

SOBRE LA INCLUSIÓN DE VARIABLES LATENTES PERCEPTUALES EN MODELOS DE ELECCIÓN DISCRETA: UNA APLICACIÓN A INFRAESTRUCTURA CICLISTA

Tomás Rossetti ^{a*}, C. Angelo Guevara ^b, Patricia Galilea ^c
y Ricardo Hurtubia ^{cd}

^a Centro de Desarrollo Urbano Sustentable e Instituto de Sistemas Complejos de Ingeniería,
terosset@uc.cl

^b Departamento de Ingeniería Civil, Universidad de Chile

^c Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile
^d Escuela de Arquitectura, Pontificia Universidad Católica de Chile

Resumen

La mayor parte del trabajo realizado en la modelación de preferencias con variables latentes se ha concentrado en las actitudinales, dejando de lado las perceptuales, presumiblemente por sus altos costos de estimación y recolección de datos. Proponemos una modificación a la especificación clásica de estos modelos que reduce los costos de recolección de datos y de estimación. Sumado a lo anterior, proveemos nuevos insumos a recientes controversias en torno a las condiciones bajo las cuales el uso de variables latentes puede entregar una mejora en la bondad de ajuste. Mostramos que se puede lograr una mejora si las variables latentes ingresan a la utilidad de maneras no lineales. Para ilustrar el enfoque propuesto, reportamos la estimación de un modelo utilizando datos de encuesta de preferencias declaradas por diseño de infraestructura ciclista.

Palabras clave: modelos integrados de elección y variables latentes; variables latentes perceptuales; ciclismo urbano

Keywords: integrated choice and latent variable models; perceptual latent variables; urban cycling

1. Introducción

Los modelos de elección discreta usualmente consideran que las características y atributos observables de los usuarios y alternativas respectivamente son las principales fuentes que explican el comportamiento de elección. Sin embargo, existen factores más complejos y difíciles de observar que pueden tener un efecto relevante sobre la forma en que se toman decisiones. Una aproximación posible para considerar estos es a través del enfoque de variables latentes, que explícitamente modela características o interpretaciones inobservables de los usuarios, tales como actitudes y percepciones (McFadden y Train, 2000; Ben-Akiva et al., 2002).

Las variables latentes pueden ser clasificadas en dos grupos: actitudinales y perceptuales. Mientras que las primeras describen las características inobservables de los tomadores de decisiones, y están ligadas a las actitudes y personalidad de los usuarios, las segundas describen cómo los usuarios perciben cierto atributo cualitativo de una alternativa, tal como calidad o belleza (Bahamonde-Birke et al., 2015). Mientras que existe bastante trabajo en torno a las variables actitudinales, muy pocos estudios han identificado variables latentes perceptuales; solo dos casos existen según nuestro mejor conocimiento (Bahamonde-Birke y Ortúzar, 2015; Palma et al., 2016). Creemos que una razón por la que esto podría pasar es que la identificación de variables latentes perceptuales es muy costosa en términos de estimación y recolección de información.

Una controversia presente en el área de los modelos con variable latente se relaciona con su contribución a un mejor ajuste a las elecciones observadas. A pesar de que varios estudios han concluido que el uso de variables latentes en el contexto de modelos de elección mejoran la bondad de ajuste, Vij y Walker (2016) recientemente demostraron que, bajo ciertos supuestos, esto es imposible.

Este trabajo realiza dos contribuciones principales. Primero, propone una modificación a la especificación clásica para estimar modelos híbridos de elección que reduce de manera importante los costos de recolección de datos y estimación. Por otra parte, y en contraste con las conclusiones de Vij y Walker (2016), este estudio presenta una aplicación con datos reales que prueban que hay casos relevantes en los que un aumento en la bondad de ajuste es posible con el uso de variables latentes.

2. Variables latentes en modelos de elección discreta

Siguiendo los principios básicos de los modelos de utilidad aleatoria, se puede asumir que un individuo n tiene cierta probabilidad de elegir la alternativa i dado un vector de atributos y características observables, X_{in} , un vector de variables latentes, X_{in}^* , un vector de parámetros β , y un vector θ_ε relacionado con el término de error de la función de utilidad, como se ve en la Ecuación (1).

$$P(y_{in}|X_{in}, X_{in}^*; \beta, \theta_\varepsilon) \quad (1)$$

Para obtener una probabilidad que no dependa de variables inobservables, es necesario calcular el valor esperado de (1). Para esto, las distribuciones de las variables latentes deben ser obtenidas. Estas pueden ser derivadas de la ecuación estructural, que se muestra en (2). Aquí, el vector γ

contiene parámetros a ser estimados que relacionan las variables latentes X_{in}^* con las variables observables X_{in} , y ω_{in} es un término de error aleatorio.

$$X_{in}^* = X^*(X_{in}; \gamma) + \omega_{in} \quad (2)$$

Dado un supuesto para la distribución del término de error, una función de densidad f puede ser obtenida para la variable latente, y la probabilidad incondicional de (1) puede ser calculada como se muestra en (3).

$$P(y_{in}|X_{in}; \beta, \gamma, \theta_\varepsilon, \theta_\omega) = \int P(y_{in}|X_{in}, X_{in}^*; \beta, \theta_\varepsilon) f(X^*|X_{in}; \gamma, \theta_\omega) dX^* \quad (3)$$

