

## La predicción de la demanda en la evaluación de proyectos

**Anna Matas**  
**José Luis Raymond**  
**Mar González-Savignat**  
**Adriana Ruíz**



**Proyecto: EVALUACIÓN SOCIOECONÓMICA Y FINANCIERA DE PROYECTOS DE TRANSPORTE**

Con la subvención del Centro de Estudios y Experimentación de Obras Públicas (CEDEX)

Ministerio de Fomento

Ref. PT-2007-001-02IAPP

[www.evaluaciondeproyectos.es](http://www.evaluaciondeproyectos.es)

# LA PREDICCIÓN DE LA DEMANDA EN LA EVALUACIÓN DE PROYECTOS

---

31/Marzo/2009

Anna Matas  
José Luis Raymond  
Mar González-Savignat  
Adriana Ruíz

## Resumen

---

La predicción de la demanda es un elemento clave en la evaluación de proyectos. Sin embargo, la evidencia nos muestra errores de predicción significativos. Este documento pretende ofrecer una relación de los diversos aspectos que debe considerar la predicción de demanda con la finalidad de evitar errores que puedan distorsionar el resultado de la evaluación. El texto se estructura en tres grandes apartados. El primero ofrece una guía de los elementos relacionados con el impacto del proyecto en la demanda. El segundo analiza los principales modelos de predicción y el último detalla formas alternativas de incorporar la incertidumbre asociada a la predicción de la demanda.



## Tabla de contenidos

---

<b>1. INTRODUCCIÓN.....</b>	<b>1</b>
<b>2. ELEMENTOS RELACIONADOS CON LA PREDICCIÓN DE LA DEMANDA DE TRANSPORTE</b>	<b>1</b>
<b>3. LOS MODELOS DE PREDICCIÓN .....</b>	<b>6</b>
3.1. PREDICCIÓN TEMPORAL DEL TRÁFICO.....	8
3.1.1. <i>El modelo tendencial</i> .....	8
3.1.2. <i>El modelo econométrico</i> .....	9
3.2. PREDICCIÓN DEL TRÁFICO DESVIADO.....	15
3.2.1. <i>La regresión logística</i> .....	16
3.2.2. <i>Modelo de elección discreta</i> .....	16
3.3. PREDICCIÓN A PARTIR DE VALORES RECOMENDADOS .....	20
<b>4. INCERTIDUMBRE ASOCIADA A LA PREDICCIÓN DE LA DEMANDA .....</b>	<b>26</b>
<b>5. CONCLUSIONES Y RECOMENDACIONES FINALES .....</b>	<b>29</b>
<b>6. REFERENCIAS.....</b>	<b>31</b>
<b>ANEXO 1: LA FORMA FUNCIONAL.....</b>	<b>34</b>
<b>ANEXO 2: EL ANÁLISIS DE COINTEGRACIÓN EN SERIES TEMPORALES.....</b>	<b>37</b>
<b>ANEXO 3: JUSTIFICACIÓN DE LA REGRESIÓN LOGÍSTICA .....</b>	<b>39</b>



## 1. INTRODUCCIÓN.

La predicción de la demanda para un determinado modo o servicio de transporte es un elemento clave para la evaluación de un proyecto. Sin embargo, no es una tarea fácil. En primer lugar, es necesario especificar un modelo que relacione la demanda de transporte con los principales factores explicativos de la misma. Dada la complejidad de las decisiones individuales en la elección entre alternativas de transporte, la interrelación entre los distintos modos y la falta de información acerca de los atributos de la oferta, el modelo debe ser considerado sólo como una aproximación a la realidad y, por consiguiente, está sujeto a un cierto nivel de error. En segundo lugar, desconocemos el valor que las variables explicativas de la demanda tomarán en el futuro; por ello, es necesario predecir estas variables, lo cual introduce una segunda fuente de incertidumbre en la predicción de la demanda.

En este documento se ofrece un análisis de los distintos aspectos que deben ser considerados en la predicción de la demanda para llevar a cabo la evaluación de un proyecto o política de transporte.

En el apartado que sigue a esta breve introducción se identifican todos los impactos sobre la demanda que genera el proyecto. En el tercer apartado se definen los modelos de predicción de la demanda más adecuados, teniendo en cuenta los datos disponibles en España. Por último, se propone un método para incorporar las distintas fuentes de incertidumbre a la predicción final.

Este análisis se realiza desde una perspectiva general de forma que sea aplicable a distintos proyectos de inversión en infraestructuras o políticas de inversión. Dado que cualquier mercado de transporte comparte un conjunto de elementos, una aproximación común es posible. No obstante, en la predicción de la demanda para cada tipo de infraestructura o servicio de transporte será preciso tomar en consideración sus características específicas.

## 2. ELEMENTOS RELACIONADOS CON LA PREDICCIÓN DE LA DEMANDA DE TRANSPORTE

La valoración de los cambios en bienestar requiere cuantificar la demanda de transporte a lo largo de la vida útil de la inversión en la situación con y sin proyecto. Para ello, es preciso responder a las siguientes preguntas: ¿cuáles son los impactos que el proyecto tendrá sobre la demanda en el conjunto de la red de transportes? y ¿cómo evolucionará la demanda a lo largo del tiempo?

A continuación se detallan los principales aspectos que deben tomarse en consideración en la predicción de la demanda.

1. Es preciso clarificar el área geográfica de influencia del proyecto. Una vez definido el ámbito territorial, la predicción debe contemplar todos los flujos de viajes en la red de transporte de esta área.
2. El análisis debe identificar cuáles son los modos de transporte y las rutas alternativas a la infraestructura o servicio sujeto a evaluación. En cada caso es necesario conocer los costes monetarios y de tiempo de dichas alternativas para poder evaluar hasta qué punto constituyen alternativas relevantes.

Para los proyectos que impliquen una infraestructura de nueva creación estos dos primeros elementos requerirán probablemente la realización de encuestas para determinar el área de captación de tráfico y la red afectada.

3. En la situación sin proyecto, el punto de partida es el *tráfico existente* en el año base. Este tráfico puede obtenerse a partir de las estadísticas elaboradas por los distintos organismos con competencias. En el caso de la carretera, por ejemplo, existen los datos del Mapa de Tráfico, que identifican el flujo de los distintos tipos de vehículos para toda la red. Para el resto de modos de transporte, ferrocarril, aéreo y marítimo, se dispone – con mayor o menor grado de desagregación – de datos de viajes por líneas o corredores de tráfico que constituyen un buen punto de partida para la predicción.
4. En la situación con proyecto, es necesario añadir al tráfico anterior aquel que procede de una reasignación entre rutas alternativas dentro del mismo modo. Las reasignaciones de tráfico son frecuentes en proyectos de inversión en carreteras. La construcción de una autovía o autopista supone un desvío de tráfico de las rutas existentes hacia las nuevas. En estas situaciones, en general, resulta suficiente realizar encuestas origen-destino que permitan identificar estos cambios. La suma de estos dos flujos es lo que se considera *tráfico existente* en el año base.
5. Cuando la inversión supone un cambio significativo en el coste generalizado de una infraestructura o servicio de transporte en relación con el de sus competidores, es necesario cuantificar el cambio modal. Probablemente, el ejemplo más claro es el impacto que la puesta en marcha de algunas líneas de alta velocidad tiene sobre los usuarios del avión. En estos casos, es preciso especificar un modelo de elección entre modos de transporte que permita evaluar el desvío de tráfico entre modos a partir de los atributos de cada alternativa. Dichos atributos incluyen el precio, los costes de tiempo y cualquier otra característica que sea determinante para la elección.

La evaluación económica contempla también los beneficios para este *tráfico desviado*; no obstante, el criterio empleado en la evaluación difiere del empleado para el tráfico existente.

6. Es recomendable comparar el resultado sobre captación de tráfico obtenido en el modelo de demanda con el que se obtendría al aplicar valores estándares de la elasticidad en relación con los costes monetarios y de tiempo. Si se observan diferencias importantes, es imprescindible revisar el modelo de demanda para evitar predicciones erróneas. La experiencia internacional muestra que en las inversiones en autopistas de peaje y en ferrocarril existe una clara tendencia a sobreestimar la demanda como consecuencia de una previsión excesivamente optimista en la captación de tráfico<sup>1</sup>. La comparación con el resultado observado en proyectos similares y valores estándares ayudará a evitar estos errores.
7. Asimismo, será necesario evaluar posibles cambios en el tráfico captado a lo largo de los años de vida útil. En particular, es preciso prever el impacto de nuevas inversiones en los modos competitivos que modifiquen la relación entre los costes generalizados.
8. Una inversión en transporte puede dar lugar también a la *generación de nuevos viajes* como consecuencia de una caída en el coste de transporte. La relevancia de este impacto es una cuestión largamente debatida en el contexto de la evolución de proyectos. Así, tradicionalmente, la evaluación de las inversiones en carretera supone una matriz de viajes fija, bajo la hipótesis de que es una aproximación a la realidad suficientemente válida. Por el contrario, dicha hipótesis no sería claramente aceptable en contextos de construcción de nuevas infraestructuras como aeropuertos o trenes de alta velocidad.

Para cuantificar el tráfico generado es necesario un modelo de demanda que relacione el tráfico con el precio y el coste del tiempo de usar la infraestructura o el servicio de transporte. En un contexto dinámico de crecimiento del tráfico es muy difícil distinguir entre los nuevos viajes resultado del propio crecimiento económico de aquellos que son generados propiamente por el proyecto. Por ello, es recomendable confiar en valores obtenidos en proyectos de similares características y/o utilizar la elasticidad de la demanda con respecto al coste generalizado. Esta forma de proceder evita, de nuevo, pecar de un exceso de optimismo en las predicciones de demanda.

---

<sup>1</sup> Para una revisión de la literatura ver Matas, Raymond, Savignat y Ruíz (2009).

9. A largo plazo, una reducción significativa de los costes de transporte puede incentivar cambios en la localización de las actividades y residencias dando lugar a un mayor número de viajes en la red y a una redistribución de los orígenes y destinos. Así, frente a una reducción del tiempo de viaje, los hogares y los negocios tienden a descentralizarse y, como consecuencia, se incrementan los kilómetros recorridos. Este tráfico debe incorporarse a la evaluación como *tráfico generado*.

La predicción del tráfico generado asociado a cambios en la localización requiere un modelo que relacione usos del suelo y transporte. Se trata, en general, de un modelo a escala regional que recoge la distribución territorial de las actividades y la accesibilidad ofrecida por los distintos modos de transporte. Actualmente, no obstante, no existen modelos de este tipo en España con rigor suficiente. La construcción de uno de ellos requiere llevar a cabo encuestas a familias y empresas que recojan de forma precisa el origen y destino de los viajes. Actualmente, tales datos no están disponibles en nuestro país.

