
- Solución:
 - 1 Ajuste de riesgo ex ante a la UPC.
 - 2 El auste de riesgo debe compensar por riesgos predecibles y socialmente aceptables.
 - 3 Mejor uso de la información.

- Descreme del mercado mediante estrategias sutiles: calidad del servicio, largas colas, tiempos prolongados para obtener citas, etc.
- Solución:
 - 1 Ajuste de riesgo ex ante a la UPC.
 - 2 El auste de riesgo debe compensar por riesgos predecibles y socialmente aceptables.
 - 3 Mejor uso de la información.

- Descreme del mercado mediante estrategias sutiles: calidad del servicio, largas colas, tiempos prolongados para obtener citas, etc.
- Solución:
 - 1 Ajuste de riesgo ex ante a la UPC.
 - 2 El auste de riesgo debe compensar por riesgos predecibles y socialmente aceptables.
 - 3 Mejor uso de la información.

- Descreme del mercado mediante estrategias sutiles: calidad del servicio, largas colas, tiempos prolongados para obtener citas, etc.
- Solución:
 - 1 Ajuste de riesgo ex ante a la UPC.
 - 2 El auste de riesgo debe compensar por riesgos predecibles y socialmente aceptables.
 - 3 Mejor uso de la información.

- Descreme del mercado mediante estrategias sutiles: calidad del servicio, largas colas, tiempos prolongados para obtener citas, etc.
- Solución:
 - 1 Ajuste de riesgo ex ante a la UPC.
 - 2 El auste de riesgo debe compensar por riesgos predecibles y socialmente aceptables.
 - 3 Mejor uso de la información.

Capacidad predictiva del gasto

- El modelo de ajuste de riesgo del Ministerio tiene un poder predictivo normal (de acuerdo a los estándares internacionales).
- Predice el 33 % del gasto del quintil de mayor gasto de salud.
- La capacidad de predecir el gasto de ciertos **riesgos predecibles** es baja: Quedan muchos incentivos a la selección de riesgos.
- Por esta razón se hace un ajuste ex-post.

Capacidad predictiva del gasto

- El modelo de ajuste de riesgo del Ministerio tiene un poder predictivo normal (de acuerdo a los estándares internacionales).
- Predice el 33 % del gasto del quintil de mayor gasto de salud.
- La capacidad de predecir el gasto de ciertos **riesgos predecibles** es baja: Quedan muchos incentivos a la selección de riesgos.
- Por esta razón se hace un ajuste ex-post.

Capacidad predictiva del gasto

- El modelo de ajuste de riesgo del Ministerio tiene un poder predictivo normal (de acuerdo a los estándares internacionales).
- Predice el 33 % del gasto del quintil de mayor gasto de salud.
- La capacidad de predecir el gasto de ciertos **riesgos predecibles** es baja: Quedan muchos incentivos a la selección de riesgos.
- Por esta razón se hace un ajuste ex-post.

Capacidad predictiva del gasto

- El modelo de ajuste de riesgo del Ministerio tiene un poder predictivo normal (de acuerdo a los estándares internacionales).
- Predice el 33 % del gasto del quintil de mayor gasto de salud.
- La capacidad de predecir el gasto de ciertos **riesgos predecibles** es baja: Quedan muchos incentivos a la selección de riesgos.
- Por esta razón se hace un ajuste ex-post.

Ajuste ex-ante

Cuadro: Ajuste fuera de muestra distribución completa

Modelo		RMSE	MAE	PR anual	PR - no anual	R^2
1.	WLS UPC	3,506,658	720,587	0.896	0.999	1.57
2.	WLS UPC + Dx	3,440,928	694,404	0.892	0.999	5.23
3.	ANN FS	3,455,366	774,190	1.064	1.179	
4.	RF FS	3,465,301	712,820	0.975	1.087	
5.	GBM FS	3,431,044	721,168	1.002	1.115	

Ajuste ex-ante

Cuadro: Ajuste fuera de muestra en el quintil superior

Modelo		RMSE	MAE	PR anual	PR no anual
1.	WLS UPC	7,749,235	1,920,486	0.291	0.335
2.	WLS UPC + Dx	7,580,659	1,983,269	0.367	0.426
3.	ANN FS	7,582,293	1,962,318	0.412	0.474
4.	RF FS	7,580,672	1,988,824	0.424	0.490
5.	GBM FS	7,517,520	1,961,026	0.430	0.500

Contenido

- 1 Main Themes
- 2 Causalidad y Predicción
- 3 Dependencia Parcial: Importancia y Causalidad
- 4 Prediction and Causality are usually Intertwined
- 5 Aplicación: Ajuste de Riesgo
 - Problemas
 - Resultados
- 6 References

References

- Machine Learning: An Applied Econometric Approach. Mullainathan, S. and J. Spiess. 2017. Journal of Economic Perspectives.
- Athey, S., and G. Imbens. 2019. Machine Learning Methods Economists Should Know About.
- Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. Prediction Policy Problems.
- Risk Adjustment Revisited using Machine Learning Techniques. Riascos, Romero, Serna. Proceeding Series of the Brazilian Society of Computational and Applied Mathematics