De esta forma, un modelo de elección discreto con variables latentes puede ser obtenido solo observando el vector X_{in} y las elecciones realizadas por los individuos, que son indicadores de las preferencias de los usuarios. Sin embargo, este indicador por sí solo probablemente no será suficiente para identificar empíricamente los efectos de las variables latentes identificadas. Por lo tanto, es útil incluir indicadores psicométricos que también informarán sobre las variables latentes de interés. Estos indicadores se incluyen a través de una ecuación de medición, como se muestra en (4), donde I_{in} es la respuesta del indicador dado por el individuo n con respecto a la alternativa i , e I es una función que relaciona la variable latente X_{in}^* con el indicador escogido a través del vector de parámetros α . Un término de error aleatorio, v_n , se incluye.

$$I_{in} = I(X_{in}^*; \alpha, v_n) \quad (4)$$

Con esto, una función de densidad $g(I_{in}|X_{in}^*; \alpha, \theta_v)$ puede ser construida, que muestra con qué probabilidad I_{in} tomará cierto valor dado X_{in}^* , α y θ_v . Con esto, una probabilidad conjunta puede ser construida, como se ve en (5). Los parámetros β , α y γ pueden ser determinados maximizando la verosimilitud para estas probabilidades.

$$P(y_{in}, I_{in}|X_{in}; \beta, \alpha, \gamma, \theta_\varepsilon, \theta_v, \theta_\omega) = \int P(y_{in}|X_{in}, X_{in}^*; \beta, \theta_\varepsilon) g(I_{in}|X_{in}^*; \alpha, \theta_v) f(X_{in}^*|X_{in}; \gamma, \theta_\omega) dX_{in}^* \quad (5)$$

Bahamonde-Birke et al. (2015) presentan una categorización de variables latentes que es útil para entender su uso y cómo deben ser modeladas. Los autores clasifican estas en dos grupos: actitudinales y perceptuales.

Variables latentes actitudinales Son específicas a cada individuo y describen la tendencia a actuar de alguna forma en particular basado en su experiencia y temperamento (Allport, 1935; Pickens, 2005). Estas variables latentes buscan modelar explícitamente actitudes que no son directamente medibles, y describen a estas en función de características observables de los usuarios.

Este es el tipo de variable latente que más frecuentemente se encuentra en la literatura. Algunos ejemplos incluyen preocupación por el estado del medio ambiente (Hess et al., 2013; Hoyos et al., 2015), actitudes hacia ciertos tipos de modos de transporte (Hurtubia et al., 2010; Yáñez et al., 2010) o la aversión al riesgo (Bartczak et al., 2016).

Variables latentes perceptuales Estas describen la forma en que los individuos procesan información relacionada con alguna alternativa en específico. Estas variables no solo dependen de las características de los usuarios, sino que también de los atributos de las alternativas. Esto significa que un cambio en los atributos de cierta alternativa cambiará el valor de la variable latente.

La exploración realizada en este tipo de variable latente ha sido bastante escasa en la literatura. Tan solo dos estudios han tratado explícitamente este tema de acuerdo a nuestro mejor conocimiento. Uno fue conducido por Bahamonde-Birke et al. (2015), que exemplificó este tipo de variable latente en un contexto de viajes interurbanos. El otro estudio es el conducido por Palma et al. (2016), que modeló preferencias por vinos. Este modelo incluye una variable latente perceptual que describe la calidad percibida de cada vino.

3. Un método alternativo para el modelado de variables latentes perceptuales

Debido al potencial que tienen las variables latentes perceptuales para explicar de mejor forma las elecciones de las personas, parece extraño que la literatura las haya utilizado de manera tan escasa. Una razón por la que esto puede suceder es porque el método clásico utilizado para identificar estas variables demanda un indicador psicométrico por cada alternativa presentada. Esto aumenta el estrés cognitivo de las personas encuestadas y la recolección de datos se vuelve más costosa, haciendo más difícil el uso de esta técnica. Además, los costos computacionales son más altos para el caso de las variables latentes perceptuales debido a la integral múltiple de la Ecuación (5).

Este estudio propone una nueva forma de identificación de estas variables utilizando indicadores recolectados por separado de los experimentos de elección. Esto permite reducir los costos de recolección, requiriendo tan solo un indicador por cada alternativa etiquetada, y en algunos casos reduciendo los costos de estimación, especialmente cuando existen sets de elección grandes. El marco de referencia para esta propuesta, utilizando la notación visual de Walker y Ben-Akiva (2002), entre otros, se ve en la Figura 1. Como se ve, esta propuesta exige la recolección de un indicador psicométrico con respecto a alguna alternativa j , con atributos X_j , que no necesariamente son iguales que los de la alternativa i presentada en el contexto de elección. Estos dos modelos se entrelazan a través de los parámetros γ del modelo de variable latente.