Nuestra recomendación es mostrarse cauto en las predicciones relacionadas con los cambios en las actividades en el territorio y confiar, nuevamente, en el resultado de proyectos similares.

10. Cuando el proyecto afecta a algún tipo de tráfico fuertemente concentrado en algunas pocas empresas – por ejemplo, determinados tráficos marítimos – es aconsejable conocer directamente la previsión de la demanda que efectúan las empresas involucradas. Esta forma de operar se utiliza habitualmente en Puertos del Estado.
11. Un aspecto adicional a tener en cuenta es la creación de espacios singulares con capacidad de generación de tráfico. Por ejemplo, la construcción de parques temáticos o grandes centros de empleo que pueden modificar los flujos de viaje.
12. Los aspectos contemplados en los anteriores puntos permiten definir los tres componentes de la demanda en el año base: tráfico existente, tráfico desviado y tráfico generado<sup>2</sup>. A la suma de los dos últimos componentes se denomina también tráfico inducido. Cabe insistir, no obstante, en que la distinción entre ellos no siempre es fácil en un contexto dinámico de aumento de la demanda.
13. Una vez identificados los distintos componentes de la demanda, es necesario predecir su evolución para el período de vida útil del proyecto. El comportamiento de la demanda dependerá de un conjunto de factores económicos y demográficos y del coste

---

<sup>2</sup> Para una definición exhaustiva de los distintos componentes de la demanda ver Hills (1996).

generalizado – básicamente tiempo y precio – del modo de transporte. Es esencial elegir un modelo de predicción acorde con las características del problema y de los datos disponibles. La incertidumbre asociada al comportamiento de las variables explicativas a largo plazo hace recomendable confiar en una ecuación relativamente simple, a no ser que la complejidad de la inversión lo desaconseje. El problema de la incertidumbre se trata de forma detallada en el apartado 4 de este documento.

14. Un supuesto frecuentemente utilizado es que el coste generalizado se mantiene constante a lo largo del tiempo y la demanda depende únicamente de las variables socioeconómicas y demográficas. Sin embargo, es bien conocido que a medida que la demanda se incrementa, también lo hace la congestión en el sistema de transporte. Por este motivo, es necesario acompañar la predicción de la demanda con una curva de oferta que modelice la relación tráfico/capacidad a lo largo de la vida útil del proyecto.

El incremento de congestión no sólo afecta a las carreteras, sino que también es perceptible en el ferrocarril, en los aeropuertos o en cualquier modo de transporte público. La curva de oferta puede integrarse a través de un modelo de asignación de tráfico a la red que proporcione las estimaciones de flujo de vehículo, velocidad y cambio en los costes. Los viajes en cada periodo deberían obtenerse a partir del punto de equilibrio para las curvas de demanda y oferta<sup>3</sup>. Es obvio que el cálculo de los equilibrios con y sin proyecto es una tarea complicada. Sin embargo, ignorar estas relaciones puede generar errores muy importantes en la evaluación de los beneficios netos de un proyecto<sup>4</sup>.

15. Para cada proyecto se deberá delimitar el nivel de detalle al que se analiza la demanda. Entre otros aspectos, es preciso decidir acerca del nivel de desagregación según tipo de tráfico (vehículos pesados y ligeros, en la carretera; pasajeros nacionales, internacionales o en tránsito en los aeropuertos; pasajeros y mercancías en el ferrocarril o tipo de mercancía transportada en los puertos) y periodo temporal (tráfico en períodos punta y valle; mes del año o día de la semana). Cuanto más alto el detalle, más complejo será hallar el equilibrio. Por ello, las decisiones de desagregación deberán tomarse con cautela y de acuerdo con las exigencias del problema y la calidad de los datos disponibles.

---

<sup>3</sup> Para un análisis de este punto puede consultarse Department of Transport (2006), TAG, Unit 3.1.

<sup>4</sup> Ver, por ejemplo, Mackie et al. (1996).

16. Otro aspecto a tomar en consideración es la desagregación de la demanda según tipo de pasajero o mercancía. En la medida que la evaluación de un proyecto utilice distintos valores del tiempo según el tipo de usuario, será necesario segmentar la demanda de acuerdo con el mismo criterio. La forma más usual de segmentar la demanda obedece al motivo de viaje (negocio u ocio) y al tipo de mercancía transportada.
17. Para evitar predicciones con un grado elevado de error, es recomendable siempre llevar a cabo análisis de sensibilidad en relación con las principales incertidumbres. En el apartado 4 de este documento se detalla cómo incorporar la incertidumbre asociada al modelo. No obstante, la posibilidad de cometer errores depende también de los supuestos acerca del comportamiento de las variables explicativas y del porcentaje de tráfico desviado y generado. En relación con estos aspectos es aconsejable establecer distintos escenarios sobre el comportamiento de las variables y proceder a la comparación de los resultados. En cualquier caso, la demanda predicha debe compararse siempre con previsiones realizadas a gran escala para la red de transportes y analizar la compatibilidad con las mismas.
18. Un último aspecto a destacar es la necesidad de mejorar las bases de datos disponibles en España para poder estimar de forma más precisa los cambios en la demanda de transporte derivados de una mejora en la red. Una de las mayores carencias es la ausencia de *matrices origen-destino* de los viajes, tanto con datos de flujos agregados de tráfico como con datos individuales y para todos los modos de transporte. El problema de falta de datos es especialmente grave para el transporte de mercancías. Para ello, es preciso la realización de encuestas en los distintos ámbitos geográficos con una identificación precisa del origen y destino de los viajes. Estas encuestas deberían tener una continuidad temporal para permitir la formación de bases de datos de panel. Respecto a los datos de flujos de viajes, es recomendable presentarlos con una mayor desagregación según tipo de usuario, con la finalidad de segmentar la demanda y mejorar la estimación.

### 3. LOS MODELOS DE PREDICCIÓN

Una vez delimitado el impacto de una inversión en la demanda, debemos elegir el modelo de predicción más adecuado a las características del proyecto. Dado el gran volumen de recursos públicos que absorbe una inversión en transporte, y su carácter habitualmente irrecuperable, en general parece recomendable utilizar todas las aproximaciones a nuestro alcance para determinar la rentabilidad de la inversión.

## CONSTRUCCIÓN DE UNA LÍNEA FERROVIARIA DIRECTA ENTRE DOS CIUDADES

### I. IDENTIFICACIÓN DEL IMPACTO DEL PROYECTO SOBRE EL FLUJO DE VIAJES

1. ¿Cuál es el área geográfica afectada?
2. ¿Qué tipo de tráfico se verá afectado?
  - a. ¿Qué tipo de mercancía?
  - b. ¿Qué tipo de pasajeros (por motivo de viaje, por ejemplo)?
3. ¿Qué rutas o modos de transporte se verán afectados?
  - a. Rutas alternativas y complementarias.
  - b. Modos sustitutos: aéreo y por carretera.
4. ¿Qué tipo de información o datos están disponibles? ¿A qué nivel de detalle? ¿Qué período temporal?

### II. CARACTERIZACIÓN Y CUANTIFICACIÓN DE LA DEMANDA

#### 1. Sin proyecto

- a. ¿Cuál es el tráfico existente en el año base?

#### 2. Con proyecto

- a. ¿Cuál es el tráfico debido a una reasignación entre las rutas existentes y la nueva ruta (dentro del mismo modo ferroviario)?
- b. ¿Es significativo el cambio en el coste generalizado en relación con los otros modos alternativos? ¿Cuál es el tráfico que puede ser captado?
  - i. Aplicación de un modelo de elección modal.
  - ii. Aplicación de valores de captación hallados en estudios de características similares.
- c. ¿Cuál es el tráfico generado?
  - i. Aplicación de un modelo de demanda considerando:
    - Precio y coste del tiempo de usar el nuevo servicio ferroviario.
    - Variables demográficas y socioeconómicas relevantes.
  - ii. Aplicación de valores de generación hallados en estudios similares.

### III. PREDICCIÓN DE LA EVOLUCIÓN DE LA DEMANDA A LO LARGO DE LA VIDA ÚTIL DEL PROYECTO (CON Y SIN PROYECTO)

#### 1. Predicción de los flujos de tráfico

- a. ¿Modelo tendencial o modelo econométrico?
- b. ¿Qué posibles cambios pueden producirse en el tráfico captado?
- c. ¿Es posible que se produzcan nuevas inversiones en los modos competitivos?
- d. ¿Es posible que en el largo plazo el proyecto ocasione una redistribución de las actividades y residencias?

#### 2. Incorporación de la incertidumbre

- a. Análisis de sensibilidad en relación con las principales fuentes de incertidumbre.
- b. Cálculo de la varianza del error de predicción para cada una de las fuentes de error.
- c. Definición de escenarios según:
  - i. Supuestos sobre el porcentaje de tráfico desviado y generado.
  - ii. Comportamiento de las variables explicativas.

De acuerdo con las observaciones efectuadas en el anterior apartado, idealmente deberíamos disponer de un modelo de tráfico integrado que contemplara la localización de las actividades en el territorio, la generación de viajes, la distribución entre destinos, la elección modal y la asignación a la red. La disponibilidad de un modelo integrado de etapas múltiples es difícil y requiere disponer de una base de datos que en este momento no existe en España<sup>5</sup>. Por otro lado, es posible que un modelo general no sea siempre aplicable a inversiones específicas. La elección del modelo más adecuado para predecir la demanda dependerá fundamentalmente de las características y magnitud del proyecto y de la disponibilidad y/o costes de obtener los datos necesarios.

Este documento contempla los modelos asociados a los principales requisitos que exige la predicción de la demanda. En concreto, se analizan, primero, los modelos cuya finalidad es predecir los flujos de tráfico y, en segundo lugar, los modelos de elección modal.

### 3.1. Predicción temporal del tráfico

La evolución de la demanda a lo largo de la vida útil del proyecto depende esencialmente del comportamiento de los costes generalizados y de los factores socioeconómicos y demográficos que la determinan. Los modelos econométricos de demanda constituyen la metodología adecuada para la predicción. Es posible, no obstante, que en determinados contextos la demanda de transporte dependa sólo de su trayectoria pasada. Si ello es así, un simple modelo tendencial permite hallar la tasa de crecimiento del tráfico a lo largo del tiempo. A continuación se describen dichos modelos.

#### 3.1.1. El modelo tendencial

El modelo tendencial o ingenuo es la vía más simple de predicción, dado que no incorpora información sobre otras variables de naturaleza económica. Obviamente, este modelo sólo recoge la evolución de la demanda existente y en ningún caso permite incorporar tráfico inducido, ya sea desviado o generado.