Con esto, la Ecuación (5) cambia. Específicamente, la probabilidad conjunta ahora no puede expresarse como una única integral, sino que como la multiplicación de dos integrales, como se ve en (6). Se logra una simplificación de la verosimilitud, que puede ser reducida a dos núcleos logit separados. Además, esta probabilidad estima simultáneamente los parámetros γ para los modelos de elección y de variable latente. Esto diferencia a este marco de referencia del enfoque secuencial, asegurando estimadores consistentes (véase Ben-Akiva et al., 2002).

$$P(y_{in}, I_{jn}|X_{in}, X_{jn}; \beta, \alpha, \gamma, \theta_\varepsilon, \theta_v, \theta_\omega) = \left(\int P(y_{in}|X_{in}, X_{in}^*; \beta, \theta_\varepsilon) f(X_{in}^*|X_{in}; \gamma, \theta_\omega) dX_{in}^* \right) \cdot \left(\int g(I_{jn}|X_{jn}^*; \alpha, \theta_v) f(X_{jn}^*|X_{jn}; \gamma, \theta_\omega) dX_{jn}^* \right) \quad (6)$$

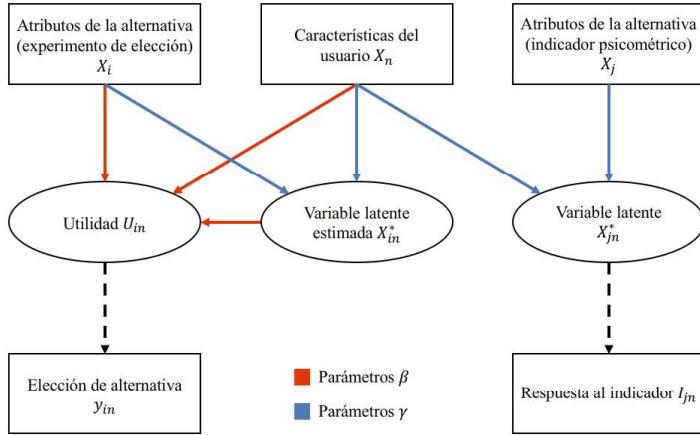


Figura 1: Marco de referencia para modelos de elección discreta con variables latentes utilizado en este estudio.

4. Sobre la contribución de variables latentes a la bondad de ajuste

En un estudio reciente, Vij y Walker (2016) prueban que modelos de elección discreta con variables latentes por lo general no presentan una mejora en el ajuste a las elecciones observadas. Esta conclusión contradice a varios estudios que reportan una mejora en el ajuste debido a variables latentes (ver, por ejemplo, Kitrinou et al., 2010; Yáñez et al., 2010; Daly et al., 2012). La demostración se realiza para un caso lineal, en el que se ve claramente que un modelo con variables latentes puede ser colapsado en uno modelo logit mixto que solo es función de variables observables.

Nosotros postulamos que, a pesar de que esto es cierto para variables latentes incluidas linealmente en la ecuación estructural del modelo de elección (es decir, la función de utilidad), esta demostración no se sostiene para casos en los que las variables latentes se incluyen de maneras no lineales. Vij y Walker afirman que, para estos casos, el mismo ajuste se obtendría simplemente incluyendo las variables observables con las mismas transformaciones. Sin embargo, supuestos de linealidad no se mantendrían, y por lo tanto existirían problemas de identificabilidad.

En los casos descritos anteriormente, el efecto neto de cada variable observable sobre la elección total no sería una constante a ser estimada sino que una función de estas variables. En términos generales, una función de utilidad lineal puede incluir una variable latente simplemente como un sumando con cualquier transformación posible, representada a través de la función h en la Ecuación (7), que a su vez puede ser expresada como una función de variables observables y del vector de parámetros γ , como se ve en la Ecuación (8).

$$U_{in}(X_{in}, X_{in}^*, \beta) = \beta \cdot X_{in} + \beta^* h(X_{in}^*) \quad (7)$$

$$U_{in}(X_{in}, \beta, \gamma) = \beta \cdot X_{in} + \beta^* h(X_{in}, \gamma) \quad (8)$$

Para calcular el efecto neto que cierta variable observable, X_{ink} , tiene sobre la elección, simplemente se debe calcular la derivada parcial de U_{in} con respecto a X_{ink} , como se muestra en la Ecuación (9).

$$\frac{\partial U_{in}}{\partial X_{ink}} = \beta_{k'} + \beta^* \frac{\partial h(X_{in}, \gamma)}{\partial X_{ink}} \quad (9)$$

El único caso en el que este efecto neto se reduce a un único parámetro es cuando h aplica una transformación lineal a X_{in}^* y la ecuación estructural de la variable latente es lineal. Para todos los otros casos, esta derivada es una función de X_{in} . Una aplicación concreta y demostración de este razonamiento se muestra en la Sección 5.3.

Considerando lo anterior, concluimos que el uso de indicadores psicométricos puede ser esencial para aumentar el ajuste a los datos de elección cuando efectos no lineales están presentes. El siguiente caso de estudio ilustra esta aseveración.

5. Caso de estudio

Esta sección presenta una aplicación de la especificación propuesta para la identificación de variables latentes perceptuales. Esta utiliza datos recolectados a través de una encuesta de preferencias declaradas que midió las preferencias por diseño de infraestructura ciclista en Santiago de Chile. La Sección 5.1 entrega detalles de esta encuesta, mientras que la Sección 5.2 presenta resultados de la estimación de distintos modelos utilizando esta especificación. Finalmente, en la Sección 5.3 se presenta una breve discusión de estos resultados.