Existen diversas expresiones matemáticas para relacionar el tiempo con la variable a predecir. Dos alternativas simples y frecuentemente usadas son la relación lineal y la semi-logarítmica:

---

<sup>5</sup> Para una revisión de los datos disponibles en España ver *González-Savignat, Matas, Raymond y Ruiz (2009)*.

$$y_t = \alpha + \beta \cdot t + u_t$$

$$\ln y_t = \alpha^* + \beta^* \cdot t + \varepsilon_t$$

La primera de ellas, al postular linealidad, implícitamente supone que la tasa de variación de los viajes es decreciente a medida que “t” aumenta (es decir,  $\frac{\Delta y_t}{y_t} = \dot{y}_t = \frac{\beta}{y_t} = \frac{\beta}{(\alpha + \beta \cdot t + u_t)}$ ), mientras que la segunda supone una tasa de variación constante igual a “ $\beta^*$ ”. Es habitual elegir entre estas alternativas de acuerdo con el grado de ajuste del modelo. Ni la suma de cuadrados de los residuos ni el  $R^2$  puede emplearse para decantar la elección debido a que la variable dependiente es distinta.

Un criterio de elección puede ser el valor de la función de verosimilitud. En la especificación lineal, la función de verosimilitud referida a la probabilidad de haberse obtenido la muestra observada de valores de la variable a explicar y viene dada por:

$$\ln L_{lineal} \propto -\frac{T}{2} \ln \left( \frac{\sum \hat{u}_t^2}{T} \right)$$

Cuando el proceso de generación de los datos se supone que obedece a la especificación semi-logarítmica, la función de verosimilitud referida a los valores de  $y$  y, por tanto, comparable con la precedente, viene dada por:

$$\ln L_{Semilogaritmica} \propto -\sum \ln Y - \frac{T}{2} \ln \left( \frac{\sum \hat{\varepsilon}_t^2}{T} \right)$$

Entre estas dos formulaciones se trata de elegir aquella con mayor valor de la función de verosimilitud.

La predicción a partir de la extrapolación de la tendencia pasada es un método común, aunque a menudo excesivamente simple. Este método supone que el comportamiento pasado continuará en el futuro, supuesto que puede resultar excesivamente ingenuo.

### 3.1.2. El modelo econométrico

En el modelo econométrico la predicción de la demanda viene determinada por un conjunto de variables explicativas de la misma. Dado que este método incorpora mayor información que el anterior debe, en principio, ofrecer mejores predicciones. La elección de las variables

explicativas se realiza en función del tipo de demanda a predecir y de los datos disponibles. Por un lado, se incorporan las variables relacionadas con el entorno socio-económico y demográfico-población, empleo, PIB o volumen de comercio exterior. Por el otro, la demanda depende del coste generalizado de uso de la infraestructura, precio y tiempo.

Frente al modelo tendencial, la principal ventaja es que permite efectuar predicciones condicionadas así como el diseño de escenarios alternativos. El principal problema de este enfoque es que desconocemos el valor de las variables explicativas a lo largo del tiempo. Por ello, es necesario predecirlas, con lo que se introduce un nuevo elemento de incertidumbre.

En forma matricial, la ecuación de demanda puede expresarse como:

$$Y = X' \beta + \varepsilon$$

Donde:

- $Y$ , es la demanda
- $X$ , las variables explicativas
- $\beta$ , los parámetros a estimar
- $\varepsilon$ , la perturbación aleatoria

La estimación de esta ecuación requiere establecer una determinada forma funcional y elegir el tipo de datos utilizados. Las formas funcionales más comunes son la doble logarítmica, la semi-logarítmica y la lineal, aunque se utilizan formas más complejas como la translog y la exponencial. En el anexo 1 se presenta una discusión sobre la elección de la forma funcional.

Respecto a los datos, la estimación de una función de demanda a efectos de predecir a largo plazo aconseja la utilización de series temporales. Este tipo de datos permite tener en cuenta que los individuos no se ajustan de forma inmediata a los cambios en la oferta. Sin embargo, cuando la inversión requiere modelizar efectos específicos de generación de tráfico en ámbitos territoriales concretos – por ejemplo, apertura de una nueva estación – los datos *de corte transversal* pueden proporcionar mejores predicciones. La utilización de datos de panel es una forma elegante y conveniente de aprovechar las ventajas de ambos tipos de datos. A continuación se desarrollan con mayor detalle las ventajas e inconvenientes de cada una de las tres posibilidades.

## Estimación con series temporales

Los datos de series temporales son la forma natural de estimar una función de demanda a efectos de predecir a largo plazo. Este tipo de datos permite modelizar el ajuste dinámico de los individuos ante cambios en los atributos de la oferta. Es conocido que los individuos no se ajustan de forma inmediata a variaciones en el precio o en el tiempo de viaje debido a la persistencia de hábitos, incertidumbre o información incompleta. Con una especificación dinámica, los datos de series temporales permiten averiguar el número de periodos necesarios para completar el cambio en la demanda y la velocidad de dicho cambio. Así, es posible distinguir entre el impacto a corto plazo y largo plazo de una inversión y determinar qué porcentaje del cambio ha finalizado en cada periodo temporal.

La estimación de un modelo de series temporales nos obliga a ser cuidadosos con el proceso de estimación econométrico. En particular, si las series no son estacionarias, como frecuentemente ocurre, una simple ecuación de regresión entre ellas puede ofrecer resultados espurios (una regresión entre dos variables independientes entre sí pero con raíz unitaria mostrará frecuentemente una relación estadísticamente significativa). Es preciso, por lo tanto, averiguar si las variables están efectivamente relacionadas o si se trata de una relación espuria. En caso de dudas, una forma heurística de proceder es verificar si al estimar la misma ecuación en niveles y en incrementos, se sigue verificando la significatividad estadística y el orden de magnitud de los coeficientes. Si la estimación en niveles ofrecía resultados sólo aparentemente satisfactorios debido al carácter no estacionario de las series (es decir, si efectivamente se trataba de una regresión espuria), al emplear incrementos la asociación debe desaparecer. No obstante, resulta más recomendable basar el análisis en el estudio de la naturaleza de las series. En concreto, en el estudio de la integración y cointegración de las variables. Una aproximación simple a este análisis se detalla en el *anexo 2*.

La estimación con series temporales, y el análisis de cointegración, exige disponer de un número elevado de observaciones. Es aconsejable trabajar con datos anuales y un periodo temporal no inferior a unos 30 años. La utilización de datos trimestrales o mensuales es una forma fácil de incrementar el número de observaciones. No obstante, debemos notar que si bien pasar de datos anuales a trimestrales o mensuales comporta multiplicar por 4 ó por 12, el tamaño de la muestra, el contenido informativo de los datos no se multiplica por igual factor dado que la variabilidad de la muestra seguirá siendo la misma. Además, el uso de variables trimestrales o mensuales dificulta la recopilación de las variables explicativas de la demanda.

En ocasiones, no obstante, es imprescindible modelizar la estacionalidad de la demanda. Por ejemplo, en transporte aéreo, las intensas puntas de demanda en periodos vacacionales determinan en gran medida la capacidad aeroportuaria. En estos casos es aconsejable bien modelizar la demanda con datos trimestrales o mensuales, bien acompañar la ecuación de demanda con un modelo ARIMA estimado con datos mensuales.

El tratamiento de la estacionalidad usualmente se aborda a través de tres alternativas:

- Empleo de variables ficticias estacionales: En el modelo de regresión, si la estacionalidad de la variable dependiente no coincide con la estacionalidad de las variables explicativas, es frecuente el empleo de variables ficticias estacionales.
- En el análisis univariante de los datos, la denominada metodología Box-Jenkins popularizó el empleo de modelos estacionales. El planteamiento era básicamente tratar la estacionalidad de las series a través de un proceso de diferenciación estacional. Por ejemplo, en el célebre modelo de tráfico aéreo que muchas veces se ha empleado para ilustrar el problema, si se dispone de datos mensuales sobre los pasajeros transportados, una forma de obviar la estacionalidad es analizar variaciones interanuales de las variables.
- Finalmente, si lo que nos interesa es examinar el comportamiento de la variable eliminando la estacionalidad, comparar tasas de crecimiento interanuales, que es un enfoque comúnmente empleado, se enfrenta al problema de que esta alternativa enmascara el perfil temporal de la serie. En efecto, con datos mensuales, el crecimiento interanual de la variable debería centrarse en el mes 6.5, o con datos trimestrales en el trimestre 2.5, si bien usualmente las tasas interanuales suelen presentarse centradas en el último período. Una alternativa es aplicar un programa estadístico de desestacionalización de la serie original. Existen distintas alternativas, siendo el programa de desestacionalización X-11 un procedimiento ampliamente utilizado.

### Estimación con datos de corte transversal

Una de las principales limitaciones de las series temporales es que sólo se disponen de un reducido número de variables explicativas. Además, la variabilidad observada en la muestra suele ser baja y dificulta la estimación eficiente de los coeficientes de la ecuación. Por ejemplo, si las empresas de transporte modifican el precio del billete de acuerdo con la inflación, la variación observada en el mismo en términos constantes será nula y, por consiguiente, será imposible determinar cuál es su efecto sobre la demanda. Los datos de corte transversal, al permitir ampliar el número de observaciones e incrementar la

variabilidad de las variables explicativas, pueden contribuir a mejorar los resultados de la estimación.

Estos datos pueden corresponder a viajes entre pares de orígenes y destinos dentro de un área geográfica o bien a datos agregados para distintas unidades territoriales observados en un determinado año. Un ejemplo del primer caso sería la estimación de una función de demanda para viajes en ferrocarril entre pares de estaciones en España, mientras un ejemplo del segundo sería la estimación de una función de demanda con datos agregados procedentes de empresas de autobús urbano de distintas ciudades españolas.

La hipótesis subyacente en la estimación de coste transversal es que los individuos se hallan en una situación de equilibrio. Esta hipótesis ha sido largamente discutida por la evidencia de ajuste no inmediato a cambios en los atributos de la oferta. Sin embargo, el hecho de trabajar con datos agregados disminuye la importancia de este supuesto dado que, al promediar entre individuos, cabe pensar que las situaciones de desequilibrio tienden a desaparecer. Obviamente, con este tipo de datos no es posible indagar acerca de la estructura dinámica del ajuste.

En particular, los datos de corte transversal son necesarios cuando la predicción requiere captar efectos de generación y atracción de nuevo tráfico. Mientras los factores de generación y atracción tienden a mantenerse constantes a lo largo del tiempo, los datos de corte transversal permiten captar la variabilidad que ofrece el hecho de contemplar distintas unidades de observación en un único momento del tiempo. Para la evaluación de inversiones en nuevas infraestructuras o de ampliación de las existentes es imprescindible estimar la capacidad de generación del área de influencia de las mismas. Un ejemplo de ello sería la construcción de una nueva estación de ferrocarril o de un nuevo aeropuerto. El hecho de disponer de una muestra formada por un conjunto de estaciones (o aeropuertos) caracterizada por un amplio número de variables explicativas permite predecir el tráfico que una nueva estación (o aeropuerto) puede generar<sup>6</sup>.

### **Estimación con datos de panel**

Una forma de aunar las ventajas de las series temporales y las observaciones de corte transversal son los datos de panel. Un panel de datos está formado por unidades de corte transversal observadas durante dos o más periodos temporales. Sus atractivos son numerosos. En primer lugar, incrementa el tamaño muestral y los grados de libertad de la

---

<sup>6</sup> En un trabajo reciente Wardman, Lythgoe y Whelan (2007) detallan las ventajas de las observaciones de corte transversal para interpretar el comportamiento de la demanda de ferrocarril

estimación; en segundo lugar, aumenta el número de variables explicativas disponibles y, generalmente, dado que la variación entre unidades en la dimensión de corte transversal es superior que la variación temporal, permite reducir la varianza de las estimaciones y aumentar su fiabilidad.

El modelo con datos de panel más general puede expresarse como:

$$y_{it} = \alpha_i + \beta_{1i}x_{1it} + \beta_{2i}x_{2it} + \dots + \beta_{ki}x_{kit} + \lambda_t + \varepsilon_{it}$$

$$i = 1, 2, \dots, n$$

$$t = 1, 2, \dots, t$$

- $\alpha_i$  recoge los efectos que varían de acuerdo con la observación corte-transversal, pero son constantes en el tiempo y se denominan efectos individuales específicos.
- $\lambda_t$  recoge los efectos que varían en la dimensión temporal, pero son constantes para los unidades de corte transversal y se denominan efectos temporales específicos.
- $\beta_{ik}$  son los coeficientes a estimar para cada una de las variables explicativas de la ecuación  $x_k$ . En la expresión más general se supone que estos coeficientes pueden variar entre unidades de corte transversal. Sin embargo, a menudo, se imponen restricciones de igualdad de acuerdo con la evidencia empírica disponible. Un modelo frecuentemente estimado es el siguiente:

$$y_{it} = \alpha_i + \beta_1x_{1it} + \beta_2x_{2it} + \dots + \beta_kx_{kit} + \lambda_t + \varepsilon_{it}$$

Esta ecuación asume que las diferencias entre unidades de corte transversal quedan recogidas por el término constante de cada ecuación, mientras que los coeficientes de las pendientes son iguales. Entre estas dos formulaciones de la ecuación es posible especificar situaciones intermedias en las que se deje variar alguno de los coeficientes entre unidades de corte transversal y se mantengan fijos aquellos para los que exista evidencia de que son iguales. Sin embargo, es aconsejable contrastar la igualdad de coeficientes mediante los tests estadísticos usuales. De esta forma disminuye el número de coeficientes a estimar y, por lo tanto, se incrementa su fiabilidad.

La contrapartida de los datos de panel es una mayor dificultad para formar la base de datos. Sin embargo, cada vez es más frecuente disponer de bases de datos históricas para los distintos modos de transporte que permiten trabajar con datos de panel. Ello es así para la

red de carreteras, los pasajeros por aeropuerto y los viajeros en las distintas líneas de ferrocarril.

### 3.2. Predicción del tráfico desviado

La inversión en un determinado modo de transporte disminuye su coste generalizado en relación con el resto de modos sustitutivos y, consecuentemente, algunos individuos cambian la alternativa seleccionada. En algunos casos, tal y como se detalla en el apartado 3.3, los viajes desviados de otros modos suponen hasta un 40% de la nueva demanda. Además, buena parte de las políticas de transporte sometidas a evaluación tiene como finalidad lograr que los individuos sustituyan modos o alternativas de transporte con elevados costes externos por otros con menores costes de congestión y medioambientales.

En la medida que sea posible, resulta aconsejable estimar una ecuación de demanda que permita determinar cómo el individuo responde a variaciones en el precio y la calidad – medida usualmente en términos de tiempo – de las alternativas de transporte. Por ello, es preciso definir una ecuación que modelice con rigor las elecciones de los individuos entre alternativas de transporte.

En este contexto, los modelos de elección discreta con datos individuales ofrecen el marco de estimación adecuado. Estos modelos gozan de una larga tradición en economía del transporte y acreditan un elevado grado de precisión en la estimación de las elasticidades de la demanda en relación con los principales atributos de los modos de transporte<sup>7</sup>. El uso de datos microeconómicos evita un posible sesgo de agregación cuando los individuos no son suficientemente homogéneos. Sin embargo, los modelos de elección discreta adolecen de dos inconvenientes importantes. Por un lado, exigen disponer de una base de datos microeconómica de un elevado tamaño muestral y que a la vez recoja todas las características de los modos de transporte para cada uno de los individuos de la muestra. Por el otro, no están exentos de dificultades para predecir flujos de tráfico agregados<sup>8</sup>.

Una posible alternativa a la elección discreta, es estimar una regresión con datos agregados en la que la variable dependiente es el porcentaje de utilización (o frecuencia) de un determinado modo de transporte. Esta aproximación puede ser válida si el proyecto afecta a

---

<sup>7</sup> Los modelos de elección discreta se utilizan para estimar el valor monetario de bienes que no tienen mercado como el tiempo, los accidentes o impactos ambientales.

<sup>8</sup> En este sentido ver Ortuzar y Willumsen (2001)

flujos de tráfico que se comportan de forma suficientemente similar a los datos agregados disponibles.

Es preciso recordar que en ambos casos la finalidad es conocer la demanda desviada. Por ello, el supuesto implícito en estos modelos es que la demanda total es fija y, consecuentemente, las elasticidades derivadas calculadas son inferiores a las obtenidas en contextos que permiten generación o supresión de tráfico.

A continuación se exponen las principales características de ambos tipos de modelos.

### 3.2.1. La regresión logística

La especificación más común con datos agregados para modelizar la distribución de los viajes entre alternativas de transporte es la regresión logística. Sean  $y_m$  e  $y_n$  la demanda para dos modos de transporte alternativos  $m$  y  $n$  y  $T$  el tráfico total. Denominando  $f_m$  el porcentaje de viajes captado por la alternativa  $m$ ,  $f_m = \frac{y_m}{T}$ , la regresión logística puede expresarse como:

$$\ln\left(\frac{f_m}{1-f_m}\right) = (\alpha_m - \alpha_n) + \beta_1(X_{m1} - X_{n1}) + \beta_2(X_{m2} - X_{n2}) + \dots + \beta_k(X_{mk} - X_{nk}) + \zeta$$

Esta especificación ha sido aplicada en múltiples estudios tanto por su facilidad de cómputo como por la evidencia de que la cuota de mercado de un determinado modo de transporte tiende a seguir una distribución en forma de S. Esta forma funcional puede justificarse bien a partir de la especificación de una función de demanda agregada semi-logarítmica para cada alternativa, o bien a partir del supuesto de que a nivel micro los individuos rigen su conducta de elección a través de una formulación logística (ver *anexo 3*).

Habitualmente este tipo de ecuaciones se estima con datos de corte transversal, dado que la cuota de mercado de una alternativa presenta una escasa variabilidad a lo largo del tiempo.

### 3.2.2. Modelo de elección discreta

Un enfoque alternativo son los modelos de elección discreta o modelos desagregados. Estos modelos se basan en datos micro – datos de las unidades individuales de decisión – y explican el comportamiento bien del individuo, bien de la empresa. La tradición de estos modelos en el ámbito de la demanda de transporte es bien conocida desde su desarrollo por

McFadden (1974). Los modelos desagregados gozan de varias ventajas, entre las que destacan un uso eficiente de la información y su fundamentación en la teoría microeconómica.

Bajo la hipótesis de que los individuos maximizan una función de utilidad estocástica, es posible establecer un modelo de elección probabilístico que describe la elección de una determinada alternativa  $i$  en términos de la utilidad esperada de cada alternativa:

$$P_i = P(V_i + \varepsilon_i \geq V_k + \varepsilon_k; k \in B) = P(\varepsilon_k - \varepsilon_i \leq V_i - V_k, k \in B)$$

Donde:

- $i, k$ , son las alternativas de transporte
- $B$ , el conjunto de alternativas factibles
- $V$ , la parte determinística de la función de utilidad
- $\varepsilon$ , la parte estocástica de la función de utilidad

Si  $V$  se asume lineal en los parámetros:

$$V_i = Z_i' \beta$$

Donde:

- $Z$ , vector de variables explicativas para cada alternativa  $i$
- $\beta$ , parámetros a estimar

En el caso de dos alternativas de transporte, 1 y 2, la probabilidad de que un individuo elija la alternativa 1 puede expresarse como:

$$P_1 = P[(\varepsilon_2 - \varepsilon_1) \leq (Z_1\beta - Z_2\beta)] = P(\phi \leq X' \beta) = G(x\beta)$$

Distintas formas funcionales para  $G$  darán lugar a distintos modelos probabilísticos.

Las formas funcionales más frecuentes son la logística y la normal que dan lugar a los modelos logit y probit, respectivamente.

Función logística:

$$G(z) = \frac{\exp(z)}{[1 + \exp(z)]}$$

Función normal:

$$G(z) = \int_{-\infty}^z \phi(v) dv$$

Donde  $\phi(z)$  es la función de densidad de una normal.

En los modelos binomiales, es decir, de sólo dos alternativas, las diferencias entre estas dos formulaciones son pequeñas y conducen a resultados similares. Por el contrario, la estimación en contextos multinomiales es claramente más complicada. Aunque es posible la generalización a un logit multinomial, este modelo exige que la propiedad de independencia de alternativas irrelevantes se cumpla. Es decir, exige que la probabilidad relativa entre dos alternativas sea independiente de la existencia de otras alternativas factibles. El cumplimiento de esta propiedad es altamente improbable en el caso de la demanda de transporte y ello obliga a buscar modelos alternativos más complejos<sup>9</sup>.

#### La elasticidad de la demanda en los modelos de elección discreta

A efectos de cuantificar el tráfico desviado, el principal parámetro de interés del modelo es la elasticidad de la demanda. Por ello, se dedica una especial atención a este aspecto.

En un modelo de elección discreta, la elasticidad de la probabilidad de elegir una alternativa  $i$  respecto a una variación en una variable continua  $x_k$  es:

$$\varepsilon_{ik} = \frac{dP_i}{dx_k} \frac{x_k}{P_i} = g(x\beta) \beta_k \cdot \frac{x_k}{P_i},$$

Donde:

$$g(z) \equiv \frac{dG}{dz}(z)$$

La elasticidad depende del coeficiente estimado  $\beta_k$ , del valor de la función de densidad y del valor que toman la probabilidad y la variable  $x_k$ . Así, pues, éste valor será distinto para cada uno de los individuos de la muestra.