5.1. Datos utilizados

Durante los últimos años, la ciudad de Santiago ha visto un aumento importante en el uso de bicicletas. Mientras que en el año 2001 solo 1,9 % de los viajes realizados durante un día laboral se hacían en este tipo de vehículo, en 2012 esta participación modal llegó a 4,0 % según la última encuesta de movilidad disponible (SECTRA, 2015). Este aumento puede ser explicado en parte por la nueva infraestructura ciclista que ha sido construida en varias calles de la ciudad durante la última década. Sin embargo, en varios casos esta infraestructura tiene un diseño deficiente, aumentando los tiempos de viaje y reduciendo la comodidad de sus usuarios (Vega et al., 2015). Además de esto, muchos ciclistas han decidido usar las veredas, infringiendo la ley y reduciendo la comodidad de los peatones. Estos usuarios usualmente argumentan que no se sienten seguros andando en la calle.

Entender cómo distintos tipos de diseño de ciclofacilidades puede inducir distintos tipos de comportamiento es necesario para generar mejor infraestructura. Más aún, es relevante entender qué elementos específicos de diseño de ciclovías al nivel de la calle pueden equiparar la percepción de seguridad que se da al nivel de la vereda.

Este estudio utiliza datos recolectados por una encuesta de preferencias declaradas, aplicada durante marzo de 2016, que busca entender qué elementos son relevantes en la elección de calles

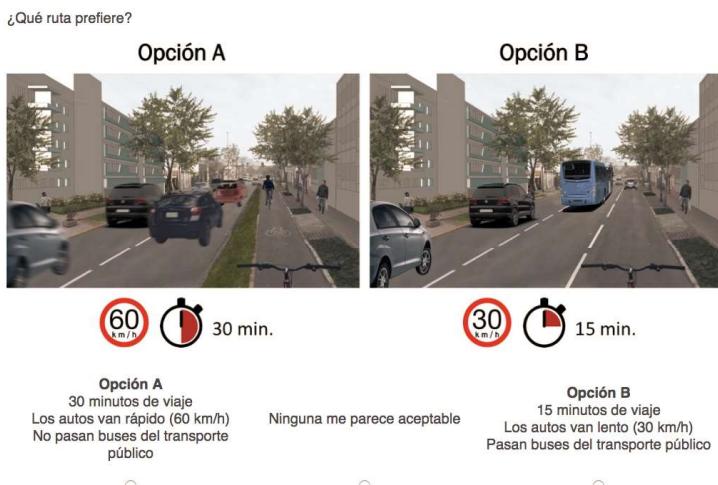


Figura 2: Ejemplo de experimento de elección presentado.

para andar en bicicleta. La encuesta se aplicó en su mayoría de manera presencial en calles con ciclofacilidades y en línea mediante cuestionarios distribuidos a través de redes sociales. Un total de 1.966 encuestas fueron respondidas, completando un total de 10.223 experimentos de elección. Detalles de la aplicación de esta encuesta se encuentran en Rossetti (2017).

Cinco variables fueron incluidas en los experimentos, con distintos niveles cada una:

1. Tipo de infraestructura ciclista presente: Esta variable buscó evaluar el peso que tiene el diseño general de la infraestructura, y se presentó en cinco niveles.
 - a) No existe infraestructura ciclista.
 - b) Infraestructura marcada con una línea al nivel de la calzada.
 - c) Infraestructura marcada con una línea y superficie pintada.
 - d) Infraestructura marcada con una línea y separadores físicos.
 - e) Infraestructura presente al nivel de la vereda.
2. Ancho: 1,5 y 3 metros.
3. Velocidad máxima para vehículos motorizados: 30 km/h y 60 km/h.
4. Presencia de buses de transporte público.
5. Tiempo de viaje: 15, 20 y 30 minutos.

Los experimentos se presentaron con el uso de imágenes para controlar de mejor forma lo que los encuestados estaban enfrentando (ver Hurtubia et al., 2015). Estos fueron construidos usando un diseño D-eficiente (ChoiceMetrics, 2012), agrupado en tres bloques con cuatro preguntas cada uno. Un ejemplo se ve en la Figura 2.

Luego de que se presentaran los experimentos de elección, una imagen al azar usada en los experimentos se mostró. Considerando esta imagen, se le pidió a los usuarios que evaluaran qué tan cómoda, segura y qué tanto les gustaba en general esta calle hipotética.

5.2. Resultados

Esta sección muestra los resultados de estimación para un modelo logit mixto. Para este, tres especificaciones se estimaron para comparar sus resultados y ajuste. Estos modelos son logit mixtos, en vez de logit multinomiales o binarios, para considerar la naturaleza de panel de estos datos. Para esto, el parámetro relacionado con la presencia de buses en la calle se mantuvo constante a lo largo de todas las respuestas entregadas por una misma persona, y cada modelo incluye la desviación estándar de este parámetro en la población encuestada. Para más información sobre modelos logit mixtos, ver McFadden y Train (2000).

Todas las relaciones estructurales del modelo de elección se modelaron linealmente, sin interacciones entre las variables. Además, todas las variables son mudas o *dummy* con la excepción del tiempo de viaje. Debido a que podría existir un sesgo a favor de las alternativas presentadas a la derecha o a la izquierda, una variable muda fue incluida para considerar este efecto. Los resultados para estos modelos se muestra en la Tabla 1. Parámetros relacionados con las ecuaciones de medición no son mostrados.

Esta tabla presenta tres modelos. El primero (“Sin VL”) es un logit mixto que no contiene variables latentes. Es el modelo más simple, y no incluye información contenida por los indicadores psicométricos en su función de verosimilitud.