La especificación del modelo logit permite derivar de forma sencilla la expresión para la elasticidad:

---

<sup>9</sup> Para un análisis de estos modelos puede consultarse Train (2003).

$$\varepsilon_i = \frac{dP_i}{dx_i} \frac{x_i}{P_i} = P_i(1-P_i) \frac{dz_i}{dx_i} \frac{x_i}{P_i}$$

Si la relación es lineal:  $z_i = \beta_0 + \beta_1 x_i$

$$\varepsilon_i = (1-P_i)\beta_1 x_i$$

Si la variable está en ln:  $z_i = \beta_0 + \beta_1 \ln(x_i)$

$$\varepsilon_i = (1-P_i)\beta_1$$

El cálculo de la elasticidad ilustra el hecho de que los modelos de elección discreta predicen la probabilidad de que un individuo de la muestra elija una determinada alternativa. No obstante, a efectos de evaluación nos interesa conocer el volumen agregado de viajes desviados. Por ello, es necesario definir un procedimiento de agregación. Siempre que la muestra sea suficientemente representativa, la forma habitual de proceder es agregar a través de los individuos que la forman. En términos de elasticidades, la elasticidad agregada se computa como la media ponderada de las elasticidades individuales:

$$E_{ik} = \frac{\sum_{n=1}^N P_{ni} \varepsilon_{nik}}{\sum_{n=1}^N P_{ni}}$$

Donde:

- $i$ , es la alternativa.
- $k$  el atributo.
- $n$  el individuo.

Los datos utilizados para estimar este tipo de modelos se obtienen de la elección del modo de transporte que utiliza el individuo y de las características de las alternativas existentes disponibles, tanto del modo elegido como del resto. Si este tipo de datos son reales, de elecciones modales observadas, son datos de preferencias reveladas.

Puede ocurrir que una de las alternativas de viaje no esté disponible pero sea necesario predecir su demanda en un futuro; en este caso, es necesario obtener datos de preferencias declaradas. Son datos derivados de contextos hipotéticos en los que se trata de inferir la elección modal del individuo si esa alternativa estuviera disponible. La principal desventaja de este tipo de información es que, al no ser observada, es difícil saber si el individuo se comportará tal y como declara. Sin embargo, presenta algunas ventajas respecto a los datos

de preferencias reveladas, por ejemplo, una mayor variabilidad en los datos, un menor grado de correlación entre las variables explicativas o ausencia de error de medición en las características de los modos de transporte.

Si existe la información o es posible obtenerla, lo mejor es estimar el modelo utilizando ambos tipos de datos. Se trata de la estimación con datos mixtos de preferencias reveladas y declaradas conjuntamente que permite explotar las ventajas de cada una y reducir el efecto de los sesgos o problemas asociados a cada una de las fuentes.

La validez de las predicciones derivadas de este tipo de modelos a largo plazo dependerá del tipo de datos utilizados y del contexto analizado. Por ejemplo, si se utilizan datos de preferencias declaradas en un contexto en el que una de las alternativas no está disponible, pero que lo estará en un futuro próximo, la validez de las predicciones será de corto plazo, hasta que los individuos hayan ajustado su patrón de viajes al nuevo contexto.

### 3.3. Predicción a partir de valores recomendados

En algunas circunstancias es difícil disponer del contexto y los datos adecuados para llevar a cabo una estimación suficientemente robusta de una función de demanda. Algunos ejemplos de ello son la predicción de la demanda para un nuevo modo de transporte o los cambios en la localización de las actividades como resultado de una inversión. Este podría ser el caso de la construcción de una nueva red de vías metropolitanas, la construcción de un nuevo aeropuerto en una ciudad o el enlace de un puerto a la red de ferrocarril. Estas dificultades se traducen en errores en la predicción. Así, por ejemplo, tal y como se detalla en González-Savignat et al. (2009), en los proyectos de transporte ferroviario se observa un sesgo positivo y sistemático en la predicción de la demanda.

En estos casos, una opción alternativa es confiar en los resultados observados en proyectos similares realizados en otros ámbitos geográficos. En particular, esta forma de proceder es útil para evaluar la generación de nuevos viajes y los cambios en la distribución modal derivados de una reducción del coste generalizado de una determinada alternativa. Asimismo, es aconsejable para determinar la tasa de crecimiento de la demanda a partir del primer año de operación.

La incorporación de la información contenida en los estudios tomados como referencia puede realizarse a través de la elasticidad estimada. Dado que la elasticidad es una medida agregada de la respuesta de la demanda a cambios en sus principales determinantes –precio, tiempo asociado al viaje o renta de los individuos – conocer el valor que toma en trabajos para inversiones similares puede ser una forma de evitar valores inflados de demanda que

no son confirmados posteriormente. Por lo tanto, puede ser conveniente y económico aplicar elasticidades derivadas de otros estudios para realizar análisis preliminares de un proyecto concreto o cuando no exista información o posibilidades de realizar una aproximación más rigurosa con un modelo de demanda apropiado. Aún así, esta práctica debe considerarse únicamente complementaria o preliminar al estudio detallado de demanda cuando se trata de decisiones importantes de políticas o inversiones de transporte.

Para evitar un uso incorrecto de valores hallados en otros estudios, es importante distinguir entre los distintos conceptos de elasticidad. En particular, debemos diferenciar entre la elasticidad de la demanda de transporte y de la demanda frente a los modos de transporte. La primera está ligada a un concepto de demanda de mercado y se refiere a la demanda de transporte en relación con otros sectores o bienes. La segunda es una elasticidad dentro de la demanda de transporte y considerando únicamente este mercado. Debido a que el nivel de agregación en los mercados y/o modos puede ser distinto, hay que tener presente que las elasticidades estimadas en distintos trabajos, pueden no ser comparables.

Una segunda distinción importante hace referencia a la elasticidad estimada en un modelo de elección entre alternativas (que estiman la distribución de un volumen fijo de viajes entre los distintos modos). Estas estimaciones derivan elasticidades directas y cruzadas entre modos de transporte, y se diferencian de las elasticidades descritas previamente en que no permiten la generación o supresión de viajes como resultado de una variación en el coste generalizado de una alternativa, sino que sólo permiten distribución entre modos<sup>10</sup>.

El contexto concreto en el que haya que predecir la demanda es el que determinará qué tipo de elasticidad es la más apropiada para cada caso. Dado que la evidencia muestra modelos de demanda muy diversos y dispares, resulta imposible recomendar valores generalizados de elasticidades. Una forma aconsejable de operar es establecer categorías de las inversiones según la elasticidad de la demanda respecto a los componentes del coste generalizado y aplicar valores promedio hallados a partir de experiencias previas<sup>11</sup>.

<sup>10</sup> Conviene mencionar que es posible establecer relaciones entre los distintos conceptos de elasticidades y, por lo tanto, entre los valores de estas elasticidades si es necesario.

<sup>11</sup> Para una revisión más extensa de elasticidades en distintos contextos se recomienda, además de los trabajos mencionados, consultar:

- *BTE, Transport Elasticities Database Online* (<http://dynamic.dotars.gov.au/btre>) que contiene 200 referencias bibliográficas y 400 tablas recogidas de literatura internacional de transporte.
- Oum, T., Waters, W. Yong, J. (1992), *A survey of recent estimates of price elasticities of demand for transport* (1990). Policy, Planning and Research working paper, World Bank ([www.worldbank.org/transport](http://www.worldbank.org/transport))
- *Passenger Demand Forecasting Handbook* (1997)

Debemos señalar que la interpretación de las elasticidades ha de estar ligada a la aproximación elegida para predecir la demanda y al tipo de datos utilizados.

A continuación, se presenta una recopilación de valores de elasticidades para dos tipos de inversión respecto a las que se ha generado un vivo debate en relación con la predicción de la demanda.

La primera elección hace referencia a las inversiones en ferrocarril y, en particular, en alta velocidad. Una muestra de la importancia de dicha inversión es la prioridad que el PEIT le otorga<sup>12</sup>. La revisión de los estudios permite observar que la variable más determinante de la demanda en tren de alta velocidad es el tiempo de viaje. La elasticidad es superior a la unidad y más alta para el segmento de viajes de negocio. Una elasticidad tiempo de -1.3 implica que una reducción del tiempo de viaje en un 1% genera un incremento de la demanda de un 1,3%. Respecto a la elasticidad-precio de la demanda de viajes en tren de alta velocidad, los valores hallados muestran también que la demanda es elástica. En este caso, sin embargo, los viajes por motivos de negocio muestran una sensibilidad a una variación del precio menor que para el resto de motivos.

Los valores de las elasticidades expuestas previamente pueden ser útiles para predecir cambios en la distribución modal de un corredor en el que el tren de alta velocidad es o puede ser una alternativa de viaje. Además del desvío de viajes entre modos, hay una cantidad importante de viajes que son nuevos y que se han generado al alterarse los costes generalizados relativos de las alternativas de transporte disponibles. La literatura consultada muestra la importancia de estimar la magnitud de los viajes generados, que suponen un crecimiento neto del número de viajes pero también muestra los escasos avances al respecto en las últimas décadas. Aunque la evidencia disponible no es muy amplia, existe alguna aportación que trata de incorporar este efecto. No obstante, son aproximaciones que exigen bases de datos muy completas y de escasa disponibilidad en la mayoría de los casos<sup>13</sup>. Por ejemplo, Yao y Morikawa (2005) estiman un modelo integral de transporte interurbano en cuatro etapas (generación de viajes, elección de destino, elección modal y elección de ruta)

---

<sup>12</sup> Muchos de los trabajos consultados calculan elasticidades para la demanda de tren convencional; sin embargo, los estudios específicos para viajes en tren de alta velocidad son más escasos. Los resultados del *cuadro 1* están referidos a alta velocidad.

<sup>13</sup> Algunos modelos estiman la demanda inducida incorporando variables que miden el incremento en la accesibilidad derivado de una inversión en infraestructura de transporte. Con variables socioeconómicas y demográficas se estima la demanda generada a partir de cambios en la localización de actividades económicas y de usos del suelo. Sin embargo, siendo consistentes en la aplicación del análisis coste-beneficio, esto es un efecto indirecto de la inversión, y se podría estar incurriendo en una doble contabilización de los beneficios sociales que ya se estarían valorando en términos de ahorros de tiempo.

con estimación del tráfico inducido en un corredor en el que se pretende introducir un segundo tren de alta velocidad. Se consideran los viajes en coche, en tren y en avión y utilizan datos desagregados y de tipo agregado. Estiman elasticidades para el tráfico inducido respecto al tiempo de viaje (-1.2 para negocio y -0.7 para ocio) y al coste (-1.7). Estos autores calculan que la puesta en marcha del tren de alta velocidad inducirá un incremento del tráfico del 16.5%.