Todos los parámetros obtenidos en este modelo muestran un efecto consistente con lo que ha concluido la literatura, tanto a través de datos de preferencias declaradas como preferencias reveladas. El resultado más relevante en términos de magnitud es que los usuarios mostraron preferencia por infraestructura con mayor separación del tráfico. En efecto, la presencia de cualquier tipo de infraestructura ciclista mostró un mayor impacto sobre las elecciones de los usuarios que cualquiera de las otras variables incluidas. Los resultados también muestran que los usuarios prefieren infraestructura más ancha, y calles con velocidades máximas bajas y sin la presencia de buses de transporte público.

El segundo modelo estimado incluyó una variable latente que explícitamente considera la seguridad percibida, e incluye esta variable en las funciones de utilidad linealmente. Los resultados se muestra en la Tabla 1, columna “VL lineal”.

La variable latente fue modelada con la ecuación estructural que se muestra en (10), donde X_i^* es la seguridad percibida para la alternativa i , y los parámetros γ_k relacionan la seguridad percibida con atributos observables, X_{ik} . ω_i es un término de error que, en este caso específico, se supuso con una distribución Normal. Luego, la función de utilidad para el modelo de elección se modeló linealmente, como se muestra en (11).

$$X_i^* = \sum_k \gamma_k X_{ik} + \omega_i \quad (10)$$

$$U_i = \sum_k \beta_k X_{ik} + \beta^* X_i^* \quad (11)$$

La ecuación de medición del indicador psicométrico se construyó de tal forma que la naturaleza ordinal de los indicadores se considerara, siguiendo recomendaciones encontradas en la literatura

Cuadro 1: Resultados para la estimación de tres modelos logit mixtos.

	Variable	Sin VL	VL lineal	VL log
		Valor (estadístico t)	Valor (estadístico t)	Valor (estadístico t)
Elección	Constante (derecha)	-0,105 (-4,48)	-0,105 (-4,47)	0,0198 (0,26)
	Infra.: Sin ciclofacilidad	0 (fijo)	0 (fijo)	0 (fijo)
	Infra.: Línea	0,824 (13,94)	-1,17 (-1,91)	-2,19 (-3,46)
	Infra.: Pintura	1,08 (16,94)	-1,16 (-1,76)	-2,41 (-3,22)
	Infra.: Separadores	1,19 (20,16)	-1,36 (-1,85)	-2,59 (-3,54)
	Infra.: Vereda	1,47 (27,74)	-1,72 (-1,91)	-2,92 (-3,64)
	Ancha	0,233 (6,28)	-0,692 (-2,34)	-0,679 (-3,11)
	Velocidad máx. baja	0,479 (16,97)		
	No hay buses (media)	0,662 (18,38)	0,137 (0,87)	-0,0772 (-0,42)
	No hay buses (desv. est.)	0,811 (19,82)	0,813 (18,98)	0,901 (16,93)
Variable latente	Tiempo de viaje	-0,0682 (-16,83)	-0,0683 (-16,27)	-0,0896 (-9,50)
	Seguridad percibida		1,46 (3,67)	7,80 (5,42)
	Infra.: Sin ciclofacilidad		0 (fijo)	0 (fijo)
	Infra.: Línea		1,37 (8,24)	1,22 (7,52)
	Infra.: Pintura		1,54 (9,41)	1,54 (10,09)
	Infra.: Separadores		1,75 (11,71)	1,74 (12,12)
	Infra.: Vereda		2,19 (13,90)	2,11 (13,34)
	Ancha		0,637 (7,11)	0,627 (7,01)
	Velocidad máx. baja		0,329 (3,76)	0,314 (7,02)
	No hay buses		0,361 (3,94)	0,432 (5,02)
	Error VL (izquierda)		1,07 (3,67)	0,940 (22,85)
	Error VL (derecha)		1 (fijo)	1 (fijo)
	<i>log-verosimilitud</i>	-6.138,47	-9.413,26	-9.407,72
	<i>log-verosimilitud (elección)</i>	-6.138,47	-6.138,51	-6.131,79
	<i>Número de elecciones</i>	10.223	10.223	10.223
	<i>Número de parámetros (elección)</i>	10	11	11
	<i>Rho-cuadrado ajustado (elección)</i>	0,13231	0,13217	0,13311

(Daly et al., 2012). El término de error asociado con esta se supuso que sigue una distribución Normal, que resulta en un modelo probit ordinal. Sin embargo, este se approximó mediante un modelo logit ordinal, que tiene una forma cerrada y diferencias no significativas con respecto al modelo probit (Lee, 1982; Ruud, 1983).

Los resultados del submodelo de variable latente muestran que los usuarios tienden a sentirse más seguros usando la bicicleta donde existe infraestructura para este vehículo, donde esta es ancha, donde las velocidades máximas son bajas y donde no hay buses de transporte público presentes. Estos resultados van en línea con conclusiones de la literatura .

La interpretación de estos resultados en el submodelo de elección no son tan sencillos como los del submodelo de variable latente. En primer lugar, el parámetro relacionado con la seguridad percibida tiene un signo positivo, indicando que los usuarios prefieren usar calles donde se sienten más seguros. Este descubrimiento, a pesar de ser intuitivo, no ha sido bien establecido en la literatura

anteriormente. Usualmente, los investigadores interpretan los resultados de un modelo de elección sin variables latentes y suponen que la seguridad es su principal causante (ver, por ejemplo, Sener et al., 2009).