**Cuadro 1: Evidencia empírica de elasticidad-precio y elasticidad-tiempo para el Tren de Alta Velocidad**

Autor	Modos	Modelo	E-precio	E-tiempo
<b>Dodgson (1991)</b>	<b>TREN</b>	Revisión de estudios	-1.0	
<b>Nash (1991)</b>	<b>TREN</b>	Revisión de estudios	-0.5 (negocio) -1.25 (ocio)	
<b>TRANSES (1991)</b>	<b>TREN</b>	Revisión de estudios		-1.6 (negocio) -0.6 (vacaciones) -1.1 (otros)
<b>Wardman (1992)</b>	<b>TREN</b>	Revisión de estudios		-1.3
<b>Nash (1997)</b>	<b>TREN</b>			-1.3
<b>Hensher (1997)</b>	<b>AVIÓN</b>	HEVLogit (avión/tav)	-3.0 / -2.3 (avión) -1,2/-0,4 (TAV)	
	<b>COCHE</b>	HEV (coche/tav)	-0,16 (coche) -0,47/-0,8 (TAV)	
<b>Nombela (2007)</b>	<b>AVIÓN</b>		-1.2	-0.2
	<b>TREN</b>	MNL (agregado)	-0.4	-1.7
	<b>BUS</b>		-0.6	-2.1
	<b>COCHE</b>		-0,06	-0.5
<b>González-Savignat (2004)</b>	<b>AVIÓN</b>	E.discreta (avión/tav)	-0.57 (negocio) -1.34 (ocio)	-1.20
	<b>COCHE</b>	E.discreta (coche/TAV)	-0,75/-1.0 (negocio) -1.14 /-1.6 (ocio)	-0.92/-1.20 (negocio) -0.68/-0.9 (ocio)

Debido a la escasa evidencia al respecto, puede ser importante revisar lo ocurrido en corredores en los que existe el tren de alta velocidad. En Francia, la experiencia del TGV muestra que un 33% de los pasajeros proceden del avión, 18% de la carretera y el 49% de los viajes es demanda inducida. Para el caso del tren de alta velocidad en Francia y Japón se obtiene una cifra entre 30-35% del tráfico desviado (King, 1996) y en Australia del 26% (Hensher, 1997). Para el caso español, en el corredor Madrid-Barcelona, De Rus y Román (2006) estiman una cifra aproximada del 31% del total de viajes (viajeros que realizan el viaje por primera vez). Vickerman (1997) señala que en el caso del TAV Madrid-Sevilla, el 32% de los viajeros proceden del avión, el 25% del coche y el 26% es tráfico generado. INRETS/IMTRANS/INTRAPLAN (2004) predicen el tráfico de pasajeros en la extensión de la futura red de alta velocidad en Europa y estiman que el 33% de la demanda será inducida. En la mayoría de las inversiones estas cifras se alcanzan entre el primer y segundo año de su puesta en servicio.

De la evidencia empírica consultada se desprende que una primera aproximación razonable a la demanda futura de una inversión en tren de alta velocidad puede ser considerar que 2/3 de la demanda procederá del desvío de viajeros de otros modos y 1/3 de nuevos viajes generados por el proyecto.

El segundo ámbito de análisis se refiere al tráfico inducido como consecuencia de una mejora en la infraestructura viaria. El *cuadro 2 (a y b)* resume los resultados de la evidencia empírica disponible para proyectos de inversión en carretera. Los estudios revisados estiman el tráfico inducido a partir de una variable proxy para la mejora de calidad en la infraestructura. Estas proxies son el número de kilómetros de red, la velocidad de circulación y el tiempo de viaje.

**Cuadro 2a: Estimaciones de tráfico inducido para proyectos de inversión en carretera. MODELOS DE DATOS DESAGREGADOS**

Estudio	Contexto	Elasticidad Kms. de red	Elasticidad tiempo de viaje o velocidad	Impacto adicional sobre la demanda
<b>Stratham et al. (2000)</b>	Datos individuales, 48 áreas EE.UU.	0.29		
<b>Barr (2000)</b>	Datos individuales, EE.UU.		-0.3 -0.5	Positivo pero inferior al hallado en otros estudios

Cuadro 2b: Estimaciones de tráfico inducido para proyectos de inversión en carretera. MODELOS DE DATOS AGREGADOS

Estudio	Contexto	Elasticidad Kms. de red	Elasticidad tiempo de viaje o velocidad	Impacto adicional sobre la demanda
<b>SACTRA (1994)</b>	Revisión 151 proyectos mejoras carretera	-	-	10.4 %
<b>SACTRA (1994)</b>	Revisión 85 proyectos mejora congestión	-	-	16.4%
<b>Goodwin (1996)</b>	Carretera	-	Tiempo de viaje: -0.5 C.P. -0.9 L.P.	10% C.P. 20% L.P.
<b>Hansen y Huang (1997)</b>	Carretera estado California	0.3–0.7 condados 0.5–0.9 área metr.	-	-
<b>Heanue (1998)</b>		-	-	6% - 22%
<b>Fulton et al. (2000)</b>	Carretera estados California, Virginia, Maryland	0.3–0.5 C.P. condados	-	-
<b>Cervero y Hansen (2002)</b>	Carreteras California	0.56 condados	-	-
<b>Noland (2001)</b>	Carreteras estatales EE.UU.	Urbanas: 0.44–0.51 C.P. 0.81–1.0 L.P.	-	21.0% - 26.5%
<b>Noland (2001)</b>	Carreteras estatales EE.UU.	Rurales: 0.23–0.41 C.P. 0.71–1.1 L.P.	-	21.0% - 26.5%
<b>Cervero (2003)</b>	Carreteras, California corredores suburbanos	-	Velocidad 0.24 C.P. 0.64 L.P.	-

C.P.: Corto plazo

L.P.: Largo plazo.

Los estudios concluyen que la inversión en carretera genera nuevos viajes. Por consiguiente, las evaluaciones basadas en la hipótesis de una matriz de viajes fija no son adecuadas. Sin embargo, la magnitud del tráfico generado varía de forma significativa entre proyectos. Con toda probabilidad, el impacto será más alto en aquellas carreteras que

operen al límite de su capacidad, cuando la elasticidad respecto del coste generalizado es elevada, y en proyectos que supongan una fuerte disminución de los costes de viaje.

Un valor promedio para las inversiones en carretera se situaría en una elasticidad de la demanda respecto a los kilómetros de red alrededor de 0.3 en el corto plazo y 0.6 en el largo plazo. Los estudios que miden el coste generalizado en términos de velocidad ofrecen resultados inferiores (Cervero, 2003). Cabe insistir que dichas elasticidades incluyen desviación de tráfico desde otros modos de transporte. Estas elasticidades suponen un impacto adicional sobre los viajes de alrededor de un 20%. El Departamento de Transporte del Reino Unido en su Transport Analysis Guide recomienda incorporar un 15% de nuevos viajes generados en las horas valle. No obstante, los estudios con datos desagregados ofrecen valores inferiores. A falta de datos para la red de tráfico en España, un valor del tráfico inducido del 20% debería tomarse como una cota máxima.

#### 4. INCERTIDUMBRE ASOCIADA A LA PREDICCIÓN DE LA DEMANDA

La incertidumbre asociada a la predicción de la demanda da lugar a errores de predicción que pueden distorsionar el resultado de la evaluación<sup>14</sup>. Es por ello recomendable buscar una forma de incorporar los posibles errores de predicción al mecanismo de evaluación.

Los errores en la predicción de la demanda obedecen a dos fuentes. En primer lugar, al desconocimiento del valor que las variables explicativas de la demanda tomarán a lo largo de la vida activa del proyecto; por ello, es necesario predecir previamente dichas variables. En segundo lugar, a los errores relacionados con la especificación de la ecuación de demanda – variables omitidas, supuestos inadecuados sobre la forma funcional y sobre la distribución estadística de los componentes aleatorios – y/o al error debido al uso de parámetros estimados en vez de valores reales.

El análisis de la varianza del error de predicción puede plantearse en los siguientes términos.

El modelo teórico o proceso de generación de los datos es:

$$y_i = x_i' \beta + u_i$$

---

<sup>14</sup> Para una revisión de la literatura relacionada con los errores en la predicción de la demanda de transporte ver González-Savignat, Matas, Raymond y Ruiz (2009).

En donde “ $y$ ” es la demanda, “ $x$ ” el vector de variables explicativas, “ $\beta$ ” los parámetros poblacionales del modelo y “ $u$ ” las perturbaciones aleatorias que captan la influencia de las variables omitidas.

El valor que adoptará la demanda en el período a predecir “ $T + h$ ” viene dado por:

$$y_{T+h} = x'_{T+h} \cdot \beta + u_{T+h}$$

El predictor utilizado viene dado por:

$$y_T(h) = \hat{x}'_{T+h} \cdot \hat{\beta}$$

en donde “ $y_T(h)$ ” significa la predicción de la demanda en el período “ $T + h$ ” efectuada en el período “ $T$ ”. Es decir, se trata de una predicción “ $h$ ” períodos hacia adelante. Los “ $\hat{\beta}$ ” son los valores estimados de los parámetros poblacionales del modelo y “ $\hat{x}_{T+h}$ ” son los valores predichos de las variables explicativas en el período “ $T + h$ ”. Denominando al error de predicción de las variables explicativas “ $v$ ”, se tiene:

$$y_T(h) = \hat{x}'_{T+h} \cdot \hat{\beta} = x'_{T+h} \cdot \beta + v'_{T+h} \cdot \hat{\beta}$$

A partir de aquí el error de predicción puede expresarse como:

$$e_T(h) = y_{T+h} - y_T(h) = x'_{T+h} \cdot (\beta - \hat{\beta}) + u_{T+h} + \hat{\beta}' \cdot v_{T+h}$$

En este contexto, la varianza del error de predicción “ $\hat{\sigma}_f$ ” puede expresarse como:

$$\hat{\sigma}_f^2 = x'_{T+h} \cdot \hat{cov}(\hat{\beta}) \cdot x_{T+h} + \hat{\sigma}_u^2 + \hat{\beta}' \cdot \hat{\Sigma}_v \cdot \hat{\beta}$$

siendo “ $\hat{cov}(\hat{\beta})$ ” la estimación de la matriz de varianzas y covarianzas de los coeficientes estimados, “ $\hat{\sigma}_u^2$ ” la varianza estimada de la perturbación aleatoria y “ $\hat{\Sigma}_v$ ” la matriz de varianzas y covarianzas de los errores predictivos en las variables explicativas.