Los otros parámetros representan efectos que no fueron capturados por la seguridad percibida, pero que tienen un efecto significativo sobre la elección de los usuarios. Por ejemplo, parámetros negativos relacionados con el tipo de infraestructura muestran que existen factores latentes que afectan a estos atributos, tales como comodidad, seguridad o agilidad. El atributo relacionado con la presencia de buses se volvió no significativo (test-t igual a 0,87), lo que significa que todo el efecto de este atributo sobre la elección puede ser explicado por la seguridad percibida.

Es importante notar que la variable latente y el modelo de elección presentaron problemas de colinealidad. Por esto, el parámetro relacionado con la velocidad se removió del modelo de elección. Este parámetro se eligió porque la evidencia indica que es más probable que la seguridad percibida explique todo el efecto de la velocidad de los autos sobre la elección.

La Tabla 1 también muestra que no hay una mejora en la bondad de ajuste de este modelo con respecto a uno sin variables latentes. Esto va en línea con lo demostrado por Vij y Walker (2016), explicado en la Sección 4, por la naturaleza lineal de la variable latente.

El tercer modelo estimado (“VL log”) transforma la variable latente (lineal) para que el efecto de la seguridad percibida sobre la elección tenga un efecto logarítmico. Esto se hizo transformando (10) en (12). Luego, el logaritmo de esta variable transformada se incluyó en las funciones de utilidad, como se muestra en (13). Debido a que \bar{X}_i^* debe ser positivo, el término de error ω_i se modeló con una distribución log-normal.

$$\bar{X}_i^* = \sum_k \lambda_k X_{ik} + 1 + \omega_i \quad (12)$$

$$U_i = \sum_k \beta_k X_{ik} + \beta^* \ln(\bar{X}_i^*) \quad (13)$$

Los resultados de este modelo muestran efectos similares a los del modelo “VL lineal”; esto es, que el parámetro relacionado con la seguridad percibida es positivo y significativo, y que los parámetros relacionados con la infraestructura en el modelo de elección son negativos.

La verosimilitud de este modelo es significativamente mayor que la del modelo sin variables latentes¹, mientras que el modelo “VL lineal” no muestra diferencias significativas². Sumado a esto, y predeciblemente debido a los tests anteriores, el ajuste del modelo “VL log” es significativamente mayor que el del modelo “VL lineal”³. Esto quiere decir que la seguridad percibida puede ser representada de mejor forma con efectos decrecientes sobre la elección. En otras palabras, los resultados muestran que un aumento en la seguridad percibida tiene un mayor efecto sobre la elección cuando la alternativa se considera insegura.

¹Test de comparación de bondad de ajuste entre dos modelos no anidados (Horowitz, 1983): $P(\bar{\rho}_2^2 - \bar{\rho}_1^2 > z) \leq 0,000222$.

²Test de comparación de bondad de ajuste entre dos modelos no anidados: $P(\bar{\rho}_2^2 - \bar{\rho}_1^2 > z) \leq 0,161$.

³Test de comparación de bondad de ajuste entre dos modelos no anidados: $P(\bar{\rho}_2^2 - \bar{\rho}_1^2 > z) \leq 0,00013$.

5.3. ¿Cuándo pueden las variables latentes entregar un mejor ajuste?

En los modelos presentados anteriormente, casi todos los atributos afectaron la elección en dos posibles formas: a través del modelo de variable latente, o a través del modelo de elección. Para analizar el efecto que cada variable observable tiene sobre la elección, considérese las Ecuación (14) a (16). Aquí, \bar{U}_i^A , \bar{U}_i^B y \bar{U}_i^C representan valores medios para las utilidades de los modelos logit mixtos sin variables latentes, con variable latente lineal y con variable latente logarítmica respectivamente, y $\hat{\beta}$ y $\hat{\lambda}$ son los estimadores de máxima verosimilitud para el modelo de elección y variable latente respectivamente.

$$\bar{U}_i^A = \sum_k \hat{\beta}_k X_{ik} \quad (14)$$

$$\bar{U}_i^B = \sum_k \hat{\beta}_k X_{ik} + \hat{\beta}^* \left(\sum_k \hat{\lambda}_k X_{ik} \right) \quad (15)$$

$$\bar{U}_i^C = \sum_k \hat{\beta}_k X_{ik} + \hat{\beta}^* \ln \left(\sum_k \hat{\lambda}_k X_{ik} + 1 \right) \quad (16)$$

El efecto total que cierta variable, $X_{ik'}$, tiene sobre alguna de estas utilidades se puede obtener simplemente calculando sus derivadas con respecto a esta variable específica, como se mencionó en la Sección 4. Las derivadas para las funciones de utilidad presentadas anteriormente se muestra en (17) a (19).

$$\frac{\partial \bar{U}_i^A}{\partial X_{ik'}} = \hat{\beta}_{k'} \quad (17)$$

$$\frac{\partial \bar{U}_i^B}{\partial X_{ik'}} = \hat{\beta}_{k'} + \hat{\beta}^* \hat{\lambda}_{k'} \quad (18)$$

$$\frac{\partial \bar{U}_i^C}{\partial X_{ik'}} = \hat{\beta}_{k'} + \frac{\hat{\beta}^* \hat{\lambda}_{k'}}{\sum_k \hat{\lambda}_k X_{ik} + 1} \quad (19)$$

Mientras que los primeros dos casos son valores independientes de los atributos de las alternativas, el tercer caso es una función de X_i . Esto quiere decir que los efectos netos en un modelo que incluye la variable latente de manera no lineal serán una función de las variables observables. Los efectos netos para los modelos presentados en este trabajo se muestran en la Tabla 2.

La Tabla muestra que los parámetros de los modelos “Sin VL” y “VL lineal” no tienen diferencias significativas. Esto se debe al efecto mostrado por Vij y Walker (2016) y descrito en la Sección 4. No es sorpresa, entonces, que las verosimilitudes de estos dos modelos no son significativamente distintas. Por otra parte, el modelo “VL log” tiene parámetros distintos que, a su vez, entregan una mejor bondad de ajuste.

Esto no quiere decir que las aseveraciones de Vij y Walker son equivocadas; en principio, si el modelo “VL log” fuera estimado observando indicadores de elección solamente, se obtendría la

Cuadro 2: Efectos netos para modelos logit mixtos estimados.

Variable	Sin VL	VL lineal	VL lineal		
			Min.	Promedio	Máx.
Constante (derecha)	-0,105	-0,105		0,0198	
Infra.: Sin ciclofacilidad	0	0		0	
Infra.: Línea	0,824	0,825	0,458	0,966	1,40
Infra.: Pintura	1,08	1,08	0,660	1,30	2,32
Infra.: Separadores	1,19	1,20	0,983	1,50	1,85
Infra.: Vereda	1,47	1,47	1,03	1,42	2,37
Ancha	0,233	0,234	0,494	0,627	0,868
Velocidad máx. baja	0,479	0,480	0,605	0,776	1,86
No hay buses (media)	0,662	0,663	0,731	0,936	2,28
No hay buses (desv. est.)	0,811	0,813		0,901	
Tiempo de viaje	-0,0682	-0,0683		-0,0896	

misma verosimilitud que la que se obtiene al incluir indicadores psicométricos. Sin embargo, en este caso específico, y en otros casos donde existen efectos lineales y no lineales, problemas de identificabilidad surgen cuando no se utilizan estos indicadores. Esto quiere decir que hay potencial para mejorar el ajuste en modelos híbridos de elección discreta que incluyen variables latentes que afectan la elección de formas no lineales. Esto es válido especialmente para variables latentes perceptuales dado que, en estos casos, es muy probable que las variables observables se presenten tanto en el submodelo de elección como en el de variable latente. No existirían problemas de identificabilidad si estos submodelos fueran funciones de variables observables distintas, que suele ser el caso para modelos con variables latentes actitudinales.

6. Conclusiones

Los modelos híbridos de elección discreta han sido ampliamente utilizados durante los últimos años para modelar explícitamente constructos latentes que afectan los patrones de decisión de las personas. Sin embargo, esta exploración ha tenido un énfasis en las variables latentes actitudinales, dejando de lado a las variables latentes perceptuales. Esto puede deberse a que, si se sigue la especificación clásica, se requiere una cantidad importante de información, lo que aumenta los costos de recolección y estimación.

Este trabajo propone una modificación a esta especificación que reducen estos costos. En vez de recolectar un indicador psicométrico por cada alternativa presentada a los usuarios encuestados, proponemos una especificación que permite recolectar tan solo un indicador por cada alternativa etiquetada por persona encuestada. Con estos indicadores, junto con las elecciones observadas, se pueden estimar modelos de elección con variables latentes conjuntamente.

Esta propuesta se probó utilizando datos de una encuesta de preferencias declaradas por diseño de infraestructura ciclista. Se le presentó a las personas encuestadas distintas calles hipotéticas, y se les pidió que mostraran por cuál preferirían usar la bicicleta. Después de esto, a los usuarios se les presentó una alternativa aleatoria y se les pidió que mostraran qué tan seguros se sentirían usando

la bicicleta en este caso. Con esta información se estimaron tres modelos logit mixtos: uno sin variables latentes, y dos que modelaron explícitamente la seguridad percibida a través de estas.

Los resultados muestran que el uso de variables latentes perceptuales es útil para comprender de mejor forma las decisiones de los usuarios. En específico, la seguridad percibida resultó ser significativa en este contexto de elección. A pesar de ser esperada, esta relación no había sido explícitamente mostrada en la literatura anteriormente.

Los modelos estimados también mostraron que las variables latentes tienen potencial para entregar un mejor ajuste a los datos. Vij y Walker (2016) demuestran que los modelos de elección discreta con variables latentes con especificaciones lineales pueden ser colapsados en modelos logit mixtos que no consideran indicadores psicométricos. Sin embargo, en este trabajo mostramos que si estas variables son incluidas en el modelo de elección de formas no lineales existe potencial para mejorar el ajuste. Esto se demostró al incluir la variable latente que describe seguridad percibida de manera logarítmica. Este caso presentó un mejor ajuste a las elecciones observadas.

El campo de los modelos de elección discreta con variables latentes perceptuales tiene varias aplicaciones interesantes, incluso fuera del contexto del transporte. Este trabajo busca llenar un vacío metodológico presenten en el área con la esperanza de aumentar el número y calidad de los estudios que utilizan este tipo de variables para comprender de mejor forma cómo deciden las personas.