En esta fórmula los componentes de la varianza del error de predicción son los siguientes:

- Varianza del error de predicción atribuible a la incertidumbre en el valor de los coeficientes del modelo:  $x'_{T+h} \cdot \hat{cov}(\hat{\beta}) \cdot x_{T+h}$

- Varianza del error de predicción atribuible a la influencia de la perturbación aleatoria:  $\hat{\sigma}_u^2$
- Varianza del error de predicción atribuible al desconocimiento de los valores futuros de las variables explicativas:  $\hat{\beta}' \cdot \hat{\Sigma}_v \cdot \hat{\beta}$

La fórmula recomendada para aislar estos componentes es la siguiente:

- La matriz de varianzas y covarianzas de los coeficientes estimados beta puede obtenerse a través de la correspondiente expresión analítica que se deduce del método de estimación empleado. Es aconsejable, a efectos de comparación, obtener esta matriz de varianzas y covarianzas por simulación (técnica del bootstrap).
- La varianza de la perturbación aleatoria puede obtenerse a través de la correspondiente expresión analítica.
- Finalmente, la matriz de varianzas y covarianzas de los errores de predicción en las variables explicativas puede aproximarse por simulación o a través de procedimientos analíticos. Por ejemplo, si se empleasen modelos VAR para predecir los valores futuros de las variables explicativas, los propios modelos VAR estimados ofrecen esta información.

En este último caso, una alternativa frecuentemente utilizada consiste en diseñar distintos escenarios que abarquen un intervalo razonable de variación para las variables exógenas, de tal forma que calculemos los beneficios y costes asociados al proyecto en cada uno de los escenarios. Esta es una forma simple de incorporar más información en el resultado de la evaluación y permite tomar decisiones sobre aceptación o rechazo del proyecto con mayor fundamento. Alternativamente, es posible predecir los valores futuros de las variables explicativas empleando modelos VAR. En este caso, la matriz de varianzas y covarianzas de los errores de predicción en las variables explicativas nos la ofrecen los propios modelos VAR estimados.

Cabría, finalmente, la posibilidad de emplear métodos puramente de simulación a efectos de calcular la distribución de los errores predictivos.

## 5. CONCLUSIONES Y RECOMENDACIONES FINALES

La predicción de la demanda de transporte asociada a un proyecto que modifique los costes de un determinado modo de transporte es ciertamente una tarea compleja y sujeta a múltiples decisiones. Se trata, sin embargo, de una tarea imprescindible y que requiere la máxima atención para acertar en la selección de proyectos.

Una parte importante de esta complejidad se deriva del hecho de que cualquier modo de transporte opera en forma de red y de la existencia de relaciones de sustituibilidad y complementariedad entre alternativas. De ahí que cualquier inversión tenga consecuencias no sólo sobre la demanda del propio modo, sino sobre la demanda del resto de rutas y modos competitivos.

Estas consecuencias son específicas para cada tipo de inversión y no existen reglas generales que puedan ser aplicadas.

Así, el primer paso en la predicción es definir el área afectada por el proyecto y determinar el origen del tráfico distinguiendo entre tráfico existente y tráfico inducido; este último comprende tanto el tráfico desviado de otros modos como el de nueva generación.

Para determinar el tráfico existente es suficiente disponer de datos de tráfico desagregados por rutas, corredores o pares de origen-destino. Estos datos están, en general, disponibles a partir de las estadísticas oficiales. En aquellos casos donde se prevé sustitución entre rutas es recomendable llevar a cabo encuestas origen y destino para establecer el tráfico desviado hacia la ruta evaluada.

Para cuantificar el tráfico desviado desde otros modos es necesario estimar un modelo de elección modal, bien de carácter agregado, bien con datos individuales. Esta última opción es recomendable en proyectos de gran envergadura y con un impacto significativo en las variables que determinan las preferencias de los individuos. Sin embargo, dada la incertidumbre asociada a los modelos de demanda y a la propia cuantificación de las variables explicativas de cada alternativa –principalmente coste monetario y tiempo de viaje– se recomienda comparar las predicciones del modelo de demanda con los resultados obtenidos en proyectos de similares características y/o usar los valores estándares de las elasticidades apropiados a cada caso.

La predicción del tráfico generado por las mejoras introducidas por el proyecto es un tema todavía más complejo. Por ello, en el contexto actual, resulta recomendable emplear los valores observados en proyectos similares para evitar una sobrepredicción del tráfico.

Los modelos econométricos disponibles para predecir la demanda cuentan con una larga tradición en la economía del transporte. Se trata de modelos estándares sobre los que en este documento se han detallado las principales características. No obstante, dada la complejidad del impacto de una inversión sobre la demanda, no es posible ofrecer una guía única. Cada proyecto deberá evaluar los métodos disponibles y elegir acorde con las características de la inversión y la información disponible.

La predicción de la demanda debe incorporar siempre un análisis de sensibilidad que tenga en cuenta la incertidumbre asociada con todo el proceso.

## REFERENCIAS

- Barr, L.C., (2000), 'Testing for the significance of induced highway travel demand in metropolitan areas', *Transportation Research Record*, 1706, 1-8.
- Bonnafous, A.,(1987), 'The Regional Impact of the TGV'. *Transportation*, 14-2, 127-137.
- Cervero, R. y M.Hansen, (2002), 'Induced travel demand and induced road investment: a simultaneous-equation analysis', *Journal of Transport Economics and Policy*, 36, 469-490.
- Cervero, R., (2003), 'Road expansion, urban growth, and induced travel', *APA Journal*, 69, 145-163.
- BTE, Transport Elasticities Database Online* (<http://dynamic.dotars.gov.au/btre>) que contiene 200 referencias bibliográficas y 400 tablas recogidas de literatura internacional de transporte.
- Department of Transport, (1994), '*Trunk roads and the generation of traffic*', SACTRA, London.
- Department of Transport, (2006), *Transport Analysis Guide (TAG)*, Reino Unido.
- de Rus, G. y Román, C., (2006), 'Análisis económico de la línea de alta velocidad Madrid-Barcelona', *Revista de Economía Aplicada*, 42,.
- Dodgson, J.S., (1991), 'Forecasting inter-modal interaction', en J.H.Rickard y J. Larkinson (eds). *Longer term Issues in Transport: A research agenda*. Avebury, Aldershot.
- Fulton, L.M. et al., (2000), 'A statistical analysis of induced travel effects in the US mid-Atlantic region', *Journal of Transportaiton and Statistics*, 3, 1-14.
- González-Savignat, M., (2004), 'Competition in Air Transport: The case of the high-speed train', *Journal of Transport Economics and Policy*, vol 38-1.
- González-Savignat, M., Matas, A., Raymond J.L. y Ruiz, A., (2009), '*Predicción de la demanda: Análisis de la incertidumbre y modelos de predicción en España*', Documento de Trabajo.

- Goodwin, P.B. (1996), 'Empirical evidence on induced traffic, a review and synthesis', *Transportation*, 23, 35-54.
- Hansen, M y Y.Huang (1997), 'Road supply and traffic in California urban areas', *Transportation Research A*, 31, 205-218.
- Heanue, K., (1998), 'Higway capacity and induced travel: issues, evidence and implications', *Transportation Research Circular*, 481, TRB.
- Hensher, D.A., (1997), 'A Practical Approach to Identifying the Market Potential for High-Speed Rail: A Case Study in the Sydney-Canberra Corridor', *Transportation Research-A*, vol 31-6, 431-446.
- Hills, P.J., (1996), 'What is induced traffic?', *Transportation*, 23, 5-16.
- INRETS/INTRAPLAN/IMTRANS, (2004), *Passenger Traffic Study 2010/2020*.
- Mackie, P., (1996), 'Induced traffic and economic appraisal', *Transportation*, 23, 103-119.
- Mandel, B., Gaudry, M. and Rothengatter, W., (1997), 'A disaggregate Box-Cox logit choice model of intercity passenger travel in Germany and its implications for high speed rail demand forecasts'. *The Annals of Regional Science*, 31, 99-120.
- Ministerio de Fomento, (2006), 'Recomendaciones para la evaluación económica, coste-beneficio de estudios y proyectos de carreteras', Servicio de Planeamiento.
- Nash, C.A., (1991), 'The case for high speed rail'. *Investigaciones Económicas*, 15, 337-354.
- Noland R.B., (2001), 'Relationships between highway capacity and induced vehicle travel', *Transportation Research, A*, 35, 47-72.
- Nombela, G, y Martín, J. (2007), 'Microeconomic impacts of investment in high speed trains in Spain', *The Annals of Regional Science*, 41-3, pg 715-733.
- Ortuzar, J.D. y L.G. Willumsen, (2001), *Modelling Transport*, 3ª edición, John Wiley & Sons, Chichester.
- Oum T.H., (1989), 'Alternative demand models and their elasticity estimates', *Journal of Transport Economics and Policy*, mayo.

- Oum, T., Waters, W. Yong, J., (1992), 'A survey of recent estimates of price elasticities of demand for transport (1990)'. Policy, *Planning and Research working paper*, World Bank ([www.worldbank.org/transport](http://www.worldbank.org/transport))
- Oum, T.H., Waters, W.G. and Yong, J.S., (1992), 'Concepts of price elasticities of transport demand and recent empirical estimates', *Journal of Transport Economics and Policy*, 139-153.
- Strathman, J.G. et al., (2000), 'Analysis of induced travel in the 1995 NPTS', *Working Paper*, Center for Urban Studies, Portland State University.
- Train, K.E., (2003), 'Discrete choice methods with simulation', *Cambridge University Press*, Reino Unido.
- TRANSES, (1991), *Economic evaluation of the European High Speed Network*. Report prepared for European Commission, Tilburg, Netherlands.
- Vickerman, R., (1997), 'High-speed rail in Europe: experience and issues for future development', *The Annals of Regional Science*, 31, 21-38.
- Wardman, M., (1992), 'The European passenger travel market: niches for high speed rail', *Working Paper 359*. Institute for Transport Studies, University of Leeds.
- Wardman, M., (1993), 'The effect of rail journey time improvements: some results and lessons of British experience relevant to high speed rail forecasting'. *Working Paper 388*. Institute for Transport Studies, University of Leeds.
- Wardman, M., Lythgoe, W. y Whelan, G. (2007), 'Rail passenger demand forecasting: cross-sectional models revisited, en Dennis y Talley (eds.)', *Railroad Economics*, Research in Transportation Economics, vol. 20, Elsevier Ltd., 119-152.
- Wooldridge, J.M., (2003), *Introductory Econometrics: A modern approach*, Thomson, South-Western.
- Yao, E. and Morikawa, T., (2005), 'A study of an integrated intercity travel demand model', *Transportation Research Part A*, 39.