Agradecimientos

Los autores quieren agradecer el apoyo entregado por el Centro de Desarrollo Urbano Sustentable (CONICYT/FONDAP 15110020), el Instituto de Sistemas Complejos de Ingeniería (ICM: P-05-004-F, CONICYT: FBO16) y CONICYT (FONDECYT 1150590). Una versión expandida de este trabajo fue enviado anteriormente a la revista *Transportation Research Part A: Policy and Practice* y está actualmente en proceso de revisión.

Referencias

- Allport, G. W. (1935). Attitudes. In *A Handbook of Social Psychology*, volume 1, pages 798–844.
- Bahamonde-Birke, F., Kunert, U., Link, H., y Ortúzar, J. D. D. (2015). About attitudes and perceptions: finding the proper way to consider latent variables in discrete choice models. *Transportation*, 42(6):1–19.
- Bahamonde-Birke, F. y Ortúzar, J. d. D. (2015). About the categorization of latent variables in hybrid choice models. *Discussion Papers*, 1527.
- Bartczak, A., Mariel, P., Chilton, S., y Meyerhoff, J. (2016). The impact of latent risk preferences on valuing the preservation of threatened lynx populations in Poland. *Australian Journal of Agricultural and Resource Economics*, 60(2):284–306.

- Ben-Akiva, M., Walker, J., Bernardino, A., Gopinath, D. A., Morikawa, T., y Polydoropoulou, A. (2002). Intergration of Choice and Latent Variable Models. *In Perpetual Motion-Travel Behavior Research Opportunities and Application Challenges*, (1):431–470.
- ChoiceMetrics (2012). Ngene 1.1.2 User Manual & Reference Guide.
- Daly, A., Hess, S., Patruni, B., Potoglou, D., y Rohr, C. (2012). Using ordered attitudinal indicators in a latent variable choice model: A study of the impact of security on rail travel behaviour. *Transportation*, 39(2):267–297.
- Hess, S., Shires, J., y Jopson, A. (2013). Accommodating underlying pro-environmental attitudes in a rail travel context: Application of a latent variable latent class specification. *Transportation Research Part D: Transport and Environment*, 25:42–48.
- Horowitz, J. L. (1983). Statistical Comparison of Non-Nested Probabilistic Discrete Choice Models. *Transportation Science*, 17(3):319–350.
- Hoyos, D., Mariel, P., y Hess, S. (2015). Incorporating environmental attitudes in discrete choice models: An exploration of the utility of the awareness of consequences scale. *Science of the Total Environment*, 505:1100–1111.
- Hurtubia, R., Atasoy, B., Glerum, A., Curchod, A., y Bierlaire, M. (2010). Considering latent attitudes in mode choice: The case of Switzerland. *12th WCTR, July 11-15, 2010-Lisbon, Portugal*, pages 1–17.
- Hurtubia, R., Guevara, C. A., y Donoso, P. (2015). Using images to measure qualitative attributes of public spaces through SP surveys. *Transportation Research Procedia*, 11:460–474.
- Kitrinou, E., Polydoropoulou, A., y Denis, B. (2010). Development of Integrated Choice and Latent Variable (ICLV) Models for the Residential Relocations Decision in Island Areas. *Choice Modelling: The State-of-the-Art and the State-of-Practice*, (May):593–618.
- Lee, L. F. (1982). Specification error in multinomial logit models. Analysis of the omitted variable bias. *Journal of Econometrics*, 20(2):197–209.
- McFadden, D. y Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, 15(November 1998):447–470.
- Palma, D., Ortúzar, J. d. D., Rizzi, L. I., Guevara, C. A., Casaubon, G., y Ma, H. (2016). Modelling choice when price is a cue for quality a case study with Chinese wine consumers. *Journal of Choice Modelling*, 19:24–39.
- Pickens, J. (2005). *Organizational Behavior in Health Care*. Jones and Bartlett Publishers, Sudbury.
- Rossetti, T. (2017). *Modelación de preferencias por diseño de infraestructura ciclista utilizando variables latentes perceptuales*. Tesis para optar al grado de magíster en ciencias de la ingeniería, Pontificia Universidad Católica de Chile.

- Ruud, P. A. (1983). Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models. *Econometrica*, 51(1):225–228.
- SECTRA (2015). Encuesta origen destino de viajes 2012. Technical report, Ministerio de Transportes y Telecomunicaciones, Santiago, Chile.
- Sener, I. N., Eluru, N., y Bhat, C. R. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation*, 36(5):511–539.
- Vega, R., Niehaus, M., Waintrub, N., Peña, C., y Galilea, P. (2015). Cycle-path design standards: a qualitative characterisation approach for cyclist flows in Santiago. In *Thredbo: International Conference Series on Competition and Ownership in Land Passenger Transport*.
- Vij, A. y Walker, J. (2016). How, when and why integrated choice and latent variable models are latently useful. *Transportation Research Part B: Methodological*, 90:192–217.
- Walker, J. y Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical social sciences*, 43(3):303–343.
- Yáñez, F., Raveau, S., y Ortúzar, J. d. D. (2010). Inclusion of latent variables in Mixed Logit models: Modelling and forecasting. *Transportation Research Part A: Policy and Practice*, 44(9):744–753.