## Anexo 1: LA FORMA FUNCIONAL

La forma funcional que relaciona la demanda con el conjunto de las variables explicativas vendrá determinada tanto por los fundamentos teóricos de la función de demanda como por la naturaleza de los datos. Una propiedad deseable de la forma funcional es que sea suficientemente flexible y que no restrinja a priori las propiedades de la función de demanda. En general, un buen punto de partida es la transformación Box-Cox, que consiste en transformar las variables de la ecuación de la siguiente forma:

$$V(\lambda) = \frac{V^\lambda - 1}{\lambda}$$

en cuyo caso

- Si  $\lambda \rightarrow 0 \Rightarrow V(\lambda) \rightarrow \ln(V)$
- Si  $\lambda \rightarrow 1 \Rightarrow V(\lambda) \rightarrow V - 1$

Por tanto, la transformación de Box-Cox anida como casos especiales a las especificaciones lineal (en realidad, para un  $\lambda$  unitario, la variable transformada pasa a ser  $V-1$ , pero restar la unidad a la variable dependiente no tiene implicaciones porque queda absorbido por el término constante de la ecuación) y la logarítmica (el valor de  $\lambda$  ha de ser positivo pero puede ser arbitrariamente pequeño, por lo que en el límite la variable transformada es indistinguible de la adopción de logaritmos)

En términos económicos la transformación Box-Cox se halla falta de una intuición económica que la respalde. Por ello suele ser frecuente contrastar a partir de tal formulación los dos valores extremos de  $\lambda$  (lambda cero y lambda uno) y seleccionar el valor de lambda no rechazado por los datos. Si ambos valores de lambda no resultan rechazados por los datos, habría que seleccionar el que ofreciese un mejor ajuste comparando las respectivas funciones de verosimilitud referidas a la misma variable dependiente. Si ambos resultasen rechazados, los datos nos estarían indicando que para el modelo especificado existe una alternativa a la transformación logarítmica o lineal que mejora la capacidad de ajuste del modelo.

La ecuación de partida es:

$$y_i(\lambda_1) = \alpha + \beta_1 X_{i1}(\lambda_2) + \beta_2 X_{i2}(\lambda_2) + \dots + \beta_n X_{in}(\lambda_2) + \varepsilon_i$$

Donde:

- $y_i$  : El flujo de viajes.

- $X_{in}$  : Las n variables explicativas.
- $\varepsilon_i$  : La perturbación aleatoria.

Es preciso notar que el parámetro  $\lambda$ , que se estima conjuntamente con el resto de parámetros de la ecuación, puede ser distinto a ambos lados de la igualdad.

Las funciones anidadas en la Box-Cox que con mayor frecuencia se utilizan son:

- a.** Ecuación lineal,  $\lambda \rightarrow 1$

$$y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in} + \varepsilon_i,$$

- b.** Ecuación doble logarítmica,  $\lambda \rightarrow 0$

$$\ln(y_i) = \alpha + \beta_1 \ln(X_{i1}) + \beta_2 \ln(X_{i2}) + \dots + \beta_n \ln(X_{in}) + \varepsilon_i$$

Una tercera forma funcional que puede resultar adecuada cuando alguna de las variables explicativas toma valores iguales o muy cercanos a cero es:

- c.** Ecuación semi-logarítmica,  $\lambda_1 \rightarrow 0$  y  $\lambda_2 = 1$ .

$$\ln(y_i) = \alpha + \beta_1 (X_{i1}) + \beta_2 (X_{i2}) + \dots + \beta_n (X_{in}) + \varepsilon_i$$

La particularidad de estas tres últimas funciones es que son lineales y, por lo tanto, fáciles de estimar. Además, no son tan restrictivas como a primera vista pudiera parecer dado que permiten incluir potencias de las variables explicativas y productos cruzados entre ellas. La función logarítmica, en general, se ajusta bien a los datos y los parámetros estimados pueden interpretarse como elasticidades. No obstante, el supuesto de elasticidad constante es discutible cuando evaluamos cambios discretos y de elevada magnitud en las variables explicativas del modelo.

Siempre es posible buscar formas funcionales más complejas, tales como la transformación translog o la exponencial. No obstante, dichas formas dificultan tanto la interpretación de los resultados como la estimación de la ecuación. En muchos casos, además, la calidad de los datos disponibles impide hallar resultados razonables<sup>15</sup>.

---

<sup>15</sup> Para un análisis de las posibles formas funcionales ver Oum (1989), Wardman,(1999) y Wardman, Lythgoe and Whelan (2007).

La elección entre formas funcionales puede hacerse según el grado de ajuste de los datos a la misma. Al igual que en el modelo tendencial, el valor de la función de verosimilitud es un criterio de selección adecuado. Es preciso, no obstante, que la función esté referida a la variable dependiente expresada en las mismas unidades. Por ejemplo, si se compara la formulación lineal con la logarítmica, la función de verosimilitud no es directamente comparable y debe ajustarse tal y como se ha detallado en el apartado anterior.

Por último, debemos tener en cuenta que el mejor ajuste dentro de la muestra no necesariamente implica la mejor capacidad predictiva postmuestral.

## Anexo 2: EL ANÁLISIS DE COINTEGRACIÓN EN SERIES TEMPORALES

La noción de cointegración contempla la posibilidad de que al relacionar series integradas, la parte no explicada sea estacionaria. En tal caso se dice que los procesos son cointegrados. La integración y la cointegración pueden ser de distintos órdenes. Con objeto de simplificar, nos centraremos en procesos integrados de primer orden dado que en la práctica posiblemente sean los más frecuentes. En tal caso, para que la relación en niveles entre series integradas de orden 1,  $I(1)$ , tenga sentido, debe verificarse que los dos procesos sean cointegrados<sup>16</sup>. Suponiendo que  $y_t$  y  $x_t$  son dos series  $I(1)$ , se dice que están cointegradas si es posible hallar un  $\beta$  tal que  $y_t - \beta x_t$  sea  $I(0)$ ; lo que implica que tenga media y varianza constante. La existencia de cointegración puede contrastarse a partir de la estimación de la ecuación  $y_t = \hat{\alpha} + \hat{\beta}x_t$ , aplicando el test de Dickey-Fuller o de Dickey-Fuller Ampliado ADF a los residuos  $\hat{u}_t = y_t - (\hat{\alpha} + \hat{\beta}x_t)$ .

Volviendo a la ecuación de demanda, bajo la hipótesis de que existe cointegración entre las variables, es posible especificar la ecuación en niveles. Tomando, por ejemplo, la ecuación logarítmica:

$$\ln(y_t) = \alpha + \beta_1 \ln(X_{t1}) + \beta_2 \ln(X_{t2}) + \dots + \beta_n \ln(X_{tn}) + \varepsilon_t$$

los coeficientes  $\beta$ , parámetros de cointegración, deben interpretarse como el impacto a largo plazo de cada una de las variables en la demanda.

Adicionalmente, la cointegración permite especificar una ecuación que toma en consideración los aspectos dinámicos a corto plazo en la relación entre las variables. Esta ecuación se conoce como Mecanismo de Corrección de Error (MCE) y para el caso de dos variables se especifica como:

$$\Delta y_t = \alpha_0 + \gamma_0 \Delta x_t + \phi \Delta y_{t-1} + \delta (y_{t-1} - \beta x_{t-1}) + u_t$$

donde  $\delta < 1$ . La ecuación puede incluir también los valores desfasados de las variables explicativas. El coeficiente  $\gamma_0$  debe interpretarse como el impacto a corto plazo de  $x_t$  sobre  $y_t$ .

<sup>16</sup> El desarrollo del concepto de cointegración se debe a Engle y Granger (1987). Para una aproximación a este tema puede consultarse Wooldridge (2003).

Si las variables  $y_t$  y  $x_t$  siguen procesos  $I(1)$  y no están cointegradas, es posible estimar un modelo dinámico en primeras diferencias:

$$\Delta y_t = \alpha_0 + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \varphi \Delta y_{t-1} + u_t$$

El coeficiente  $\beta_0$  refleja el impacto a corto plazo de  $\Delta x_t$  sobre  $\Delta y_t$  y  $\frac{(\hat{\beta}_0 + \hat{\beta}_1)}{(1-\varphi)}$  el impacto a largo plazo.

### Anexo 3: JUSTIFICACIÓN DE LA REGRESIÓN LOGÍSTICA

La regresión logística asume implícitamente las siguientes funciones de demanda:

$$\ln y_m = \alpha_m + \beta_1 X_{m1} + \beta_2 X_{m2} + \dots + \beta_k X_{mk} + \varepsilon_m$$

$$\ln y_n = \alpha_n + \beta_1 X_{n1} + \beta_2 X_{n2} + \dots + \beta_k X_{nk} + \varepsilon_n$$

De manera que,

$$\ln y_m - \ln y_n = (\alpha_m - \alpha_n) + \beta_1(X_{m1} - X_{n1}) + \beta_2(X_{m2} - X_{n2}) + \dots + \beta_k(X_{mk} - X_{nk}) + (\varepsilon_m - \varepsilon_n)$$

$$\ln\left(\frac{y_m}{y_n}\right) = (\alpha_m - \alpha_n) + \beta_1(X_{m1} - X_{n1}) + \beta_2(X_{m2} - X_{n2}) + \dots + \beta_k(X_{mk} - X_{nk}) + (\varepsilon_m - \varepsilon_n)$$

Si T es el tráfico total:

$$y_n = T - y_m$$

Denominando  $f_m$  al porcentaje de viajes captado por la alternativa  $m$ ,  $f_m = \frac{y_m}{T}$ , y sustituyendo en la anterior expresión se obtiene la regresión logística:

$$\ln\left(\frac{f_m}{1 - f_m}\right) = (\alpha_m - \alpha_n) + \beta_1(X_{m1} - X_{n1}) + \beta_2(X_{m2} - X_{n2}) + \dots + \beta_k(X_{mk} - X_{nk}) + \zeta$$

siendo,  $\zeta = \varepsilon_m - \varepsilon_n$

Una vía alternativa de justificación es suponer que a nivel micro los individuos rigen su conducta de elección a través de una formulación logística del tipo:

$$p_i = \frac{\exp(X_i' \beta)}{1 + \exp(X_i' \beta)}$$

en donde “ $p_i$ ” es la probabilidad de elegir la alternativa “ $m$ ”.

La probabilidad de elegir la alternativa “ $n$ ” viene dada por:

$$p_i = \frac{1}{1 + \exp(X_i' \beta)}$$

dado que ambas probabilidades deben sumar la unidad.

Calculando, la ratio se tiene:

$$\frac{p_i}{1-p_i} = \exp(X_i'\beta)$$

O bien:

$$\ln \frac{p_i}{1-p_i} = X_i'\beta$$

No obstante, esta ecuación no puede estimarse dado que las probabilidades individuales se desconocen. Utilizando las frecuencias observadas como un estimador de las probabilidades, se tiene:

$$f_i = p_i + \mathcal{G}_i$$

Al sustituir las probabilidades por sus estimadores se deduce:

$$\ln \frac{f_i}{1-f_i} = X_i'\beta + \xi_i$$